AD_____

Award Number: DAMD17-03-1-0023

TITLE: Intensity Modulated Radiation Treatment of Prostate Cancer Guided by High Field MR Spectroscopic Imaging

PRINCIPAL INVESTIGATOR: Lei Xing, Ph.D

CONTRACTING ORGANIZATION: Leland Stanford Junior University
Stanford CA 94305-4125

REPORT DATE: May 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-05-2006	Final	01 May 03 – 30 Apr 06

4. TITLE AND SUBTITLE
Intensity Modulated Radiation Treatment of Prostate Cancer Guided by High Field MR Spectroscopic Imaging

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0023

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lei Xing, Ph.D

E-Mail: lei@reyes.stanford.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Leland Stanford Junior University
Stanford CA 94305-4125

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
None provided.

15. SUBJECT TERMS
None provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	229	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction....4
Body....4
Key Research Accomplishments....8
Reportable Outcomes....8
Conclusions....10
References....11
Appendices....13

I. INRTODCUCTION

This Idea Award (DAMD17-03-1-0023, entitled "Intensity Modulated Radiation Treatment of Prostate Cancer Guided by High Field MR Spectroscopic Imaging") was awarded to the principal investigator (PI) for the period of May 1, 2003—April 30, 2006. This is the final report for the grant. The goal of this project is to establish biologically conformal - as opposed to anatomically conformal - IMRT as a viable modality through integration with 3T magnetic resonance spectroscopic imaging (MRSI) to more effectively kill prostate tumor cells. The underlying hypothesis driving this work is that the MRSI-guided IMRT will provide substantially improved dose distributions required to achieve greater local tumor control while maintaining, or reducing, complications to sensitive structures. The specific aims of the project are: (1) To establish a robust procedure for registering and mapping of MR spectroscopic data to CT/MRI images for prostate irradiation. (2) To develop an inverse planning system for MRSI-guided IMRT prostate treatment and demonstrate the feasibility of concurrent dose escalation to intraprostatic lesion(s) through a set of phantom studies and at least two previously treated prostate cases who had undergone CT/MRSI scans. Under the generous support from the U.S. Army Medical Research and Materiel Command (AMRMC), the PI has contributed significantly to prostate cancer research by applying physics and engineering knowledge to prostate cancer research. A number of significant conference abstracts and refereed papers have been resulted from the support. The preliminary data obtained under the support of the grant has also enabled the PI to start new research initiatives, in particularly, in adaptive prostate radiation therapy. The past year's research activities of the PI are highlighted in the following.

II.RESEARCH AND ACCOMPLISHMENTS

In current clinical practice, radiation treatment planning, performed under the guidance of MRI or CT images, is aimed at delivering a uniform dose to the whole prostate gland. This treatment scheme tacitly assumes that biology distribution in the prostate target is spatially uniform, mainly because of the lack of an effective imaging tool to differentiate regions of cancer and normal prostatic epithelium[1-3]. In reality, because radiation is toxic, the conventional treatment method often leads to clinically significant side effects and complications. Furthermore, the approach seriously limits the dose deliverable to the cancerous cells because the risk of developing later gastrointestinal and genitourinary complications increases significantly as the dose to the prostate is escalated. Therefore, many patients are treated with sub-optimal doses despite of the well-known fact that the survival probability of the patients can be greatly enhanced if radiation dose could be escalated. New treatment schemes are necessary to safely escalate radiation dose to the prostate without damaging the sensitive structures.

Recent developments of functional/metabolic imaging techniques are making it increasingly possible to identify the cancerous region(s) in the prostate[4-6]. Simultaneously, a new modality of radiation therapy, intensity modulated radiation therapy (IMT) [7, 8], has recent emerged, which provides us with unprecedented means to produce customized 3D dose distributions with sub-centimeter resolution. The goal of this project is to develop enabling tools for integrating MRSI and IMRT to selectively escalate radiation doses to the intraprostatic lesions. Successful completion of the project will lay the foundation for the next generation IMRT treatment of prostate cancer and provide radiation oncologists with a significantly improved means of delivering biologically conformal doses of ionizing radiation to the prostate gland while maintaining or reducing the dose to the adjacent bladder and rectum. When fully implemented, the system will allow us to destroy tumor cells more intelligently without unnecessarily compromising the tolerance of the adjacent normal structures. It will thus have widespread impact on prostate cancer management.

Toward improving prostate radiotherapy we have made significant progress in integration of various novel imaging modalities into treatment planning process and contributed greatly in the battle against prostate cancer. The stated goals of the proposed study have been fulfilled completely and several new initiatives are resulted. The research accomplishments are summarized below.

II.1 Refinement of 3T endorectal-coil based MRSI data acquisition techniques: Robust prostate MRSI plays an important role in biologically conformal radiation therapy (BCRT)[9]. We have made effort in enhancing the performance of 3T MRSI data acquisition developed by our group[6]. Our method is based on a 2D J-resolved acquisition to better identify the citrate and choline resonances. These radio frequency (RF) pulses, consisting of a match 90-180 spectral spatial pair, are optimized for choline imaging while minimizing peak RF power constraints that typically limit 3T body MRI. We have systematically and quantitatively evaluated the 3T volumetric spiral MRSI pulse sequences and optimized the sequence for high spatial resolution, minimum acquisition time, and maximum information content. The sequence for imaging choine and citrate has been tested on several subjects. The new imaging tool showed significant potential for improved treatment target definition. This work has been published in *Magnetic Resonance Imaging in Medicine* (Apendex-1).

II.2 Development of image registration techniques*:* Image registration is at the foundation of multi-modality image-guided radiation therapy because it is a necessary step to map imaging information acquired by a new imaging modality onto simulation CT images for treatment panning. We have provided

viable solutions to this longstanding problem and substantially improved the current rigid and deformable image registration techniques. Progress includes: (i) control-volume based image registration[10]; (ii) multistage image registration[11, 12]; and (iii) deformable registration with incorporation of *a priori* system knowledge[13]. Several important papers on the subject have been published or submitted for publication (Appendix-3, -5, -6, -7, and –10). In (i), we proposed a technique that allows us to automate the selection of control point pairs in conventional landmark based registration. In this approach, instead of attempting to find the correspondences of the control regions in the reference image through user interaction, we map each control volume to the corresponding parts of the reference image by using an automated image registration algorithm. The conventional automated image registration algorithm is then used to complete the image registration process with the guidance of auto-determined control points. The approach is robust and has great potential for clinical MRI/MRSI-CT registration (Apendix-6). In (ii), a multiscale image registration technique is developed for the registration of medical images that contain significant levels of noise (Apendix-5 and –10). Experiments using mean squares, normalized correlation, and mutual information optimal linear registration are presented that determine the noise levels at which registration using these techniques fails. Further experiments in which classical denoising algorithms are applied prior to registration are presented, and it is shown that registration fails in this case for significantly high levels of noise, as well. In (iii), we investigated a strategy of using *a priori* knowledge of the system to reduce the dimensionality of the deformable image registration problem and to speed up the registration calculation. Conventional deformable registration treats all image volume equally and the calculations are "brute-force" in nature. In reality, some regions can be mapped between the moving and fixed images with higher confidence than others. This knowledge is incorporated to greatly facilitate the BSpline (or other models) deformable calculation. In the process of warping the moving image to optimally match the two input images, only those deformations that do not modify the pre-established associations of the control volumes are permissible. This significantly reduces the search space and improves the convergence behavior of the gradient-based iterative optimization calculation. The proposed algorithm is evaluated by using digital phantoms and patient CT images. The deformable registration algorithm developed in this project was also applied to several other image-guided radiation therapy projects, such as radiation dose reduction in 4D CT[14, 15], image interpolation in 4D CT[16], onboard CT-based prostate IMRT dose validation (this work will be published in Jan. 2007 issue of *Physics in Medicine and Biology*, see Apendix-9 for detail) and electron density mapping from simulation CT to CBCT[17, 18].

II.3 Inverse planning for biologically conformable radiation therapy (BCRT)*:* It is well known that the spatial biology distribution in most tumors and sensitive structures is heterogeneous[1-3]. Recent progress in biological imaging including MRSI is making the mapping of this distribution increasingly possible. We have established a theoretical framework to quantitatively incorporate the spatial biology data into IMRT inverse planning (Appendix-2). We, for the first time, derived a general formula for determining the desired dose to each tumor voxel for a known biology distribution of the tumor based on a linear-quadratic (LQ) model[9, 19]. The desired target dose distribution was then used as the prescription for inverse planning. An objective function with the voxel-dependent prescription was constructed with incorporation of the nonuniform dose prescription. The functional unit density distribution in a sensitive structure is also considered phenomenologically when constructing the objective function. Our calculations revealed that it is technically feasible to produce deliberately nonuniform dose distributions with consideration of biological information. A comparison of the new planning scheme with that of existing techniques suggested that the new BCRT technique significantly improves the tumor control probability (TCP) while reducing or keeping the normal tissue complication probability (NTCP). Contractually, this study completes the tasks 3.1 and 3.2 (together with study II.4 described below). BCRT incorporates patient specific biological information and provides an outstanding opportunity for us to truly individualize radiation treatment and to intelligently escalate dose in prostate radiation therapy.

II.4 Image guided adaptive IMRT for improved prostate cancer treatment**:** With the development of highly conformal prostate radiation therapy techniques such as IMRT and BCRT, how to accurately deliver the high radiation dose to the prostate target(s) and verify the dose distribution while sparing the sensitive structures becomes a major concern. To ensure what is planned on computer can be realized in a clinical setting, we have (a) performed a series of dosiemtric measurements to validate the spatially non-uniform doses generated using BCRT inverse planning, and (b) developed an adaptive strategy for improved prostate radiotherapy (Appendix-8, –9, -11 and -12). Newly emerged onboard cone beam CT (CBCT) is used to acquire the volumetric anatomical information of a patient prior to treatment on a routine basis. The IMRT treatment plan is then adaptively modified with consideration of organ deformation and delivered doses. An inverse planning system based on dynamic optimization algorithm has been established [20, 21]. Image transfer from the Varian Trilogy accelerator with onboard CBCT capability has been established and dosimetric validation strategy has been developed[23, 24]. With the image registration technique described above, the CBCT can be registered to the simulation CT so that the prostate and seminal vesicle (SV) targets, the involved sensitive structures, and biologically active tumor

regions identified by MRSI can be mapped to the CBCT for adaptive replanning. The registration technique (Apendix-6 and –7) is also employed to establish a voxel-to-voxel correspondence between the CT/CBCTs for the cumulative dose calculation. With the margins used in current prostate IMRT, we found that SV, bladder and rectum doses benefited most from the adaptive therapy basis[23, 24]. In addition, our study suggested that correcting the patient's daily setup just through the translation and rotation is often not enough and accounting for the organ deformation is important, especially if the target margin is to be reduced for dose escalation or for hypofractionated treatment[22-24].

III. KEY RESEARCH ACCOMPLISHMENTS

- Refined the endorectal coil-based 3T MRSI technique for prostate imaging.
- Provided viable solutions to the longstanding deformable image registration problem and made it possible to accurately map MRI/MRSI information to treatment planning CT images.
- Setup a novel framework for BCRT inverse planning.
- Developed inverse planning system for onboard CBCT based adaptive IMRT planning.
- Established a procedure for BCRT dose verification and image guided adaptive prostate IMRT.

IV. REPORTABLE OUTCOMES

The following is a list of publications resulted from the grant support during the last funding cycle. Copies of the publication materials are enclosed with this report.

Refereed publications:

1. Kim D, Mayer D., Xing L, Daniel B, Margolis, D., Spielman D., "In vivo detection of citrate for prostate cancer at 3 Tesla", *Magnetic Resonance Imaging in Medicine*, 53, 1177-1182, 2005.
2. Yang Y and Xing L: "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution ". *Medical Physics* 32, 1473-84, 2005.
3. Schreibmann E and Xing L: "Narrow band deformable registration of prostate MRI/MRSI and CT studies". *International Journal of Radiation Oncology, Biology, Physics* 62, 595–605, 2005.
4. Schreibmann E and Xing L: "Dose-volume based ranking of incident beam directions and its utility in facilitating IMRT beam placement". *International Journal of Radiation Oncology, Biology, Physics* 63, 584-593, 2005.
5. Paquin D, Levy D, Schreibmann E., Xing L, Multistage image registration, *Mathematical Biosciences and Engineering* 3, 389-418, 2006.
6. Schreibmann E, and Xing L: "Image registration with auto-mapped control volumes". *Medical Physics* 33, 1165-79, 2006.
7. Schreibmann E, Chen G, and Xing L: "Image interpolation in 4D CT". *International Journal of Radiation Oncology, Biology, Physics* 64, 1537-1550, 2006.
8. Pawlicki T., Kim G, Hsu A, Cortutz C, Boyer A, Xing L, King C.R., Luxton G, Investigation of Linac-based Image-guided Hypofractionated Prostate Radiotherapy, Medical Dosimetry 31, 91-122, 2006.

9. Loo B, Draney MT, Sivanadan R, Ruehm S, Pawlicki T, Xing L, Herfkens RJ, Q.T. Le: "Indirect MR lymphangiography using conventional gadolinium contrast", *International Journal of Radiation Oncology, Biology, Physics*, in press, 2006.
10. Xing, L, Thorndyke B, Schreibmann E, Li T, Yang Y, Kim G., Luxton G, Koong, A, Overview of image guided radiation therapy (IGRT), *Medical Dosimetry* (invited review) 31, 91-122, 2006.
11. Thorndyke B, Schreibmann E., Koong A., and Xing L: "Enhancing the performance of 4D PET imaging by retrospective stacking". *Medical Physics* 51, in press, 2006.
12. Li T, Munro P, McGuiness C, Yang Y, Chao M, Loo B, and Koong A, Xing L, Four-dimensional cone-beam computed tomography using an onboard imager, *Medical Physics* 33, submitted, 2006.
13. Chao M, Schreibmann E., Li T, Xing L, Automated contour mapping for IGRT, *Medical Physics*, submitted, 2006.
14. Paquin D, Levy D, Xing L, Multistage deformable image registration, *IEEE Trans Med Imag*, submitted, 2006.
15. Yang Y, Schreibmann E., Li T, Xing L, Evaluation of dosimetric accuracy of kV cone beam CT-based dose calculation, *Physics in Medicine and Biology*, in press (to appear in Jan. 2007).

Book Chapters:

1. Xing L, Yang Y, and Spielman D, Molecular/Functional Image-Guided Radiation Therapy, in IMRT Handbook and *Clinical Applicatio*ns, T. Bortfeld, R. Schmidt-Ulrich, We De Neve (editors), Springer-Verlag Heidelberg, Berlin, 187-198, 2006.
2. Li J and Xing L, Radiation Dose Planning, Computer-Aided, in *Encyclopedia of Medical Devices and Instrumentation*, John G. Webster (editor), John Wiley & Sons, in press.

Conference abstract:

1. Xing, L. and Spielman D, Integration of MRI/MRSI into Radiation Therapy Treatment Planning, 2005 AAPM Annual Meeting, Seattle, WA (invited talk).
2. Schreibmann E. and Xing L., Registration of prostate MRI/MRSI and CT studies using narrow band approach, 2005 AAPM Annual Meeting, Seattle, WA.
3. Schreibmann E. and Xing L., EUD-based beam orientation optimization, 2005 AAPM Annual Meeting, Seattle, WA.
4. Xing L, Levy, D. and Yang Y., Incorporating clinical outcome data into inverse treatment planning, 2005 AAPM Annual Meeting, Seattle, WA.
5. Yang Y. and Xing L, Prescription for biologically conformal radiation therapy, 2005 AAPM Annual Meeting, Seattle, WA.
6. Schreibmann E, and Xing L: "Image registration with auto-mapped control volumes". 2005 ASTRO annual meeting, Denver, CO.
7. Yang Y and Xing L: "Optimization of radiation dose-time-fractionation scheme with consideration of tumor specific biology". 2005 ASTRO annual meeting, Denver, CO.
8. Yang Y, Levy D, and Xing, L, Relationship between EUD-based and dose-based inverse planning, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
9. Y Yang, N Yue, W Fu, X Li, D Heron, M Huq, L Xing, Time-Resolved 4D Dynamic Arc Therapy, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
10. M Chao, E Schreibmann, T. Li, L. Xing, Knowledge-Based Auto-Contouring in 4D Radiation Therapy, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
11. Q Xu, Z He, R Hamilton, L Xing, Registration of X-Ray Portal Images with 4DCT DRRs for Patient Setup Verification, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
12. A de la Zerda, B Armbruster, L Xing, A New Dose Optimization Algorithm for Adaptive Radiation Therapy, 2006 AAPM Annual Meeting, Orlando, FL, 2006.

13. M Chao,T Li, L Xing, Enhanced 4D CBCT Imaging for Slow-Rotating On-Board Imager, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
14. E Schreibmann, B Thorndyke, L Xing, Intra- and Inter-Modality Registration of Four-Dimensional (4D) Images, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
15. B Armbruster, A de la Zerda, L Xing, A New 4D IMRT Algorithm and Its Performance Analysis, 2006 AAPM Annual Meeting, Orlando, FL, 2006.
16. A Hsu, B Thorndyke, T Pawlicki, L Xing, Accuracy of Gated IMRT Delivery On the Varian Linac Using the Real-Time Position Management System**,** 2006 AAPM Annual Meeting, Orlando, FL, 2006.
17. T Li, L Xing, P Munro, Y Yang, B Loo, A Koong, 4D Cone-Beam CT (CBCT) Using An On-Board Imager**,** 2006 AAPM Annual Meeting, Orlando, FL, 2006.
18. C Wang, L Xing, Evaluation of Kv CBCT-Based Dose Verification**,** 2006 AAPM Annual Meeting, Orlando, FL, 2006.
19. E Schreibmann, W Cai, X Chen, L Xing, Voxel-Based MicroCT-MicroPET Image Registration for Molecular Imaging Study**,** 2006 AAPM Annual Meeting, Orlando, FL, 2006.
20. S Kamath, E Schreibmann, L Xing, Deformable Image Registration with Auto-Mapped Control Volumes**,** 2006 AAPM Annual Meeting, Orlando, FL, 2006.
21. Daly M, Lieskovsky Y, Pawlicki T, Yau J, Pinto H, Kaplan M, Fee W, Koong A, Goffinet D, Xing L, Le Q, Evaluation of patterns of failure and subjective salivary function in patients treated with IMRT for head and neck squamous cell carcinomas, 2006 ASCO Annual Meeting, Atlanta, GA, 2006.
22. Yang Y, Levy D, and Xing, L, Voxel-based penalty scheme for IMRT inverse planning, 2006 World Congress on Medical Physics and Bioengineering, Seoul, Korea, 2006.
23. Xing, L. and Spielman D, MRI/MRSI and Radiation Therapy Treatment Planning, 2006 World Congress on Medical Physics and Bioengineering, Seoul, Korea, 2006 (invited talk).
24. Xing, L., Li T, Thorndyke B, Schreibmann E., Chao M., and Spielman D, Integration of Molecular Imaging into Radiation Therapy Treatment Planning, 2006 AAPM Annual Meeting, Orlando, FL (invited talk).
25. Chao M, Schreibmann, Li T, and Xing L: "Automated contour mapping for IGRT". 2006 ASTRO annual meeting, Philadelphia, PA.
26. Kamath S, Schreibmann E, and Xing L: "Automated contour mapping for IGRT". 2006 ASTRO annual meeting, Philadelphia, PA.
27. A de la Zerda,, B Armbruster, and Xing L: "Inverse planning for adaptive radiation therapy". 2006 ASTRO annual meeting, Philadelphia, PA.
28. Xing L, A de la Zerda, B Armbruster, Li T, Chao M, Hancock S, King C.: "Adaptive radiation therapy for improved prostate radiation therapy". 2006 ASTRO annual meeting, Philadelphia, PA.
29. Xing L, "Image guided radiation therapy", 2006 ASTRO annual meeting, Philadelphia, PA (invited talk).

V. CONCLUSION

In this project we have successfully completed all tasks listed in the Statement of Work of the proposal. We have shown that biologically conformable radiation therapy (BCRT) is technically feasible and advantageous as compared with current treatment strategy. Coupled with biological imaging techniques such MRSI, the approach has significant potential to substantially improve the existing prostate radiation therapy. The technical tools developed in this project greatly facilitate the planning, delivery, and quality assurance of the MRSI-guided prostate treatment. Translation of the newly developed techniques to

routine clinical practice is being planned at our Hospital and it is expected that these techniques will lead to positive impact to clinical prostate management worldwide.

Reference:

1. Ling, C. C.; Humm, J.; Larson, S.; Amols, H.; Fuks, Z.; Leibel, S.; Koutcher, J. A., Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *International Journal of Radiation Oncology, Biology, Physics* **2000,** 47, (3), 551-560.
2. Alber, M.; Paulsen, F.; Eschman, S. M.; Machulla, H. J., On biologically conformal boost dose optimization. *Physics in Medicine & Biology* **2003,** 48, (10), N31-N35.
3. Xing, L.; Cotrutz, C.; Hunjan, S.; Boyer, A. L.; Adalsteinsson, E.; Spielman, D. M., Inverse Planning for Functional Image-Guided IMRT. *Physics in Medicine & Biology* **2002,** 47, (10), 3567-3578.
4. Kurhanewicz, J.; Vigneron, D. B.; Hricak, H.; Narayan, P.; Carroll, P.; Nelson, S. J., Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm3) spatial resolution. *Radiology* **1996,** 198, (3), 795-805.
5. Zaider, M.; Zelefsky, M. J.; Lee, E. K.; Zakian, K. L.; Amols, H. I.; Dyke, J.; Cohen, G.; Hu, Y.; Endi, A. K.; Chui, C.; Koutcher, J. A., Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging. *International Journal of Radiation Oncology Biology Physics* **2000,** 47, (4), 1085–1096.
6. Kim, D. H.; Margolis, D.; Xing, L.; Daniel, B.; Spielman, D., In vivo prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 Tesla. *Magnetic Resonance in Medicine* **2005,** 53, 1177-82.
7. IMRT Collaborative Working Group, Intensity-modulated radiotherapy: current status and issues of interest. *International Journal of Radiation Oncology, Biology, Physics* **2001,** 51, (4), 880-914.
8. Xing, L.; Thorndyke, B.; Schreibmann, E.; Yang, Y.; Li, T. F.; Kim, G. Y.; Luxton, G.; Koong, A., Overview of image-guided radiation therapy. *Med Dosim* **2006,** 31, (2), 91-112.
9. Yang, Y.; Xing, L., Towards biologically conformal radiation therapy (BCRT): Selective IMRT dose escalation under the guidance of spatial biology distribution. *Medical Physics* **2005,** 32), 1473-1484.
10. Schreibmann, E.; Xing, L., Image registration with auto-mapped control volumes. *Med Phys* **2006,** 33, (4), 1165-79.
11. Paquin, D.; Levy, D.; Schreibmann, E.; Xing, L., Multistage image registration. *Mathematical Biosciences and Engineering* **2006,** 3, 389-418.
12. Paquin, D.; Levy, D.; Schreibmann, E.; Xing, L., Multistage deformable image registration. *IEEE Transactions on Medical Imaging* **2006**, submitted.
13. Kamath, S.; Paquin, D.; Levy, D.; Schreibmann, E.; Xing, L. In *Incorporating priori knowledge into deformable image registration*, 2006 Annual Meeting of ASTRO, Philadelphia, PA.
14. Li, T.; schreibmann, E.; Yang, Y.; Xing, L., Model based image reconstruction for 4D computed tomography. *Physics in Medicine and Biology* **2006,** 51, 253-267.
15. Li, T.; Schreibmann, E.; Thorndyke, B.; Xing, L., Radiation Dose Reduction in 4D Computed Tomography. *Medical Physics* **2005,** 32, (6), 2094.
16. Schreibmann, E.; Chen, G. T.; Xing, L., Image interpolation in 4D CT using a BSpline deformable registration model. *Int J Radiat Oncol Biol Phys* **2006,** 64, (5), 1537-50.
17. Yang, Y.; Schreibmann, E.; Li, T.; Xing, L., Dosimetric evaluation of kV cone-beam CT (CBCT) based dose calculation. *Medical Physics* **2006**.
18. Yang, Y.; Li, T.; Schreibmann, E.; Boyer, A.; Xing, L., Is Cone Beam CT Suitable for Dose Verification? *Medical Physics* **2005,** 32, (6), 2163.
19. Yang, Y.; Xing, L., Optimization of radiation dose-time-fractionation scheme with consideration of tumor specific biology. *Medical Physics* **2005,** 32, 3666-3677.

20. de la Zerd, A.; Armbrush, B.; Xing, L., Inverse Planning for Adaptive Radiation Therapy. *Medical Physics* **2006**, submitted.
21. de la Zerd, A.; Armbrush, B.; Xing, L. In *Inverse Planning for Adaptive Radiation Therapy*, 2006 Annual Meeting of ASTRO, Philadelphia, PA.
22. Pawlicki T., Kim G, Hsu A, Cortutz C, Boyer A, Xing L, King C.R., Luxton G, Investigation of Linac-based Image-guided Hypofractionated Prostate Radiotherapy, Medical Dosimetry 31, 91-122, **2006**.
23. T. Li, E. Schreibmann, A. Koong, Q. Xu, R. Hamilton and L. Xing**,** Verification of Gated Radiation Therapy Using Pre-Treatment Four-Dimenstional Cone-Beam CT, *International Journal of Radiation Oncology*Biology*Physics, Volume 66, Issue 3, Supplement 1,* ***2006,*** *Page S604.*
24. L. Xing**,** A. de la Zerd, M. Cao, T. Li, B. Armbrush, Y. Yang, P. Lee, T. Pawlicki, S. Hancock and C. King, On-Board Volumetric CT-based Adaptive IMRT For Improved Prostate Cancer Treatment, *International Journal of Radiation Oncology*Biology*Physics, Volume 66, Issue 3, Supplement 1,* ***2006****, Pages S624-S625.*

Appendix I. manuscripts published or submitted for publication

Manuscripts:

1. Kim D, Mayer D., Xing L, Daniel B, Margolis, D., Spielman D., "In vivo detection of citrate for prostate cancer at 3 Tesla", *Magnetic Resonance Imaging in Medicine*, 53, 1177-1182, 2005.
2. Yang Y and Xing L: "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution ". *Medical Physics* 32, 1473-84, 2005.
3. Schreibmann E and Xing L: "Narrow band deformable registration of prostate MRI/MRSI and CT studies". *International Journal of Radiation Oncology, Biology, Physics* 62, 595–605, 2005.
4. Schreibmann E and Xing L: "Dose-volume based ranking of incident beam directions and its utility in facilitating IMRT beam placement". *International Journal of Radiation Oncology, Biology, Physics* 63, 584-593, 2005.
5. Paquin D, Levy D, Schreibmann E., Xing L, Multistage image registration, *Mathematical Biosciences and Engineering* 3, 389-418, 2006.
6. Schreibmann E, and Xing L: "Image registration with auto-mapped control volumes". *Medical Physics* 33, 1165-79, 2006.
7. Schreibmann E, Chen G, and Xing L: "Image interpolation in 4D CT". *International Journal of Radiation Oncology, Biology, Physics* 64, 1537-1550, 2006.
8. Xing, L, Thorndyke B, Schreibmann E, Li T, Yang Y, Kim G., Luxton G, Koong, A, Overview of image guided radiation therapy (IGRT), *Medical Dosimetry* (invited review) 31, 91-122, 2006.
9. Yang Y, Schreibmann E., Li T, Wang C, Xing L, Evaluation of dosimetric accuracy of kV cone beam CT-based dose calculation, *Physics in Medicine and Biology*,in press, 2007.
10. Paquin D, Levy D, Xing L, Multistage deformable image registration, *IEEE Trans Med Imag*, submitted, 2006.

Book Chapter:

Xing L, Wu Q, Yong Y and Boyer AL: Physics of IMRT and Inverse Treatment Planning, in Intensity Modulated Radiation Therapy: *A Clinical Perspective*, Mundt AF and Roeske JC. (editors), page 20-52, BC Decker Inc. Publisher, Hamilton, Canada, 2005.

Xing L, Yang Y, and Spielman D, Molecular/Functional Image-Guided Radiation Therapy, in IMRT Handbook and *Clinical Applicatio*ns, T. Bortfeld, R. Schmidt-Ulrich, We De Neve (editors), Springer-Verlag Heidelberg, Berlin, 187-198, 2006.

Appendix ---Reprints and preprints

1. Reprint - Kim D, Mayer D, Xing L, Daniel B, Margolis, D, Spielman D: "In vivo detection of citrate for prostate cancer at 3 Tesla", *Magnetic Resonance Imaging in Medicine* 53, 1177-1182 (2005).
2. Reprint- Yang Y and Xing L: "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution ". *Medical Physics* 32, 1473-84, 2005.
3. Reprint - Schreibmann E and Xing L: "Narrow band deformable registration of prostate MRI/MRSI and CT studies". *International Journal of Radiation Oncology, Biology, Physics*, 62, 595-605, 2005 (figures featured in the cover of the issue of the journal).
4. Reprint - Schreibmann E and Xing L: "Dose-volume based ranking of incident beam directions and its utility in facilitating IMRT beam placement". *International Journal of Radiation Oncology, Biology, Physics* 63, 584-593, 2005.
5. Reprint - Paquin D, Levy D, Schreibmann E., Xing L, Multistage image registration, *Mathematical Biosciences and Engineering* 3, 389-418, 2006 (figures featured in the cover of the issue of the journal).
6. Reprint - Schreibmann E, and Xing L: "Image registration with auto-mapped control volumes". *Medical Physics* 33, 1165-79, 2006.
7. Reprint- Schreibmann E, Chen G, and Xing L: "Image interpolation in 4D CT using a BSpline deformable model". *International Journal of Radiation Oncology, Biology, Physics* 64, 1537-1550, 2006.
8. Reprint- Xing, L, Thorndyke B, Schreibmann E, Li T, Yang Y, Kim G., Luxton G, Koong, A, Overview of image guided radiation therapy (IGRT), *Medical Dosimetry* 31, 91-122, 2006.
9. Preprint - Yang Y, Schreibmann E., Li T, Xing L, Evaluation of dosimetric accuracy of kV cone beam CT-based dose calculation, *Physics in Medicine and Biology*, in press.
10. Preprint - Paquin D, Levy D, Xing L, Multistage deformable image registration, *IEEE Transactions on Medical Imaging*, submitted, 2006.
11. Book chapter - Xing L, Wu Q, Yong Y and Boyer AL: Physics of IMRT and Inverse Treatment Planning, in Intensity Modulated Radiation Therapy: *A Clinical Perspective*, Mundt AF and Roeske JC. (editors), page 20-51, BC Decker Inc. Publisher, Hamilton, Canada, 2005.
12. Book chapter - Xing L, Yang Y, and Spielman D, Molecular/Functional Image-Guided Radiation Therapy, in IMRT Handbook and *Clinical Applicatio*ns, T. Bortfeld, R. Schmidt-Ulrich, We De Neve (editors), Springer-Verlag Heidelberg, Berlin, 2006.

In Vivo Prostate Magnetic Resonance Spectroscopic Imaging Using Two-Dimensional J-Resolved PRESS at 3 T

Dong-hyun Kim,[1]* Daniel Margolis,[1] Lei Xing,[2] Bruce Daniel,[1] and Daniel Spielman[1]

In vivo magnetic resonance spectroscopic imaging of the prostate using single-voxel and multivoxel two-dimensional (2D) *J*-resolved sequences is investigated at a main magnetic field strength of 3 T. Citrate, an important metabolite often used to aid the detection of prostate cancer in magnetic resonance spectroscopic exams, can be reliably detected along with the other metabolites using this method. We show simulations and measurements of the citrate metabolite using 2D *J*-resolved spectroscopy to characterize the spectral pattern. Furthermore, using spiral readout gradients, the single-voxel 2D *J*-resolved method is extended to provide the spatial distribution information as well all within a reasonable scan time (17 min). Phantom and in vivo data are presented to illustrate the multivoxel 2D *J*-resolved spiral chemical shift imaging sequence. Magn Reson Med 53:1177–1182, 2005. © 2005 Wiley-Liss, Inc.

Key words: magnetic resonance spectroscopic imaging; prostate cancer; spiral readout gradients; 2D *J*-resolved spectroscopy; citrate; polyamine

In addition to the morphologic information provided by magnetic resonance imaging, the additional information gained using magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) increases the specificity of the examination for prostate cancer (PCa). In these examinations, the ratio of (choline + creatine) to citrate is often regarded as a marker for PCa (1). To date, MRSI protocols for PCa detection have been well established at a main magnetic field strength of 1.5 T (2).

The advent of higher field strength scanners provides the potential for improvement over 1.5-T systems due to the inherent increase in the signal-to-noise ratio (SNR). For PCa exams using MRS/MRSI methods, this advantage can be exploited in various forms, which include using higher spatial resolution acquisitions to increase the accuracy of localization of the cancerous tissues (3). Scan times can also be made shorter compared to 1.5 T for the same SNR, thereby reducing the overall MR examination time. The extension of 1.5-T MRS/MRSI protocols for usage in 3-T PCa can therefore have potential merits.

[1]Department of Radiology, Stanford University, Stanford, California, USA.

[2]Department of Radiation Oncology, Stanford University, Stanford, California, USA.

Grant sponsor: Lucas Foundation; Grant sponsor: NIH; Grant numbers: CA 48269, RR09784, and 1R01CA098523–01A1; Grant sponsor: Department of Defense; Grant number: DAMD17–03-1–0023.

*Correspondence to: Dong-hyun Kim, Radiological Science Laboratory at the Lucas MRS/I Center, Department of Radiology, Stanford University, 1201 Welch Road, Stanford, CA 94305-5488, USA. E-mail: dhkim@mrsc.ucsf.edu

Received 4 August 2004; revised 10 December 2004; accepted 15 December 2004.

DOI 10.1002/mrm.20452
Published online in Wiley InterScience (www.interscience.wiley.com).

© 2005 Wiley-Liss, Inc.

However, the process of advancing to higher field strength requires several considerations. For clinical prostate examinations using spectroscopic techniques, one of the issues that arise involves the detection of the citrate metabolite. Strong coupling of the AB system of citrate induces echo-time-dependent modulations of the signal response, which differs significantly with field strength (4,5). One method which exploits the echo time dependencies is the 2D *J*-resolved spectroscopic sequence (6–8). Acquisitions at incremental echo times can be gathered to obtain the coupling information of such metabolites. Information of uncoupled metabolites can also be gathered. Another advantage of using 2D *J*-resolved sequences in the case of prostate is the potential to separate the polyamine metabolite from the creatine and choline peaks (9,10). The three metabolites resonate at similar frequencies, which make them hard to differentiate using normal acquisitions. But, since the polyamines are also strongly coupled, the 2D *J*-resolved method can provide additional information, which can be used to distinguish between creatine and choline. In addition, 2D *J*-resolved spectroscopy has been used to reduce sideband artifacts for applications in the brain and regions outside of the brain such as the breast (11,12). Finally, the acquisitions from multiple echo times can also help determine the T_2 values of metabolites of interest in addition to water.

To take full advantage of this method, collecting the spatial distribution information of the metabolites will be preferred over single-voxel techniques. Although phase encoded MRSI methods have been used in obtaining the spatial information, this can be problematic when combining with the 2D *J*-resolved method since the minimum total scan time will be increased proportional to the number of echo time steps used. Therefore, a different approach that can reduce the minimum total scan time needs to be established. One of the spatial encoding methods that achieve such characteristic is the spiral readout MRSI (13). Using spiral MRSI, the spatial coverage can be controlled with high efficiency, thereby significantly decreasing the minimum total scan time compared to the phase encoded method. The additional time available can therefore be used to gather the 2D *J*-resolved spectroscopic data (14,15).

This work involves the study of 2D J-resolved single-voxel and multivoxel spectroscopic acquisition methods targeted for clinical application of PCa detection. We first explore the use of a 2D *J*-resolved single-voxel spectroscopic sequence to illustrate the detection of the citrate metabolite and to observe the characteristics of the *J*-coupled spectral pattern. Multivoxel 2D *J*-resolved acquisitions are performed using spiral-based MRSI.

METHODS

Simulations and phantom measurements were conducted assuming a 2D *J*-resolved acquisition using a PRESS excitation scheme. For the simulations, we solved the full density matrix of strongly coupled two-spin systems with nonselective 180° pulses. The timing of the PRESS sequence was assumed to be 90°–[t_{int}]–180°–[$t_1/2$]–180°–[$t_1/2 - t_{int}$]–t_2 (acquire), where t_{int} was 10 ms. The *J*-coupling constant was assumed to be 15.4 Hz with a chemical shift value of 0.12 ppm (= 16.6 Hz at 3 T) (16). Also, the T_2 was assumed to be 200 ms with a line width of 10 Hz. For the actual single-voxel 2D *J*-resolved measurements, a phantom composed of citrate, creatine, and choline metabolites was used to emulate the existence of cancerous tissue. In both cases, the echo time spacing was adjusted to be 7.8 ms for a total of 64 steps from 35 to 534 ms in the F1 domain (Fourier domain corresponding to t_1 dimension). This resulted in a 2-Hz spectral resolution with a bandwidth of 128 Hz in the F1 domain. The spectral bandwidth in the F2 domain (Fourier domain corresponding to t_2 dimension) was 5000 Hz with 2048 data point acquisitions.

In vivo data were collected from a patient suspicious for recurrent PCa using the single-voxel 2D *J*-resolved technique. Single-voxel 2D *J*-resolved spectroscopic data using the PRESS sequence were acquired from two different regions near the peripheral zone of the prostate. All PRESS sequences were preceded by CHESS water suppression and very selective saturation pulses for spatial saturation (17). The voxel size chosen was 1 × 1.12 × 1.08 cm = 1.2 cc. Four acquisitions were averaged per incremental echo time for a total scan time of 8 min (TR = 2 s) for each voxel.

Multivoxel 2D *J*-resolved data using spiral MRSI were also acquired. A detailed description of the spiral MRSI sequence can be found in Ref. 14. First, phantom data were collected to demonstrate the feasibility of obtaining spatial as well as the coupling information followed by in vivo data collection. Spiral readout gradients were applied to a PRESS excitation sequence. The spirals were designed using the formula given by Glover (18). A 32 × 32 spatial matrix covering a 24-cm field of view was used. Sixteen spatial interleaves of the spirals were used to acquire the required *k*-space. The number of spirals repeated during each acquisition was 256 lobes, which resulted in a 950-Hz spectral bandwidth in the F2 domain. Sixteen different echo times starting from 35 to 285 ms with 15.6-ms intervals were used to collect F1 domain information, which accounted for F1 spectral resolution of 4 Hz and a bandwidth of 64 Hz. Collected data were processed by first gridding in the k_x, k_y, and t_2 domains followed by a four-dimensional FT into the *x*, *y*, F1, and F2 domains. In summary, each of the 32 × 32 reconstructed voxels had 16 spectra, which covers the F1 dimension from *J*(−32) Hz to *J*(28) Hz. The nominal voxel size was 0.59 cc while the total scan time to acquire the necessary *k*-space and the t_1 space data was 17 min (TR = 2 s).

For the multivoxel 2D *J*-resolved experiment, phantom data were acquired using the head coil while in vivo data were collected with the body coil for excitation followed by a rigid endorectal coil for signal reception. A phantom comprised of creatine, choline, and citrate solution was built for this study, which was surrounded by lipids to confirm the lipid suppression capabilities. For the in vivo exams, a high-resolution (512 × 512) T_2-weighted anatomic image was obtained and a region of interest covering most of the prostate was selected for the PRESS sequence. To date, seven patients who were suspicious of prostate cancer were referred to by a pathologist and imaged using

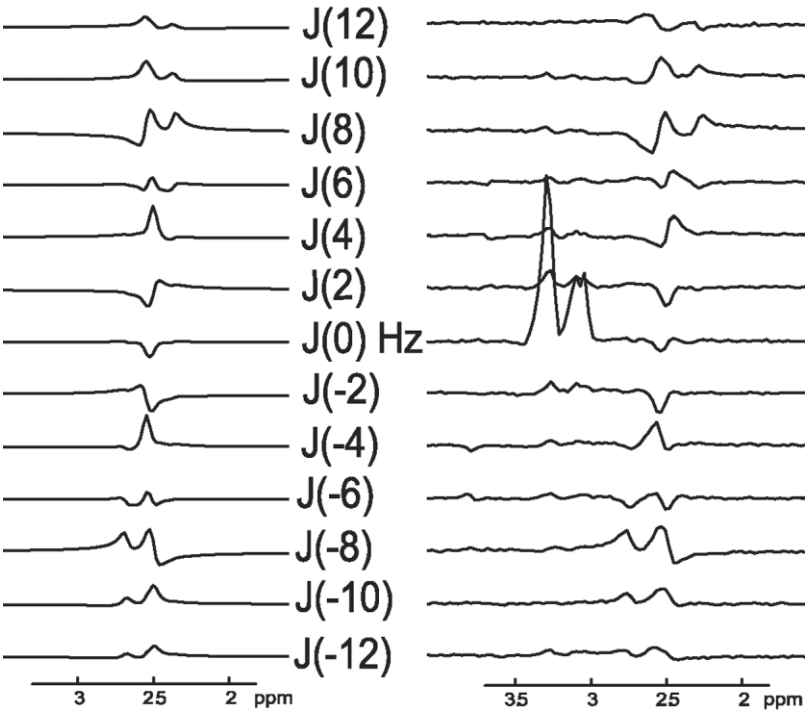


FIG. 1. 2D *J*-resolved spectra obtained from simulations (left) and phantom measurements (right) at 3 T. The echo time interval was 7.8 ms starting from 35 ms for 64 steps. Choline and creatine metabolites were added for the phantom measurement. Due to the modulations as a function of echo time, the 2D *J*-resolved spectra show a strong signal from the citrate metabolite at reconstructed lines beyond the *J*(0) Hz line for both simulated and measured data with similar spectral patterns. In this respect, the detection of the citrate resonance can be made outside of the *J*(0) Hz line using the 2D *J*-resolved acquisition.

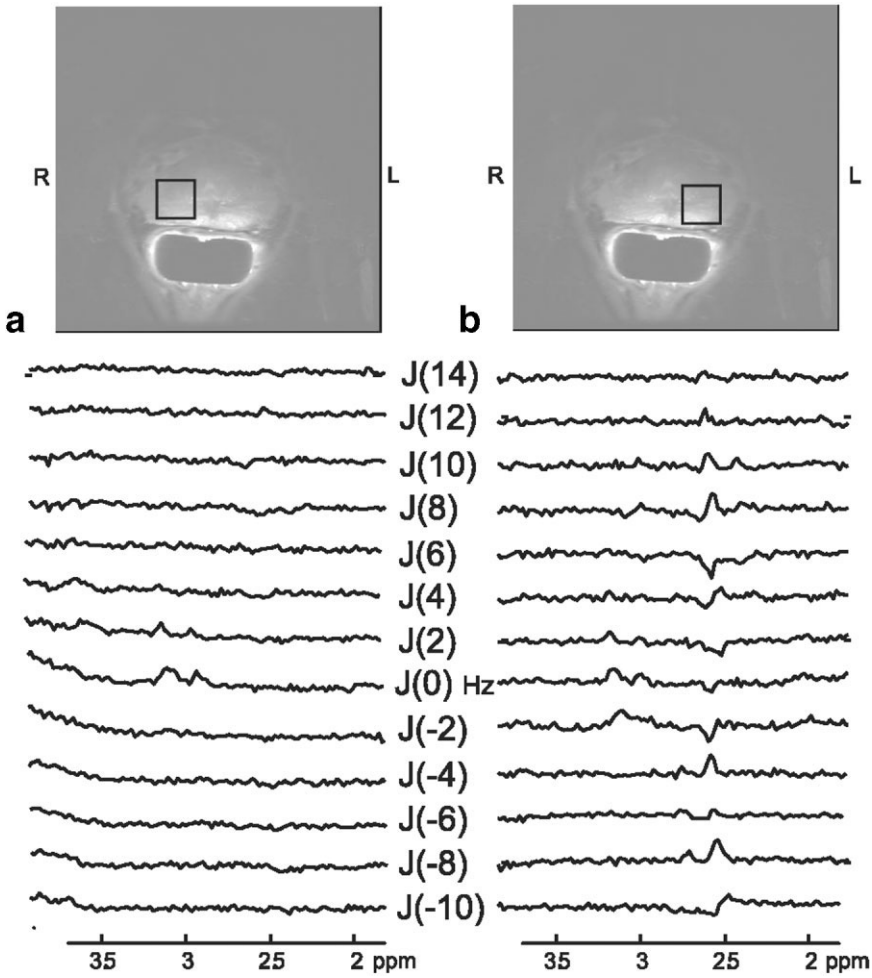


FIG. 2. Single-voxel 2D *J*-resolved spectroscopy results obtained in vivo from a subject suspicious of recurrent PCa. Two voxels were selected for the examination as shown in the T_2-weighted images. The reconstructed spectra of several F1 lines are shown. In (**a**), even though the presence of creatine and choline metabolites is evident, there is no visible citrate. As for the region shown in (**b**), the citrate is visible (2.6 ppm region from $J(-10)$ to $J(12)$ Hz line) in the spectra while other metabolites are also present. This shows that the 2D *J*-resolved spectroscopy can be useful for in vivo detection of citrate.

a 3 T GE Signa scanner (GE Health Care, Waukesha, WI). All in vivo studies were conducted under IRB guidelines and with informed consent.

RESULTS

Figure 1 shows the simulated 2D *J*-resolved citrate spectra (left) along with the reconstructed 2D *J*-resolved data acquired with a phantom (right) using the single-voxel 2D *J*-resolved acquisition. Spectra corresponding to the F1 domain in the range of $J(-12)$ to $J(12)$ Hz were extracted where most of the energy is concentrated. In both cases, due to the modulations occurring as a function of echo time, resonances are clearly seen beyond the $J(0)$ line for the citrate metabolite. Individual spectra from each F1 line reveal the similarity of the patterns between the simulated and measured results of the citrate. The $J(0)$ line, also referred to as the TE-averaged line, has a slight negative peak at the citrate position, which is due to the strong negative peaks at echo times ranging from 60 to 120 ms.

Two single-voxel 2D *J*-resolved spectra from an in vivo subject are presented in Fig. 2. The patient had a prior history of prostatic adenocarcinoma, which was treated by external beam radiation. The two regions that were selected are shown in the anatomic T_2-weighted images along with the resulting *J*-resolved spectra. The spectra obtained from the right side of the subject (Fig. 2a) displays negligible citrate metabolite intensity compared to the creatine and choline resonances located near the 3.0-ppm region. In comparison, the spectra from the left side of the subject (Fig. 2b) reveal the presence of citrate as seen from the modulations occurring in the reconstructed F1 lines along with the creatine and choline metabolites. These two comparisons show that with the 2D *J*-resolved acquisition method, the strongly coupled citrate metabolite can be resolved while the presence of other metabolites can be established. Even though the number of radiofrequency (RF) phase cycling steps has been reduced to 4 in this case, strong residual signal from outside of the PRESS box is not observed.

In Fig. 3, results obtained from the multivoxel 2D *J*-resolved sequence via spiral MRSI are shown. In Fig. 3a, an image of the phantom that was used for the experiment is given. In Fig. 3b, the metabolite spectra corresponding to the voxel selected in Fig. 3a are given. We extracted the TE-averaged line from each reconstructed voxel and manually phased them. The TE-averaged spectra show the well-resolved spatial distribution of the metabolites with a slight negative peak of the citrate as in the case of the single-voxel experiment. Spatial saturation pulses eliminated most of the lipids, as can be seen from the absence of any sidebands arising from the lipids. In Fig. 3c, lines from

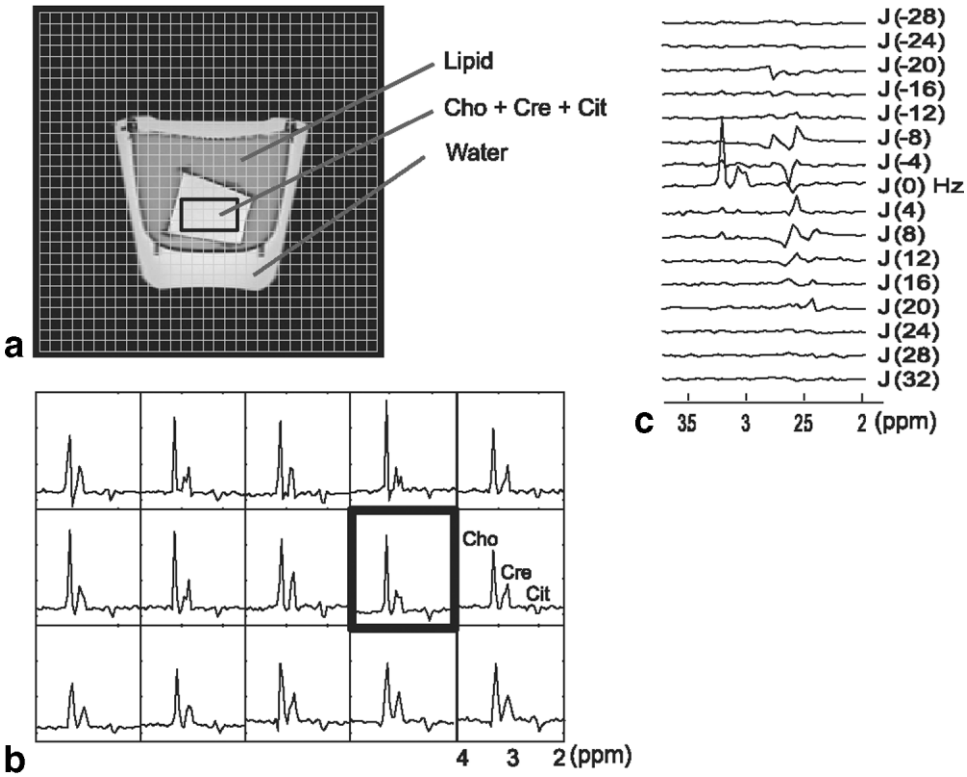


FIG. 3. Phantom results using spiral based multivoxel 2D *J*-resolved MRSI. A metabolite phantom surrounded by lipid and water was used (**a**). In (**b**), the metabolite spectra corresponding to *J*(0) Hz are displayed from the selected voxels shown in (a). The TE-averaged spectra show the well-resolved spatial distribution of the metabolites. In (**c**), all the reconstructed F1 lines for the voxel highlighted in (b) are shown. The F1 domain lines clearly depict the presence of citrate near the 2.5-ppm region.

J(−28) to *J*(32) Hz that were reconstructed for the voxel highlighted in Fig. 3b are shown. The *J*-resolved F1 domain lines clearly depict the presence of citrate located near the 2.5-ppm region.

Figure 4 shows spiral readout 2D *J*-resolved MRSI results obtained from an in vivo subject who had been reported as having a Gleason score of 3 + 3. The T_2-weighted image is shown in Fig. 4a with a grid representing the displayed voxels shown in Fig. 4b and c. Reconstructed spectra corresponding to the TE-averaged lines are given in Fig. 4b. As with the case of the phantom experiment, the TE-averaged line largely represents spectra from metabolites that are uncoupled. This is illustrated by the existence of choline and creatine metabolites that can be seen near the middle region of the displayed voxels. In Fig. 4c, the spectra corresponding to *J*(8) Hz are displayed where the citrate metabolite can be resolved. This is illustrated near the upper left region of the prostate where several voxels show a peak near the 2.5 ppm region, which corresponds to the citrate metabolite. For several voxels, however, lipid contamination can be visible near the 2.5-ppm region, which compromised the detection of citrate.

Figure 5 shows spiral readout 2D *J*-resolved MRSI results obtained from another in vivo subject who had been reported as having adenocarcinoma of the prostate with a Gleason score of 3 + 4. The TE-average lines in Fig. 5a show voxels of signal contributing from creatine and choline as well as polyamines residing in *J*(0) Hz. In Fig. 5c, the spectra corresponding to *J*(8) Hz line are displayed. The spectra show clear visualization of the citrate metabolite. In addition, polyamines are seen to be resolved as well. Of the seven patients examed, five patients had observable signal from any of the metabolites of interest.

DISCUSSION

We have shown the application of a 2D *J*-resolved PRESS sequence, which can aid the detection of PCa at the field strength of 3 T. In the first part of the paper, simulations and phantom measurements showing the characteristics of the citrate metabolite resonance for 2D *J*-resolved single-voxel acquisitions were illustrated, which showed the strong dependency with echo time due to its strong coupling. In the second part spiral readout gradients were applied to the 2D *J*-resolved PRESS sequence to obtain additional spatial distribution information. The efficiency of the spiral *k*-space trajectory makes it possible to cover the whole *k*-space within a reasonable scan time.

For a truly feasible clinical protocol to be implemented, several prerequisites need to be established, however. It is important for a good spatial suppression pulse to be used so that there are no aliasing or ringing artifacts. This problem has been demonstrated from the in vivo exam where lipid sidebands interfered with the detection of metabolites. This result is in comparison with the phantom results where good lipid suppression was accomplished. For the phantom experiment, a head coil with a relatively homogeneous RF profile and low power was used, whereas this was not the case for the in vivo exam, which used the body coil for excitation. Another important prerequisite is the main field homogeneity. The presence of air inside the endorectal coil or near the prostate region can degrade the homogeneity, leading to line broadening and potential overlap of the choline and creatine metabolites as seen in the in vivo example.

In this study, we addressed the issue of strong coupling citrate peak using a 2D *J*-resolved spectroscopic acquisi-

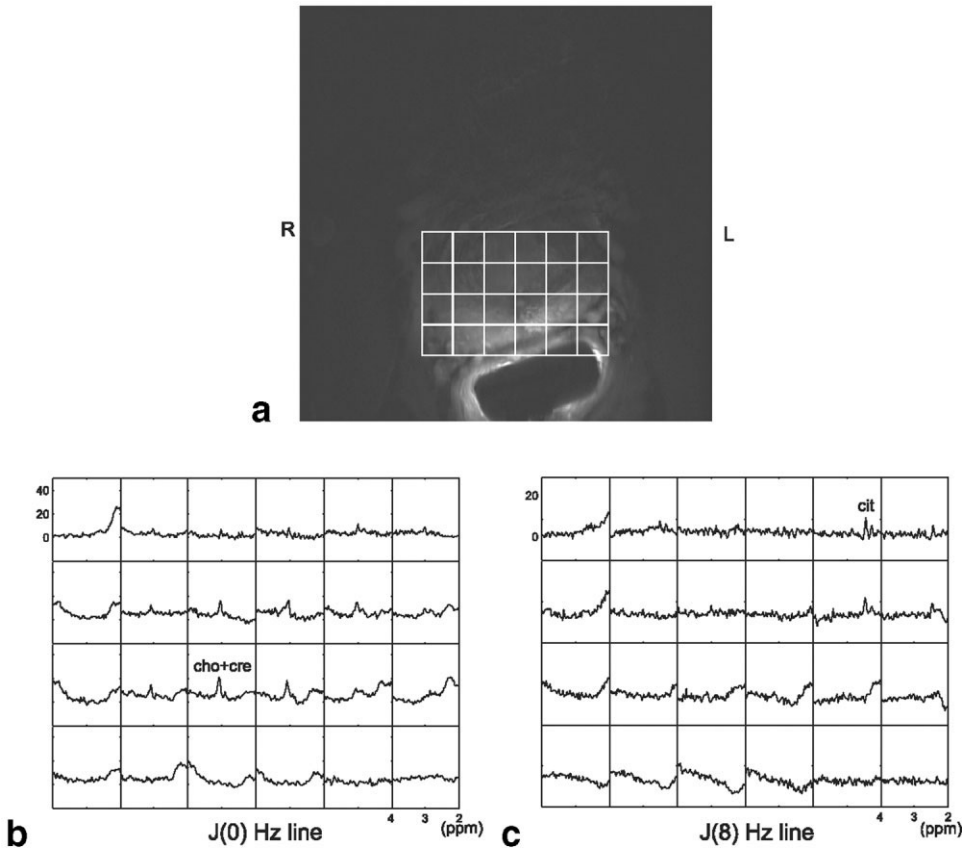


FIG. 4. Results obtained from a patient diagnosed with prostate cancer with a Gleason score 3 + 3 using spiral-based 2D *J*-resolved MRSI. (**a**) T_2-weighted image with a grid representing the displayed voxels. (**b**) Reconstructed spectra corresponding to the *J*(0) Hz lines. The existence of choline and creatine metabolites can be seen near the middle region of the displayed voxels. (**c**) Spectra corresponding to *J*(8) Hz line from F1 domain. Several voxels show a peak near the 2.5-ppm region at the upper left region, which corresponds to citrate. Lipid contamination can be visible near the 2.5-ppm region for several voxels.

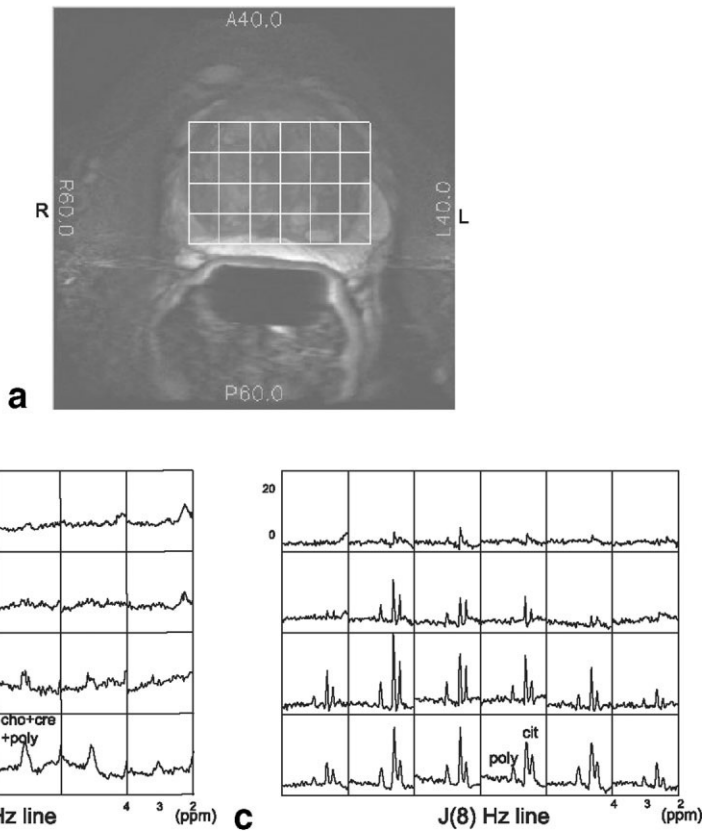


FIG. 5. Results obtained from a patient diagnosed with adenocarinoma of the prostate with a Gleason score 3 + 4 using spiral-based 2D *J*-resolved MRSI. (**a**) T_2-weighted image with a grid representing the displayed voxels. (**b**) Reconstructed spectra corresponding to the *J*(0) Hz lines. The coexistence of choline, creatine, and polyamine metabolites can be seen. (**c**) Spectra corresponding to the *J*(8) Hz line. Several voxels show a clear peak near the 2.5-ppm region corresponding to citrate. Polyamines can also be observed for several voxels around the 3.1-ppm region.

tion sequence. This can be dealt with in a different way as recently shown using a *J*-refocused sequence (19). One of the advantages of using the 2D *J*-resolved method includes the potential to detect changes in the citrate coupling constant, which can be used as another marker for PCa. This coupling constant is believed to be related to the zinc concentration, which is directly related to presence of PCa (20). In addition, any information obtained from the second spectral dimension, for example, from the polyamines as seen, can add to the physiologic information of the prostate tissue (9). On the negative side, a *J*-refocused scheme would require many fewer acquisitions and can achieve better SNR due to a shortened TE.

CONCLUSION

Single-voxel and multivoxel 2D *J*-resolved spectroscopy methods have been demonstrated for in vivo prostate at field strength of 3 T. Using the 2D *J*-resolved method, strong coupling of citrate can be well resolved. For multivoxel 2D *J*-resolved spectroscopic imaging, spiral-based readout sequences are used, which enable data acquisition within a reasonable scan time.

REFERENCES

1. Scheidler J, Hricak H, Vigneron DB, Yu KK, Sokolov DL, Huang RL, Zaloudek CJ, Nelson SJ, Carroll PR, Kurhanewicz J. 3D ^{1}H-MR spectroscopic imaging in localizing prostate cancer: clinico-pathologic study. Radiology 1999;213:473–480.
2. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional ^{1}H MR spectroscopic imaging of the in situ human prostate with high (0.24–0.7 cm^3) spatial resolution. Radiology 1996; 198:795–805.
3. Vigneron DB, Chen A, Cunningham C, Xu D, Hurd R, Sailasuta N, Pauly J, Nelson S, Kurhanewicz J. High resolution 3D MR spectroscopic imaging and J-resolved MRS of the prostate at 3 Tesla. In: Proceedings of the 12th Annual Meeting of ISMRM, Kyoto, Japan, 2004. p 386.
4. Wilman AH, Allen PS. The response of the strongly coupled AB system of citrate to typical ^{1}H MRS localization sequences. J Magn Reson B 1995;107:25–33.
5. Mulkern RV, Bowers JL, Peled S, Kraft RA, Williamson DS. Citrate signal enchancement with a homonuclear J-refocusing modification to double-echo PRESS sequences. Magn Reson Med 1996;36:775–780.
6. Aue WP, Karhan J, Ernst RR. Homonuclear broadband decoupling and two-dimensional J-resolved NMR spectroscopy. J Chem Phys 1976;64: 4226.
7. Thomas MA, Ryner L, Mehta M, Turski P, Sorenson JA. Localized 2D J-resolved ^{1}H MR spectroscopy of human brain tumors in vivo. J Magn Reson Imaging 1996;6:453–459.
8. Dreher W, Leibfritz D. On the use of two-dimensional-J NMR measurements for in vivo proton MRS: Measurements of homonuclear decoupled spectra without the need for short echo time. Magn Reson Med 1995;34:331–337.
9. Swanson MG, Vigneron DB, Tran TC, Sailasuta N, Hurd RE, Kurhanewicz J. Single-voxel oversampled J-resolved spectroscopy of in vivo human prostate tissue. Magn Reson Med 2001;45:973–980.
10. Yue K, Marumoto A, Binesh N, Thomas MA. 2D JPRESS of human prostates using an endorectal receiver coil. Magn Reson Med 2002;47: 1059–1064.
11. Hurd RE, Gurr D, Sailasuta N. Proton spectroscopy without water suppression: the oversampled J-resolved experiment. Magn Reson Med 1998;40:343–347.
12. Bolan PJ, DelaBarre L, Baker EH, Merkle H, Everson LI, Yee D, Garwood M. Eliminating spurious lipid sidebands in ^{1}H MRS of breast lesions. Magn Reson Med 2002;48:215–222.
13. Adalsteinsson E, Irarrazzabal P, Topp S, Meyer C, Macovski A, Spielman DM. Volumetric spectroscopic imaging with spiral-based k-space trajectories. Magn Reson Med 1998;39:889–898.
14. Adalsteinsson E, Spielman DM. Spatially resolved two-dimensional spectroscopy. Magn Reson Med 1999;41:8–12.
15. Hiba B, Serduc R, Provent P, Farion R, Remy C, Ziegler A. 2D J-resolved spiral spectroscopic imaging at 7 T: application to mobile lipid mapping in a rat glioma. Magn Reson Med 2004;52:658–662.
16. Schick F, Bongers H, Kurz S, Jung W, Pfeffer M, Lutz O. Localized proton MR spectroscopy of citrate in vitro and of the human prostate in vivo at 1.5T. Magn Reson Med 1993;29:38–43.
17. Tran TC, Vigneron D, Sailasuta N, Tropp J, LeRoux P, Kurhanewicz J, Nelson SJ, Hurd R. Very selective suppression pulses for clinical MRSI studies of brain and prostate cancer. Magn Reson Med 2000;43:23–33.
18. Glover GH. Simple analytic spiral k-space algorithm. Magn Reson Med 1999;42:412–415.
19. Cunningham CH, Marjanska M, Chen AP, Xu D, Pauly JM, Sailasuta N, Hurd RE, Kurhanewicz J, Garwood M, Vigneron DB. Sequence design incorporating the LASER technique for prostate MRSI at high field. In: Proceedings of the 12th Annual Meeting of ISMRM, Toronto, Canada, 2003. p 682.
20. Van der Graaf M, Heerschap A. Effect of cation binding on the protonchemical shifts and the spin-spin coupling constant of citrate. J Magn Reson B 1996;112:58–62.

Towards biologically conformal radiation therapy (BCRT): Selective IMRT dose escalation under the guidance of spatial biology distribution

Yong Yang and Lei Xing[a)]
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305-5847

(Received 28 September 2004; revised 17 March 2005; published 11 May 2005)

It is well known that the spatial biology distribution (e.g., clonogen density, radiosensitivity, tumor proliferation rate, functional importance) in most tumors and sensitive structures is heterogeneous. Recent progress in biological imaging is making the mapping of this distribution increasingly possible. The purpose of this work is to establish a theoretical framework to quantitatively incorporate the spatial biology data into intensity modulated radiation therapy (IMRT) inverse planning. In order to implement this, we first derive a general formula for determining the desired dose to each tumor voxel for a known biology distribution of the tumor based on a linear-quadratic model. The desired target dose distribution is then used as the prescription for inverse planning. An objective function with the voxel-dependent prescription is constructed with incorporation of the nonuniform dose prescription. The functional unit density distribution in a sensitive structure is also considered phenomenologically when constructing the objective function. Two cases with different hypothetical biology distributions are used to illustrate the new inverse planning formalism. For comparison, treatments with a few uniform dose prescriptions and a simultaneous integrated boost are also planned. The biological indices, tumor control probability (TCP) and normal tissue complication probability (NTCP), are calculated for both types of plans and the superiority of the proposed technique over the conventional dose escalation scheme is demonstrated. Our calculations revealed that it is technically feasible to produce deliberately nonuniform dose distributions with consideration of biological information. Compared with the conventional dose escalation schemes, the new technique is capable of generating biologically conformal IMRT plans that significantly improve the TCP while reducing or keeping the NTCPs at their current levels. Biologically conformal radiation therapy (BCRT) incorporates patient-specific biological information and provides an outstanding opportunity for us to truly individualize radiation treatment. The proposed formalism lays a technical foundation for BCRT and allows us to maximally exploit the technical capacity of IMRT to more intelligently escalate the radiation dose. © *2005 American Association of Physicists in Medicine*. [DOI: 10.1118/1.1924312]

Key words: inverse planning, biological model, TCP, NTCP, IMRT

I. INTRODUCTION

Intensity modulated radiation therapy (IMRT) has been used clinically to provide a highly conformal radiation dose to the target volume while reducing the doses to the surrounding sensitive structures.[1–13] The current IMRT inverse planning is typically aimed at producing a homogeneous target dose under the assumption of uniform biology within the target volume. In reality, it is known that the spatial biology distributions in most tumors and normal tissues are rarely homogeneous. To maximize the efficacy of IMRT, it is desirable to take the inhomogeneous biological information into account and to produce customized nonuniform dose distributions on a patient specific basis. This type of radiation treatment is referred to as biologically conformal radiotherapy (BCRT).[14–19] The simultaneous integrated boost (SIB) to elective volumes recently appearing in the literature[11,17,20] represents a simple example of BCRT. However, an underlying deficiency of the current SIB approach is that the boost doses are based on previous experience, not patient-specific biological information characterizing the spatial tumor burden distribution.

To establish the BCRT treatment planning scheme, three major aspects must be addressed: (i) Determination of the distribution of biological properties of the tumor and critical structures; (ii) Prescription of the desired dose distribution for inverse planning; and (iii) Inverse planning to generate most faithfully the prescribed nonuniform dose distribution. Recently spurred efforts in biological imaging, such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance spectroscopy imaging (MRSI), are aimed at providing solutions to the first problem.[21–31] To give a few examples, the clonogen density in malignant glioma can be obtained based on the choline/creatine ratio through MRSI,[29,30] tumor hypoxia can be quantified using PET imaging with fluorinated misonidazole (FMISO),[27,28] tumor proliferation rate can be obtained based on the voxel activity level in DNA proliferation imaging (e.g., fluoro-L-thymidine PET),[25,26,32] and lung functional importance distributions can be obtained by perfusion imaging.[33] While the development of molecular imaging techniques is critically important in mapping out biology distributions, the successful integration of this information

into IMRT planning through steps (ii) and (iii) is also indispensable to fully exploit the obtained biology information to improve patient care. In this study we focus our efforts on the last two problems, with the optimistic assumption that spatial biology distributions within a patient have already been determined from biological imaging or other means. Our goal is to establish a theoretical framework for quantitatively incorporating the biological data into IMRT inverse optimization, and to show the advantage of the selective dose escalation scheme in enhancing tumor control probability (TCP) and reducing the normal tissue complication probability (NTCP). In conjunction with the rapid development of molecular imaging techniques, this study lays a technical foundation for BCRT and provides a basis for clinically realizing the new treatment strategy in the future.

II. METHODS AND MATERIALS

A. Biological characterization and nonuniform target dose prescription

We assume that biological properties influencing radiation treatment are characterized phenomenologically by three radiobiology parameters: clonogen density (ρ), radiosensitivity (α), and proliferation rate (γ). Generally, these parameters are voxel dependent. In this work we concentrate on their spatial variation within tumor, and ignore the time dependence of the last two parameters.

To accomplish BCRT, an important step is to derive the desired dose distribution that maximizes the cell killing based on (ρ, α, γ) metrics. In the case of uniform biology, it is well known that the target dose should be uniformly distributed. It is, however, not clear at all what form of dose distribution should be used to maximize the cell killing for an arbitrary biology distribution. We start from a linear quadratic (LQ) model[34–36] with inclusion of the tumor cell proliferation. According to this model, the tumor clonogen survival S_i in a voxel of volume V_i after an irradiating dose D_i is given by

$$S_i = \rho_i V_i \exp(-\alpha_i' D_i + \gamma_i \Delta T), \tag{1}$$

where $\alpha_i' = \alpha_i[1 + d_i/(\alpha_i/\beta_i)]$, ρ_i is the initial clonogen density, d_i is the fractional dose, α_i and β_i are the linear-quadratic coefficients of the cell survival curve, $\gamma_i = \ln 2/T_p$ is the cell proliferation rate, T_p is the potential cell doubling time, and ΔT is the overall treatment time. The TCP of a voxel i can be expressed as

$$\mathrm{TCP}_i = \exp[-\rho_i V_i \exp(-\alpha_i' D_i + \gamma_i \Delta T)]. \tag{2}$$

The TCP for the whole tumor is the product of the TCP_i of all voxels within the tumor volume, i.e.,

$$\mathrm{TCP} = \prod_i \mathrm{TCP}_i. \tag{3}$$

For a given set of $\{\rho, \alpha, \gamma\}$, the task is to find the dose distribution that maximizes the TCP. Because of the limitation of normal tissue dose tolerances, an arbitrarily high dose to the tumor cannot be achieved and certain constraints need to be imposed.[36–41] In line with previous researchers,[36,37,40] we restrict the integral dose to the tumor volume to a constant. Mathematically, the constraint is written as

$$\sum_i m_i D_i = E_t, \tag{4}$$

where m_i is the mass of voxel i, and E_t is the integral target dose.

With the above formulation, the task becomes the maximization of the TCP under the constraint (4). The Lagrange multiplier method is employed to solve the problem. In this approach, a function

$$L(\mathrm{TCP}_1, \ldots \mathrm{TCP}_i, \ldots) = \prod_i \mathrm{TCP}_i + \lambda\left(\sum_i m_i D_i - E_t\right), \tag{5}$$

is introduced, where λ is the Lagrange multiplier, and the solution is obtained by solving the equations

$$\frac{\partial L}{\partial \mathrm{TCP}_i} = 0. \tag{6}$$

When mass and volume are equal for all voxels in the target, using a process similar to Ebert and Hoban[40] (see the Appendix), we obtained a general formula for determining the desired dose, $D_0^T(i)$, at the voxel i

$$D_0^T(i) = \frac{\alpha_{\mathrm{ref}}'}{\alpha_i'} D_{\mathrm{ref}} - \frac{1}{\alpha_i'}(\gamma_{\mathrm{ref}} - \gamma_i)\Delta T - \frac{1}{\alpha_i'} \ln\left(\frac{\alpha_{\mathrm{ref}}' \rho_{\mathrm{ref}}}{\alpha_i' \rho_i}\right), \tag{7}$$

where D_{ref} is the reference dose for the voxel with reference radiobiological parameters $(\rho_{\mathrm{ref}}, \alpha_{\mathrm{ref}}, \gamma_{\mathrm{ref}})$. In general, D_{ref} should be set to a value that yields a clinically sensible TCP at the reference voxel. For a given disease site, the radiation dose used in current clinical practice with "intent to cure" can be used as a good starting point in selecting the value of D_{ref}. Using Eq. (7), it is straightforward to determine the desired target prescription dose once the radiobiological parameter (ρ, α, γ) metrics and D_{ref} are known. Note that the desired dose distribution represents an ideal situation without considering the specific dosimetric tolerances of the sensitive structures. In reality, this dose distribution may or may not be exactly realizable. Nevertheless, it sets a landmark and serves as the prescription dose in inverse planning to guide the dose optimization process.

The fractional dose, d_i, is required to obtain the parameter α_i' in Eq. (7). On the other hand, d_i is not known until $D_0^T(i)$ is known. We use a simple iterative method to solve the dilemma. First, the fractional dose is initially set to $d_i = D_{\mathrm{ref}}/N_f$, N_f being the fractional number. Second, $D_0^T(i)$ is calculated using Eq. (7) and $d_i = D_0^T(i)/N_f$ is updated. The new $D_0^T(i)$ is then obtained using the updated d_i. We find that $D_0^T(i)$ converges to the solution in less than five iterations. In this study we set $\alpha/\beta = 10$ Gy for all target voxels. The formalism proposed here is, however, general and can be extended to deal with nonuniform distributions of the α/β ratio.

B. Inverse planning with spatially nonuniform dose prescription

The next logical step after obtaining the calculated prescription dose is to use inverse planning to derive the optimal beam profiles that will produce the prescribed dose distribution. To proceed, we construct an objective function to take the known biological information into account. In addition to the voxel-specific prescription as determined by Eq. (7), the nonlinear dose responses of tumor and normal structures are considered using the concept of equivalent volume[42–48] of a voxel, which is defined as

$$(\Delta V_{\text{eff}})_i = V_i \phi(i)(D(i)/D_t)^{1/n}, \tag{8}$$

where $(\Delta V_{\text{eff}})_i$ is the effective volume for voxel i with volume V_i and dose $D(i)$, D_t is the desired dose for a target voxel or the $\text{TD}_{5/5}$ of the corresponding organ, and $\phi(i)$ is the functional unit density. The value of n characterizes the dose-volume effect of an organ and reflects its architecture (serial or parallel) of the sensitive structure. It is obtained by fitting to clinical dose-volume data. For a sensitive structure, n is a positive number $(n>0)$ while for a target, n should be assigned with a small negative value $(-1<n<0)$. $\phi(i)\equiv 1$ for a target voxel.

A general form of the inverse planning objective function in the voxel domain is written as

$$F = \sum_{\tau=1}^{t_\tau} r_\tau \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \{1 + [D_c(i)/D_0^T(i)]^{1/n_\tau}\}[D_c(i) - D_0^T(i)]^2 + \sum_{\sigma=1}^{s_\sigma} r_\sigma \frac{1}{N_\sigma} \sum_{i=1}^{N_\sigma} \{1 + \phi_\sigma(i)[D_c(i)/TD_{\sigma,5/5}]^{1/n_\sigma}\} D_c(i)^2, \tag{9}$$

where r_τ and r_σ are the structure specific importance factors of target τ and sensitive structure σ, respectively, t_τ and s_σ the number of targets and sensitive structures, N_τ and N_σ the total number of voxels of target τ or sensitive structure σ, n_τ and n_σ the n parameter of target τ and sensitive structure σ, $D_c(i)$ the calculated dose in voxel i, $D_0^T(i)$ the prescription dose in a target voxel i given by Eq. (7), and $\text{TD}_{\sigma,5/5}$ the $\text{TD}_{5/5}$ of sensitive structure σ. The objective function becomes the conventional quadratic objective function if the term in the bracket inside each summation is set to unity (this is true when the dose-volume effect is negligible, i.e., when $n_\tau = n_s = +\infty$). More detailed information about the optimization algorithm can be found in Ref. 49.

C. Implementation

A software module for optimizing the objective function (9) is implemented in the platform of the PLUNC treatment planning system (University of North Carolina, Chapel Hill, NC). The dose calculation engine and a variety of image/beam/plan display and evaluation tools of PLUNC are used to review and compare the optimization results. The ray-by-ray iterative algorithm (SIITP) reported earlier[50,51] is employed to obtain the optimal beam intensity profiles. The dose-volume histograms (DVHs) of the involved organs are displayed at the end of each iterative step to visually monitor the optimization process.

TABLE I. Dose-volume parameters of various sensitive structures used for calculating NTCP in this study.

Sensitive structures	n	m	$D_{50/5}$ (Gy)
Bladder	0.50	0.11	80
Rectum	0.12	0.15	80
Femoral head	0.25	0.12	65

D. Plan review tools

It is desirable to extend the currently used plan review tools to deal with a biologically heterogeneous system. For a target, we define the effective dose at a voxel as the physical dose normalized by the desired dose determined by Eq. (7). The effective-dose volume histogram (EDVH), which is obtained by replacing the dose with the effective dose in conventional DVH, is a useful tool for assessing BCRT plans. For a sensitive structure we replace the fractional volume by $\phi_i V_i$ to construct a functional dose volume histogram (FDVH), similar to that proposed by Lu *et al.*[52] and Marks *et al.*[33] After including the heterogeneous biological information into the EDVH or FDVH, the wisdom used in interpreting a conventional DVH can be applied to assess the BCRT plans. In addition to the effective dose and the EDVH or FDVH, a cluster of DVHs, each corresponding to a given set of biological parameters $\{\rho, \alpha, \gamma\}$, is also useful to assess dosimetric behavior of the system as a function of the biological status of the system.

Besides the dosimetric evaluation tools, we also used the TCP and NTCPs for plan evaluation. In calculating TCP and NTCP, the heterogeneous biology distributions need to be taken into account. TCP is calculated using Eqs. (2) and (3) and NTCP is assessed using Lyman's model. The Kutcher–Burman effective-volume DVH reduction method[44] is extended to include the nonuniform functional unit density distribution using Eq. (8) when transforming a nonuniform dose distribution into a uniform irradiation of an effective partial volume. Model parameters from Burman *et al.*[53] are listed in Table I for the NTCP calculation.

E. Case studies

A prostate case with two different hypothetical distributions of radiobiological parameters is used to test the proposed BCRT inverse planning scheme. In each study, the target consists of the prostate gland with a few intraprostatic lesions. The sensitive structures include the rectum, bladder, and femoral heads. Figures 1(a)and 3(a) show the geometric shapes and locations of the structures in the two examples.

In the first example the target includes four biologically different regions, and the functional unit density distributions in the sensitive structures are uniform. Region 1 represents the basis reference target volume with typical parameters[54,55] $\rho_{0i} = 5\times 10^5$ clonogen/cm^3, $\alpha_i = 0.26$ Gy^{-1}, and $\gamma_i = \ln 2/40$ day^{-1}. The radiobiological parameters of the intra-

TABLE II. Radiological parameters for the target regions in the two examples.

Targets		ρ_{0i} (clonogen/cm^3)	α_i (Gy^{-1})	γ_i (day^{-1})
Example 1	Region 2	5×10^8	0.26	ln 2/40
	Region 3	5×10^5	0.13	ln 2/40
	Region 4	5×10^5	0.26	ln 2/10
Example 2	Region 2	5×10^6	0.20	ln 2/10
	Region 3	5×10^3	0.10	ln 2/60

prostatic lesions are listed in Table II. The parameters n_σ characterizing the dose-volume effect of the sensitive structures in the objective function (9) can be found in Table I. The parameter n_τ is chosen to be -0.2. For comparison, five IMRT plans, indexed by plan 1, -2, -3, -4, and -5, are generated. Plan 1 is obtained using the BCRT optimization scheme described above with $D_{\text{ref}}=70$ Gy. Plan 2 is obtained by prescribing the whole target a uniform dose of 70 Gy. Plan 3 and -4 are similar to plan 2 except that the dose is escalated to 81 and 91 Gy,[12,14] respectively. Plan 5 is the SIB IMRT plan with the same prescribed doses as that of the BCRT. In plan 1 to -4, the objective function expressed in Eq. (9) is used and in plan 5 the conventional dose-based quadratic objective function is adopted. The optimization parameters (maximum dose constraints and importance factors) in the dose-based method were adjusted by trial and error to obtain the "optimal" plan. The same beam configuration (five equally spaced 15 MV photon beams with gantry angles of 0°, 72°, 144°, 216°, and 288° in IEC convention) is used in generating the five plans.

In the second example we hypothetically introduced a higher functional unit density region in the rectum [R_region 2 as shown in Fig. 3(a)] in addition to three biologically different target regions. The functional unit density of the R_region 1 is assigned a value of 1 and that of the R_region 2 is set to be 4. The same set $(\rho_{0i}, \alpha_i, \gamma_i)$ as the previous example and a reference dose of 70 Gy are assigned to the prostate gland. The parameters for other target regions are listed in Table II. Once again, five IMRT plans are generated: Plan 1 is obtained using the proposed selective dose escalation scheme, plan 2, -3, and -4 are generated using different uniform prescription doses (70, 81, and 91 Gy) and plan 5 is SIB plan with the same prescription as plan 1 but is optimized using the conventional quadratic objective function. In generating these five plans, seven equally spaced 15 MV photon beams (0°, 51°, 103°, 154°, 206°, 257°, and 309°) are employed.

III. RESULTS

A. Example 1: Prostate case with four biologically different regions

In the first example, based on Eq. (7) and the parameters listed in Table II, the prescription doses to the target region 2, 3, and 4 are determined to be 85, 119, and 75 Gy, respectively. In order to examine the capability of the BCRT inverse planning system in producing an extremely nonuniform dose within a target volume, we have used an "extreme" combination of $\{\rho, \alpha, \gamma\}$, which leads to an exceedingly high prescription dose in region 3 (119 Gy). Figures 1(b)–1(d) show the isodose distributions of plan 1 in a transverse slice and two sagittal slices. The EDVH of the target and the DVHs of the sensitive structures are plotted in Fig. 2 for plan 1 in solid curves. For comparison, the corresponding EDVHs and DVHs of plan 2, -3, -4, and -5 are also shown in the figures as dashed, dotted, dash-dotted, and dash-dot-dotted curves, respectively. As seen from Fig. 1, all regions in the prostate are well covered by their prescription doses and the sensitive structures are well spared. Even in this extreme case, it seems that the inverse planning system can satisfy the biological requirement. A steep dose gradient is found at the interface between the target and the rectum. A comparison of the target EDVH in Fig. 2(a) indicates that above 98.5% of the target voxels achieved their desired doses in plan 1 and plan 5. However, for the uniform dose escalation scheme, the desired doses in some regions (region 2, 3 and part of region 4 in plan 2; region 2 and 3 in plan 3; and region 3 in plan 4) are not achieved. We found that, in plan 1, the doses to the surrounding sensitive structures are not significantly increased compared with those of plan 2, despite the fact that some voxels in region 4 receive a dose as high as 119 Gy. In plan 1, the rectum, bladder, and femoral heads are better spared in comparison with plan 3 and -4. However, by comparing the DVHs of plan 1 and -5, it is noticed that, although the target coverage in plan 5 is similar to that in plan 1, the sensitive structures in plan 5 receive much higher doses than plan 1, indicating that the proposed approach can improve the sensitive structure sparing compared with the conventional dose-based quadratic objective function. In addition, as can be expected, the target doses in plan 1 and -5 are less uniform in the target volume in comparison with that of plan 2, -3, and -4. This is more pronounced in the target region 1, where about 50% of the volume receives a dose larger than 85 Gy as shown in Fig. 2(b), resulting in an effective dose above 120% in ~50% of the target voxels [see Fig. 2(a)]. However, the increase of dose inhomogeneity is desirable here provided that the NTCPs are not compromised.

Table III lists the calculated TCPs for the targets and NTCPs for the sensitive structures with consideration of heterogeneous biology in all plans. It is seen that the overall TCPs for the three plans with uniform target dose prescriptions (plan 2, -3, and -4) are all less than that of the BCRT plan (plan 1) and SIB plan (plan 5). This is understandable because, in plan 2, -3 and -4, some target regions (such as target region 3) receive doses much less than the desired doses. For example, in plan 4, the TCP for target region 3 is only 0.461. Even if the TCPs for region 1, 2, and 4 are all close to 1.00, the resultant total TCP for plan 4 is 0.461. In contrast, the TCPs of plan 1 and plan 5 are 0.984 and 0.981, respectively. Furthermore, we found that the NTCPs of the sensitive structures in plan 1 are very close to plan 2, significantly less than plan 3, -4, and -5. For example, the rectum NTCPs are 0.21% for plan 1 and 0.20% for plan 2. These are

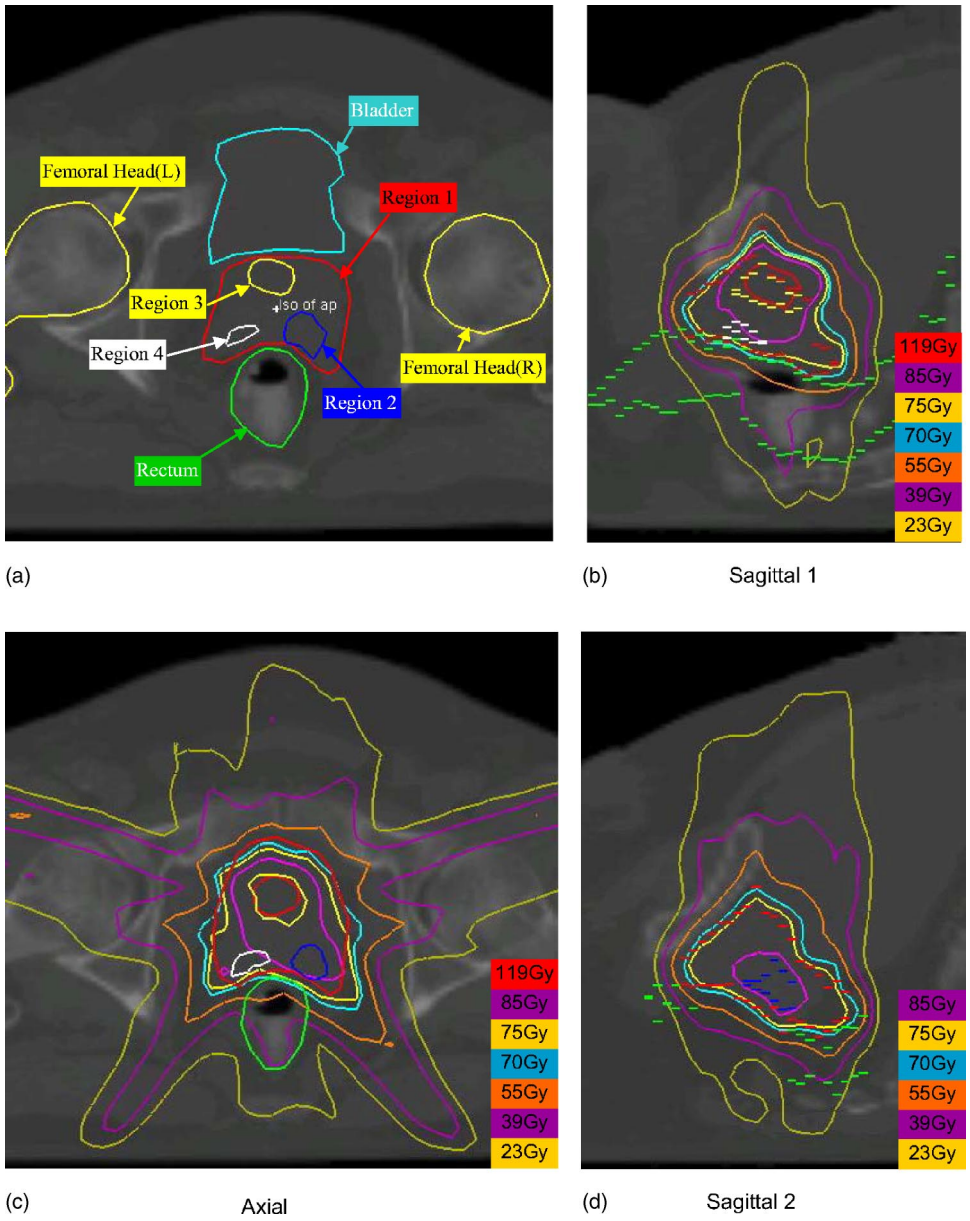


FIG. 1. A hypothetical prostate case with four biologically different regions (example 1). (a) Geometric shapes and locations of the targets and sensitive structures; (b)–(d): Isodose distributions in an axial slice and two sagittal slices for plan 1, generated by optimizing the objective function with a nonuniform dose prescription derived from Eq. (7).

increased to 0.65%, 1.84%, and 0.89% for Plan 3, -4, and -5, respectively. Again, although similar overall TCPs are achieved for the BCRT and dose-based SIB IMRT plans when the same dose prescriptions are used, the rectum NTCPs are significantly reduced using the proposed formulism. This is consistent with our previous study of the objective function in the context of conventional IMRT aiming to deliver a uniform dose to the target volume.[49]

B. Example 2: Prostate case with three biologically different regions and nonuniform importance in rectum

In the second hypothetical example, there are three biologically different regions in the prostate and two functionally different regions in the rectum. The prescription doses for the three target regions are 70 (reference dose), 99, and 121 Gy, as determined by Eq. (7) with the biological parameters listed in Table II. Figures 3(b)–3(f) show the isodose distributions of plan 1 in three transverse slices and two sagittal slices. The EDVHs and DVHs of the target and sensitive structures for plan 1 to plan 5 are plotted in Fig. 4 as solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves, respectively. Similar to the previous example, in plan 1, all regions in the prostate are well covered by a dose comparable to the prescription and the sensitive structures are well spared. The dose gradient at the interface between the target and the rectum is very sharp for all the plans. From Fig. 4(a) we find that above 98% of the target voxels achieved their desired doses in plan 1. As a consequence of incorporating functional unit density information in inverse planning, the rectum sparing is even better than that of plan 2, much better than that of plan 3, -4. However, we notice that the sparing of the femoral heads in plan 1 is not as good as that in plan 2, -3, and -4. This is because high-intensity beamlets that pass

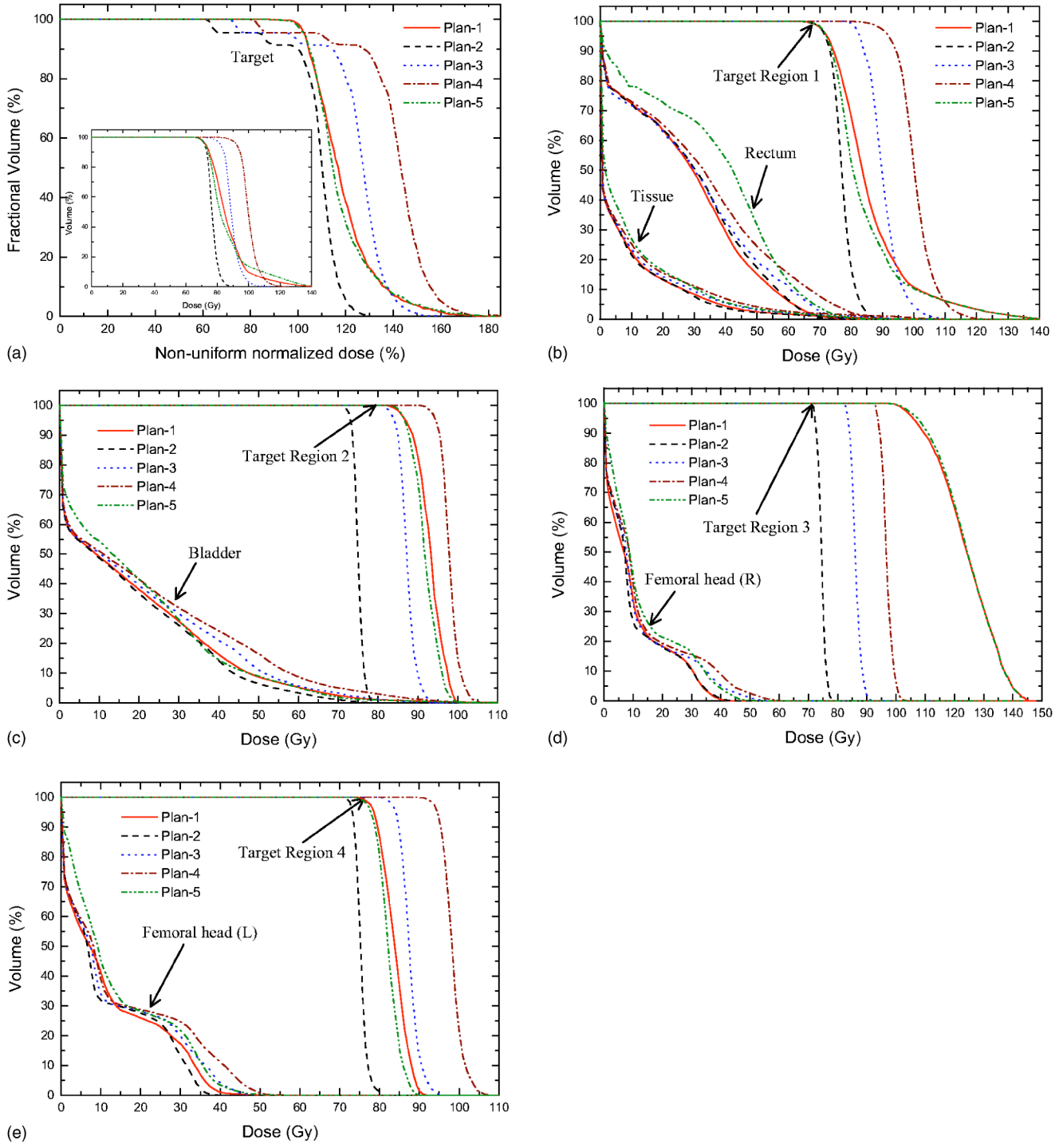


FIG. 2. Comparison of EDVHs and DVHs of the BCRT plan (plan 1), three uniform IMRT plans (plan 2: 70 Gy, plan 3: 81 Gy, and plan 4: 91 Gy) and the SIB plan (plan 5) in example 1. (a) Target EDVHs for the five plans (insert is the regular DVHs of the prostate target). The normalized doses to the target region 1, 2, 3 and 4 are 70, 85, 119, and 75 Gy, respectively; (b)-(e) DVHs of different target regions and sensitive structures for the five plans. The solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves represent the results of plan 1, 2, 3, 4, and 5, respectively. The effective dose is defined as the physical dose at a voxel normalized by its desired dose determined by Eq. (7).

through the femoral heads are needed to adequately irradiate the target region 3, as seen from Figs. 3(b) and 3(c). In addition, similar to the first example, the target coverage in plan 5 is close to that in plan 1, but the doses to the sensitive structures in plan 5 are much higher than that in plan 1.

Table IV lists the calculated TCPs and NTCPs for all plans. Once again, we found that the TCP of the target in the proposed BCRT technique is much higher and the NTCP of the rectum is lower compared with those obtained using the conventional uniform dose escalation schemes. Remarkably,

TABLE III. Comparison of TCP and NTCP for the five IMRT plans for example 1.

		Plan 1 (BCRT plan)	Plan 2 (70 Gy Uniform)	Plan 3 (81 Gy Uniform)	Plan 4 (91 Gy Uniform)	Plan 5 (SIB plan)
TCP	Region 1	0.997	0.995	1.000	1.000	0.994
	Region 2	0.998	0.642	0.995	1.000	0.999
	Region 3	0.989	0.000	0.002	0.461	0.989
	Region 4	1.000	0.997	1.000	1.000	0.998
	Overall	0.984	0.000	0.002	0.461	0.981
NTCP (%)	Rectum	0.212	0.196	0.652	1.84	0.885
	Bladder	1.6×10^{-5}	1.4×10^{-5}	2.3×10^{-5}	4.2×10^{-5}	3.6×10^{-5}
	Femoral head (R)	2.0×10^{-5}	2.1×10^{-6}	2.6×10^{-4}	1.75×10^{-4}	3.9×10^{-5}
	Femoral head (L)	1.2×10^{-5}	2.0×10^{-6}	7.0×10^{-4}	5.26×10^{-4}	6.9×10^{-5}

the overall TCP for the target is increased from 0.823 to 0.982 and the NTCP of the rectum is reduced from 3.1% to 0.40% when plan 4 is replaced by the selective dose escalation scheme (plan 1). Again, we found that, for similar overall TCPs, the rectum NTCPs of the BCRT plan are much lower in comparison with that obtained using dose-based SIB scheme.

IV. DISCUSSION

Equation (7) provides a general formula for determining the desired target dose distribution based on the known biology information of the system, and represents one of the main results of this study. A few special cases are worth discussing here. First, when the biology distribution is uniform in the target, a uniform dose of D_{ref} is desired. This is consistent with previous studies[37] and existing clinical knowledge.

When the clonogen density ρ is nonuniform while the values of α and γ are constant across the target, we have

$$D_0^T(i) = D_{\text{ref}} - \frac{1}{\alpha'} \ln\left(\frac{\rho_{0\text{ref}}}{\rho_{0i}}\right), \tag{10}$$

which is identical to the formula obtained by Webb and Nahum.[36] Equation (10) indicates that the desired dose depends on the tumor cell density logarithmically and is thus relatively insensitive to a variation in the clonogen density. For $\alpha_{\text{ref}}=0.312$, for example, even if the clonogen density in a tumor voxel is 10 times higher than that of the reference situation, the desired dose is only about 7 Gy higher than the reference value. A detailed discussion of this special situation has been presented by Webb and Nahum.[36]

Another special case is that the tumor clonogen density and the proliferation rate are constant and the radiosensitivity α is spatially nonuniform. Equation (7) now becomes

$$D_0^T(i) = \frac{\alpha'_{\text{ref}}}{\alpha'_i} D_{\text{ref}} - \frac{1}{\alpha'_i} \ln\left(\frac{\alpha'_{\text{ref}}}{\alpha'_i}\right). \tag{11}$$

The desired dose is approximately inversely proportional to the parameter α'_i and is thus sensitively dependent on the value of parameter α'_i. This is similar to the conclusions of Ebert and Hoban[40] and Levin-Plotnik and Hamilton.[41] For example, if α'_i is reduced from 0.312 (corresponding to $\alpha=0.26$, fractional dose $d_i=2.0$ Gy, and α/β ratio=10 Gy) to 0.18 (corresponding to $\alpha=0.15$, $d_i=2.0$ Gy, and $\alpha/\beta=10$ Gy), the desired dose is increased by about 70% (from 70 Gy to about 118 Gy).

If we keep the tumor clonogen density and radiosensitivity α' constant and only allow the proliferation rate γ to vary spatially, then

$$D_0^T(i) = D_{\text{ref}} + \frac{1}{\alpha'_i}(\gamma_i - \gamma_{\text{ref}})\Delta T. \tag{12}$$

Thus, the desired dose increases linearly with the proliferation rate. In this work the potential cell-doubling times, T_p, used by King *et al.*[54] are adopted. Since T_p for a prostate tumor is relatively longer, its influence on the desired dose is

TABLE IV. Comparison of TCP and NTCP for the five IMRT plans for example 2.

		Plan 1 (BCRT plan)	Plan 2 (70 Gy Uniform)	Plan 3 (81 Gy Uniform)	Plan 4 (91 Gy Uniform)	Plan 5 (SIB Plan)
TCP	Region 1	0.997	0.995	1.000	1.000	0.968
	Region 2	0.989	0.000	0.587	0.981	0.987
	Region 3	0.996	0.006	0.408	0.839	0.990
	Overall	0.981	0.000	0.239	0.823	0.946
NTCP (%)	Rectum	0.397	0.414	1.46	3.12	1.25
	Bladder	1.5×10^{-5}	1.2×10^{-5}	1.8×10^{-5}	4.3×10^{-5}	3.9×10^{-5}
	Femoral head (R)	3.7×10^{-5}	1.5×10^{-5}	1.8×10^{-5}	5.3×10^{-5}	2.3×10^{-5}
	Femoral head (L)	4.9×10^{-5}	1.1×10^{-5}	3.0×10^{-5}	4.5×10^{-5}	3.6×10^{-5}

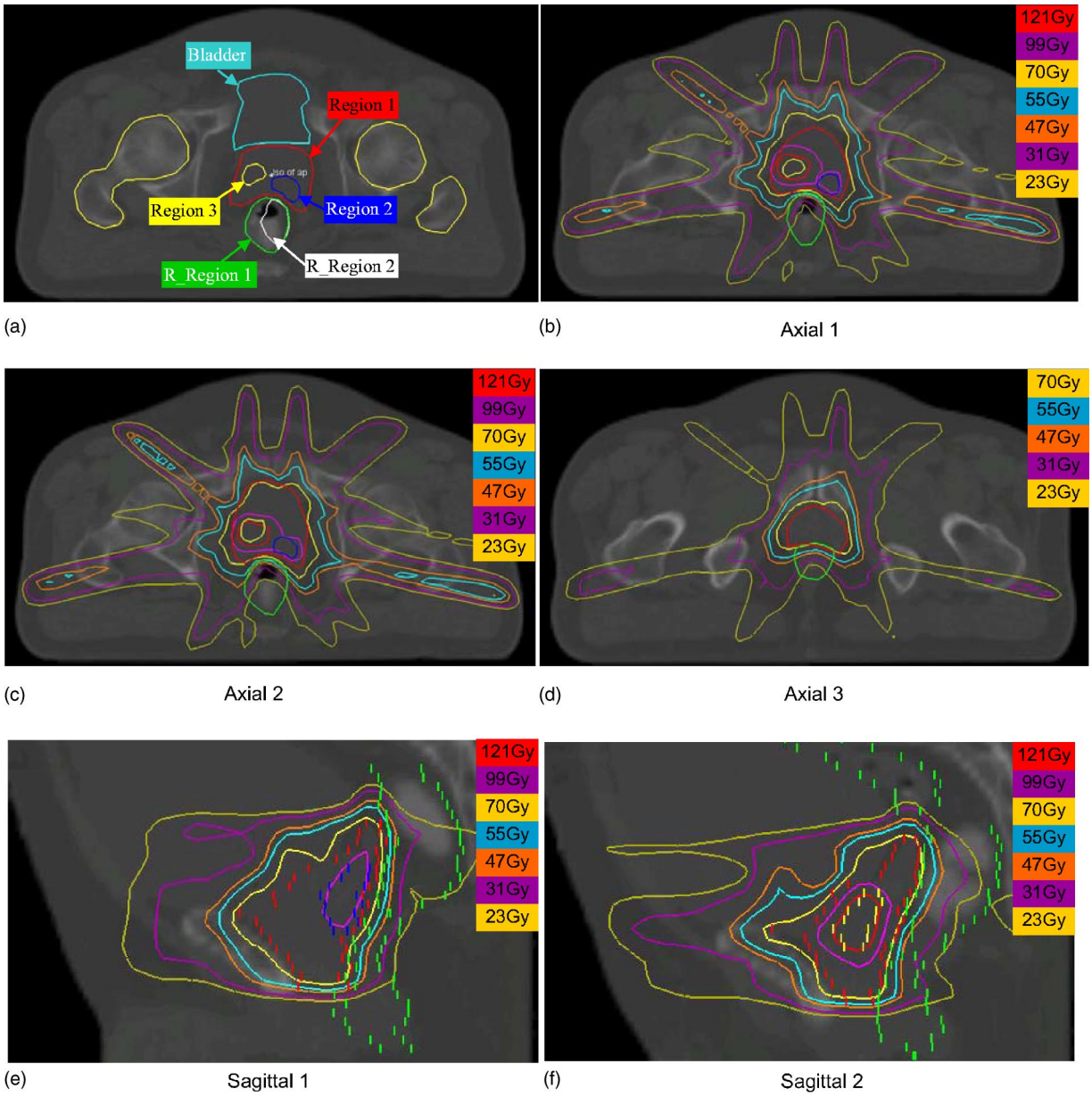


FIG. 3. A hypothetical prostate case with three biologically different regions and nonuniform importance in the rectum (example 2). (a) Geometric shapes and locations of the targets and sensitive structures; (b)–(d) Isodose distributions in three axial slices and two sagittal slices for plan 1, generated by optimizing the objective function with nonuniform dose prescription derived from Eq. (7).

not very significant. However, for other more rapidly proliferating tumors, the proliferation rate may play an important role. In such situations, reducing the overall treatment time ΔT (e.g., using an accelerated scheme) is helpful to minimize the influence of the proliferation rate.

We emphasize that the quadratic term in the linear-quadratic model plays an important role in accounting for the fractionation effect. If only the linear term is kept, the total dose $D_0^T(i)$ in Eq. (7) is no longer entangled with the fractional dose d_i. When the quadratic term is "switched on," the value of $D_0^T(i)$ depends not only on the total reference dose but also on the fractional dose. For a large fractional dose, the total dose will be less, and vice versa. In other words, the total dose received by a voxel is determined by two contributing factors, one being the local biological parameters $\{\rho, \alpha, \gamma\}$, and the other being the coupling between the fractional dose and the total dose. The latter is responsible for the phenomenon that the total dose needs to be decreased when the number of fractions is reduced. If the quadratic term were ignored, according to Eq. (7), the dose required at a voxel would be much higher. For example, the desired doses for target region 3 in example 1 are determined to be 119 and 135 Gy with and without inclusion of the quadratic term, respectively.

We also would like to emphasize that in this study, the radiosensitivity α' and proliferation rate γ are assumed to be constants during the whole treatment course. In reality, both α' and γ may change with time due to such biological pro-

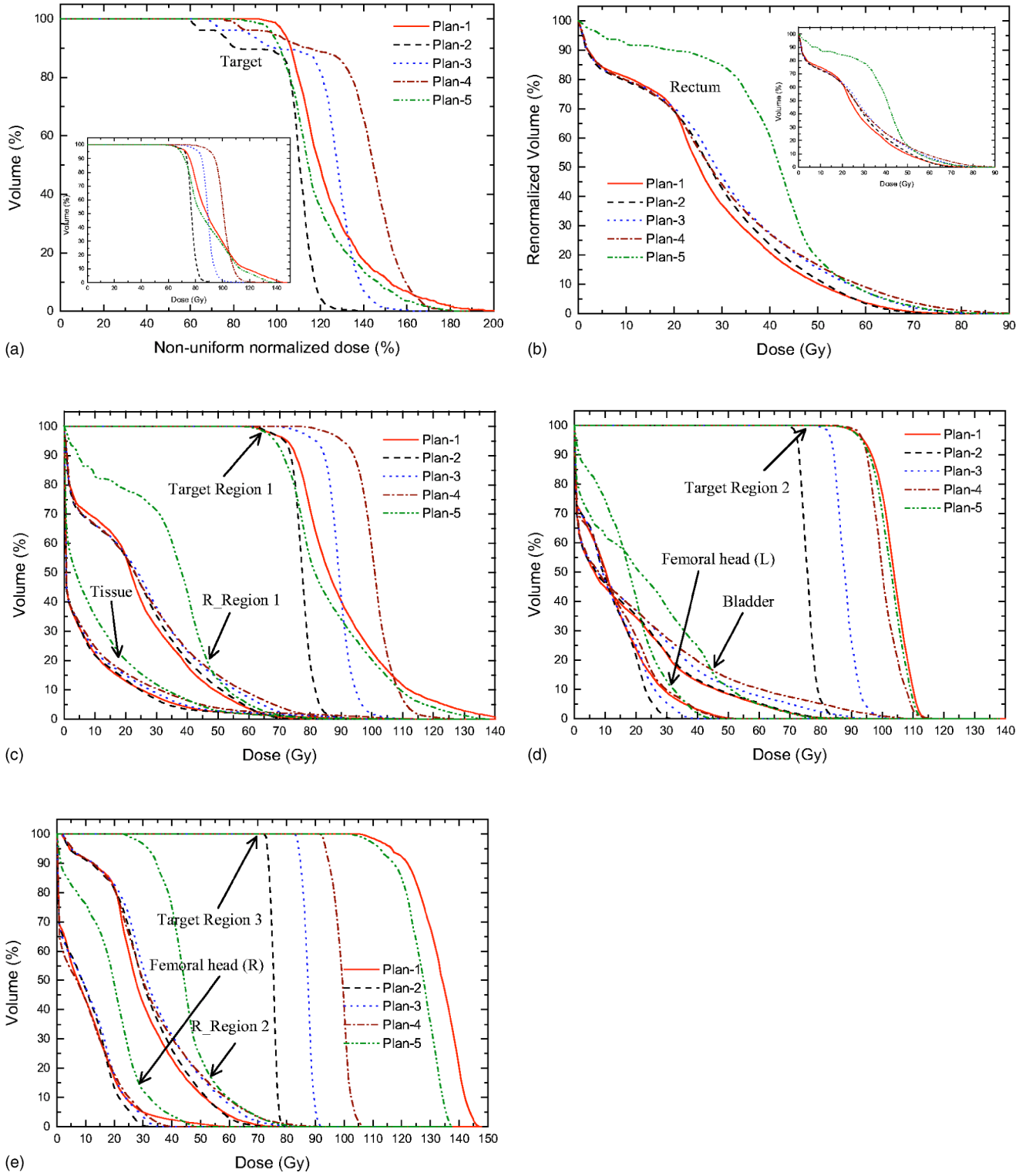


FIG. 4. Comparison of EDVHs, FDVHs, and DVHs of the BCRT plan (plan 1), three uniform IMRT plans (plan 2: 70 Gy, plan 3: 81 Gy, and plan 4: 91 Gy) and the SIB plan (plan 5) in example 2. (a) The target EDVHs for the five plans (insert is the regular DVHs of the prostate target). The normalized doses to the target region 1, 2, and 3 are 70, 99, and 121 Gy, respectively; (b) The rectum FDVHs for the five plans (insert is the regular DVHs of the rectum); (c)–(e): DVHs of different target regions and sensitive structures for the five plans. The solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves represent the results of plan 1, 2, 3, 4, and 5, respectively.

cesses as tumor cell redistribution[56] and reoxygenation.[57] The time dependence of these factors may result in a reduction of the desired prescription dose, and this effect should be investigated in the future.

Comparing with the uniform dose escalation scheme, our study clearly suggests that deliberately incorporating an inhomogeneous dose distribution significantly enhances the TCP and reduces the NTCP. Physically, we believe that the

significant improvement arises from the more effective use of radiation in the newly proposed treatment scheme. A great deal of dose is "wasted" in the conventional uniform dose escalation scheme. For example, in the first example the increased doses in the target region 1 and 4 have almost no contributions to the enhancement of the TCP when plan 2 (70 Gy uniform dose to the prostate gland) is replaced by plan 3 (81 Gy) or plan 4 (91 Gy). Even though part of the prostate receives high doses in the selective dose escalation scheme (for example, 119 Gy in target region 3 of the first example), the total deposited energy in the targets is still less than that of plan 3 or -4. It is thus not difficult to understand why deliberately nonuniform dose distributions can, in general, reduce the radiation side effects and represent a more intelligent way to irradiate the tumor target.

A similar deficiency also exists in the current SIB approach. Although it is clear that the regions with different tumor burdens should be given different doses, the specific values for the regions are determined in an *ad hoc* manner. The empirical boost dose could be too low, in which case the tumor control is sacrificed, or too high, in which case other parts of the system are compromised. The problem is aggravated when the tumor burden varies continuously from point to point. In the proposed BCRT approach, the prescribed dose is voxel dependent and is determined based on the tumor biology distribution. In addition, a more sophisticated objective function is developed to take the dose-volume effect and functional density information of the sensitive structures into account, resulting in better sparing of the sensitive structures.

Finally, it should be recognized that our knowledge of radiobiological parameters for tumors or normal tissues is still very crude and the validity of the model is still under establishment. Therefore, the LQ model and the parameters adopted in the paper are fine for a proof of principle but they should not be taken as more than that.

V. CONCLUSION

In the presence of nonuniform biology distributions, IMRT inverse planning is complicated by the fact that it is not clear what represents the appropriate spatial dose prescription, which is generally used as a landmark to guide the dose optimization process. In this work, we have described a technique for deriving the prescription dose based on an LQ model with consideration of the cell proliferation. The relation is quite general and can be used as prescription dose to guide an arbitrary inverse planning objective function aimed to produce customized dose distribution in accordance with the spatial biology information. For a given patient, IMRT inverse planning now consists of two steps: Derivation of the prescription dose, and beam profile optimization that produces as closely as possible this prescription dose. The formalism proposed here lays a technical foundation for future BCRT development, allowing us to escalate tumor dose more intelligently and effectively. When combined with state-of-the-art biological imaging techniques, which promise to reveal detailed patient-specific biology distribution information, this study may have significant implication for the management of cancer in the future.

ACKNOWLEDGMENTS

The authors wish to thank Dr. B. Thorndyke for carefully reading the manuscript. We would also like to thank Dr. S. Hancock, Dr. A. Koong, Dr. C. King, Dr. Q. T. Le, Dr. A. Boyer, and Dr. D. Spielman for useful discussion. This research is supported in part by the National Cancer Institute (5 R01 CA98523-01) and Department of Defense (DAMD17-03-1-0019 and 0023).

APPENDIX: DERIVATION OF EQUATION (7)

We present the detailed derivation process for Eq. (7) under the condition of equal mass and volume for all target voxels.

Substituting Eq. (5) into Eq. (6), we obtain

$$\lambda \mathrm{TCP}_i \frac{\partial(m_i D_i)}{\partial \mathrm{TCP}_i} = \lambda \mathrm{TCP}_{\mathrm{ref}} \frac{\partial(m_{\mathrm{ref}} D_{\mathrm{ref}})}{\partial \mathrm{TCP}_{\mathrm{ref}}} = -\mathrm{TCP}. \tag{A1}$$

Since $\lambda \neq 0$ otherwise, TCP becomes zero according to the requirement of Eq. (6), which corresponds to a minimum. If we assumed that mass for all target voxels is equal, then Eq. (A1) becomes

$$\mathrm{TCP}_i \frac{\partial D_i}{\partial \mathrm{TCP}_i} = \mathrm{TCP}_{\mathrm{ref}} \frac{\partial D_{\mathrm{ref}}}{\partial \mathrm{TCP}_{\mathrm{ref}}}. \tag{A2}$$

From Eq. (2) we have

$$D_i = \frac{-1}{\alpha_i'} \ln\left\{ -\gamma_i \Delta T + \ln\left[\frac{-1}{\rho_i V_i} \ln \mathrm{TCP}_i \right] \right\}. \tag{A3}$$

Substituting the expressions from Eq. (A3) for both D_i and D_{ref} into Eq. (A2), we have

$$\alpha_i' \ln(\mathrm{TCP}_i) = \alpha_{\mathrm{ref}}' \ln(\mathrm{TCP}_{\mathrm{ref}}). \tag{A4}$$

The desired doses, $D_0^T(i)$, producing maximum TCP with the constraint of constant integral dose, can be obtained by substituting TCP_i and $\mathrm{TCP}_{\mathrm{ref}}$ expressed in Eq. (2) into Eq. (A4)

$$D_0^T(i) = \frac{\alpha_{\mathrm{ref}}'}{\alpha_i'} D_{\mathrm{ref}} - \frac{1}{\alpha_i'}(\gamma_{\mathrm{ref}} - \gamma_i)\Delta T - \frac{1}{\alpha_i'} \ln\left(\frac{\alpha_{\mathrm{ref}}' \rho_{\mathrm{ref}} V_{\mathrm{ref}}}{\alpha_i' \rho_i V_i} \right). \tag{A5}$$

When volume for all target voxels is equal, Eq. (A5) becomes Eq. (7).

[a]Author to whom correspondence should be addressed. Electronic mail: lei@reyes.stanford.edu

[1]S. Webb, *Intensity-Modulated Radiation Therapy* (Institute of Physics, Bristol, 2001), Vol. XV.

[2]C. C. Ling, C. Burman, C. S. Chui, G. J. Kutcher, S. A. Leibel, T. LoSasso, R. Mohan, T. Bortfeld, L. Reinstein, S. Spirou, X. H. Wang, Q. Wu, M. Zelefsky, and Z. Fuks, "Conformal radiation treatment of prostate cancer using inversely-planned intensity-modulated photon beams produced with dynamic multileaf collimation," Int. J. Radiat. Oncol., Biol., Phys. **35**, 721–730 (1996).

[3]S. Y. Woo, W. H. Grant 3rd, D. Bellezza, R. Grossman, P. Gildenberg, L.

S. Carpentar, M. Carol, and E. B. Butler, "A comparison of intensity modulated conformal therapy with a conventional external beam stereotactic radiosurgery system for the treatment of single and multiple intracranial lesions," Int. J. Radiat. Oncol., Biol., Phys. **35**, 593–597 (1996).
[4]N. Lee, P. Xia, J. M. Quivey, K. Sultanem, I. Poon, C. Akazawa, P. Akazawa, V. Weinberg, and K. K. Fu, "Intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: An update of the UCSF experience," Int. J. Radiat. Oncol., Biol., Phys. **53**, 12–22 (2002).
[5]Q. T. Le, L. Xing, and A. L. Boyer, "Head and Neck IMRT," in *3D Conformal Radiation Therapy & Intensity Modulated Radiation Therapy in the Next Millennium*, edited by J. G. Purdy and J. Palta (Medical Physics Publishing, Madison, WI, 1999), pp. 190–195.
[6]E. Huang, B. S. Teh, D. R. Strother, Q. G. Davis, J. K. Chiu, H. H. Lu, L. S. Carpenter, W. Y. Mai, M. M. Chintagumpala, M. South, W. H. Grant 3rd, E. B. Butler, and S. Y. Woo, "Intensity-modulated radiation therapy for pediatric medulloblastoma: Early report on the reduction of ototoxicity," Int. J. Radiat. Oncol., Biol., Phys. **52**, 599–605 (2002).
[7]J. G. Rosenman, J. S. Halle, M. A. Socinski, K. Deschesne, D. T. Moore, H. Johnson, R. Fraser, and D. E. Morris, "High-dose conformal radiotherapy for treatment of stage IIIA/IIIB non-small-cell lung cancer: Technical issues and results of a phase I/II trial," Int. J. Radiat. Oncol., Biol., Phys. **54**, 348–356 (2002).
[8]A. J. Mundt, A. E. Lujan, J. Rotmensch, S. E. Waggoner, S. D. Yamada, G. Fleming, and J. C. Roeske, "Intensity-modulated whole pelvic radiotherapy in women with gynecologic malignancies," Int. J. Radiat. Oncol., Biol., Phys. **52**, 1330–1337 (2002).
[9]M. J. Krasin, B. T. Crawford, Y. Zhu, E. S. Evans, M. R. Sontag, L. E. Kun, and T. E. Merchant, "Intensity-modulated radiation therapy for children with intraocular retinoblastoma: Potential sparing of the bony orbit," Clin. Oncol. (R Coll. Radiol) **16**, 215–222 (2004).
[10]J. Meyer, M. H. Phillips, P. S. Cho, I. Kalet, and J. N. Doctor, "Application of influence diagrams to prostate intensity-modulated radiation therapy plan selection," Phys. Med. Biol. **49**, 1637–1653 (2004).
[11]A. Eisbruch, "Intensity-modulated radiotherapy of head-and-neck cancer: Encouraging early results," Int. J. Radiat. Oncol., Biol., Phys. **53**, 1–3 (2002).
[12]M. J. Zelefsky, Z. Fuks, M. Hunt, Y. Yamada, C. Marion, C. C. Ling, H. Amols, E. S. Venkatraman, and S. A. Leibel, "High-dose intensity modulated radiation therapy for prostate cancer: Early toxicity and biochemical outcome in 772 patients," Int. J. Radiat. Oncol., Biol., Phys. **53**, 1111–1116 (2002).
[13]S. M. Zhou, L. B. Marks, G. S. Tracton, G. S. Sibley, K. L. Light, P. D. Maguire, and M. S. Anscher, "A new three-dimensional dose distribution reduction scheme for tubular organs," Med. Phys. **27**, 1727–1731 (2000).
[14]C. C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J. Koutcher, "Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality," Int. J. Radiat. Oncol., Biol., Phys. **47**, 551–560 (2000).
[15]A. Brahme, "Individualizing cancer treatment: Biological optimization models in treatment planning and delivery," Int. J. Radiat. Oncol., Biol., Phys. **49**, 327–337 (2001).
[16]L. Xing, C. Cotrutz, S. Hunjan, A. L. Boyer, E. Adalsteinsson, and D. Spielman, "Inverse planning for functional image-guided intensity-modulated radiation therapy," Phys. Med. Biol. **47**, 3567–3578 (2002).
[17]K. S. Chao, S. Mutic, R. L. Gerber, C. A. Perez, and J. A. Purdy, "A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guided intensity-modulated radiation therapy," Int. J. Radiat. Oncol., Biol., Phys. **49**, 1171–1182 (2001).
[18]R. Macklis, M. Weinhous, and G. Harnisch, "Intensity-modulated radiotherapy: Rethinking basic treatment planning paradigms," Int. J. Radiat. Oncol., Biol., Phys. **48**, 317–318 (2000).
[19]M. Alber, F. Paulsen, S. M. Eschmann, and H. J. Machulla, "On biologically conformal boost dose optimization," Phys. Med. Biol. **48**, N31–35 (2003).
[20]Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, "Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose," Int. J. Radiat. Oncol., Biol., Phys. **52**, 224–235 (2002).
[21]S. R. Cherry, "*In vivo* molecular and genomic imaging: New challenges for imaging physics," Phys. Med. Biol. **49**, R13–48 (2004).
[22]J. Kurhanewicz, M. G. Swanson, S. J. Nelson, and D. B. Vigneron, "Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer," J. Magn. Reson Imaging **16**, 451–463 (2002).
[23]C. H. Contag and B. D. Ross, "It's not just about anatomy: *In vivo* bioluminescence imaging as an eyepiece into biology," J. Magn. Reson Imaging **16**, 378–387 (2002).
[24]S. J. DiBiase, K. Hosseinzadeh, R. P. Gullapalli, S. C. Jacobs, M. J. Naslund, G. N. Sklar, R. B. Alexander, and C. Yu, "Magnetic resonance spectroscopic imaging-guided brachytherapy for localized prostate cancer," Int. J. Radiat. Oncol., Biol., Phys. **52**, 429–438 (2002).
[25]J. G. Tjuvajev *et al.*, "Imaging of brain tumor proliferative activity with iodine-131-iododeoxyuridine," J. Nucl. Med. **35**, 1407–1417 (1994).
[26]A. F. Shields, J. R. Grierson, B. M. Dohmen, H. J. Machulla, J. C. Stayanoff, J. M. Lawhorn-Crews, J. E. Obradovich, O. Muzik, and T. J. Mangner, "Imaging proliferation *in vivo* with [F-18] FLT and positron emission tomography," Nat. Med. **4**, 1334–1336 (1998).
[27]J. S. Rasey, W. J. Koh, M. L. Evans, L. M. Peterson, T. K. Lewellen, M. M. Graham, and K. A. Krohn, "Quantifying regional hypoxia in human tumors with positron emission tomography of [18F]fluoromisonidazole: A pretherapy study of 37 patients," Int. J. Radiat. Oncol., Biol., Phys. **36**, 417–428 (1996).
[28]W. J. Koh *et al.*, "Evaluation of oxygenation status during fractionated radiotherapy in human nonsmall cell lung cancers using [F-18]fluoromisonidazole positron emission tomography," Int. J. Radiat. Oncol., Biol., Phys. **33**, 391–398 (1995).
[29]B. L. Miller, L. Chang, R. Booth, T. Ernst, M. Cornford, D. Nikas, D. McBride, and D. J. Jenden, "*In vivo* 1H MRS choline: Correlation with *in vitro* chemistry/histology," Life Sci. **58**, 1929–1935 (1996).
[30]R. K. Gupta, T. F. Cloughesy, U. Sinha, J. Garakian, J. Lazareff, G. Rubino, L. Rubino, D. P. Becker, H. V. Vinters, and J. R. Alger, "Relationships between holine magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma," J. Neuro-Oncol. **50**, 215–226 (2000).
[31]R. J. Hamilton, M. J. Blend, C. A. Pelizzari, B. D. Milliken, and S. Vijayakumar, "Using vascular structure for CT-SPECT registration in the pelvis," J. Nucl. Med. **40**, 347–351 (1999).
[32]A. F. Shields, D. A. Mankoff, J. M. Link, M. M. Graham, J. F. Eary, S. M. Kozawa, M. Zheng, B. Lewellen, T. K. Lewellen, J. R. Grierson, and K. A. Krohn, "Carbon-11-thymidine and FDG to measure therapy response," J. Nucl. Med. **39**, 1757–1762 (1998).
[33]L. B. Marks, G. W. Sherouse, M. T. Munley, G. C. Bentel, D. P. Spencer, and C. Scarfone, "Incorporation of functional status into dose-volume analysis," Med. Phys. **26**, 196–199 (1999).
[34]T. E. Schultheiss, G. K. Zagars, and L. J. Peters, "An explanatory hypothesis for early- and late-effect parameter values in the LQ model," Radiother. Oncol. **9**, 241–248 (1987).
[35]J. F. Fowler, "The linear-quadratic formula and progress in fractionated radiotherapy," Br. J. Radiol. **62**, 679–694 (1989).
[36]S. Webb and A. E. Nahum, "A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density," Phys. Med. Biol. **38**, 653–666 (1993).
[37]S. Webb, P. M. Evans, W. Swindell, and D. J. O., "A proof that uniform dose gives the greatest TCP for fixed integral dose in the planning target volume," Phys. Med. Biol. **39**, 2091–2098 (1994).
[38]A. Brahme and A. K. Agren, "Optimal dose distribution for eradication of heterogeneous tumours," Acta Oncol. **26**, 377–385 (1987).
[39]N. A. Stavreva, P. V. Stavrev, and W. H. Round, "A mathematical approach to optimizing the radiation dose distribution in heterogeneous tumours," Acta Oncol. **35**, 727–732 (1996).
[40]M. A. Ebert and P. W. Hoban, "Some characteristics of tumour control probability for heterogeneous tumours," Phys. Med. Biol. **41**, 2125–2133 (1996).
[41]D. Levin-Plotnik and R. J. Hamilton, "Optimization of tumour control probability for heterogeneous tumours in fractionated radiotherapy treatment protocols," Phys. Med. Biol. **49**, 407–424 (2004).
[42]T. E. Schultheiss, C. G. Orton, and R. A. Peck, "Models in radiotherapy: Volume effects," Med. Phys. **10**, 410–415 (1983).
[43]J. T. Lyman and A. B. Wolbarst, "Optimization of radiation therapy. III. A method of assessing complication probabilities from dose-volume histograms," Int. J. Radiat. Oncol., Biol., Phys. **13**, 103–109 (1987).
[44]G. J. Kutcher and C. Burman, "Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method," Int. J. Radiat. Oncol., Biol., Phys. **16**, 1623–1630 (1989).

[45]R. Mohan, G. S. Mageras, B. Baldwin, L. J. Brewster, G. J. Kutcher, S. Leibel, C. M. Burman, C. C. Ling, and Z. Fuks, "Clinically relevant optimization of 3-D conformal treatments," Med. Phys. **19**, 933–944 (1992).
[46]J. O. Deasy, K. S. Chao, and J. Markman, "Uncertainties in model-based outcome predictions for treatment planning," Int. J. Radiat. Oncol., Biol., Phys. **51**, 1389–1399 (2001).
[47]M. K. Martel, W. M. Sahijdak, R. K. Ten Haken, M. L. Kessler, and A. T. Turrisi, "Fraction size and dose parameters related to the incidence of pericardial effusions," Int. J. Radiat. Oncol., Biol., Phys. **40**, 155–161 (1998).
[48]L. Jones and P. Hoban, "A comparison of physically and radiobiologically based optimization for IMRT," Med. Phys. **29**, 1447–1455 (2002).
[49]Y. Yang and L. Xing, "Clinical knowledge-based inverse treatment planning," Phys. Med. Biol. **49**, 5101–5117 (2004).
[50]L. Xing and G. T. Y. Chen, "Iterative algorithms for inverse treatment planning," Phys. Med. Biol. **41**, 2107–2123 (1996).
[51]L. Xing, R. J. Hamilton, D. Spelbring, C. A. Pelizzari, G. T. Chen, and A. L. Boyer, "Fast iterative algorithms for three-dimensional inverse treatment planning," Med. Phys. **25**, 1845–1849 (1998).
[52]Y. Lu, D. R. Spelbring, and G. T. Chen, "Functional dose-volume histograms for functionally heterogeneous normal organs," Phys. Med. Biol. **42**, 345–356 (1997).
[53]C. Burman, G. J. Kutcher, B. Emami, and M. Goitein, "Fitting of normal tissue tolerance data to an analytic function," Int. J. Radiat. Oncol., Biol., Phys. **21**, 123–135 (1991).
[54]C. R. King, T. A. DiPetrillo, and D. E. Wazer, "Optimal radiotherapy for prostate cancer: Predictions for conventional external beam, IMRT, and brachytherapy from radiobiologic models," Int. J. Radiat. Oncol., Biol., Phys. **46**, 165–172 (2000).
[55]A. E. Nahum, B. Movsas, E. M. Horwitz, C. C. Stobbe, and J. D. Chapman, "Incorporating clinical measurements of hypoxia into tumor local control modeling of prostate cancer: Implications for the alpha/beta ratio," Int. J. Radiat. Oncol., Biol., Phys. **57**, 391–401 (2003).
[56]S. L. Scott, P. H. Gumerlock, L. Beckett, Y. Li, and Z. Goldberg, "Survival and cell cycle kinetics of human prostate cancer cell lines after single- and multifraction exposures to ionizing radiation," Int. J. Radiat. Oncol., Biol., Phys. **59**, 219–227 (2004).
[57]R. F. Kallman and M. J. Dorie, "Tumor oxygenation and reoxygenation during radiation therapy: Their importance in predicting tumor response," Int. J. Radiat. Oncol., Biol., Phys. **12**, 681–685 (1986).

Int. J. Radiation Oncology Biol. Phys., Vol. 62, No. 2, pp. 595–605, 2005
Copyright © 2005 Elsevier Inc.
Printed in the USA. All rights reserved
0360-3016/05/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2005.02.001

PHYSICS CONTRIBUTION

NARROW BAND DEFORMABLE REGISTRATION OF PROSTATE MAGNETIC RESONANCE IMAGING, MAGNETIC RESONANCE SPECTROSCOPIC IMAGING, AND COMPUTED TOMOGRAPHY STUDIES

EDUARD SCHREIBMANN, PH.D., AND LEI XING, PH.D.

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA

Purpose: Endorectal (ER) coil–based magnetic resonance imaging (MRI) and magnetic resonance spectroscopic imaging (MRSI) is often used to obtain anatomic and metabolic images of the prostate and to accurately identify and assess the intraprostatic lesions. Recent advancements in high-field (3 Tesla or above) MR techniques affords significantly enhanced signal-to-noise ratio and makes it possible to obtain high-quality MRI data. In reality, the use of rigid or inflatable endorectal probes deforms the shape of the prostate gland, and the images so obtained are not directly usable in radiation therapy planning. The purpose of this work is to apply a narrow band deformable registration model to faithfully map the acquired information from the ER-based MRI/MRSI onto treatment planning computed tomography (CT) images.

Methods and Materials: A narrow band registration, which is a hybrid method combining the advantages of pixel-based and distance-based registration techniques, was used to directly register ER-based MRI/MRSI with CT. The normalized correlation between the two input images for registration was used as the metric, and the calculation was restricted to those points contained in the narrow bands around the user-delineated structures. The narrow band method is inherently efficient because of the use of *a priori* information of the meaningful contour data. The registration was performed in two steps. First, the two input images were grossly aligned using a rigid registration. The detailed mapping was then modeled by free form deformations based on B-spline. The limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (*L-BFGS*), which is known for its superior performance in dealing with high-dimensionality problems, was implemented to optimize the metric function. The convergence behavior of the algorithm was studied by self-registering an MR image with 100 randomly initiated relative positions. To evaluate the performance of the algorithm, an MR image was intentionally distorted, and an attempt was then made to register the distorted image with the original one. The ability of the algorithm to recover the original image was assessed using a checkerboard graph. The mapping of ER-based MRI onto treatment planning CT images was carried out for two clinical cases, and the performance of the registration was evaluated.

Results: A narrow band deformable image registration algorithm has been implemented for direct registration of ER-based prostate MRI/MRSI and CT studies. The convergence of the algorithm was confirmed by starting the registration experiment from more than 100 different initial conditions. It was shown that the technique can restore an MR image from intentionally introduced deformations with an accuracy of approximately 2 mm. Application of the technique to two clinical prostate MRI/CT registrations indicated that it is capable of producing clinically sensible mapping. The whole registration procedure for a complete three-dimensional study (containing 256 × 256 × 64 voxels) took less than 15 min on a standard personal computer, and the convergence was usually achieved in fewer than 100 iterations.

Conclusions: A deformable image registration procedure suitable for mapping ER-based MRI data onto planning CT images was presented. Both hypothetical tests and patient studies have indicated that the registration is reliable and provides a valuable tool to integrate the ER-based MRI/MRSI information to guide prostate radiation therapy treatment. © 2005 Elsevier Inc.

Narrow band registration model, Endorectal coil, Prostate imaging, Treatment planning.

Reprint requests to: Lei Xing, Ph.D., Stanford University School of Medicine, Department of Radiation Oncology, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Tel: (650) 498-7896; Fax: (650) 498-4015; E-mail: lei@reyes.stanford.edu

This work is supported in part by a Research Grant from the Prostate Cancer Research Program of the U.S. Department of Defense (DAMD17-03-1-0023) and the National Cancer Institute (1 R01 CA98523-01).

Acknowledgments—The authors thank Drs. Steve L. Hancock, Christopher King, Arthur Boyer, Yong Yang, Cristian Cotrutz, Brian Thorndyke, Dong-Hyun Kim, Daniel Spielman, and Bruce Daniel for useful discussions.

Received Nov 12, 2004, and in revised form Feb 4, 2005. Accepted for publication Feb 4, 2005.

INTRODUCTION

Endorectal (ER) coil-based magnetic resonance imaging (MRI) and magnetic resonance spectroscopic imaging (MRSI) provide high-resolution anatomic and metabolic information and allow us to better assess the extent of prostate cancer and the involvement of seminal vesicles (1–12). The new imaging information can also be used to guide the dose optimization process to selectively boost the high tumor burden regions (13–16). To use the imaging data for treatment planning, an indispensable step is the registration of ER-based MRI/MRSI and the treatment planning CT images. Presently, manual and/or automated image fusion tools implemented in most commercial treatment planning systems are based on geometric translations and rotations of the images and are generally not suitable for handling the problem because the endorectal coil displaces and distorts the prostate and surrounding tissue (17). A deformable registration procedure must be applied to map the ER-based MRI/MRSI data onto the planning CT images acquired in a normal treatment position without the insertion of ER probe (18, 19).

There have been several relevant image registration techniques reported in the literature. Zaider *et al.* (13) and Court and Dong (20) used a rigid transformation for the correction of tissue displacement. A deformable procedure based on the finite element model (FEM), in which images are described as blocks of elastic materials on which forces apply, was proposed by Bharath *et al.* (21). In this approach, the parameters that control the behavior of the elastic material and are responsible for the conversion of forces into local deformations of the elastic material are Young's elastic modulus and Poisson's ratio. Although powerful, the model has the drawback that values of the elasticity and density constant for various tissues are not readily available and have to be found by a trial and error procedure. The method also relies on using complicated software to generate a FEM mesh and masks of the involved structures. Recently, a simpler technique based on spline interpolation was proposed (18). When the first and second derivatives of the cost function can be analytically deduced, this approach reduces the problem to solving a set of system equations (22). Wu *et al.* (23) used a free-form intramodality registration between the MRI images obtained with and without inflatable ER probe acquired at the end of the study. Their study suggested that the free-form registration is stable and accurate for dealing with the problem. However, the approach is valid only for intramodality registration and entails extra MRI scans of the pelvic region with body coil. In addition, a term was introduced in their cost function to constrain the optimization to smooth deformations, posing a new problem of how to objectively select the relative weightings of different terms in the cost function.

An improvement to this method can be achieved by using a spline model with the smoothness of the deformation field assured by the interpolation between a grid of fixed control points. In this setup, the cost function is composed of a single term and no weighting factors are required. A simple method along this line is to deduce the spline coefficients from a set of user-defined control points, as was done by Fei *et al.* (18) in a study of warping and rigid registration of MR volumes. This method was also applied to directly register prostate CT and MRI/MRSI and validated by using a series of phantom measurements (24, 25). Coselmon *et al.* (26) used a similar technique to study the accuracy of mutual-information–based CT registration of the lung at exhale and inhale respiratory states.

In general, image registration can be formulated as an optimization problem where the variables are a group of transformation parameters that lead to the best match between the input images. The match is quantified in mathematical terms by the use of a metric, which ranks a potential matching based on the image histograms, resolution, or pixel values of the involved organs. There is a vast literature on using image intensity information such as mean squares, normalized correlation, and mutual information (27) to construct metrics for guiding the registration process. The mutual information represents a popular choice when dealing with multimodality image registration. Briefly, mutual information is a criterion from information theory and is related to entropy, which is a measure of uncertainty, variability, and complexity. When each of the two images provides the most information about the other, the mutual information metric is maximal and the two images are considered to be coregistered. Methods using signed distance information from edge features have also been reported (28), in which the signed distance field of an object surface is used as the shape representation. The signed distance field is a continuous scalar function defined throughout the volume, and its value is simply the distance to the nearest point on the primary surface, with opposite signs on opposite sides of the surface. The central idea is that if the data shapes are registered correctly, the signed distance fields should match in the common coordinate system.

To register the prostate MRI/MRSI with CT images, theoretically, all we need to do is to use an optimization algorithm to minimize an mutual information metric or alike. Such setups have been applied before on the problem of CT–positron emission tomography (PET) registration (29), with splines modeling the deformations and a limited memory quasi-Newton (*L-BFGS*) algorithm (30) optimizing the system. In reality, the mutual information metric is noisy, and a deterministic optimization algorithm is generally not suitable to find the global solution. More powerful algorithms capable of statistically escaping from local minima, such as simulated annealing or genetic algorithms, are often used to attack the problem. Even with these powerful stochastic algorithms, finding the true solution of the problem can still be challenging because of the wide variety of possible pixel intensities within an organ. A narrow band formalism was devised to improve the convergence the calculation. A narrow band is composed of all points within two isosurfaces defined by the signed distance values

of $\pm d$. The method is a hybrid of the techniques based on image intensity distributions and the signed distance information (28). In essence, the technique is a two-step image registration (31) in which an organ is first represented by a data structure containing the signed distance values from objects, followed by an image registration using a pixel-based metric. A signed distance field of an object is represented through the narrow band, since the concept prevents self-intersection problems and seamlessly handles changes in topology. Previous studies have suggested that the technique improves the convergence behavior of the calculation and reduces the computational efforts (31), because sophisticated statistical considerations can be replaced with simpler pixel-based metrics computed only in the regions of clinical interest.

In this work we apply the narrow band deformable registration model to map the ER-based MRI/MRSI data onto CT images and report our implementation of the algorithm. In particular, we present our experience with the selection of model parameters, optimization algorithm, and the validation of the technique. The general reference drawn from this study is that the narrow band technique is robust and accurate for mapping information between different types of images. The approach is quite general and, with minor modifications, it should be applicable to many other deformable image registration problems in radiation therapy.

METHODS AND MATERIALS

Software platform

All calculations are implemented using an open-source software toolkit named the Insight Toolkit (ITK) (32), which consists of template-based codes for a large number of image visualization, segmentation, and registration classes. The programs contained in ITK are easily extendable, making it an ideal platform for the development of image registration methods. Concise and clear descriptions on the use of the available subroutines are provided in the ITK manual. For convenience, in the following we outline the methods used in our calculation with attention paid to the issues specific to radiation therapy image registration and to the implementation of the methods.

Registration framework

The overall image registration process of ER-based MRI/MRSI and treatment planning CT images is shown in Fig. 1. The input to

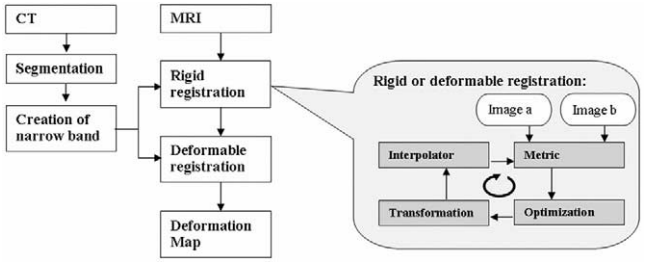
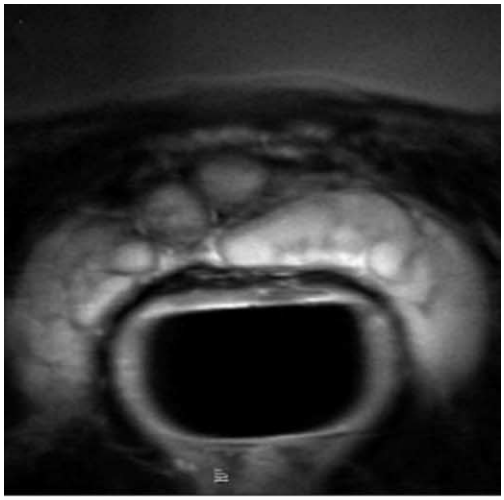
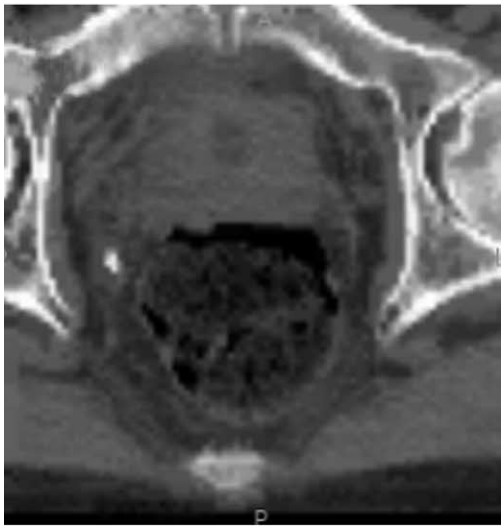


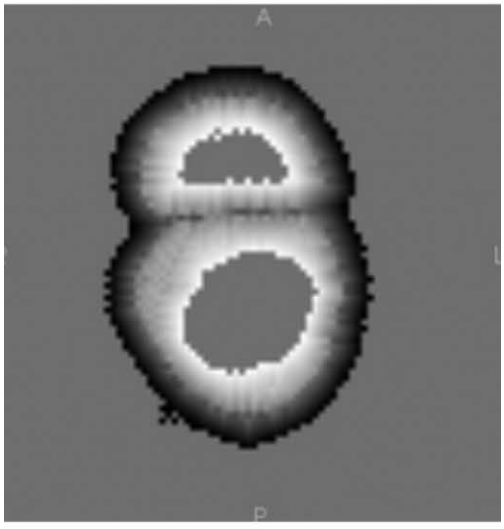
Fig. 1. Narrow band-based image registration procedure. A two-step registration is employed here to successively deal with translational and deformation miss-matches. (CT = computed tomography; MRI = magnetic resonance imaging.)



(a)



(b)



(c)

Fig. 2. (a) Magnetic resonance image zoomed in the prostate region, (b) computed tomography image of the pelvic region, and (c) the narrow band representation of rectum and prostate.

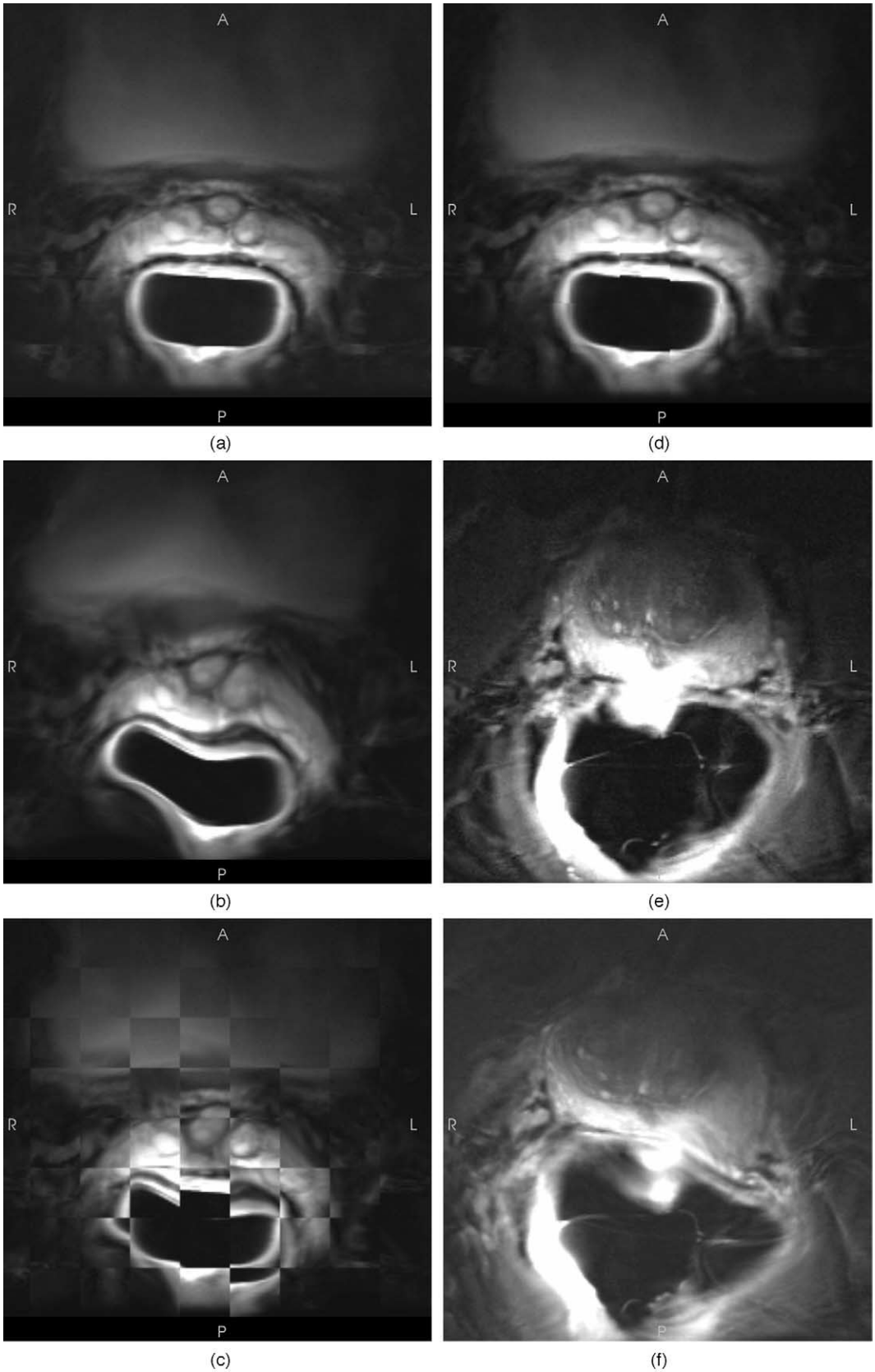
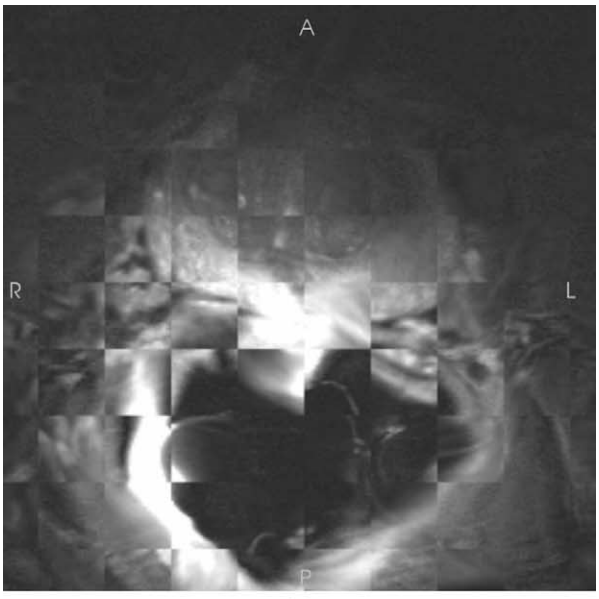
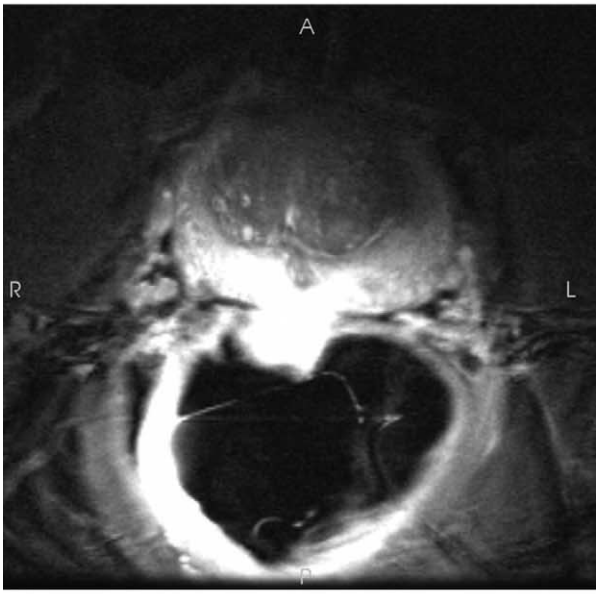


Fig. 3. Checkerboard visualization of algorithm's capability to recover from intentionally introduced deformations for two cases. The test is to register the warped image to the "gold standard" represented by the original form of the deformed magnetic resonance image. For the first case, the original and intentionally deformed images are shown in (a) and (b), respectively. The checkerboard images before and after deformable registration are shown in (c) and (d), respectively. The registration process and the results for the second case are shown in (e) to (h). Figure continues on p. 599.



(g)



(h)

Fig. 3. *(Continued)*

the registration software is the images to be registered: a fixed image and a floating image, described by their intensity distributions $I_a(\mathbf{x})$ and $I_b(\mathbf{x})$, respectively. In our problem, the former one is assigned to be the treatment planning CT image and the later the ER-based MRI/MRSI. To facilitate the computational process, we separated the image registration calculation into two steps. A rigid registration is first performed to grossly align the two images and the deformable registration then follows.

For convenience, a patient is divided into a grid with N^3 cells. The corner of a lattice cell is referred to as a node and is indexed by i ($i = 1, 2, \ldots N^3$). The displacement of a node i is specified by a vector $\mathbf{x}_i$, and the displacement vectors, $\{\mathbf{x}_i\}$, of a collection of nodes characterize the tissue deformation. The displacement at a location $\mathbf{x}$ on the image is deduced by fitting a polynomial expressed using the basis spline (BSpline) (33, 34) to the grid nodes $\mathbf{x}_i$. Unlike other spline models, the BSplines are locally controlled. That is, the displacement of an interpolation point is influenced only by that of the closest grid points, and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations.

A signed narrow band is defined on the CT image for prostate and rectum based on the physician-delineated contour. A narrow band is composed of a number of nodes with known distances to the contours, which introduces additional positional information in the representation. Typical narrow bands for rectum and prostate are shown in Fig. 2c as a gray-level image. Pixel intensities correspond to the signed distances with dark/white values representing the negative/positive distances. The zero level set is represented with a medium gray intensity. The gray background in this image is not part of the narrow band structure.

The task of standard image registration is to find the transformation matrix, $T(\mathbf{x})$, that maps an arbitrary point $\mathbf{x}$ on the fixed image to the corresponding point $\mathbf{x}'$on the floating image (or vice versa) so that the best possible match, as measured by the registration metric, is achieved. The choice of the metric for ranking different possible matching differentiates one algorithm from the others and is fundamental to the success of the image registration. In our calculation, each organ in the floating image is represented by the data structure of the signed distance or by the narrow band surrounding the edge of the organ. The narrow band acts as a shape representation model of an anatomic structure. It is the fixed or target object in the registration process so that the band does not need to be regenerated for each evaluation of the metric (31). As shown in Fig. 1, the image registration proceeds in an iterative fashion. The matrix coefficients of $T(\mathbf{x})$ are the node displacements and are adjusted iteratively to minimize the normalized correlation between the two images defined as:

$$f = \frac{\sum_{i=1} I_a(\mathbf{x}_i) I_b(T\mathbf{x}_i)}{\sqrt{\sum_{i=1} I_a^2(\mathbf{x}_i) \sum_{j=1} I_b^2(T\mathbf{x}_j)}} \tag{1}$$

where i and j are the node indices within the narrow band on the fixed image, $I_a(\mathbf{x}_i)$ is the intensity of the node at $\mathbf{x}_i$ on the fixed image a, and $I_b(T\mathbf{x}_i)$ is the intensity of the image b at where the node $\mathbf{x}_i$ is mapped. The metric uses both voxel intensity and delineated structures as complementary information for the registration.

The L-BFGS algorithm for optimization

Optimization of the normalized correlation function with respect to the displacements of the nodes, $\{\mathbf{x}_i\}$, yields the transformation coefficients $T(\mathbf{x})$ that map the points on image a to image b. Since the two images do not necessarily have the same size, an interpolation may be needed to compute intensity at a mapped point $\mathbf{x} = T(\mathbf{x})\mathbf{x}$. To facilitate the optimization it is preferable that both the deformable model and the metric are differentiable (29). This condition is satisfied for the system that we are dealing with, as demonstrated in a previous mathematical work (35).

We used the limited memory *BFGS* algorithm *L-BFGS* (30), which is known for its superior performance in dealing with high-dimensionality problems, to optimize the system here. *L-BFGS*, as compared with a conventional gradient search method such as the Newton's algorithm, does not require the exact inverse Hessian matrix. For an N-dimensional problem, only $4mN$ operations are needed in *L-BFGS* to calculate the descent direction at an

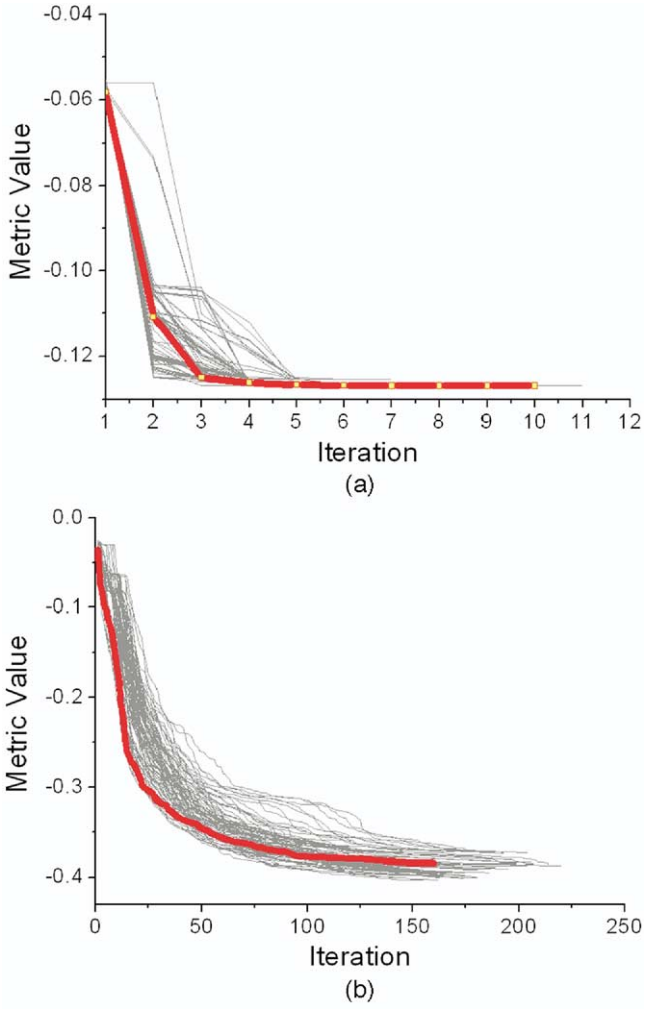


Fig. 4. Narrow band metric as a function of iteration step for calculations with 100 different initial values. Both rigid (a) and deformable (b) registration calculation are plotted. In both cases, a typical convergence is displayed in red. Rigid registration convergence is achieved after less than 10 iterations. Deformable registration converges in approximatively 100 iterations.

iteration k, indirectly from the m previous values of $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ of the system variable $\mathbf{x}$ (30, 36). A value of $m = 5$ is recommended. For a large N system, the number of operations becomes significantly less than $O(N^2)$, as involved in *BFGS* calculation. Starting from a positive definitive approximation of the inverse Hessian $\mathbf{H}_0$ at $\mathbf{x}_0$, *L-BFGS* derives the optimization variables by iteratively searching through the solution space. At an iteration k, the calculation proceeds as follows:

(a) Determine the descent direction $\mathbf{p}_k = -\mathbf{H}_k \nabla f(\mathbf{x}_k)$;
(b) Line search with a step size $\alpha_k = \arg\min f(\mathbf{x}_k + \alpha \mathbf{p}_k)$, where $\alpha \geq 0$ is the step size defined in the *L-BFGS* software package;
(c) Update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$;
(d) Compute $\mathbf{H}_{k+1}$ with the updated $\mathbf{H}_k$.

At each iteration a backtracking line search is used in *L-BFGS* (30) to determine the step size of movement to reach the minimum of f along the ray $\mathbf{x}_k + \alpha \mathbf{p}_k$. For convergence α has to be chosen such that a sufficient decrease criterion is satisfied, which depends on the local gradient and function value and is specified in *L-BFGS* by the Wolfe conditions (30).

During an image registration process, the above iterative calculation based on *L-BFGS* algorithm continues until either a pre-set maximum number of iterations (typically, 500 iterations) is reached or the following stopping criterion is fulfilled:

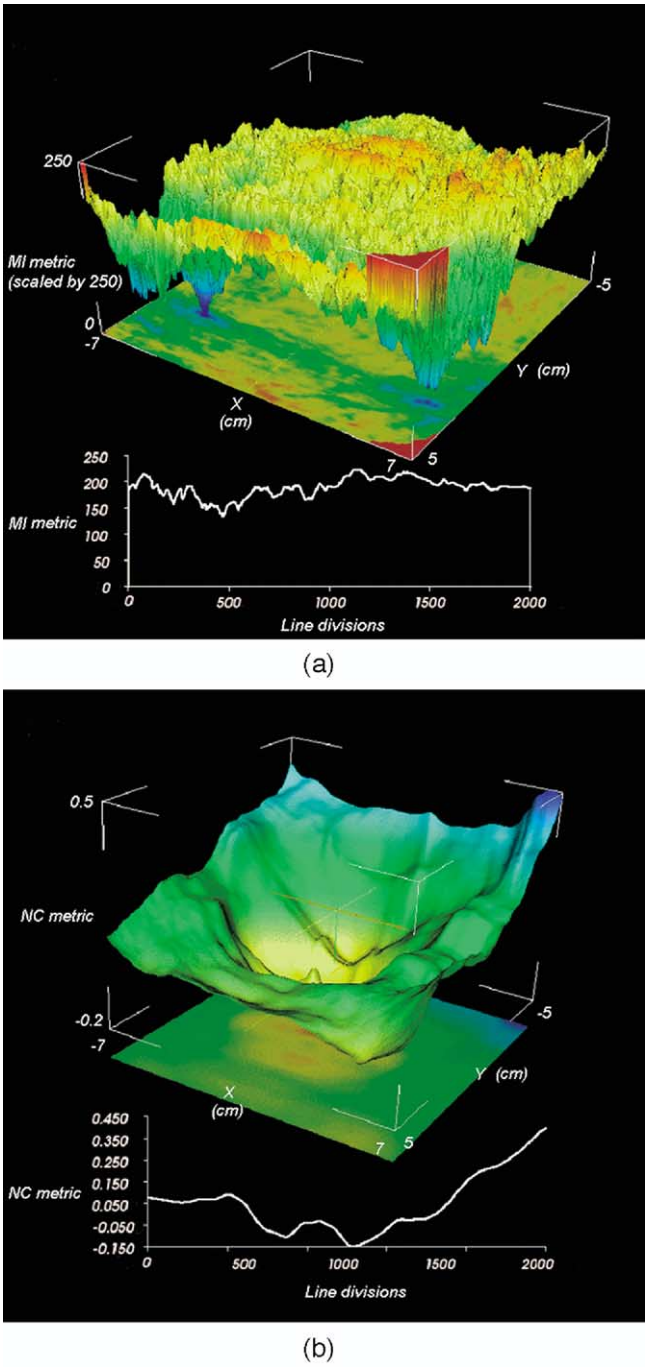
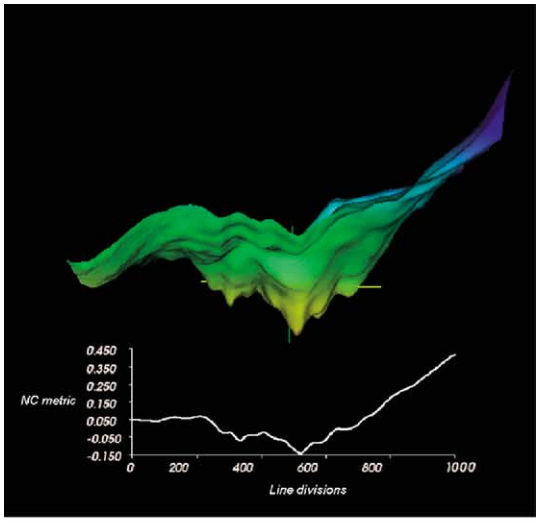
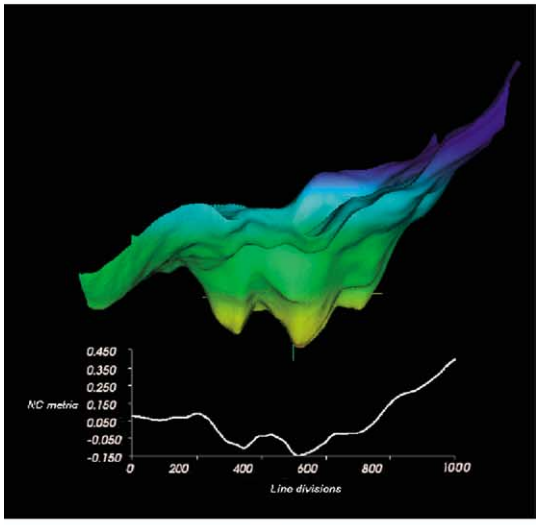


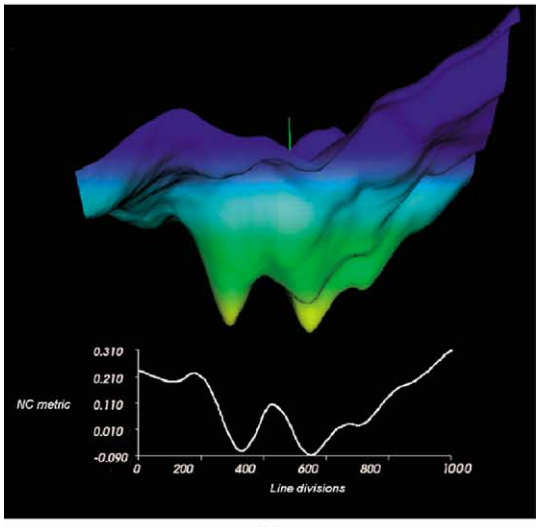
Fig. 5. The mutual information metric (a) and the narrow band metric (b) as a function of relative displacement in the xy plane. Two identical magnetic resonance images (from Fig. 2a) were used as the input fixed and floating images. The insert in each panel shows the corresponding metric when the displacement is confined to the x-direction. The mutual information metric is seen to be very "noisy" and multiple maxima exist. The narrow band metric, on the other hand, is much smoother and has a single minimum.



(a)



(b)



(c)

$$\frac{\|\nabla f(\mathbf{x}_k)\|_2}{\max(1, \|\mathbf{x}_k\|_2)} < \varepsilon \tag{2}$$

In our calculations, we chose $\varepsilon = 10^{-6}$. It is possible that *L-BFGS* may produce unrealistic deformations in finding the minimum of the metric. Similar to previous investigators (29), we used the bounded version of *L-BFGSB* (37), where the variables representing deformation vectors are restricted to within certain limits. The convergence behavior of the *L-BFGS* algorithm is studied by starting the registration calculation with different random initial positions. Intermediate values obtained during the optimization calculations are recorded and plotted for comparison.

Image acquisition

The MRI images (marked as FSE-xlT2 fat) were acquired using a 3-Tesla MR scanner (Signa; GE Medical Systems, Milwaukee, WI). The radiofrequency (RF) excitation was achieved by using the whole body birdcage resonator, and the MR signal was received using a 4-element phased-array antenna (GE Medical Systems) combined with a rigid single loop receiver-only surface coil with a fixed geometry that enables optimal tuning and matching for use at 3T. The coil dimensions are similar to transrectal ultrasound transducers used for routine sonographically guided prostate imaging and biopsy. Patient CT images were acquired using a PQ5000 CT Scanner (Philips Medical Systems, Cleveland, OH).

Behavior of the metric function

Monitoring the change of the metric function under the relative displacement of two input images provides a useful examination of the behavior of the solution space for image registration. To better understand the narrow band based normalized correlation metric, we used two identical images as input and computed the function by successively translating them along the x- and y-directions. In this case, it is intuitively conceivable that the value of the metric function is minimum when they are perfectly aligned and increases as the two images are displaced away from each other. A desirable metric function should exhibit a single distinct global minimum and vary smoothly as the two images slide away from each other. For comparison, the same calculation was also performed for conventional mutual information metric (27). The two types of functions are presented and reviewed by using 3D graphs, where surface height represents the metric value.

Registration accuracy

In this study, the input images involved an MRI as the fixed image and the same image with intentionally introduced deformations as the floating images. The ability of the narrow band algorithm to restore from the deformation was tested. The original image in this study serves as the "gold standard" for the evaluation.

Fig. 6. The narrow band metric as a function of relative displacement in the xy plane and in the x-direction when the narrow band width takes a value of 3 mm (a), 5 mm (b), and 15 mm (c). For a small width, the narrow band has a small capture radius with the minima of the metric accentuated. Increasingly narrow band width enlarges the capture radius. The computation becomes more demanding in this case owing to the increased number of narrow band nodes. A width of approximately 6 mm seems to balance the two limiting factors.

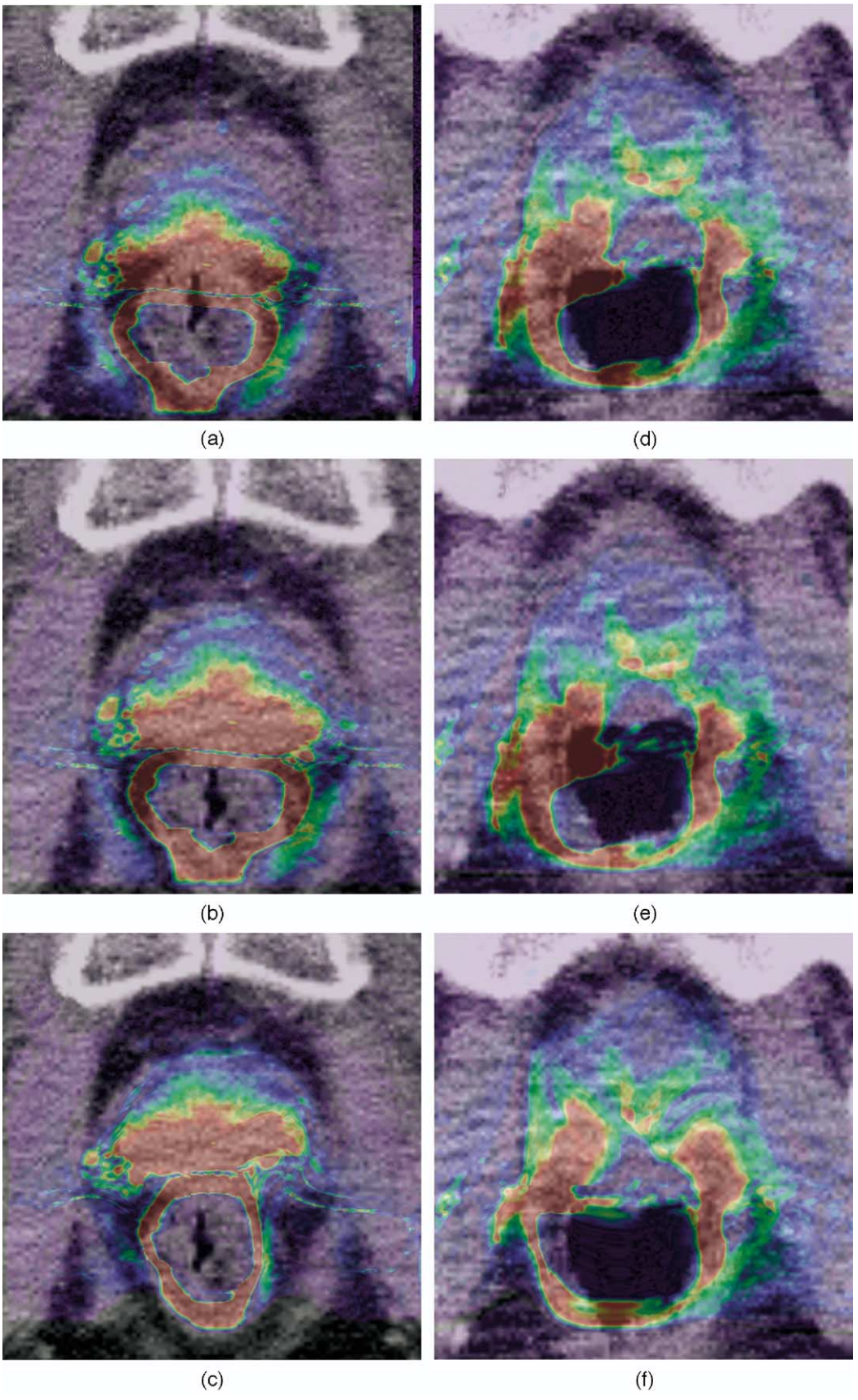


Fig. 7. Registration result for two clinical cases. The first row shows the colored overlay of the endorectal-based magnetic resonance and computed tomography images before and after rigid and after deformable image registration for the first case. The computed tomography image is used as the background. The second row shows the same for the second clinical case.

Figures 3a and 3b show the original and deformed MR images for the first patient. The deformation was generated by assigning the nodes in the original image with a set of random displacement vectors. Eight grid points with spacing of 2.4 cm were used in the calculation. We used the proposed algorithm to restore the distortions and evaluated the algorithm's ability to recover the gold standard. The difference between the original and restored images was quantified in terms of displacements of visible anatomic landmarks. The images before and after restoration were reviewed by using the popular checkerboard display tool, where images to be compared are merged together in a chess-like pattern. The original and deformed images for the second case are shown in Figs. 3e and 3f. Similar calculation and analysis were performed for this case.

Case study

The narrow band based image registration algorithm was applied to register ER-based MRI and CT prostate images for two clinical cases. The MRI and CT images were acquired using the protocols described earlier. The colored overlay of the two types of images was generated before and after the narrow band image registration calculations for assessment of the results. For each case, the convergence of the calculation was also studied.

RESULTS AND DISCUSSION

Convergence analysis

The algorithm's ability to obtain the same result with different starting conditions was examined. In Fig. 4, we plot the metric value as a function of iteration step starting from 100 different initial transform parameters, $\{\mathbf{x}_i\}$, for the first case shown in Fig. 3. The input images in this study included an ER-based MRI as the floating image and the treatment planning CT as the fixed image. Two types of registrations were studied here: a rigid (Fig. 4a) and a deformable (Fig. 4b) registration. For the rigid registration, the normalized correlation converges to -0.130, after 11 iterations. The final value of the metric is found to be independent of the initial starting conditions. For deformable registration, the minimum metric value was found to be -0.399. The longest run in obtaining this solution was 218 iterations. Once again, the result was independent of the initial starting conditions. For each type of registration, a typical evolution of the metric as a function of the iteration step is displayed in Fig. 4 in the red curve.

A narrow band is a compact representation of a structure because only points adjacent to its border are used. This permits a reduction of memory requirements for the optimization process, with computation times reduced by 1 or 2 orders of magnitude. For this particular study, *L-BFGS* needed approximatively 100 iterations to converge. Compared with some other potentially more powerful evolutionary algorithms, an analysis not shown here revealed that a typical 1 + 1 optimizer (38) needs as much as 100,000 iterations for convergence. This is reflected in the computation times, where 15 min are needed by *L-BFGS* for a full study of $256 \times 256 \times 30$ voxels, vs. more than 1 h needed by a 1 + 1 optimizer.

Behavior of mutual information and narrow band metric functions

In Fig. 5, we plot the mutual information and narrow band metric values when the two identical input images are translated on the x-y plane. Because the attempt here is to register an MRI image (Fig. 5a) with itself, the position of global minimum is known to correspond to a null translation. The variation of the metrics as a function of the translation in the x-direction is plotted as inset in Fig. 5. The mutual information metric space, presented in Fig. 5a, was obtained with settings of 100 spatial samples and 0.4 standard deviation (27). While the general trend toward a global solution located at the center can be deduced, the metric space is noisy and many local minima are present. This metric is thus not suitable to be used with gradient-based optimizers.

As can be seen from Fig. 5b, the search space corresponding to the normalized correlation metric is much smoother than that of the mutual information metric (27) and has a distinct minimum at the expected location. The local minimum at the left is resulted from the alignment of the rectum with the prostate. However, this minimum is shallower, indicating a poor fit, owing to the difference in rectum/prostate shapes. We note that with the selection of different narrow band widths it is possible to change the capture radius of the metric function (Fig. 6). As the narrow band width increases, the metric capture radius increases; and at the same time, the computation time needed to find the solution also increases since more nodes are involved. In this trade-off, we found a narrow band width of approximatley 6 mm to be optimal for practical application.

The narrow band metric uses both voxel intensity and delineated structures information for the registration. For prostate MRI-CT registration, an intensity-based metric such as the mutual information is intractable owing to the lack of clearly differentiable anatomic structures in the CT images (Fig. 2b) and the presence of a broad range of voxel intensities in the MRI image (Fig. 2a). The narrow band metric compensates for the inconveniences by restricting the calculation to selected regions of interest. The narrow band not only represents the shape of the contour but also a transition zone around it. This results in a large capture radius in the optimization calculation. As compared with control-points based methods, where the registration accuracy depends heavily on the specification of the control points, the narrow band based image registration entails little information from the user other than the delineation of the involved organs, which one has to do anyway for the purpose of radiation treatment planning.

Evaluation of the algorithm's ability of restoring from an intentionally introduced deformation

For the deformation depicted in Fig. 3b, ideally, the application of a deformable image registration technique should be able to restore its original shape when the undeformed image (Fig. 3a) is used as a reference. To better visualize the intentionally introduced deformation, in Fig.

3c we show the checkerboard image of the original and distorted image. The initial deformations up to 10 mm in the original image are clearly visible in the checkerboard transition zone. To use the narrow band based registration tool, the organs in the reference image were outlined and narrow bands were generated. The restored image is shown in Fig. 3d along with the original reference image in a checkerboard fashion. Recovered and gold standard reference images differ by less than 2 mm, as assessed by comparing displacements of the anatomic landmarks between neighborhood squares of the checkerboard. The original and restored images are highly similar, with virtually no difference visible in the checkerboard. Similar level of accuracy was found for the second case studied (Figs. 3e–3h). In both cases, we found that the calculation converged to the anticipated solution with fewer than 220 iterations.

Case study

In Fig. 7, we show the prostate MRI-CT registration results for two clinical cases. The same colored overlay is constructed before (Fig. 7a for Case 1 and Fig. 7d for Case 2) and after registration calculation (Fig. 7c for Case 1 and Fig. 7f for Case 2). In the colored overlay, the original CT image is used as the image background. The colored overlay represents the MRI image, with a colored scheme corresponding to different MRI pixel intensities, red for high intensities and blue for low intensities. Because of the use of the ER probe, the shapes of both rectum and prostate were deformed. As can be seen from Figs. 7a or 7d, the anterior surface of the rectum was flattened. After the rigid registration calculation, we noticed that the gross misalignment was removed (Figs. 7b and 7e). As expected, there was still substantial deformation after the initial rigid registration that needs to be corrected by a deformable image registration algorithm. In the next stage of the registration calculation, the deformable model was "switched on," which restored the shapes of the rectum and prostate to their regular forms as observed in the CT images. This can be seen from Figs. 7c for the first case or Fig. 7f for the second case.

In our calculation, we observed that the regular grid of BSpline control points could be mapped to a region outside the narrow band. While it seems that this does not directly affect the accuracy of the method, it may prolong the calculation by computing the displacements in regions where no metric information is available. Setups have been proposed to adapt the splines control mesh to regions where deformation is found to be significant (39), and the extension of the method would allow us to use the BSpline control points defined only in the regions within the narrow band. Implementation of this type of technique should further reduce the computation time required to find the optimal solution.

CONCLUSIONS

We have applied the narrow band algorithm for the registration of ER-based prostate MRI and treatment planning CT images. The narrow band is a compact representation of a structure because only pixels close to the structure boundaries are considered. A distance map around structures permits the use of simple registration metrics without relying on introducing control points. Both validation and patient studies have indicated that the registration is reliable and provides a valuable tool for integrating the ER-based MRI information into radiation therapy treatment planning. Because the contours of the organs are needed anyway for treatment planning purposes, the approach involves little additional work in constructing the narrow bands. Our study indicated that the narrow band metric has a smooth metric space, which permits us to use deterministic algorithms such as *L-BFGS* to optimize the system. The narrow band registration procedure was found to be fast, stable, and capable of providing practically acceptable accuracy. Finally, the method works with CT images of any quality, as important information is contained within the narrow band representation specified by the user. This might have a significant practical implication with the prevalence of KeV and MeV cone-beam CT in clinical practice.

REFERENCES

1. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24–0.7-cm^3) spatial resolution. *Radiology* 1996;198:795–805.
2. Scheidler J, Hricak H, Vigneron DB, *et al.* Prostate cancer: Localization with three-dimensional proton MR spectroscopic imaging—clinicopathologic study. *Radiology* 1999;213:473–480.
3. Pickett B, Kurhanewicz J, Coakley F, Shinohara K, Fein B, Roach M. Use of MRI and spectroscopy in evaluation of external beam radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2004;60:1047–1055.
4. Schnall MD, Imai Y, Tomaszewski J, Pollack HM, Lenkinski RE, Kressel HY. Prostate cancer: Local staging with endorectal surface coil MR imaging. *Radiology* 1991;178:797–802.
5. Schnall MD, Lenkinski RE, Pollack HM, Imai Y, Kressel HY. Prostate: MR imaging with an endorectal surface coil. *Radiology* 1989;172:570–574.
6. Presti JC, Hricak H, Narayan PA, Shinohara K, White S, Carroll PR. Local staging of prostatic carcinoma: Comparison of transrectal sonography and endorectal MR imaging. *AJR Am J Roentgenol* 1996;166:103–108.
7. Hricak H, Dooms GC, Jeffrey RB, *et al.* Prostatic carcinoma: Staging by clinical assessment, CT, and MR imaging. *Radiology* 1987;162:331–336.
8. Perrotti M, Kaufman RP Jr, Jennings TA, *et al.* Endo-rectal coil magnetic resonance imaging in clinically localized prostate cancer: Is it accurate? *J Urol* 1996;156:106–109.
9. Futterer JJ, Scheenen TW, Huisman HJ, *et al.* Initial experience of 3 Tesla endorectal coil magnetic resonance imaging and 1H-spectroscopic imaging of the prostate. *Invest Radiol* 2004;39:671–680.
10. Jager GJ, Severens JL, Thornbury JR, de La Rosette JJ, Ruijs

SH, Barentsz JO. Prostate cancer staging: Should MR imaging be used? A decision analytic approach. *Radiology* 2000;215:445–451.
11. Quinn SF, Franzini DA, Demlow TA, *et al.* MR imaging of prostate cancer with an endorectal surface coil technique: Correlation with whole-mount specimens. *Radiology* 1994; 190:323–327.
12. Kim DH, Margolis D, Xing L, Daniel B, Spielman D. *In vivo* prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 Tesla. *Magn Reson Med* 2005; in press.
13. Zaider M, Zelefsky MJ, Lee EK, *et al.* Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging. *Int J Radiat Oncol Biol Phys* 2000;47:1085–1096.
14. Xing L, Cotrutz C, Hunjan S, Boyer AL, Adalsteinsson E, Spielman D. Inverse planning for functional image-guided intensity-modulated radiation therapy. *Phys Med Biol* 2002; 47:3567–3678.
15. Das SK, Miften MM, Zhou S, *et al.* Feasibility of optimizing the dose distribution in lung tumors using fluorine-18-fluorodeoxyglucose positron emission tomography and single photon emission computed tomography guided dose prescriptions. *Med Phys* 2004;31:1452–1461.
16. Ling CC, Humm J, Larson S, *et al.* Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys* 2000;47:551–560.
17. Hirose M, Bharatha A, Hata N, *et al.* Quantitative MR imaging assessment of prostate gland deformation before and during MR imaging–guided brachytherapy. *Acad Radiol* 2002;9: 906–912.
18. Fei B, Kemper C, Wilson DL. A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes. *Comput Med Imag Graph* 2003;4:267–281.
19. Christensen GE, Carlson B, Chao KS, *et al.* Image-based dose planning of intracavitary brachytherapy: Registration of serial-imaging studies using deformable anatomic templates. *Int J Radiat Oncol Biol Phys* 2001;51:227–243.
20. Court LE, Dong L. Automatic registration of the prostate for computed-tomography-guided radiotherapy. *Med Phys* 2003; 30:2750–2757.
21. Bharath A, Hirose M, Hata N, *et al.* Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Med Phys* 2001;28: 2551–2560.
22. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol* 2004;49:3067–3087.
23. Wu X, Dibiase SJ, Gullapalli R, Yu CX. Deformable image registration for the use of magnetic resonance spectroscopy in prostate treatment planning. *Int J Radiat Oncol Biol Phys* 2004;58:1577–1583.
24. Lian J, Hunjan S, Dumoulin C, *et al.* Integrating deformable MRI/MRSI and CT image registration into the prostate IMRT treatment planning [Abstract]. *Int J Radiat Oncol Biol Phys* 2003;57(Suppl. 2):S207.
25. Lian J, Xing L, Hunjan S, Spielman B, Daniel B. Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study. *Med Phys* 2004; 31:3087–3094.
26. Coselmon MM, Balter JM, McShan DL, Kessler ML. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med Phys* 2004;31:2942–2948.
27. Viola P, Wells WM. Alignment by maximization of mutual information. *Int J Comput Vision* 1997;24:137–154.
28. Borgefors G. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transact Pattern Anal Machine Intel* 1988;10:849–865.
29. Mattes D, Haynor RD, Vesselle H, Lewellen KT, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Transact Med Imag* 2003;22:120–128.
30. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Mathemat Progr* 1989;45:503–528.
31. Ng L, Ibanez L. Narrow band to image registration in the Insight Toolkit. *Lecture Notes Comput Sci* 2003;2717:271–280.
32. Ibanez L, Schroeder W, Ng L, Cates J. ITK software guide. Clifton Park, NY: Kitware; 2003.
33. Lee S, Wolberg G, Chwa KY, Shin SY. Image metamorphosis with scattered feature constraints. *IEEE Transact Visualiz Comput Graph* 1996;2:337–354.
34. Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. *IEEE Transact Visualiz Comput Graph* 1997;3:228–244.
35. Thevenaz P, Unser T. Spline piramids for inter-modal image registration using mutual information. *IEEE Transact Med Imag* 1997;16:187–198.
36. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C. 2nd ed. Cambridge, England: Cambridge University Press; 1992.
37. Byrd RH, Lu P, Nocedal J, Zhu CY. A limited memory algorithm for bound constrained optimization. *SIAM Jon Sci Comput* 1995;5:1190–1208.
38. Wienholt W. Entwurf neuronaler Netze: Ein daten-basiertes Verfahren zur Modellierung nichtlinearer Systeme mit Hilfe evolutionarer Algorithmen und unscharfer Logik. Frankfurt am Main: Harri Deutsch; 1996.
39. Camara O, Colliot O, Delso G, Bloch I. 3D Nonlinear PET-CT image registration algorithm with constrained free-form deformations. *Proc 3rd IASTED Int Conf Visualiz Imaging Image Process* 2003;1:516–521.

Int. J. Radiation Oncology Biol. Phys., Vol. 63, No. 2, pp. 584–593, 2005
Copyright © 2005 Elsevier Inc.
Printed in the USA. All rights reserved
0360-3016/05/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2005.06.008

PHYSICS CONTRIBUTION

DOSE–VOLUME BASED RANKING OF INCIDENT BEAM DIRECTION AND ITS UTILITY IN FACILITATING IMRT BEAM PLACEMENT

EDUARD SCHREIBMANN, PH.D., AND LEI XING, PH.D.

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA

Purpose: Beam orientation optimization in intensity-modulated radiation therapy (IMRT) is computationally intensive, and various single beam ranking techniques have been proposed to reduce the search space. Up to this point, none of the existing ranking techniques considers the clinically important dose–volume effects of the involved structures, which may lead to clinically irrelevant angular ranking. The purpose of this work is to develop a clinically sensible angular ranking model with incorporation of dose–volume effects and to show its utility for IMRT beam placement.

Methods and Materials: The general consideration in constructing this angular ranking function is that a beamlet/beam is preferable if it can deliver a higher dose to the target without exceeding the tolerance of the sensitive structures located on the path of the beamlet/beam. In the previously proposed dose-based approach, the beamlets are treated independently and, to compute the maximally deliverable dose to the target volume, the intensity of each beamlet is pushed to its maximum intensity without considering the values of other beamlets. When volumetric structures are involved, the complication arises from the fact that there are numerous dose distributions corresponding to the same dose–volume tolerance. In this situation, the beamlets are not independent and an optimization algorithm is required to find the intensity profile that delivers the maximum target dose while satisfying the volumetric constraints. In this study, the behavior of a volumetric organ was modeled by using the equivalent uniform dose (EUD). A constrained sequential quadratic programming algorithm (CFSQP) was used to find the beam profile that delivers the maximum dose to the target volume without violating the EUD constraint or constraints. To assess the utility of the proposed technique, we planned a head-and-neck and abdominal case with and without the guidance of the angular ranking information. The qualities of the two types of IMRT plans were compared quantitatively.

Results: An effective angular ranking model with consideration of volumetric effect has been developed. It is shown that the previously reported dose-based angular ranking represents a special case of the general formalism proposed here. Application of the technique to a abdominal and a head-and-neck IMRT case indicated that the proposed technique is capable of producing clinically sensible angular ranking. In both cases, we found that the IMRT plans obtained under the guidance of EUD-based angular ranking were improved in comparison with that obtained using the conventional uniformly spaced beams.

Conclusions: The EUD-based function is a general approach for angular ranking and allows us to identify the potentially good and bad angles for clinically complicated cases. The ranking can be used either as a guidance to facilitate the manual beam placement or as prior information to speed up the computer search for the optimal beam configuration. Thus the proposed technique should have positive clinical impact in facilitating the IMRT planning process. © 2005 Elsevier Inc.

Intensity-modulated radiation therapy, Inverse planning, Dose optimization, Beam orientation.

INTRODUCTION

Several previous studies have indicated that the selection of beam directions plays an important role in intensity-modulated radiation therapy (IMRT) planning (1–8). For a given patient, a practical challenge is how to obtain the optimal beam configuration within a clinically acceptable time. The influence of a candidate beam configuration to the final dose distribution is not known until an inverse planning with the configuration is done, which necessitates a trial-and-error process for the determination of a clinically sensible IMRT beam configuration (9–11). The use of computer optimization promises to automate the beam selection process (12–15). In practice, however, beam orientation optimization is computationally intensive because of the interplay between beam directions and beamlet intensities. Many researchers

Reprint requests to: Lei Xing, Ph.D., Department of Radiation Oncology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Tel: (650) 498-7896; Fax: (650) 498-4015; E-mail: lei@reyes.stanford.edu

Supported in part by a research grant from the U.S. Department of Defense (DAMD17-03-1-0023) and National Cancer Institute (1 R01 CA98523-01).

Acknowledgments—We would like to thank the useful discussion with Drs. A. Pugachev, Y. Yang, and C. Cotrutz.

Received July 21, 2004, and in revised form May 19, 2005. Accepted for publication June 3, 2005.

have attempted to use single beam ranking to reduce the search space (16–23). Although such technique ignores the beam interplay and does not yield the final beam configuration, it affords useful information on which are potentially good/bad directions and is thus valuable to aid the beam placement (24–26). Furthermore, the information can be used as *a priori* knowledge to greatly improve the convergence behavior of the beam orientation optimization and to speed up the calculation process (21).

Single beam ranking functions can be divided into geometry based and dosimetry based models. The underlying difference between these models lies in what fundamental quantities are used to define the optimality. The former approach is essentially an extension of beam's eye view volumetrics (24, 27, 28) and ranks a beam direction based on the fractional volume of sensitive structures falling into the incident beam (16, 22, 26). The dosimetry-based technique, on the other hand, ranks a beam direction based on patient geometry as well as the *a priori* dose tolerance information of the involved sensitive structures (17, 18, 20). In line of the latter approach, Pugachev and Xing proposed a heuristic beam's eye view dosimetric (BEVD) model in which the ranking of an incident beam is gauged by the maximally deliverable dose to the planning target volume (PTV) without exceeding the tolerance of the sensitive structure located on the path of the beam (20). While the idea of their technique is quite general, practical implementation is hindered when the involved sensitive structure is a parallel organ whose tolerance cannot be described by a single dose value but a dose–volume relation. In this situation, the method described in Pugachev *et al.* (20) is not directly applicable and the evaluation of the single beam ranking becomes very nontrivial.

The purpose of this work is to develop a clinically sensible angular ranking model with incorporation of dose–volume effect and to show its utility for IMRT beam placement. Instead of working in the dose domain, we establish a more adequate formalism based on the concept of equivalent uniform dose (EUD) (29, 30). As is well known, the dose–volume effect of a structure is well described (31–33) by using the EUD, which is phenomenologically defined as:

$$EUD = \left(\frac{1}{N} \sum_i D_i^a \right)^{\frac{1}{a}}, \tag{1}$$

where N is the number of voxels in the structure, and a is the tumor or normal tissue-specific parameter that describes the dose–volume effect. When $a = \infty$, the EUD is equal to the maximal dose, and when $a = -\infty$, the EUD is equal to the minimum dose. Tumors generally have large negative values of a, whereas serial critical structures (e.g., spinal cord) have large positive values and parallel critical structures that exhibit a large volumetric effect (e.g., liver, parotids, and lungs) have small positive values. A virtue of the EUD-based approach is that the volumetric status of an organ can be characterized by a single parameter instead of two (dose and volume), making the incorporation of dose–volume effect in the single beam ranking more tractable. A few recent works (34–36) have attempted to relate EUD-based formalism to dose–volume constraints, dose-based inverse planning, and multiobjective optimization, and significant insights have been obtained in this regard.

METHODS AND MATERIALS

Background

The figure of merit of a beam direction is generally measured by how much dose can be delivered to the target and is calculated using the *a priori* dosimetric and geometric information of the given patient. For computational purposes, a beam direction is divided into a grid of beamlets. For a serial organ, each beamlet crossing the target is an independent element and is assigned to the maximum intensity (20). After a forward dose calculation using the maximum beam intensity profile, the score of the given beam direction (indexed by b) is obtained according to

$$S_b = \frac{1}{N_T} \sum_{n \in \text{Target}} \left(\frac{d_{nb}}{D_T^P} \right)^2, \tag{2}$$

where d_{nb} is the dose delivered to the voxel n by the beam from the direction indexed by b, N_T is the number of voxels in the target, and D_T^P is the target prescription.

When the tolerance of the sensitive structure is more accurately assessed by volumetric information, a modification of the above procedure is needed to rank the beam direction. The main computational difficulty here is that the beamlets cannot be dealt with independently as there are usually multiple dose distributions for a given dose–volume tolerance. A practical and efficient approach is to use the EUD model described earlier to solve the problem.

EUD as a general means for characterizing the dose–volume tolerance

The dose volume tolerance of a sensitive structure is more conveniently described by a single variable, namely, the EUD tolerance. The value of EUD tolerance (and the value of a in the EUD formula [Eqn. 1]) for a given organ can be obtained from the literature (37). For a given case, the figure of merit of a beam direction should now be measured by what the beam could achieve dosimetrically without exceeding the EUD tolerances of the organs located on the path of the beam. Although the score can still be evaluated by using Eq. 1 based on the integral dose to the target, the method of determining the maximum intensity profile for a given beam direction must be modified. In this situation, the beamlet intensities of the maximum intensity profile can no longer be determined individually because of the degenerate nature of a dose–volume tolerance (i.e., there are more than one dose distributions for a given dose–volume tolerance). Generally, there are multiple intensity profiles, each corresponding to a possible dose distribution

of the given EUD tolerance. A computer algorithm is necessary to find the one that delivers the maximal integral dose to the target volume. For this purpose, we mathematically formulated the problem into a constrained optimization, in which the objective is to maximize S_b defined in Eq. 2 with the EUDs of the sensitive structures below their tolerance values.

Calculation of beam ranking

The constrained optimization problem was solved using a Lagrange multiplier method (38). For the system that we are dealing with here, the Lagrange function takes the form of

$$L(\mathbf{x}, \lambda) = S_b(\mathbf{x}_b) + \sum_i \lambda_{b^i}^i \left[EUD_{tol}^i(\mathbf{x}_b) - EUD^i(x_b) \right], \tag{3}$$

where $\mathbf{x}$ is the system variable vector (i.e., the beamlet intensities for a given beam), i is the structure index, $EUD_{tol}^i(\mathbf{x})$ and $EUD^i(\mathbf{x})$ are the EUD tolerance and computed EUD, respectively, and $\lambda_{b^i}^i$ is the Lagrange multiplier. A constrained sequential quadratic programming algorithm (CFSQP) (39) with the Armijo type arc search mode (40) was employed to optimize the system described above. Variables to be optimized included the intensities of the beamlets passing through the PTV. A major advantage of the algorithm is that it is capable of dealing with nonlinear inequality constraints (40). The calculation starts with an initial intensity profile, in which each beamlet is assigned with a small but random value, and then iteratively maximizes the angular ranking function while satisfying the constraints. Briefly, the k-th iteration can be described by the following steps:

1. Calculate the inverse Hessian matrix, $\mathbf{H_k}$, from the gradient of $L(\mathbf{x},\lambda)$;
2. Determine the descent direction $\mathbf{p_k} = -\mathbf{H_k} \bigtriangledown f(\mathbf{x_k})$;
3. Line search with a step size $\alpha_k = \underset{\alpha>0}{\operatorname{argmin}} f(\mathbf{x_k} + \alpha \mathbf{p_k})$, where α is the step size defined in the CFSQP software package;
4. Update $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_k \mathbf{p_k}$;
5. Check stopping criteria. If not satisfied, repeat from 1.

CFSQP calculation stops either when a maximum number of 200 iterations is reached or when the difference in the values of the ranking function between two successive iterations becomes less than 10^{-6}. All calculations are implemented in an in-house treatment planning platform (6).

It is important to emphasize that the goal of the above calculation is not to find the optimal IMRT solution. Instead, we are searching for the beam profiles that deliver the highest achievable dose in the target without violating the dose–volume or EUD tolerances of the involved sensitive structures. In the final solution, any increase in the beamlet weights would lead to a dose exceeding the volumetric tolerance of a sensitive structure.

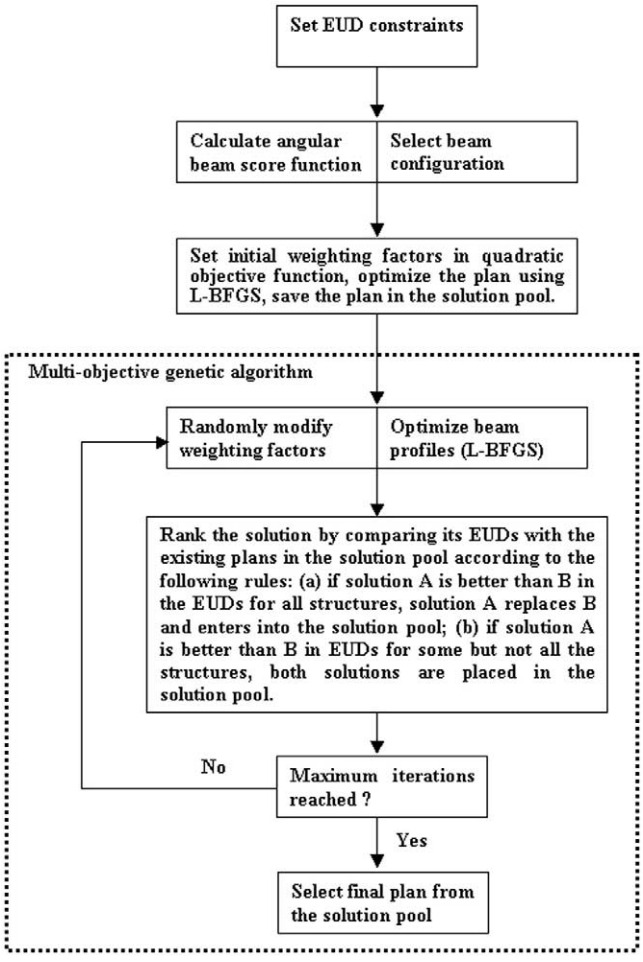


Fig. 1. Flow chart of the overall plan optimization process. EUD = equivalent uniform dose.

Optimization of IMRT beam profiles

The flow chart of the plan optimization process is shown in Fig. 1. We used a hybrid of a gradient-based algorithm (LBFGS) and a genetic-like algorithm (NSGA-IIc) (41) to obtain the optimal IMRT plan after the beam configuration is selected. The LBFGS is specially designed for dealing with systems with a large number of variables (42). A least-squares cost function was used as the objective for each structure, with the prescription dose set to 0 for sensitive structures to continuously lower the doses to the sensitive structures if there is room for improvement. The details of the LBFGS algorithm and our experience in implementing the algorithm were reported in a previous work (43). The EUDs of the involved structures were employed to generate the solution pool through the combinational use of a genetic-like algorithm (NSGA-IIc) (41) (see Fig. 1 flow chart for details). At each iteration, the EUDs of the new solution are compared with the plan within the solution pool and ranked according to the rules stated in the flow chart. A solution that violates less the EUD constraints has a higher probability to survive in the next generation and thus a larger chance to be selected as the final solution. In this sense, the EUDs here act as "soft constraints" in

conventional optimization or the decision-function in the two-step optimization proposed by Xing *et al.* (44). This setup has the advantage that the EUD constraints do not directly interfere with the objective function of the system. This model of implementing the constraints was found to be efficient in a previous study (6).

Case studies

The performance of the above algorithm and the advantage of the proposed angular ranking technique are illustrated using two clinical IMRT cases: an abdominal tumor with the kidney, spinal cord, and liver delineated as sensitive structures, and a head-and-neck case adjacent to the parotid and spinal cord. For each case, we computed the coplanar angular score function (from 0° to 360°) in an increment of 5°. The photon beam energy was 15 MV for both cases.

To understand the behavior of the technique in modeling different degrees of volumetric effect, we computed the score function for a series of values of parameter a in each case. As a increases to a large positive value, the organ behaves as a serial structure and the score function obtained by using the proposed technique should converge to the results obtained using the method outlined by Pugachev and Xing (20).

To illustrate the advantage of the technique, we generated a five-field IMRT treatment plan with the beam configuration obtained under the guidance of the computed angular score function for each patient. The plans were then compared with that obtained with five uniformly spaced beams (the beam directions are 0°, 72°, 144°, 216°, and 288°). In practice, 5~7 beams are often used for IMRT treatment because a good compromise can be achieved between delivery efficiency and the treatment quality. After the beam configuration was determined, the beamlets were optimized using a multiobjective approach described above (see Fig. 1) to yield the final IMRT plan. The optimization result is a database of plans representing the best achievable tradeoff

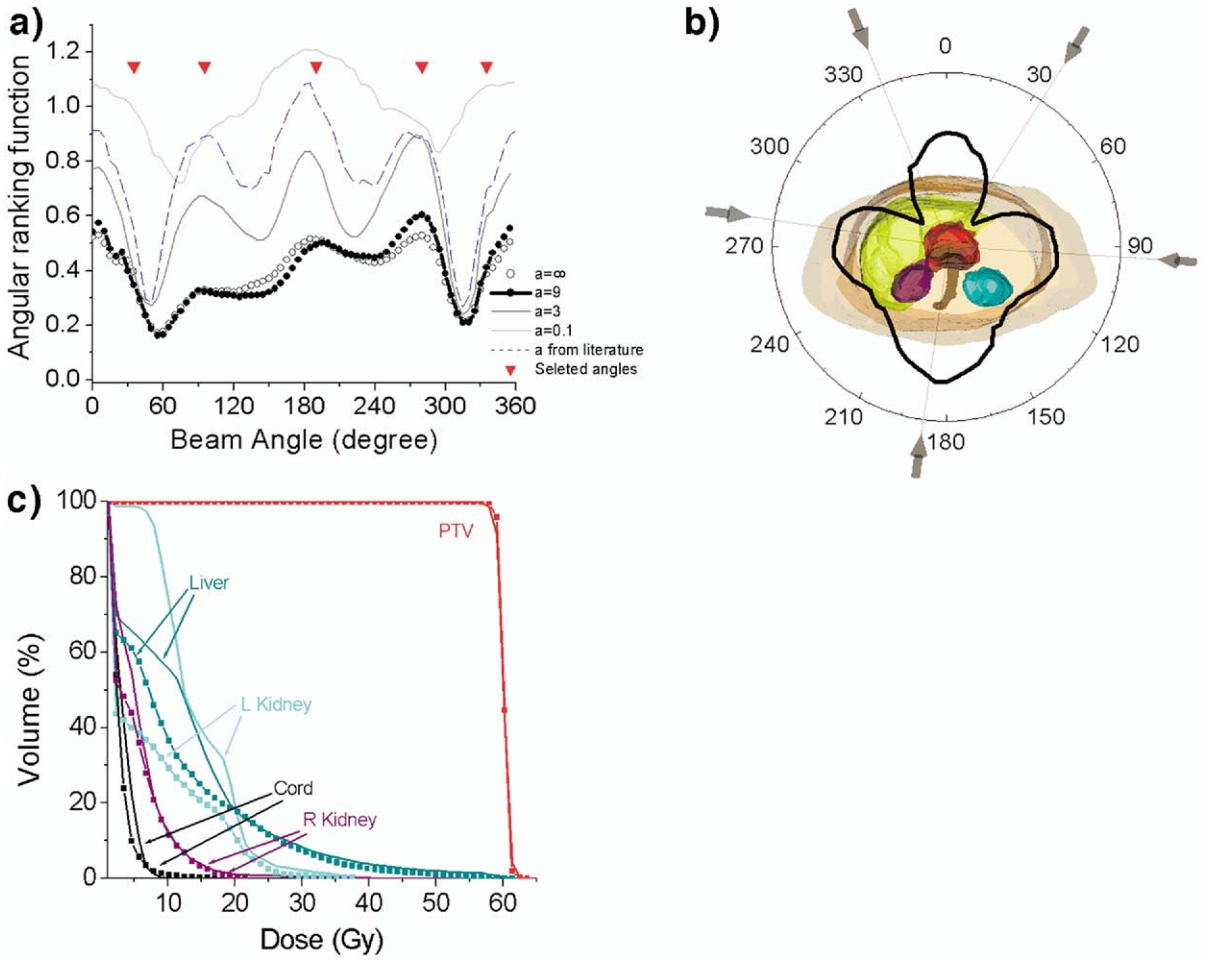


Fig. 2. (a) Angular ranking function for a series of a parameters. The selected five directions for intensity-modulated radiation therapy planning are labeled by arrows. The curve depicted by the open circles represents the result obtained using the approach described in Pugachev *et al.* (20). (b) Angular score obtained with published equivalent uniform dose tolerances and model parameters superimposed on the patient's geometry. Angles selected for intensity-modulated radiation therapy planning are shown by arrows. (c) Dose–volume histograms corresponding to the intensity-modulated radiation therapy plans obtained with (dot-dashed curves) and without (solid) the guidance of the angular ranking function.

between optimization objectives. The dose–volume histograms (DVHs) of the target and sensitive structures were used for the plan comparison.

To better assess the performance of the proposed technique, two additional experiments were carried out for the head-and-neck case. First, we intentionally selected the set of beams with the worst (corresponding to the dips of the beam orientation ranking function) and the best (corresponding to the peaks of the beam orientation ranking function, as chosen above) individual performances. The resulting IMRT plans were then compared quantitatively. This study highlights the importance of beam selection and the possible influence of individually selected beams. For the abdominal patient, we also investigated the differences in the final IMRT treatment plans obtained using the new method with EUD-based angular score function and the previous non-EUD based score function.

RESULTS

Abdominal case

The angular score functions corresponding to different values of the *a* parameter are presented in Fig. 2a along with the score obtained using the method reported in Ref. 20 for serial organs. As can be intuitively anticipated, the EUD-based angular score function approaches to the latter curve (denoted by the circles) as *a* increases. The values of *EUD* tolerances for the involved organs were taken from the literature (Table 1) (45). The score function corresponding to the *a* parameters extracted from Sang and Dunscombe (45) is also presented in Fig. 2b, where a polar graph has been overlayed on the patient's geometry to better interpret the curve. The influence of the sensitive structures is clearly reflected in the score function. The lowest scores occur at the angles close to 50° and 320°, from which directions the beam passes through the kidneys and the liver. The score function peaks at directions where a large portion of PTV can be irradiated with minimal intersection of sensitive structures.

A general way for beam placement after the BEVD score is obtained is to start with an equispaced beam configuration with one of the beams placed at the angle with the highest BEVD score value. The angular score described above is obtained under the assumption of a single incident beam and thus reflects only one facet of the beam configuration selection problem. The angles of the rest of the beams are adjusted in such a way that the final beam configuration balances the requirement of the single beam BEVD score (which favors placing beams at the peak positions of the BEVD score function) and the interplay between the beams (which prefers a beam configuration with the beams away from each other). In other words, to obtain the final beam configuration, it is required to consider the interplay between the incident beams. The merit of the proposed technique is that it is capable of identifying potentially "good" and "bad" directions even when the sensitive structures are volumetric in their responses to radiation, and thus provides a useful guidance for IMRT beam placement. In particular, it permits us to avoid placing beams in the "bad" beam directions (the dips of the score function). One would otherwise have to choose a beam configuration entirely based on empiric "guessing." Based on the computed score function, we placed the five incident beams at 35°, 95°, 190°, 280°, and 335°, as indicated by the arrows in Figs. 2a and 2b. The DVHs of the corresponding IMRT plan are plotted in Fig. 2c as dot-dashed lines. The DVHs of a conventional IMRT plan with five equispaced beams are also plotted in Fig. 2c as solid curves.

The computed EUD values of the involved structures for IMRT plans with equispaced beams and newly proposed EUD-based and non-EUD-based BEVD-selected beams are also listed in Table 1. We observed that the IMRT plan with the beam configuration selected under the guidance of the EUD-based scoring function represents the best treatment when judged using the EUDs of the involved structures. This is not surprising because the EUDs of various organs were fully considered during the beam configuration selection. The IMRT plan with beams derived from the non-EUD-based score function (the five beam directions derived from this score function were 20°, 120°, 175°, 260°, and 350°) was inferior to the plan mentioned above but better than that of the equispaced beam configuration. As seen from the table and Fig. 2c, the liver and both kidneys sparing are improved for the same PTV coverage. For example, the fractional volume of the right kidney receiving a dose of 15 Gy is reduced from 37% to 23% and the

Table 1. EUD tolerances (45) and calculated EUD values for IMRT plans with five equi-angled and BEVD-selected beams for the abdominal case

	Left kidney	Right kidney	Liver	Cord
α	5.1	5.1	0.59	7.4
EUD tolerance	28	28	39	43
EUD (equispaced beams)	16.82	11.96	8.77	5.31
EUD (BEVD-selected beams)	14.24	8.34	6.97	6.68
EUD (beams with non-EUD-based BEVD)	15.23	9.58	7.03	6.12

Abbreviations: BEVD = beam's eye view dosimetric; EUD = equivalent uniform dose; IMRT = intensity-modulated radiation therapy.

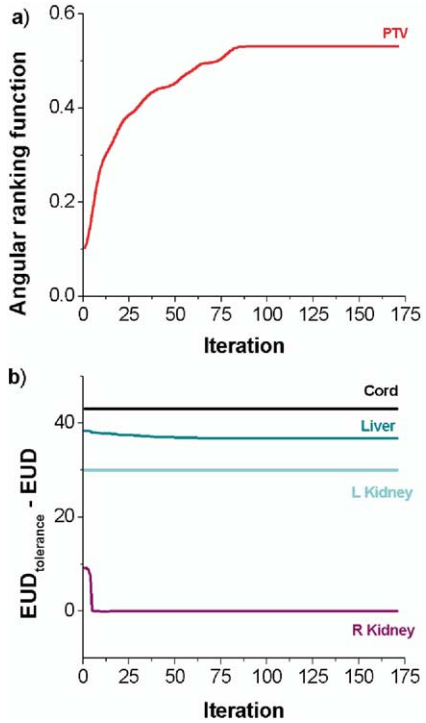


Fig. 3. Convergence behavior of the CFSQP algorithm for the 255° beam direction. Presented are evolutions of the angular ranking (a) and the sensitive structure constraints (b) as function of the iteration step. Only the right kidney and the liver influence the algorithm's convergence because other structures are not on the path of the beam.

reductions in lower doses are even significant. In the IMRT plan with BEVD-selected beam configuration, however, the maximum dose to the spinal cord is increased slightly. This is because a relatively high tolerance EUD was assigned to the cord, which therefore imposed less constraint on the angular ranking when the beam passes through it. For this case, the most dose-limiting organ is the left kidney when one attempts to escalate the target dose. Hence, the slight increase in the cord dose has little clinical impact. In a sense, the increase in the cord dose is an indication that the newly introduced ranking function can more effectively balance the requirements of the involved structures.

Convergence behavior of the CFSQP algorithm is demonstrated by plotting the angular score as a function of iteration step (Fig. 3a) for the 255° direction. The EUDs of the sensitive structures at each iteration step are shown in Fig. 3b. With the chosen initial beamlet intensities (small but random values), the angular score is progressively increased while constraints are progressively saturated, limited by the tolerances of the sensitive structures. Constraints of the sensitive structures that are not on the path of the beam remain to be constant throughout the iterative calculation. In this particular situation, the right kidney is the most dose-limiting organ. Of course, being dose-limiting for one of the beams does not mean that the organ is dose-limiting when all the beams are considered. Indeed, as seen from Fig. 2b and Table 1, it turns out that the left kidney is most dose-limiting for the chosen beam configuration and receives the highest dose or EUD. Generally speaking, this is not a symmetric system from the dosimetric point of view, even though the two kidneys have the same EUD tolerance. The asymmetric behavior is caused primarily by the presence of the liver near the right kidney. When the dose to the liver is limited to its EUD tolerance, a by-consequence is that the dose to the right kidney is also lowered. On the other hand, the left kidney is the sole dose-limiting organ on the left side and a much higher dose is delivered to it to deliver the prescribed dose to the target volume. The asymmetry here can also be seen from the angular ranking function (Fig. 2a).

Head-and-neck case

Dependence of EUD score on the a parameter is presented in Fig. 4a. As a increases, the angular score curve approaches to the curve computed using the method outlined in Pugachev *et al.* (20) (denoted by the open circles). This calculation provides a useful check of the new algorithm. It is interesting to note that the change in the peak positions of the angular function can be as large as 20° when the sensitive structures are changed from serial (corresponding to a high a value) to parallel (corresponding to a low a value). The change of the score function in amplitude is also striking (from ~0.2 to ~0.8 at 80° and 280°). As the structure becomes more volumetric (e.g., when $a = 0.1$), the score function becomes less "spiky."

When the published values of *EUD* tolerances (45) are used (Table 2), the lateral directions passing through parotid are disadvantageous, whereas the anterior-posterior directions passing through the cord have slightly higher rankings. The polar score function overlayed on the patient's geometry is presented in Fig. 4b. It is clearly seen that low scores occur at ~90° and ~270°, from which directions the beams pass through the parotids. Moderate scores appear at 0° and 180°, from which directions the beams pass through the cord. In this case, as can be intuitively imagined, high angular rankings appear in the oblique directions when the beam is less intercepted by the sensitive structures.

Under the guidance of the ranking function depicted in Fig. 4a, we selected 45°, 125°, 190°, 245°, and 310° for the five-beam IMRT treatment. The selected directions are indicated in Figs. 4a and 4b by arrows. The DVHs of the involved structures are shown in Fig. 4c (dot-dashed lines) along with the results obtained by using five uniformly spaced beams. As a result of BEVD guidance for beam placement, the IMRT plan is improved significantly. Together with a dose improvement in dose coverage of the PTV, all sensitive structures are better spared. The improvement in brainstem sparing is most distinct as compared with the conventional plan. The resulting EUD of the brainstem was reduced by ~65% (see Table 2). The reduction in the parotid EUD was found to be ~16%. The fractional volume of the spinal cord receiving a dose of 14 Gy is reduced from 32% to 22%. Clinically, the improvement in the dose homoge-

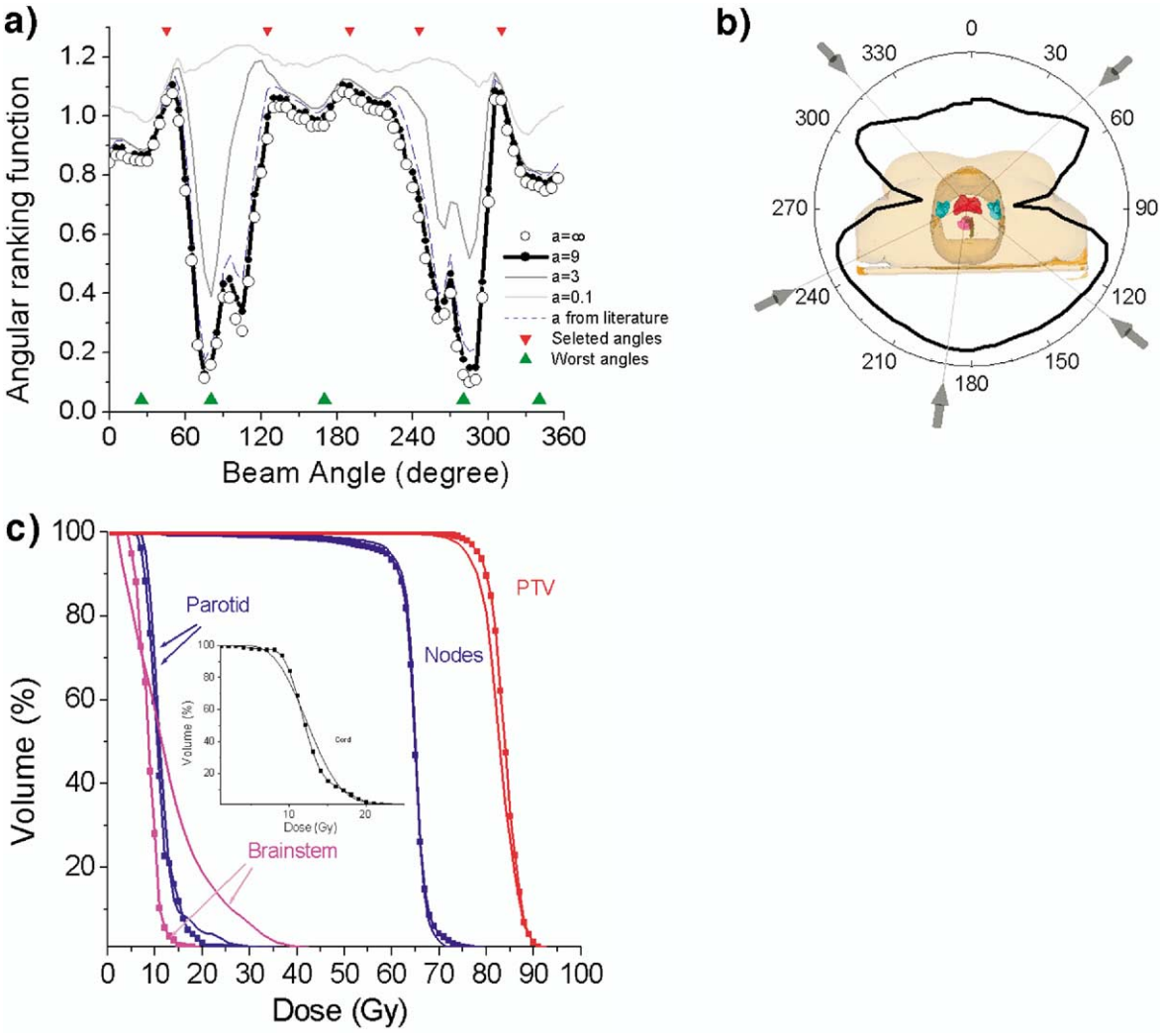


Fig. 4. (a) Angular ranking function for a series of values of equivalent uniform dose parameters. The selected five directions for intensity-modulated radiation therapy planning are labeled by arrows. The curve depicted by the open circles represents the result obtained using the approach described in Pugachev *et al.* (20). (b) Angular score obtained with published equivalent uniform dose tolerances and model parameters superimposed on the patient's geometry. Angles selected for intensity-modulated radiation therapy planning are shown by arrows. (c) Dose–volume histograms corresponding to the intensity-modulated radiation therapy plans obtained with (dot-dashed curves) and without (solid) the guidance of the angular ranking function.

neity within the PTV may have some practical implications in enhancing the tumor control probability. In addition, better sparing of the involved sensitive structures yields room for possible boost treatment of the patient, or, in the case of recurrence, for the retreatment. The convergence behavior of the optimization calculation for the 180° direction is presented in Fig. 5. Once again, it is seen that the calculation converges rapidly. The con-

Table 2. EUD tolerances (45) and calculated EUD values for IMRT plans with five equi-angled and BEVD-selected beams for the head-and-neck case

	Parotid	Brainstem	Cord	Score
α	5.0	4.6	7.4	
EUD tolerance	30	49	43	
EUD (equispaced beams)	16.67	27.55	20.99	1.30
EUD (BEVD-selected beams)	14.00	9.50	19.56	1.33
EUD (beams with worst BEVD scores)	18.71	22.82	20.29	1.27

Abbreviations as in Table 1.

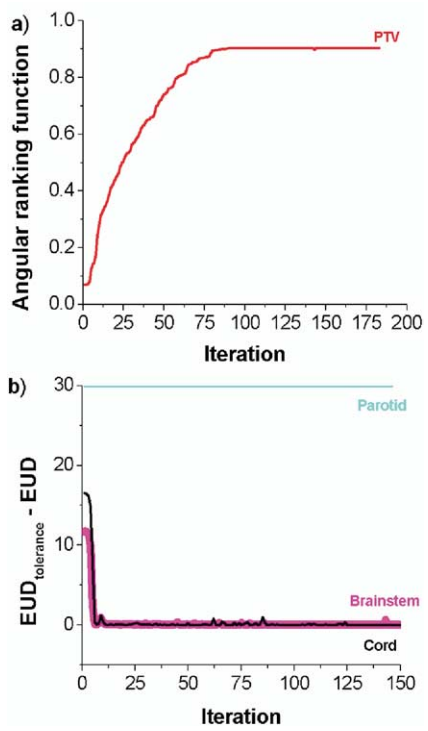


Fig. 5. Convergence behavior of the CFSQP algorithm for the 180° beam direction. Presented are evolutions of the angular ranking (a) and the sensitive structure constraints (b) as function of the iteration step. The cord and brainstem limit the dose deliverable to the planning target volume in this beam direction.

straints imposed by the parotid glands are satisfied from the beginning of the iterative calculation because the beam does not pass them.

In Fig. 6 we compare the DVHs for IMRT plans obtained with five beams placed at the best and worst individual performance directions as revealed by the BEVD score function. The worst beam directions were identified to be 25°, 80°, 170°, 280°, and 340°. As seen from Fig. 6, the PTV dose was deteriorated uniformly when the worst beam configuration was used for IMRT planning. Simultaneously, the doses to all the sensitive structures are increased as compared with that with best individual performance beams. The computed EUDs listed in Table 2 also support the above observation. This comparison study underscores the importance of beam orientation selection and provides additional evidence that the BEVD is a useful tool to facilitate IMRT planning.

DISCUSSION

Increasing the effectiveness of radiation therapy depends on improving our understanding of dose–volume factors affecting tolerance and local control and developing techniques for incorporating the volumetric data into IMRT treatment plan optimization. To rank an incident beam direction in the presence of volumetric organs, a conceptually simple approach is to sample the tolerance dose distributions of the structures according to the dose–volume tolerance requirement and to compute the score for each sampling (46, 47). With multiple sampling, we can obtain a "band" of scores rather than a single score curve. The best achievable target dose corresponds to the top of the band, which can be used for guiding the beam placement. While conceptually simple, sampling requires a large amount of computation and is difficult to implement. In this work we proposed a more practical and efficient approach to solve the problem. A more natural and computationally efficient method to consider the dose–volume effect in the angular ranking is to use the EUD model, which was first developed by Niemierko to characterize the dose–volume effect. A major feature of the EUD formalism is that it allows us to use a single parameter to capture the dosimetric or volumetric characteristics of an organ. Several studies (34, 36, 48–50) have shown the utility of the formalism in plan optimization.

One of the technical hurdles in calculating the EUD-based angular score ranking is that, for a given set of EUD tolerances, there are multiple intensity profiles and the one that yields the highest score function (2) needs to be found using a constrained optimization algorithm. We found that the CFSQP implemented in this work was capable of quickly converging to a stable solution for the cases studied. As described in the previous section, the IMRT plans obtained with beam configurations chosen based on the EUD-based score compares favorably with IMRT plans obtained using other approaches, including equispaced beams and beams obtained under the guidance of a non-EUD-based score function. We note that, because the non-EUD-based score function represents a special case of the general EUD-based function, there may be cases where the EUD-based score yields the same IMRT plan as the non-EUD-based

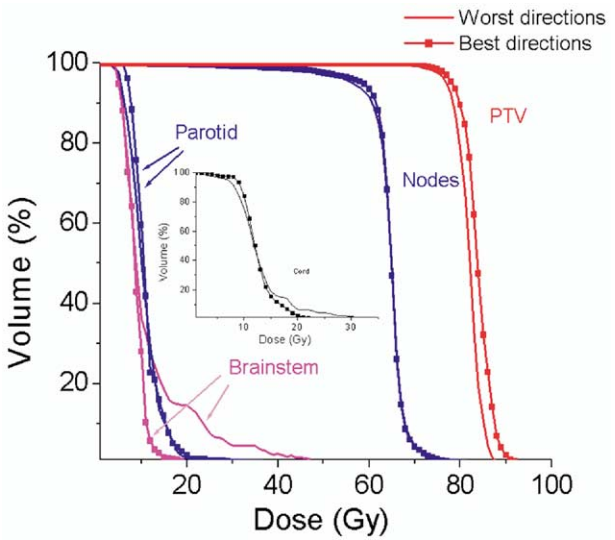


Fig. 6. Comparison of dose–volume histograms of intensity-modulated radiation therapy plans with the best and worst individual performance directions as identified by the angular ranking function.

score (when all the organs involved are serial). However, it is important to keep in mind that the EUD-based function is a more general approach and is needed when volumetric organs are involved.

Generally speaking, the optimal beam configuration needs to balance the single beam angular ranking and the beam interplay. The former favors placing beams at the peak positions of the BEVD score function, whereas the latter prefers a beam configuration with the beams away from each other. Although the two factors are not always conflicting with each other, the beams other than the one placed at the highest peak often need to be moved away for BEVD score peaks to separate the beams. We admit that there exists a certain uncertainty in the placements of these beams. However, the uncertainty in this process is far less when compared with the situation where no BEVD information is available. In other words, the introduction of angular score function significantly reduces the size of the beam configuration search space and enables us to quickly obtain a clinically sensible plan. The results of this study demonstrate the effectiveness of a newly developed algorithm that incorporates the volumetric information of the involved organs. Further investigation into this technique should be performed to investigate other types of clinical cases and to develop a full beam orientation optimization method to take advantage of the *a priori* knowledge about the angular search space as derived from the BEVD score evaluation (21).

CONCLUSIONS

An important issue in IMRT beam orientation ranking is how to incorporate the existing clinical endpoint data into the calculation. In this work we have described an effective method for incorporating the dose–volume effects into the evaluation of single beam ranking. In the proposed technique, the volumetric effect of an organ is characterized by the EUD and the ranking of a beam direction is assessed by the dose delivered to the PTV without exceeding the EUD tolerances or dose–volume limits of the sensitive structures. Because of the degenerate nature of the volumetric tolerance, an optimization algorithm is needed to find the maximum beam intensity profile from the pool of candidate beams that produce the same EUD. The EUD-based angular score allows a fast evaluation of all accessible beam angles based on the *a priori* volumetric and geometric information of the patient. Dosimetric or geometrical scoring schemes proposed previously by different investigators represent special cases of the general formalism here. Application of the reported technique to two clinical IMRT cases indicates that the method is capable of identifying potentially good and bad directions and is useful to guide the beam placement process. The information can also be used as prior knowledge to facilitate the full beam orientation optimization process (21). Considering the relative simplicity of the model and the efficiency of the calculation, the proposed technique should be a useful addition to the armamentarium of IMRT planning.

REFERENCES

1. Stein J, Mohan R, Wang XH, *et al.* Number and orientations of beams in intensity-modulated radiation treatments. *Med Phys* 1997;24:149–160.
2. Pugachev AB, Boyer AL, Xing L. Beam orientation optimization in intensity-modulated radiation treatment planning. *Med Phys* 2000;27:1238–1245.
3. Das S, Cullip T, Tracton G, *et al.* Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. *Int J Radiat Oncol Biol Phys* 2003;55:215–224.
4. Djajaputra D, Wu Q, Wu Y, Mohan R. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Phys Med Biol* 2003;48:3191–3212.
5. Meedt G, Alber M, Nüsslin F. Non-coplanar beam direction optimization for intensity-modulated radiotherapy. *Phys Med Biol* 2003;48:2999–3019.
6. Schreibmann E, Lahanas M, Xing L, Baltas D. Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Phys Med Biol* 2004;49:747–770.
7. Gaede S, Wong E, Rasmussen H. An algorithm for systematic selection of beam directions for IMRT. *Med Phys* 2004; 31:376–388.
8. Li Y, Yao J, Yao D. Automatic beam angle selection in IMRT planning using genetic algorithm. *Phys Med Biol* 2004;49: 1915–1932.
9. Nutting CM, Rowbottom CG, Cosgrove VP, *et al.* Optimisation of radiotherapy for carcinoma of the parotid gland: A comparison of conventional, three-dimensional conformal, and intensity-modulated techniques. *Radiother Oncol* 2001;60:163–172.
10. Earl MA, Shepard DM, Naqvi S, Li XA, Yu CX. Inverse planning for intensity-modulated arc therapy using direct aperture optimization. *Phys Med Biol* 2003;48:1057–1089.
11. Stavrev P, Hristov D, Warkentin B, Sham E, Stavreva N, Fallone BG. Inverse treatment planning by physically constrained minimization of a biological objective function. *Med Phys* 2003;30:2948–2958.
12. Wang C, Dai J, Hu Y. Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Phys Med Biol* 2003;48:4065–4076.
13. Gokhale P, Hussein EM, Kulkarni N. Determination of beam orientation in radiotherapy planning. *Med Phys* 1994;21:393–400.
14. Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: Some theoretical considerations. *Med Phys Biol* 1993;38:291–304.
15. Hosseini-Ashrafi ME, Bagherebadian H, Yahaqi E. Pre-optimization of radiotherapy treatment planning: An artificial neural network classification aided technique. *Phys Med Biol* 1999;44:1513–1528.
16. Haas OC, Burnham KJ, Mills JA. Optimization of beam orientation in radiotherapy using planar geometry. *Phys Med Biol* 1998;43:2179–2193.
17. Rowbottom CG, Webb S, Oldham M. Improvements in prostate radiotherapy from the customization of beam directions. *Med Phys* 1998;25:1171–1179.
18. Rowbottom CG, Oldham M, Webb S. Constrained customi-

zation of non-coplanar beam orientations in radiotherapy of brain tumors. *Phys Med Biol* 1999;44:383–399.
19. Woudstra E, Storchi PRM. Constrained treatment planning using sequential beam selection. *Phys Med Biol* 2000;45:305–328.
20. Pugachev A, Xing L. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 2001;51:1361–1370.
21. Pugachev A, Xing L. Incorporating prior knowledge into beam orientation optimization. *Int J Radiat Oncol Biol Phys* 2002;54:1565–1574.
22. Schreibmann E, Lahanas M, Uricchio R, *et al.* A geometry based optimisation algorithm for conformal external beam orientation. *Phys Med Biol* 2003;48:1825–1841.
23. Lee KE, Fox T, Crocke I. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Ann Op Res* 2003;119:165–181.
24. Chen GT, Spelbring DR, Pelizzari CA, *et al.* The use of beam's eye view volumetrics in the selection of non-coplanar radiation portals. *Int J Radiat Oncol Biol Phys* 1992;23:153–163.
25. Das SK, Marks LB. Selection of coplanar or noncoplanar beams using three-dimensional optimization based on maximum beam separation and minimized nontarget irradiation. *Int J Radiat Oncol Biol Phys* 1997;38:643–655.
26. Cho BJ, Roa WH, Robinson D, Murry B. The development of target-eye-view maps for selection of coplanar or noncoplanar beams in conformal radiotherapy treatment planning. *Med Phys* 1999;26:2367–2372.
27. Myrianthopoulos LC, Chen GT, Vijayakumar S, Halpern HJ, Spelbring DR, Pelizzari CA. Beam's eye view volumetrics: An aid in rapid treatment plan development and evaluation. *Int J Radiat Oncol Biol Phys* 1992;23:367–375.
28. McShan DL, Kessler ML, Fraass BA. Advanced interactive planning techniques for conformal therapy: High level beam descriptions and volumetric mapping techniques. *Int J Radiat Oncol Biol Phys* 1995;33:1061–1072.
29. Niemierko A. A generalized concept of equivalent uniform dose (EUD). *Med Phys* 1999;26:1100.
30. Niemierko A. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Med Phys* 1997;1:103–110.
31. Kavanagh BD, Timmerman RD, Benedict SH, *et al.* How should we describe the radiobiologic effect of extracranial stereotactic radiosurgery: Equivalent uniform dose or tumor control probability? *Med Phys* 2003;30:321–324.
32. Ebert MA. Viability of the EUD and TCP concepts as reliable dose indicators. *Phys Med Biol* 2000;45:441–457.
33. Wang JZ, Li XA. Evaluation of external beam radiotherapy and brachytherapy for localized prostate cancer using equivalent uniform dose. *Med Phys* 2003;30:34–40.
34. Yang Y, Xing L. Clinical knowledge-based inverse treatment planning. *Phys Med Biol* 2004;49:5101–5117.
35. Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models. *Phys Med Biol* 2004;49:1991–2013.
36. Thieke C, Bortfeld T, Niemenko A, Nill S. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. *Med Phys* 2003;9:2332–2339.
37. Emami B, Lyman J, Brown A, *et al.* Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys* 1991;21:109–122.
38. Crooks SM, Xing L. Application of constrained least-squares techniques to IMRT treatment planning. *Int J Radiat Oncol Biol Phys* 2002;4:1217–1224.
39. Lawrence CT, Tits AL. Computationally efficient feasible sequential quadratic programming algorithm. *SIAM J Optimiza* 2001;11:1092–1118.
40. Panier E, Tits E. On combining feasibility, descent, and superlinear convergence in inequality constrained optimization. *Math Prog* 1993;59:261–276.
41. Deb K, Goel T. Controlled elitist non-dominated sorting genetic algorithms for better convergence. *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science* 2001; 1993:67–81.
42. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program* 1989;45:503–528.
43. Lahanas M, Schreibmann E, Baltas D. Constrained free gradient-based optimization algorithms for multiobjective inverse planning in intensity modulated radiotherapy. *Phys Med Biol* 2003;48:2843–2871.
44. Xing L, Li JG, Donaldson S, Le QT, Boyer AL. Optimization of importance factors in inverse planning. *Phys Med Biol* 1999;44:2525–2536.
45. Song W, Dunscombe P. EUD-based margin selection in the presence of set-up uncertainties. *Med Phys* 2004;31:849–859.
46. Langer M, Brown R, Urie M, Leong J, Stracher M, Shapiro J. Large scale optimization of beam weights under dose–volume restrictions. *Int J Radiat Oncol Biol Phys* 1990;18:887–893.
47. Starkschall G, Pollack A, Stevens CW. Treatment planning using a dose–volume feasibility search algorithm. *Int J Radiat Oncol Biol Phys* 2001;49:1419–1427.
48. Wu Q, Mohan R, Niemierko A, Schmidt-Ullrich R. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys* 2001; 52:224–235.
49. Thomas E, Chapet O, Kessler ML, Lawrence TS, Haken RKT. Benefit of using biologic parameters (EUD and NTCP) in IMRT optimization for treatment of intrahepatic tumors. *Int J Radiat Oncol Biol Phys* 2005;62:571–578.
50. Lian J, Xing L. Incorporating model parameter uncertainty into inverse treatment planning. *Med Phys* 2004;31:2711–2720.

MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **3**, Number **2**, April **2006**

http://www.mbejournal.org/

pp. **389–418**

MULTISCALE IMAGE REGISTRATION

Dana Paquin

Department of Mathematics, Stanford University
Stanford, CA 94305-2125

Doron Levy

Department of Mathematics, Stanford University
Stanford, CA 94305-2125

Eduard Schreibmann

Department of Radiation Oncology, Stanford University
Stanford, CA 94305-5947

Lei Xing

Department of Radiation Oncology, Stanford University
Stanford, CA 94305-5947

(Communicated by Yang Kuang)

Abstract. A multiscale image registration technique is presented for the registration of medical images that contain significant levels of noise. An overview of the medical image registration problem is presented, and various registration techniques are discussed. Experiments using mean squares, normalized correlation, and mutual information optimal linear registration are presented that determine the noise levels at which registration using these techniques fails. Further experiments in which classical denoising algorithms are applied prior to registration are presented, and it is shown that registration fails in this case for significantly high levels of noise, as well. The hierarchical multiscale image decomposition of E. Tadmor, S. Nezzar, and L. Vese [20] is presented, and accurate registration of noisy images is achieved by obtaining a hierarchical multiscale decomposition of the images and registering the resulting components. This approach enables successful registration of images that contain noise levels well beyond the level at which ordinary optimal linear registration fails. Image registration experiments demonstrate the accuracy and efficiency of the multiscale registration technique, and for all noise levels, the multiscale technique is as accurate as or more accurate than ordinary registration techniques.

1. **Introduction.** Often in image processing, images must be spatially aligned to allow practitioners to perform quantitative analyses of the images. The process of aligning images taken, for example, at different times, from different imaging devices, or from different perspectives, is called image registration. More precisely, image registration is the process of determining the optimal spatial transformation that maps one image to another. Typically, two images are taken as input, and the registration process is then the optimization problem which determines the

2000 *Mathematics Subject Classification.* Primary: 68U10; Secondary: 92C55, 62P10, 94A08.

Key words and phrases. image registration, multiscale analysis, noise, CT, MRI, mutual information.

geometric mapping that brings one image into spatial alignment with the other image. In practice, the particular type of transformation as well as the notion of optimal will depend on the specific application.

Examples of applications in which image registration is particularly important include astro- and geophysics, computer vision, remote sensing, and medicine. In this paper, we will focus on medical image registration. Image registration plays an important role in the analysis of medical images. For example, images taken from different sensors often contain complementary information. By bringing the two images into alignment so that anatomical features of one modality can be detected in the other modality, the information from the different modalities can be combined. In neurosurgery, for example, tumors are typically identified and diagnosed using magnetic resonance images (MRI), but stereotaxy technology (the use of surgical instruments to reach specified points) generally uses computed tomography (CT) images. Registration of these modalities allows the transfer of coordinates of tumors from the MRI images to the CT images. See [14] for a discussion of the applications of multimodality imaging to problems in neurosurgery. As another example, medical image data acquired prior to diagnosis can be compared with data acquired during or after treatment to determine the effectiveness of the treatment. To compare images taken at different times, however, the images must first be brought into spatial alignment so that actual differences in the data can be distinguished from differences that result from the image acquisition process.

In the context of medical imaging, the goal of the registration process is to remove artificial differences in the images introduced by patient movement, differences in imaging devices, etc., but at the same time, to retain real differences due to actual variations of the objects. Medical images, however, often contain significant levels of noise due to instrumentation imperfections, data acquisition techniques, image reconstruction methods, transmission and/or compression errors, and other factors. Although numerous successful image registration techniques have been published, we will see that ordinary image registration algorithms can fail to produce meaningful results when one or both of the images to be registered contains significant levels of noise.

Since noise is generally present in digital images, image denoising is a fundamental problem in image processing. Indeed, many approaches to image denoising have been presented. Thus a simple solution to the problem of image registration in the presence of noise would be to first apply a denoising algorithm to the noisy image(s), and then use existing image registration techniques to register the denoised images. However, common denoising algorithms, most notably spatial filtering algorithms, have the disadvantage that while they are successful in removing noise, they often remove edges as well. Additionally, most denoising procedures require a priori knowledge of the noise level, variance, and/or model, information not typically known in practice. For these and other reasons, we will demonstrate that ordinary image registration of noisy images fails to produce acceptable results even when classical denoising algorithms are applied to the noisy images prior to registration (for significantly high levels of noise). Moreover, even more advanced denoising techniques such as anisotropic diffusion (which was designed to remove noise while preserving edges) will be shown to fail to register noisy images. Thus, we seek a technique that enables successful image registration when one or both of the images to be registered is noisy.

Generally, we would like to consider an image f consisting of *coarse* and *fine* scales. The general shape and main features of an image are considered the coarse scales, and details and textures, such as noise, are the fine scales of the image. Separating the coarse and fine scales of an image, therefore, is an effective tool in denoising. Indeed, several denoising algorithms have been proposed using separation of the coarse and fine scales of an image, most notably [19], [18], [11], and [20]. The method presented in [20] presents a multiscale technique in which an image f is decomposed in a hierarchical expansion $f \sim \Sigma_j u_j$, where the u_j (called the components of f relative to the decomposition) resolve edges of f with increasing scales. More precisely, for small k, the sum $\Sigma_j^k u_j$ is a coarse representation of the image f, and as k increases, the sum captures more and more detail (and hence, noise) of the image.

In this paper, we present a multiscale image registration technique based on the multiscale decomposition of [20] that is particularly effective when one or both of the images to be registered contains significant levels of noise. Since the hierarchical expansion $f \sim \Sigma_j u_j$ decomposes the image f into components which contain increasingly fine scales, we expect a component-wise registration algorithm to produce accurate results for noisy images. That is, given a noisy image f, for small values of k, the component $\Sigma_j^k u_j$ retains the general shape of the image f while removing the details and noise of the image. Thus, if we wish to register the image f with another image, say g, we expect that registration of the components $\Sigma_j^k u_j$ with g will provide an accurate estimation of the actual transformation that brings the two images into spatial alignment with one another, for sufficiently small values of k. Similarly, if both f and g are noisy, we expect decomposing both images and performing component-wise registrations should accurately estimate the optimal transformation. We will demonstrate that multiscale image registration enables successful image registration for images that contain levels of noise that are significantly higher than the levels at which ordinary registration fails.

This paper is organized in the following way. In Section 2, we discuss the image registration problem and review standard image registration techniques. In Section 3, we present the problem of image registration in the presence of noise, and illustrate the failure of current techniques when one or both of the images to be registered contains high levels of noise. In Section 4, we briefly discuss classical and modern denoising techniques, and illustrate the failure of ordinary image registration of noisy images even when the images are denoised prior to registration. In Section 5, we review the multiscale image decomposition of [20], and illustrate the results of the hierarchical multiscale decomposition obtained upon applying the algorithm to noisy images. In Section 6, we present image registration techniques based upon the multiscale decomposition described in Section 5, and in Section 7, we present the results of our multiscale image registration experiments.

2. **The registration problem.**

Given a *fixed* and *moving* image, the registration problem is the process of finding an *optimal transformation* that brings the moving image into spatial alignment with the fixed image. While this problem is easy to state, it is difficult to solve. The main source of difficulty is that the problem is ill-posed, which means, for example, that the problem may not have a unique solution. Additionally, the notion of optimality may vary for each application: for example, some applications may require consideration only of rigid transformations, while other applications require non-rigid transformations. Finally, computation time

and data storage constraints place limitations on the complexity of models that can be used for describing the problem. This following discussion of image registration follows the presentation in [12].

2.1. **The mathematical setting.** A two-dimensional gray-scale image f is a mapping which assigns to every point $x \in \Omega \subset \mathbb{R}^2$ a gray value $f(x)$ (called the intensity value of the image at the point x). We will consider images as elements of the space $L^2(\mathbb{R}^2)$. Color images can be defined, for example, in terms of vector-valued functions $\mathbf{f} = (f_1, f_2, f_3)$ representing the RGB-color scales. For the medical imaging applications that we are interested in, images are in fact given in terms of discrete data, and the function f must be obtained by interpolation. We will not discuss this construction here, but we will assume that an interpolation method has been chosen.

Image registration is necessary, for example, for two images of the same object that are not spatially aligned. This occurs when the images are taken at different times, from different perspectives, or from different imaging devices. The basic input data to the registration process are two images: one is defined as the fixed image $f(x)$ and the other as the moving image $m(x)$. The goal is then to find a transformation ϕ such that the fixed image $f(x)$ and the deformed moving image $m_\phi(x) := m(\phi(x))$ are similar. To solve this problem in a mathematical way, the term *similar* needs to be defined appropriately. For example, if the images to be registered are taken from different devices, there may not be a correspondence between the intensities $f(x)$ and $m_\phi(x)$ for an optimal ϕ. Additionally, we may consider measures of similarity between the images which are not related to the intensities. Thus, the registration problem necessarily involves a discussion of the distance measures, or metrics, used to compare images. The general problem can then be stated as follows.

Given a distance measure $D : (L^2(\mathbb{R}^2))^2 \to \mathbb{R}$ and two images $f(x), m(x) \in L^2(\mathbb{R}^2)$, the solution ϕ of the registration problem is given by the following minimization problem:

$$\phi = \underset{\psi:\mathbb{R}^2\to\mathbb{R}^2}{\operatorname{argmin}} D(f, m_\psi). \tag{1}$$

In many applications, the set of allowable transformations to be considered in the minimization problem (1) is restricted to a strict subset of the set of all maps $\psi : \mathbb{R}^2 \to \mathbb{R}^2$. For example, we may require the transformation ϕ to be smooth, or we may impose specific parametric requirements, such as requiring the transformation to be rigid, affine, polynomial, etc.

2.2. **Landmark-based registration.** Landmark-based registration is an image registration technique based on a finite set of image features. The problem is to determine the transformation such that for a finite set of features, any feature of the moving image is mapped onto the corresponding features of the fixed image. More precisely, let $F(f, j)$ and $F(m, j)$, $j = 1, \ldots, m$ be given features of the fixed and moving images, respectively. The solution ϕ of the registration problem is then a map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$F(f, j) = \phi(F(m, j)), \ j = 1, \ldots m. \tag{2}$$

For a more general notion of landmark-based registration, we define the following distance measure:

$$D^{LM}(\phi) := \sum_{j=1}^{m} ||F(f,j) - \phi(F(m,j)||_l^2, \tag{3}$$

where $||\cdot||_l$ denotes a norm on the landmark, or feature, space. For example, if the features are locations of points, then $||\cdot||_l = ||\cdot||_{\mathbb{R}^2}$. We can then restate (2) as the minimization problem in which the solution $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ of the registration problem is given by:

$$\phi = \underset{\psi:\mathbb{R}^2\to\mathbb{R}^2}{\operatorname{argmin}} D^{LM}(\psi). \tag{4}$$

To solve this minimization problem, the transformation either is chosen to be an element of an n-dimensional space spanned, for example, by polynomials, splines, or wavelets, or it is required to be smooth in some sense. In the first case, the features to be mapped are the locations of a number of user-supplied landmarks. Let χ_k, $k = 1, \ldots n$ be the basis functions of the space. Then the minimization of

$$D^{LM}(\phi) := \sum_{j=1}^{m} ||F(f,j) - \phi(F(m,j)||_l^2$$

can be obtained upon expanding $\phi = (\phi_1, \phi_2)$ in terms of the basis functions χ_k and solving the resulting least squares problems.

In the case in which we require the transformation ϕ to be smooth, we introduce a functional which imposes smoothness restrictions on the transformation. That is, we look for a transformation ϕ which interpolates the features $F(f,j)$ and $F(m,j)$, and which is smooth in some sense. Such a transformation is called a minimal norm solution, and it turns out (see [8]) that the solution can be expressed in terms of radial basis functions.

Landmark-based registration is simple to implement, and the numerical solution requires only the solution of a linear system of equations. However, the main drawback of the landmark-based approach is that the registration process depends on the location of the landmarks. As the detection and mathematical characterization of landmarks (for example, anatomical landmarks in medical images) is not fully automated, the landmarks must be user-supplied, and this can be a time-consuming and difficult process, even for a medical expert; see, for example, [17]. Additionally, landmark-based registration does not always results in a physically meaningful registration. See [12, p. 44], for a simple example of a situation in which landmark-based registration fails to produce meaningful results.

2.3. **Principal-axes-based registration.** Principal-axes image registration is based on the idea of landmark-based registration, but it uses features that can be automatically detected. These features are constructed as follows. For an image $f : \mathbb{R}^2 \to \mathbb{R}$, and a function $g : \mathbb{R}^2 \to \mathbb{R}$, we define the expectation value of g with respect to f by

$$\mathbb{E}_f(g) := \frac{\int_{\mathbb{R}^2} g(x) f(x)\, dx}{\int_{\mathbb{R}^2} g(x)\, dx}. \tag{5}$$

If $u : \mathbb{R}^2 \to \mathbb{R}^{m\times n}$, we set $\mathbb{E}_f(u) := \mathbb{E}_f[u_{j,k}] \in \mathbb{R}^{m\times n}$. The center of an image f is defined by

$$C_f := \mathbb{E}_f[x] \in \mathbb{R}^2, \tag{6}$$

and the covariance by

$$\mathrm{Cov}_f := \mathbb{E}_f[(x - C_f)(x - C_f)^T] \in \mathbb{R}^{2\times 2}. \tag{7}$$

Given fixed and moving images, $f(x)$ and $m(x)$, the centers c_f and c_m and eigendecompositions of the covariance matrices Cov_f and Cov_m are used as the features F_i, and the registration problem is to compute $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that $F_i(m(\phi)) = F_i(f)$ for the features F_i.

This method is described in detail in [1]. The principal-axes method of image registration has the advantages that it is computationally fast and simple and requires few registration parameters, but has the disadvantages that it is not suitable for images of multiple modalities and that the solutions may be ambiguous. In particular, the principal-axes-based method cannot distinguish between images with the same center and covariance, even though images with very different structure and orientation may have the same center and/or covariance.

2.4. **Optimal parametric registration.** An alternative approach to registration is to use methods that are based on the minimization (or maximization) of some distance measure, or metric, D. The transformation ϕ is restricted to some parameterized space, and the registration can be obtained by minimizing (or maximizing) the distance D over the parameterized space. In particular, we will discuss metrics based on intensity, correlation, and mutual information. Given a metric D, a fixed image f, and a moving image m, optimal parametric registration is the problem of finding a transformation ϕ in some pre-specified parameterizable space such that $D(f, m(\phi))$ is minimized (or maximized in certain cases). Examples of commonly used parameterizable spaces in image registration are polynomial and spline spaces. We will primarily be interested in rigid and affine linear transformations. An affine linear map is a map of the form $\phi(x) = Ax + b$, $A \in \mathbb{R}^{2\times 2}$, $\det A > 0$, $b \in \mathbb{R}^2$. Such a map allows rotations, translations, scales, and shears of the coordinates. A translation (or rigid) transformation is a special case of an affine transformation which allows only rotations and translations of the coordinates, and in this case, the matrix A is required to be orthogonal with determinant 1. Optimal parametric registration is probably the most commonly used image registration technique.

To minimize $D(f, m(\phi))$, we must choose an optimization technique. That is, an optimal parametric registration technique is described by a metric to be minimized (or maximized) and an optimizer which controls the minimization (or maximization). The implementation of the registration algorithm works in the following way: at each iteration, the distance D between the two images is computed. An affine transformation is then applied to the moving image, and the distance between the images is recomputed. In theory, this process continues until the distance is minimized (or maximized), though in practice there is some stopping criterion.

At each stage, the optimizer determines the parameters of the transformation that will be applied to the moving image. Examples of commonly used optimizers include gradient descent and regular step gradient descent. Gradient descent optimization advances the parameters of the transformation in the direction of the gradient, where the step size is governed by a user-specified learning rate. Regular step gradient descent optimization advances the parameters of the transformation in the direction of the gradient where a bipartition scheme is used to compute the step size.

2.4.1. *The mean squares metric.* The mean squares metric computes the mean-squared pixel-wise difference in intensity between two images f and m:

$$MS(f,m) := \frac{1}{N}\sum_{i=1}^{N}(f_i - m_i)^2, \tag{8}$$

where N is the total number of pixels considered, f_i is the i^{th} pixel of image f, and m_i is the i^{th} pixel of image m. Note that the optimum value of the mean squares metric is 0, and poor matches between the images f and m result in large values of $MS(f,m)$. This metric has the advantage that it is computationally simple. It is based on the assumption that pixels in one image should have the same intensity as (spatially) corresponding pixels in the second image. Thus, the mean squares metric is restricted in practice to images of the same modality.

2.4.2. *The normalized correlation metric.* The normalized correlation metric computes pixel-wise cross-correlation and normalizes it by the square root of the auto-correlation function:

$$NC(f,m) := -\frac{\sum_{i=1}^{N}(f_i \cdot m_i)}{\sqrt{\sum_{i=1}^{N} f_i^2 \cdot \sum_{i=1}^{N} m_i^2}}, \tag{9}$$

where N, f_i, and m_i are as defined for the mean squares metric. The negative sign in (9) causes the optimum value of the metric to occur when the minimum is reached. Thus the optimal value of the normalized correlation metric is -1. As with the mean squares metric, the normalized correlation metric is restricted to images of the same modality.

2.4.3. *The mutual information metric.* Mutual information is an information-theoretic approach to image registration that was proposed independently by Viola and Wells [22] and Collignon et al. [4] in 1995. The idea is that mutual information computes the amount of information that one random variable (here, image intensity) gives about another random variable (here, intensity values of another image). More precisely, given a fixed image $f(x)$ and a moving image $m(x)$, we wish to compute the transformation ϕ which *maximizes* the mutual information; i.e.,

$$\phi = \arg\max_{\psi} I(f(x), m(\phi(x))). \tag{10}$$

Maximization of the mutual information criterion assumes that the statistical dependence between corresponding image intensity values is maximized when the images are geometrically aligned.

The mutual information $I(f(x), m(\phi(x))$ is defined in terms of entropy, where we consider x as a random variable over coordinate locations in the coordinate system of the fixed image. Let $h(\cdot)$ denote the entropy of a random variable: $h(x) := -\int p(x)\ln p(x)\,dx$, where $p(x)$ is the probability density function of the random variable x. Note that it is not clear how to construct $p(x)$; we will discuss methods for estimating the probability densities. The joint entropy of two random variables x and y is given by $h(x,y) = -\int p(x,y)\ln p(x,y)\,dx\,dx$, where $p(x,y)$ is the joint probability density function of the random variables x and y. Entropy can be considered as a measure of the uncertainty or complexity of a random variable.

If x and y are independent, then $p(x,y) = p(x)p(y)$, so $h(x,y) = h(x) + h(y)$. However, if there is any dependency (as would be the case if x and y are intensity values of images of the same object), then $h(x,y) < h(x) + h(y)$. The difference is defined to be *mutual information*:

$$I(f(x), m(\phi(x))) = h(f(x)) + h(m(\phi(x)) - h(f(x), m(\phi(x))). \tag{11}$$

The terms in (11) can be interpreted in the following way. The first term, $h(f(x))$, is the entropy of the fixed image and is independent of the transformation ϕ. The second term, $h(m(\phi(x)))$, is the entropy of $m(\phi(x))$, so maximization of mutual information encourages transformations ϕ for which $m(\phi(x)$ has a high level of complexity or uncertainty. The third term $-h(f(x), m(\phi(x)))$ is the negative joint entropy of $f(x)$ and $m(\phi(x))$, so maximization of mutual information is related to minimization of the joint entropy of $f(x)$ and $m(\phi(x))$. A detailed overview of mutual information based registration can be found in [16].

Mutual information has the following properties. Let $u(x)$ and $v(x)$ denote two images.

1. $I(u(x), v(x)) = I(v(x), u(x))$. Mutual information is symmetric.
2. $I(u(x), u(x)) = h(u(x))$. The information an image contains about itself is equal to the entropy of the image.
3. $I(u(x), v(x)) \leq h(u(x))$ and $I(u(x), v(x)) \leq h(v(x))$. The information that the images contain about each other can not be greater than the information contained in the individual images.
4. $I(u(x), v(x)) \geq 0$.
5. $I(u(x), v(x)) = 0$ if and only if $u(x)$ and $v(x)$ are independent. If the images $u(x)$ and $v(x)$ are independent, no information about one image is gained when the other image is known.

The entropies in equation (11) are defined in terms of integrals over the probability densities associated with the images $f(x)$ and $m(x)$. However, in a typical medical image registration problem, the probability densities are not directly accessible, and thus must be estimated from the image data. Parzen windowing, described in [5] and used in [22], is a common technique for density estimation. In this method, continuous density functions are constructed by a super-position of kernel functions $K(\cdot)$ centered at the elements of a sample of intensities taken from the image. The estimation of the probability density $p(z)$ is thus given by

$$p(x) \cong P^{\star}(z) = \frac{1}{N_S} \sum_{z_j \in S} K(z - z_j), \tag{12}$$

where N_S is the number of spatial samples in S and K is an appropriately chosen kernel function. The kernel function K must be smooth, symmetric, have zero mean, and unit mass. Examples of suitable candidates for K include the Gaussian density and the Cauchy density. In [22], Viola and Wells use a Gaussian density function with standard deviation σ to estimate the probability density functions. The optimal value of σ depends on the particular images to be registered.

Upon estimating the probability densities using the Parzen windowing technique, the entropy integral $h(z) = -\int p(z) \ln(p(z))\, dz$ must be evaluated, for example, by using a sample mean:

$$h(z) \cong -\frac{1}{N_R} \sum_{z_j \in R} \ln(P^\star(z_j)), \tag{13}$$

where R is a second sample of intensities taken from the image. That is, two separate intensity samples S and R are taken from the image. The first is used to estimate the probability density, and the second is used to approximate the entropy.

The main advantage of the mutual information measure is that was shown to be generally applicable for multi-modality registration, whereas intensity-based measures are typically not applicable for multimodality registration. Mutual information registration has been successfully used for a number of complex applications. Most notably, mutual information has been shown to be highly accurate for MRI-CT registration; see, for example, [9], [15], and [21].

2.5. **Non-parametric image registration.** All of the image registration techniques that we have discussed so far have been based on certain parameters. For example, either the transformation ϕ can be expanded in terms of basis functions that span a specified finite-dimensional space, or the registration is controlled by a specified set of external features. Non-parametric techniques do not restrict the transformation to a parameterizable set. Given two images, a fixed image $f(x)$ and a moving image $m(x)$, we seek a transformation ϕ such that $m(\phi(x))$ is similar to $f(x)$ in a certain sense. Upon defining a suitable distance measure D, the registration problem is then to minimize the distance between $m(\phi(x))$ and $f(x)$. However, a direct minimization is often not possible in the non-parametric case. The problem is ill-posed: small changes in the input data may lead to large changes in the output. Additionally, the solution is not unique. Given these constraints, a stable numerical implementation is often impossible. To circumvent these problems, a regularizing, or smoothing, term S is introduced, and the registration problem becomes the minimization of the distance between $m(\phi(x))$ and $f(x)$ plus a smoothing term $S(\phi)$. That is, the registration is based on a regularized minimization of the distance between the images.

In the discussion of non-parametric image registration, the transformation $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is split into the trivial identity part and the deformation or displacement part u; i.e.,

$$\phi(x) = x - u(x). \tag{14}$$

Upon decomposing ϕ in this way, we have $m(\phi(x)) = m(x - u(x)) := m_u(x)$. Given a distance D and a smoother S, the elastic registration problem is then the minimization of $D(f(x), m_u(x)) + \alpha S(u)$, where $\alpha \in \mathbb{R}$ is a positive regularizing parameter.

The choice of smoother S typically depends on the particular application. Examples of non-parametric image registration techniques include elastic registration [3], fluid registration [2], and diffusion registration [6]. Elastic registration uses linear elasticity theory to model the deformation of an elastic body. In this case, the regularizing term $S(u)$ is the linearized elastic potential of the displacement u. In fluid registration, the regularization is based on the linearized elastic potential of the time derivative of u. Finally, diffusion registration uses a regularization that is based on spatial derivatives of the displacement.

Remark. *In this section, we presented a brief overview of the major image registration techniques currently used in image registration. In practice, the best*

registration method for a given set of images will depend on the particular features of the images themselves. However, numerous studies comparing the accuracy and performance of different image registration techniques for various applications have been presented. The most extensive of these is [24]*, which originally consisted of a comparison of 16 methods but has since been substantially expanded.*

3. **Registration in the presence of noise.** In this section, we study the effect of noise on image registration, and we determine the approximate noise level at which registration fails. This study is conducted on the brain proton density slice images shown in Figure 1 below. The image on the right is the result of translating the image on the left by 13 mm to the right in the X-direction and 17 mm downward in the Y-direction. Let I denote the original image, and let T denote the translated image.

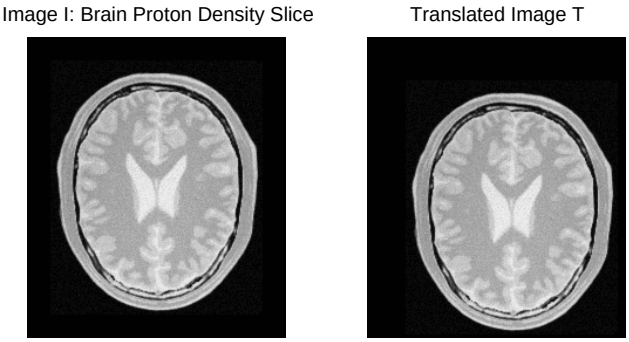


FIGURE 1. Original image I and translated image T.

Initially, we will consider the registration problem in which one of the images (here, the fixed image) is noisy. We will add increasing levels of noise to the image I and register the non-noisy translated image T with the noisy images. Our goal is to determine the approximate noise levels at which various image registration techniques fail, and to develop an algorithm that will enable registration beyond these levels. Since we know the exact transformation that brings T into spatial alignment with I, we can effectively evaluate and compare the accuracy of various registration techniques. We will demonstrate that our multiscale registration technique enables accurate registration of the translated image T with images that contain significant levels of noise. Eventually, we will also apply our techniques to the case in which both the fixed and the moving images contain high levels of noise. Before we present these results, we discuss the notion of noise in some detail.

REMARK. *In this paper we present the results only for registration experiments using the images I and T shown in Figure 1. We have performed numerous experiments using other images, and we obtained results similar to those presented in this paper. For the sake of brevity, we limit the results presented in this paper to the experiments using the images in Figure 1.*

3.1. **Noise.** Digital images are often degraded by random noise. In imaging, the term noise refers to random fluctuations in intensity values that occur during image capture, transmission, or processing, and that may distort the information given by

the image. Image noise is not part of the ideal signal and may be caused by a wide range of sources, such as detector sensitivity, environmental radiation, transmission errors, discretization effects, etc. Noise is generally classified as either independent noise or noise which is dependent on the image data.

Independent noise can often be described by an additive noise model, in which the observed image $f(x)$ is the sum of the true image $s(x)$ and the noise $n(x)$:

$$f(x) = s(x) + n(x). \tag{15}$$

Within this framework of additive noise, the noise $n(x)$ is commonly modeled by Gaussian noise of mean m and variance v. A multiplicative noise model describes noise that is dependent on the image data. This is often referred to as speckle noise:

$$f(x) = s(x) + s(x)n(x) = s(x)(1 + n(x)). \tag{16}$$

In this case, $n(x)$ is uniformly distributed random noise with mean m and variance v. Impulse noise, or salt-and-pepper noise, is noise that resembles salt and pepper granules randomly distributed over the image. Impulse noise is typically defined by the following model. We let $s(x)$ denote the actual image, and $f(x)$ denote the observed image. Then

$$f(x) = \begin{cases} s(x), & \text{with probability } 1 - \delta, \\ \eta(x), & \text{with probability } \delta, \end{cases} \tag{17}$$

where $\eta(x)$ is an identically distributed, independent random process. With this model, an arbitrary pixel $x \in \Omega \subset \mathbb{R}^2$ is affected by noise with probability δ, and not affected with probability $1 - \delta$. We will refer to δ as the impulse noise density, as adding impulse noise of density δ to an image $f(x)$ affects approximately $\delta \cdot \text{size}(f)$ pixels. The random process $\eta(x)$ is typically such that the corrupted pixels are either set to the maximum value, have single bits flipped over, or are set alternatively to zero or to the maximum value. This last case results in a salt-and-pepper appearance. Note that unaffected pixels always remain unchanged.

In Figure 2, we add additive Gaussian noise of mean 0 and variance 0.2, multiplicative speckle noise of mean 0 and variance 0.2, and impulse noise of density 0.2 to the brain proton density slice image I.

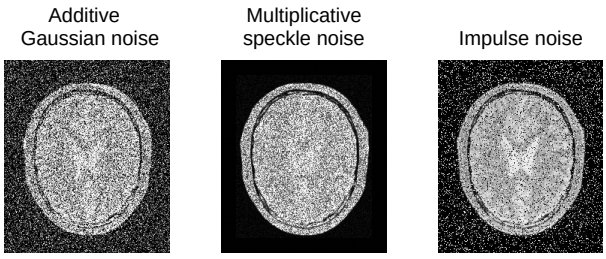


FIGURE 2. An illustration of the addition of various types of noise to the image I.

In this paper, we will study the problem of image registration in the presence of high levels of impulse noise. We will determine the impulse noise density level at which ordinary registration methods fail, and we will present a multiscale registration algorithm that enables accurate registration for noise levels higher than those at which ordinary methods fail. To study the effect of varying noise densities on the registration process, we add impulse noise of increasing densities δ to the brain proton density slice image I, and register the (non-noisy) translated image T with the noisy images. Let I_δ denote the image I with added impulse noise of density δ. In Figure 3, we illustrate the noisy images I_δ for increasing values of δ.

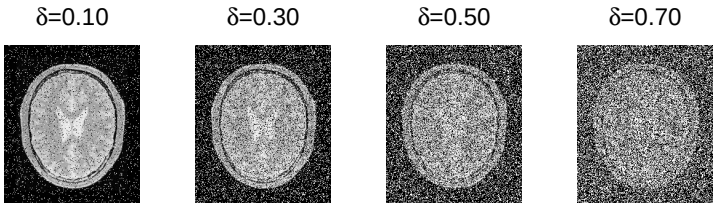


FIGURE 3. An illustration of adding impulse noise of increasing densities δ to the image I.

REMARK. Although in this paper we present the results of image registration experiments using only impulse noise, we have also conducted numerous experiments using other types of noise, including additive Gaussian noise and speckle noise. The results obtained with all other types of noise are similar.

3.2. **Registration results.** For increasing noise densities δ, we register T with I_δ using various registration methods. Recall that the image T is the result of translating the original image I 13 units in X and 17 units in Y, and that I_δ is the result of adding uniform impulse noise of density δ to the image I. Since T is a rigid transformation of I, we will restrict the registration process to linear transformations; i.e., we will consider optimal linear registrations. The optimal transformation ϕ produced by the optimal linear registration process will consist of two parameters, namely X- and Y-translation values. We will let ϕ_X and ϕ_Y denote the X- and Y-translation parameters, respectively, of the optimal transformation ϕ. For comparison purposes, we will perform the optimal linear registration using the mean squares, normalized correlation, and mutual information metrics.

We use the following parameters for the registration algorithms. For the mean squares and normalized correlation registration algorithms, we use the regular step gradient descent optimizer. Due to the stochastic nature of the metric computation in the mutual information algorithm, the regular step gradient descent optimizer does not work well in the case of mutual information. Instead, we use the gradient descent optimizer with a user-specified learning rate of 20.0. See [7] for a detailed

TABLE 1. The results obtained upon registering the translated image T with the noisy image I_δ, where δ is the impulse noise density; ϕ_X and ϕ_Y denote the X- and Y-translation values of the optimal transformation ϕ produced by the registration algorithm, and n is the number of iterations until convergence. The actual translation values are 13 units in X and 17 units in Y.

	Mean Squares			Normalized Correlation			Mutual Information		
δ	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
0.00	12.99	17.00	18	13.01	17.00	18	12.75	17.03	200
0.10	12.99	17.01	28	12.99	17.01	20	12.83	16.88	200
0.20	13.03	16.98	17	13.04	16.98	19	12.98	16.64	200
0.30	12.97	17.03	28	13.02	17.02	11	13.02	17.02	200
0.40	18.89	7.16	15	8.05	1.30	13	11.08	9.72	200
0.50	2.16	7.06	19	9.09	2.18	8	9.72	7.12	200
0.60	29.81	3.19	40	4.08	0.24	7	4.57	5.17	200
0.70	2.08	1.14	13	3.11	2.13	12	3.08	2.86	200

discussion of these parameters. Finally, we set the maximum number of iterations for each algorithm to 200. As we shall see, mean squares and normalized correlation registrations typically converge very quickly to the optimal value. Mutual information, on the other hand, often does not actually reach the true optimal solution but instead oscillates within one or two pixels of the optimal solution (generally after 100-150 iterations). By reducing the learning rate, we can increase the likelihood of convergence, but this increases the computation time significantly without improving the accuracy of the solution.

For each of these three registration algorithms, and for each δ we record the X- and Y-translation parameters, denoted by ϕ_X and ϕ_Y, respectively, of the optimal transformation ϕ produced by the registration process. We also record the number of iterations n until convergence. The results are shown in Table 1. Recall that the actual translation values are 13 units in X and 17 units in Y. We also record the number of iterations until convergence, which we denote by n.

The results presented in Table 1 indicate that optimal linear registration in the presence of impulse noise fails when the impulse noise density in the fixed image reaches approximately 0.40, regardless of the metric used.

4. **Denoising.**

4.1. **Denoising techniques.**

In this section, we discuss various denoising techniques. Image denoising is a fundamental problem in image processing, and there has been much research and progress on the subject. As our primary interest is not denoising but the problem of image registration of noisy images, we do not focus on the general problem of image denoising. Instead, we present a few of the most common and computationally simple denoising techniques. We will then apply these techniques to one of our noisy images and study the effect of denoising on the image registration techniques. In particular, in Section 3, we saw that ordinary optimal linear registration of noisy images failed when the impulse noise density was greater than 0.40. Also in this section, we shall determine whether or not denoising prior

to registration enables successful registration of noisy images for which registration failed previously.

Spatial filtering is the traditional approach to removing noise from images. Spatial filters use the assumption that noise occupies the higher regions of the frequency spectrum, and thus they attenuate high spatial frequencies. Local spatial filtering is a process in which the value of a given pixel in the filtered image is computed by applying some algorithm to the pixel values in a neighborhood of the given pixel. Typical implementations of spatial filters include mean filtering, median filtering, and Gaussian smoothing. Mean filtering computes the value of each output pixel by computing the statistical mean of the neighborhood of the corresponding input pixel. Thus, applying a mean filter to a noisy image reduces the amount of variation in gray-level intensity between pixels. Although this filter is computationally easy to implement, it is sensitive to the presence of outliers. Median filtering, which computes the value of each output pixel by computing the statistical median of the neighborhood of the corresponding input pixel, is more robust to the presence of outliers, and is thus commonly used for removing impulse noise from images. Convolution with a Gaussian kernel is another commonly used spatial filtering technique. See [23] for an overview of classical spatial filtering techniques.

In Figure 4, we illustrate the effect of applying a mean, median, and Gaussian convolution filter to the noisy image $I_{0.70}$, the brain proton density slice image with impulse noise of density 0.70. As is indicated in Figure 4, spatial filters smooth the data to remove noise but also blur edges.

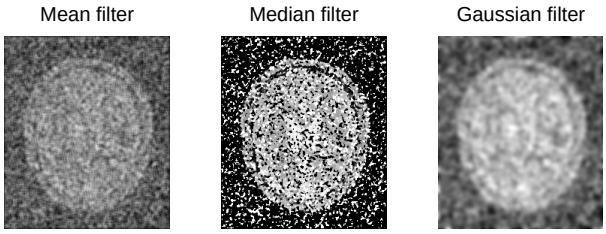


FIGURE 4. The results of applying mean, median, and Gaussian filters to the brain proton density slice image with impulse noise of density 0.70.

More advanced denoising techniques that remove noise more effectively while preserving edges include wavelet-based methods [10], total variation methods [19], and PDE-based anisotropic diffusion methods [13], to name a few. Total variation denoising reduces the total variation of the image, and thus removes noise, textures, and fine-scale details while preserving edges. In Figure 5, we illustrate the effect of applying these denoising techniques to the noisy image $I_{0.70}$, the brain proton density slice image with impulse noise of density 0.70.

4.2. **Registration results after denoising.** In this section, we register the translated image T with the denoised images illustrated in Figures 4 and 5. As in Section

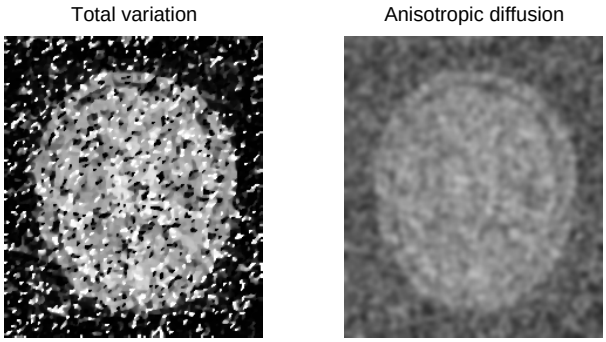


FIGURE 5. The results of applying the Osher-Rudin total variation and the Perona-Malik anisotropic diffusion denoising algorithms to the brain proton density slice image with impulse noise of density 0.70.

TABLE 2. The results obtained upon registering the translated image T with the denoised images obtained upon applying median, mean, and Gaussian convolution filters to the noisy image $I_{0.70}$. ϕ_X and ϕ_Y are the X- and Y-translation values of the optimal transformation ϕ produced by the registration algorithm, n is the number of iterations until convergence. The actual translation values are 13 units in X and 17 units in Y.

Denoising Technique	Mean Squares			Normalized Correlation			Mutual Information		
	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
Mean Filtering	31.83	1.15	46	16.88	1.11	29	5.39	5.30	200
Median Filtering	18.87	1.26	31	2.38	6.90	34	4.39	4.06	200
Gaussian Filtering	18.86	-0.76	31	2.19	0.25	11	7.38	7.37	200
Total Variation	6.11	4.26	19	5.29	9.15	15	6.30	6.23	15
Anisotropic Diffusion	2.10	1.13	11	4.09	6.22	10	10.62	14.77	200

3, we use mean squares, normalized correlation, and mutual information optimal linear registration. For each registration method, we let ϕ denote the optimal transformation produced by the registration algorithm, and we let ϕ_X and ϕ_Y the X- and Y-translation parameters of the optimal transformation ϕ. We denote by n the number of iterations of each registration algorithm until convergence. We record the results in Table 2. The moving image in each case is the translated image T; recall that the actual translation values are 13 in X and 17 in Y.

The results presented in Table 2 indicate that the application of some of the classical as well as modern denoising techniques prior to registration does not enable successful registration of the noisy image $I_{0.70}$ with the translated image T. Although the more advanced denoising techniques such as total variation and anisotropic diffusion result in translation values that are closer to the actual values, particularly when mutual information registration is used, we conclude from Table 2 that denoising prior to registration does not produce accurate registration results for images that contain high levels of noise.

5. **Multiscale decomposition.** In this section, we present the multiscale image representation using hierarchical (BV, L^2) decompositions of [20]. The multiscale decomposition will provide a hierarchical expansion of an image that separates the essential features of the image (such as large shapes and edges) from the fine scales of the image (such as details and noise). The decomposition is hierarchical in the sense that it will produce a series of expansions of the image that resolve increasingly finer scales, and hence will include increasing levels of detail. We will eventually apply the multiscale decomposition algorithm to the problem of image registration in the presence of noise, and will demonstrate the accuracy of the multiscale registration technique for noisy images such as those that were considered in Sections 3 and 4.

We will use the following mathematical spaces in the decomposition algorithm. The space of functions of bounded variation, BV, is defined by

$$BV = \left\{ f \quad \middle| \quad ||f||_{BV} := \sup_{h \neq 0} |h|^{-1} ||f(\cdot + h) - f(\cdot)||_{L^1} < \infty \right\}.$$

We will also use the Sobolev space $W^{-1,\infty}$ with norm given by:

$$||f||_{W^{-1,\infty}} := \sup_{g} \left[\int \frac{f(x)g(x)}{||g||_{W^{1,1}}} \, dx \right],$$

where $||g||_{W^{1,1}} := ||\nabla g||_{L^1}$.

5.1. **The hierarchical decomposition.** Define the J-functional $J(f, \lambda)$ as follows:

$$J(f, \lambda) := \inf_{u+v=f} \lambda ||v||_{L^2}^2 + ||u||_{BV}, \tag{18}$$

where $\lambda > 0$ is a scaling parameter that separates the L^2 and BV terms. This functional $J(f, \lambda)$ was introduced in the context of image processing by Rudin, Osher, and Fatemi [19]. They suggested the following. Let $[u_\lambda, v_\lambda]$ denote the minimizer of $J(f, \lambda)$. The BV component, u_λ, captures the coarse features of the image f, while the L^2 component, v_λ, captures the finer features of f such as noise. This model is effective in denoising images while preserving edges, though it requires prior knowledge of the noise scaling λ.

Tadmor, et al. propose in [20] an alternative point of view in which the minimization of $J(f, \lambda)$ is interpreted as a decomposition $f = u_\lambda + v_\lambda$, where u_λ extracts the edges of f and v_λ extracts the textures of f. This interpretation depends on the scale λ, since texture at scale λ consists of edges when viewed under a refined scale (2^λ, for example). We refer to $v_\lambda = f - u_\lambda$ as the residual of the decomposition. Upon decomposing $f = u_\lambda + v_\lambda$, we proceed to decompose v_λ as follows:

$$v_\lambda = u_{2\lambda} + v_{2\lambda},$$

where

$$[u_{2\lambda}, v_{2\lambda}] = \underset{u+v=v_\lambda}{\operatorname{arginf}} J(v_\lambda, 2\lambda).$$

Thus, we obtain a two-scale representation of f given by $f \cong u_\lambda + u_{2\lambda}$, where now $v_{2\lambda} = f - (u_\lambda + u_{2\lambda})$ is the residual. Next we decompose $v_{2\lambda}$ and continue this process, which results in the following hierarchical multiscale decomposition of f. Starting with an initial scale $\lambda = \lambda_0$, we obtain an initial decomposition of the image f:

$$f = u_0 + v_0, \quad [u_0, v_0] = \underset{u+v=f}{\operatorname{arginf}} J(f, \lambda_0).$$

We then refine this decomposition to obtain

$$v_j = u_{j+1} + v_{j+1}, \quad [u_{j+1}, v_{j+1}] = \underset{u+v=v_j}{\operatorname{arginf}} J(v_j, \lambda_0 2^{j+1}), \quad j = 0, 1, \ldots$$

After k steps of this process, we have:

$$f = u_0 + v_0 = u_0 + u_1 + v_1 = u_0 + u_1 + u_2 + v_2 = \ldots = u_0 + u_1 + \ldots + u_k + v_k,$$

which is a multiscale image decomposition $f \sim u_0 + u_1 + \ldots + u_k$, with a residual v_k. As k increases, the u_k components resolve edges with increasing scales $\lambda_k = \lambda_0 2^k$.

5.2. **Implementation.**

5.2.1. *Initialization.* As described in [20], the initial scale λ_0 should capture the smallest oscillatory scale in f, given by

$$\frac{1}{2\lambda_0} \leq ||f||_{W^{-1,\infty}} \leq \frac{1}{\lambda_0}. \tag{19}$$

However, in practice, we may not be able to determine the size of $||f||_{W^{-1,\infty}}$, so we determine the initial choice of λ_0 experimentally. Following [20], for the applications presented in this paper, we will use $\lambda_0 = 0.01$ and $\lambda_j = \lambda_0 2^j$.

5.2.2. *Numerical discretization.* We follow the numerical algorithm of [20] for the construction of our hierarchical decomposition. In each step, we use finite-difference discretization of the Euler-Lagrange equations associated with the $J(v_j, \lambda_{j+1})$ to obtain the next term, u_{j+1}, in the decomposition of the image f. The Euler-Lagrange equation associated with the minimization of the functional $J(f, \lambda)$ given in equation (18) is

$$u_\lambda - \frac{1}{2\lambda} \operatorname{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right) = f,$$

with the Neumann boundary conditions:

$$\left.\frac{\partial u_\lambda}{\partial n}\right|_{\partial\Omega} = 0, \tag{20}$$

where $\partial\Omega$ is the boundary of the domain Ω and n is the unit outward normal.

We thus obtain an expansion $f \sim \sum_{j=0}^{k} u_j$, where the u_j are constructed as approximate solutions of the recursive relation given by the following elliptic PDE:

$$u_{j+1} - \frac{1}{2\lambda_{j+1}}\text{div}\left(\frac{\nabla u_{j+1}}{|\nabla u_{j+1}|}\right) = -\frac{1}{2\lambda_j}\text{div}\left(\frac{\nabla u_j}{\nabla |u_j|}\right). \tag{21}$$

Note that $J(f,\lambda)$ contains a singularity when $|\nabla u_\lambda| = 0$. To remove this singularity, we replace $J(f,\lambda)$ by the regularized functional

$$J^\epsilon(f,\lambda) := \inf_{u+v=f}\left\{\lambda||v||_{L^2}^2 + \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2}\, dx\, dy\right\}, \tag{22}$$

and at each step, we find the minimizer u_λ of J^ϵ. The Euler-Lagrange equation for the regularized J^ϵ functional is

$$u_\lambda - \frac{1}{2\lambda}\text{div}\left(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2 + |\nabla u_\lambda|^2}}\right) = f \in \Omega,$$

with Neumann boundary conditions.

To numerically implement the method, we cover the domain Ω with a grid ($x_i := ih, y_j := jh$), and discretize the elliptic PDE of equation (21) as follows:.

$$\begin{aligned}
u_{i,j} &= f_{i,j} + \frac{1}{2\lambda}D_{-x}\left[\frac{1}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j})^2 + (D_{0y}u_{i,j})^2}}D_{+x}u_{i,j}\right] \\
&+ \frac{1}{2\lambda}D_{-y}\left[\frac{1}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j})^2 + (D_{+y}u_{i,j})^2}}D_{+y}u_{i,j}\right] \\
&= f_{i,j} + \frac{1}{2h^2}\left[\frac{u_{i+1,j} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j})^2 + (D_{0y}u_{i,j})^2}}\right. \\
&\left. - \frac{u_{i,j} - u_{i-1,j}}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j})^2 + (D_{0y}u_{i-1,j})^2}}\right] \\
&+ \frac{1}{2h^2}\left[\frac{u_{i,j+1} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j})^2 + (D_{+y}u_{i,j})^2}}\right. \\
&\left. - \frac{u_{i,j} - u_{i,j-1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1})^2 + (D_{-y}u_{i,j})^2}}\right],
\end{aligned} \tag{23}$$

where D_+, D_-, and D_0 denote the forward, backward, and centered divided differences, respectively. To solve the discrete regularized Euler-Lagrange equations (24), we use the Gauss-Siedel iterative method to obtain:

$$u_{i,j}^{n+1} = f_{i,j} + \frac{1}{2h^2}\left[\frac{u_{i+1,j}^n - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j}^n)^2 + (D_{0y}u_{i,j}^n)^2}} - \frac{u_{i,j}^{n+1} - u_{i-1,j}^n}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j}^n)^2 + (D_{0y}u_{i-1,j}^n)^2}}\right] + \frac{1}{2h^2}\left[\frac{u_{i,j+1}^n - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j}^n)^2 + (D_{+y}u_{i,j}^n)^2}} - \frac{u_{i,j}^{n+1} - u_{i,j-1}^n}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1}^n)^2 + (D_{-y}u_{i,j}^n)^2}}\right]. \tag{24}$$

To satisfy the Neumann boundary conditions (20), we first reflect f outside Ω by adding grid lines on all sides of Ω. As the initial condition, we set $u_{i,j}^0 = f_{i,j}$. We iterate this numerical scheme for $n = 0, 1, \ldots N$ until $||u^{n_\infty} - u^{n_\infty - 1}||$ is less than some preassigned value so that $u_{i,j}^{n_\infty}$ is an accurate approximation of the fixed point steady solution u_λ.

Finally, we denote the final solution $u_\lambda := \{u_{i,j}^{n_\infty}\}_{i,j}$. To obtain the hierarchical multiscale decomposition, we reiterate this process, each time updating f and λ in the following way:

$$\begin{aligned} f_{new} &\leftarrow f_{current} - u_\lambda, \\ \lambda_{new} &\leftarrow 2\lambda_{current}. \end{aligned} \tag{25}$$

That is, at each step, we apply the $J(f_{current} - u_\lambda, 2\lambda)$ minimization to the residual $f_{current} - u_\lambda$ of the previous step. Taking $\lambda_j = \lambda_0 2^j$, we obtain after k steps a hierarchical multiscale decomposition $f = u_{\lambda_0} + u_{\lambda_1} + \ldots + u_{\lambda_k} + v_{\lambda_k}$, where we write $u_{\lambda_j} = u_j$. We call the u_j, $j = 1, 2, \ldots, k$ the components of f and the v_k the residuals.

EXAMPLE 1. ***Decomposition of a noisy image.*** *We apply the hierarchical multiscale decomposition of* [20] *as described in Section 5 to the noisy image* $I_{0.70}$ *in Figure 3, using the following parameters:* $m = 12$ *hierarchical steps,* $\lambda_0 = 0.01$, $\lambda_j = \lambda_0 2^j$, $\epsilon = 0.001$, $n = 10$, *and* $h = 1$. *In Figures 6 and 7, we illustrate the components* u_{λ_j} *and the residuals* v_{λ_j} *for this decomposition. Note that in each hierarchical step, an additional amount of texture is seen in the components. Further, the noise is not seen in the first few components, while most of the texture is kept, and the noise only reappears as the refined scales reach the same scales as the noise itself. Our goal is to use this multiscale decomposition to register the noisy image* $I_{0.70}$ *with the translated image* T.

6. **Multiscale registration.** Consider again the noisy images I_δ shown in Figure 3 with impulse noise of increasing densities δ. Recall that in Section 3, we demonstrated that registration of the the translated image T with the noisy image I_δ

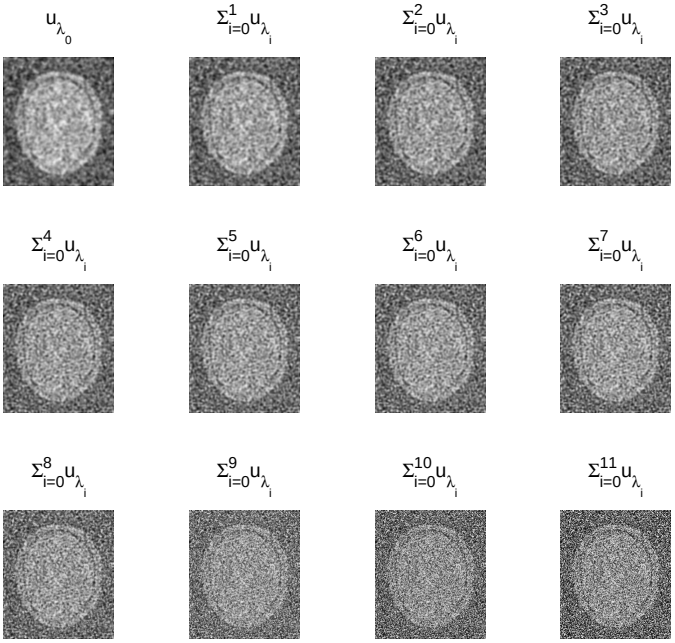


FIGURE 6. Multiscale decomposition of the noisy image $I_{0.70}$ shown in Figure 3.

failed when $\delta \geq 0.40$, regardless of the metric used in the optimal linear registration process. Moreover, registration using these classical methods failed even after denoising the noisy image using various standard denoising techniques, as demonstrated in Section 4. In this section, we present new methods for image registration that allow for a successful registration of the translated image T with the noisy images I_δ for values of the noise density δ significantly greater than the levels at which classical registration and registration after denoising fail. These registration techniques will be based on the hierarchical multiscale decomposition described in Section 5.

Consider two images A and B, and suppose that we want to register image B with image A. Suppose that one or both of the images contains a significant amount of noise. If only one of the images is noisy, we assume that it is image A. We propose the following multiscale registration method. First, we apply the multiscale hierarchical decomposition to both images. Let m denote the number of hierarchical steps used for the multiscale decompositions. For ease of notation, given an image f, we let

$$C_k(f) := \sum_{i=0}^{k} u_{\lambda_k} \tag{26}$$

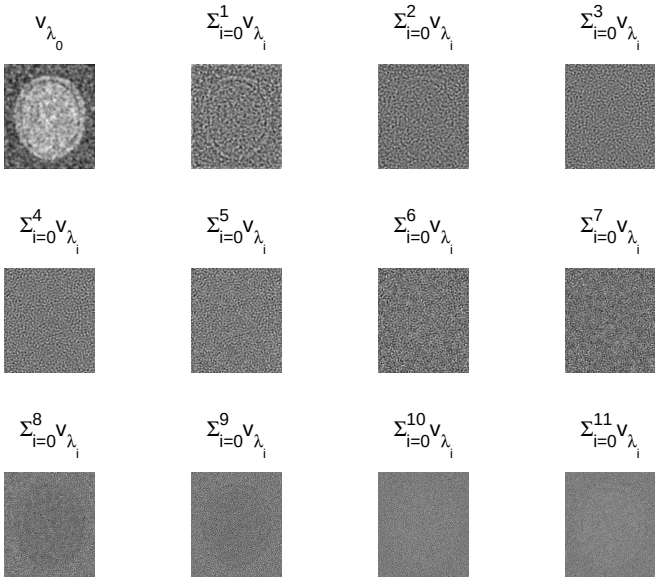


FIGURE 7. The residuals of the multiscale decomposition of the noisy image $I_{0.70}$ shown in Figure 3.

denote the k^{th} component of the image f, $k = 0, 1, \ldots, m-1$, obtained as in Section 5. Thus $C_k(A)$ will denote the k^{th} component of the image A, and $C_k(B)$ will denote the k^{th} component of image B.

We will present two algorithms; in the first, we register image B with the components of image A, and in the second, we register the components of image B with the components of image A.

6.1. **Algorithm I: One-node multiscale registration.** In our first multiscale registration algorithm, we register image B with the k^{th} component of A, for $k = 0, 1, \ldots, m-1$. This is illustrated by the schematic in Figure 8.

We refer to this algorithm as a *one-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, the moving image is always the image B. We only use the multiscale components of the fixed image A for the one-node algorithm.

Let ϕ_k denote the optimal transformation produced by the registration algorithm upon registering B with $C_k(A)$, $k = 0, 1, \ldots, m-1$. Recall that $C_0(A)$ contains only the coarsest scales of the image A, and as k increases, $C_k(A)$ contains increasing levels of detail (and hence, noise) of the image A. Thus, we expect that registration of image B with $C_k(A)$ should give an improvement compared to ordinary registration for the first few values of k. As k increases, however, we expect that eventually the component $C_k(A)$ will become too noisy to give successful registration.

Upon determining the transformations ϕ_k with a suitable registration algorithm (e.g., an optimal linear registration), we have several options for defining the optimal transformation Φ that should bring the image B into spatial alignment with the

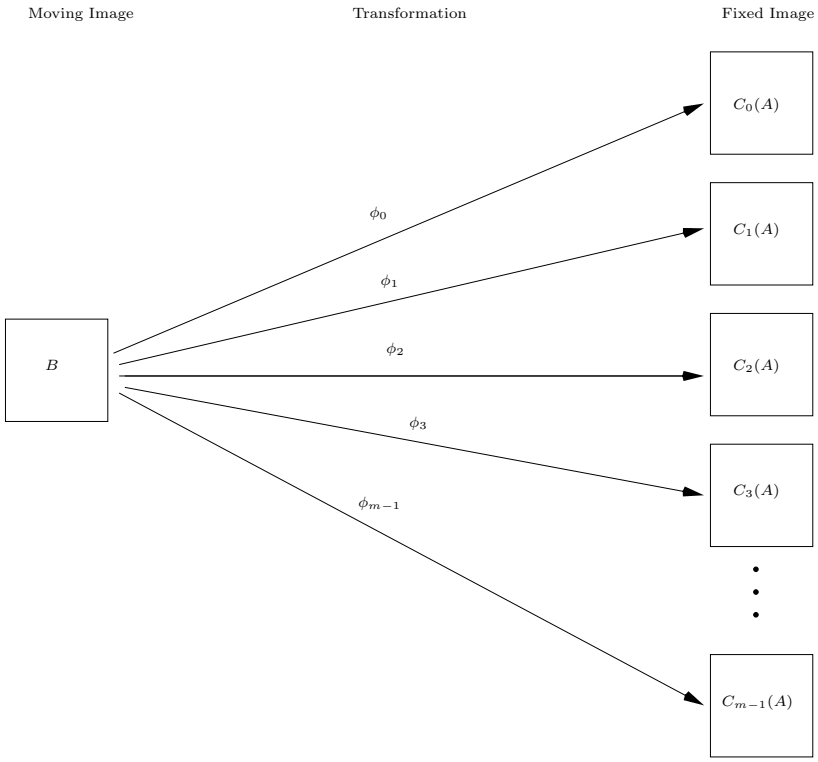


FIGURE 8. This schematic represents a one-node multiscale image registration algorithm in which we register the moving image B with the k^{th} component of the fixed image A, for $k = 0, 1, \ldots, m-1$, where m is the number of hierarchical steps used for the multiscale decompositions.

image A. The first option would be to take into account the registration parameters corresponding to the coarse scales only, i.e., the first few values of k, for which we expect a more accurate registration. Upon determining the number of registrations that we wish to take into account, we could then estimate Φ by averaging the registration parameters corresponding to those coarse scale registrations. A second option would be to define Φ as a weighted average of the ϕ_k; i.e.,

$$\Phi := \frac{1}{m} \sum_{k=0}^{m-1} a_k \phi_k, \tag{27}$$

where the weights a_k are appropriately chosen non-negative real numbers such that $\Sigma a_k = m$. For example, we could perform a statistical analysis on the registration parameters corresponding to the ϕ_k, and use the mean and standard deviation (or the mean and standard deviation of the first several values) to determine the weights a_k.

6.2. **Algorithm II: Multi-node multiscale registration.** In our second multiscale registration algorithm, we register the k^{th} component of image B with the

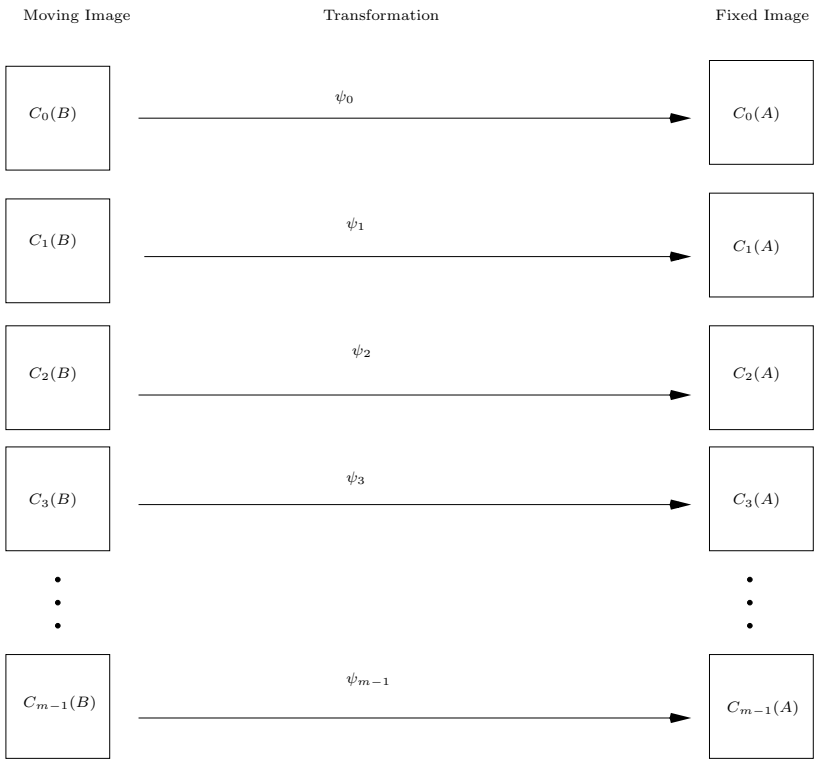


FIGURE 9. This schematic represents a multi-node multiscale image registration algorithm in which we register the k^{th} component of the moving image B with the k^{th} component of the fixed image A, for $k = 0, 1, \ldots, m-1$, where m is the number of hierarchical steps used for the multiscale decompositions.

k^{th} component of image A, for $k = 0, 1, 2, \ldots m-1$, as illustrated by the schematic in Figure 9.

We refer to this algorithm as a *multi-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, we consider both the components of the fixed image A and the components of the moving image B.

Let ψ_k denote the optimal transformation produced by the registration algorithm upon registering $C_k(B)$ with $C_k(A)$, $k = 0, 1, \ldots, m-1$. As before, we expect that registration of $C_k(B)$ with $C_k(A)$ should give an improvement compared to ordinary registration for the first few values of k. As k increases, however, we expect that eventually the components $C_k(A)$ and $C_k(B)$ will become too noisy to register successfully. Since this algorithm considers components of both images, we expect that it will be particularly successful in the case in which both images are noisy.

As in the case of the one-node multiscale registration algorithm, we can define the optimal transformation Ψ that should bring image B into spatial alignment with image A either by taking into account only the first few registration results corresponding to registration of the coarse scales, and averaging the registration parameters corresponding to these first few registrations, or by computing a weighted average:

TABLE 3. The registration results upon registering T with $I_{0.70}$ using Algorithm I. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The transformation parameters ϕ_X and ϕ_Y are the X- and Y-translation parameters of the optimal transformation ϕ produced by the registration algorithm. The actual translation values are 13 in X and 17 in Y. The moving image in all simulations is the translated image T.

	Mean Squares			Normalized Correlation			Mutual Information		
Fixed Image	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
$I_{0.70}$	4.57	5.17	200	2.08	1.14	7	4.08	0.24	7
$C_0(I_{0.70})$	12.65	16.36	200	3.08	1.11	12	3.11	0.17	9
$C_1(I_{0.70})$	12.69	16.78	200	2.08	3.08	14	2.13	2.12	12
$C_2(I_{0.70})$	12.56	16.79	200	2.11	3.08	14	2.14	3.11	15
$C_3(I_{0.70})$	12.53	16.76	200	3.08	2.11	14	3.11	2.14	7
$C_4(I_{0.70})$	12.48	16.76	200	24.88	1.16	36	18.86	1.18	30
$C_5(I_{0.70})$	12.46	16.78	200	40.80	1.07	52	0.21	1.18	11
$C_6(I_{0.70})$	12.43	16.80	200	28.86	0.15	46	27.84	2.19	42
$C_7(I_{0.70})$	12.43	16.79	200	-2.87	4.11	15	0.18	3.14	12
$C_8(I_{0.70})$	12.43	16.74	200	25.89	3.12	40	-1.84	4.12	14
$C_9(I_{0.70})$	9.33	9.41	200	6.05	4.12	12	7.99	2.08	16
$C_{10}(I_{0.70})$	8.44	8.32	200	-3.92	8.12	21	4.09	3.15	16
$C_{11}(I_{0.70})$	6.96	6.46	200	8.97	6.13	13	3.65	1.17	27

$$\Psi := \frac{1}{m} \sum_{k=0}^{m-1} b_k \psi_k, \tag{28}$$

where the weights b_k are appropriately chosen non-negative real numbers such that $\Sigma b_k = m$.

7. **Examples of multiscale registration.**

7.1. **A noisy fixed image.** In this section, we use the multiscale registration technique described in Section 6 to register the translated (non-noisy) image T with the noisy image $I_{0.70}$. Recall that $I_{0.70}$ is the image obtained upon adding impulse noise of density 0.70 to the brain proton density slice image I. As before, let $C_k(I_{0.70})$ denote the k^{th} component in the multiscale decomposition of $I_{0.70}$, for $k = 0, 1, \ldots m$, obtained as in Section 6. We perform the multiscale decomposition using $m = 12$ hierarchical steps, $\lambda_0 = 0.01$, and $\lambda_j = \lambda_0 2^j$. In Table 3, we present the results of $m = 12$ registration simulations, obtained upon registering T with $C_k(I_{0.70})$, $k = 0, 1, \ldots, 11$, using Algorithm I of Section 6.1. For each registration, we let ϕ denote the optimal transformation produced by the registration algorithm, and we let ϕ_X and ϕ_Y the X- and Y-translation parameters of the optimal transformation ϕ. The moving image in each registration is the translated image T. For reference, we also include in the first line of Table 3 the parameters obtained using ordinary registration.

It is clear from the results presented in Table 3 that the results obtained using mean squares and normalized correlation methods are completely inaccurate. Thus, the one-node multiscale algorithm did not produce meaningful results for these metrics. For mutual information, however, the X and Y translation parameters are clustered around 12.5 units in X and 16.8 units in Y for $k = 0, 1, \dots 8$, but then are significantly different for the remaining values of k. We expected that the multiscale registration results would be an accurate approximation of the actual transformation Φ for small values of k, but then would deviate as k became sufficiently large, because as k becomes large, increasing scales of detail (and hence, noise) appear in the component C_k. Thus, even without knowing the actual values of the X- and Y-translations, it makes sense to take into account only the parameters corresponding to the first 9 registrations ($k = 0, 1, \dots 8$). Averaging the translation parameters for the first 9 registrations, we obtain $\Phi_X = 12.52$ and $\Phi_Y = 16.73$. Since the actual values are 13 in X and 17 in Y, we see that multiscale mutual information registration produced very accurate results in this case, and indeed is a significant improvement compared to ordinary registration as well as to classical and modern denoising followed by registration.

Next, we provide the results obtained with Algorithm II by registering the multiscale components of the translated image T with the multiscale components of the noisy image $I_{0.70}$. Let $C_k(T)$ and $C_k(I_{0.70})$ denote the multiscale components of T and $I_{0.70}$, respectively, obtained through the multiscale decomposition presented in Section 5. As before, we use $m = 12$ hierarchical steps, $\lambda_0 = 0.01$, and $\lambda_j = \lambda_0 2^j$ to perform the decomposition. In Table 4, we present the results of $m = 12$ registration simulations, obtained upon registering $C_k(T)$ with $C_k(I_{0.70})$, $k = 0, 1, \dots, 11$. For each registration, we let ψ denote the optimal transformation produced by the registration algorithm, and let ψ_X and ψ_Y denote the X- and Y-translation parameters of the optimal transformation ψ. For reference, we also include in the first line of Table 4 the parameters obtained using ordinary registration.

To estimate the transform parameters Ψ_X and Ψ_Y, we note that for mutual information, the translation parameters ψ_x and ψ_y are clustered together for the first 9 registrations, and for mean squares and normalized correlation, the values are clustered together for the first 2 registrations. Thus for mutual information we determine Ψ by averaging the parameters corresponding to the first 9 registrations, and for mean squares and normalized correlation, we average the first 2 values. In Table 5, we present the X- and Y-translation values corresponding to these averages.

REMARK. *Since the actual translation values are 13 in X and 17 in Y, we see that the multinode multiscale registration of the translated image T with the noisy image $I_{0.70}$ produces very accurate results for each of the three optimal linear registration metrics considered here (mean squares, normalized correlation, and mutual information). The main difference between the results obtained with Algorithm I and Algorithm II is the accurate registration of the coarse scales obtained with Algorithm II.*

7.2. **Noisy fixed and moving images.** In this section, we consider the registration problem in which both the fixed and moving images are noisy. Consider the noisy images $I_{0.40}$ and $T_{0.40}$, where T, as before, is the result of translating I 13 units in X and 17 units in Y, and A_δ denotes the image obtained by adding impulse noise of density δ to the image A. The noisy images are shown in Figure 10.

TABLE 4. The registration results obtained with Algorithm II. Here we register the k^{th} multiscale component $C_k(T)$ of the translated image T with the k^{th} multiscale component $C_k(I_{0.70})$ of the noisy image $I_{0.70}$ obtained via the multiscale decomposition discussed in Section 5. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The transformation parameters ψ_X and ψ_Y are the X- and Y-translation parameters of the optimal transformation ψ produced by the registration algorithm. The actual translation values are 13 in X and 17 in Y.

Fixed and Moving Images	Mean Squares		Normalized Correlation		Mutual Information	
	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y
$I_{0.70}$ and T	4.57	5.18	2.08	1.14	4.08	0.24
$C_0(I_{0.70})$ and $C_0(T)$	12.69	16.66	12.29	17.72	12.96	17.08
$C_1(I_{0.70})$ and $C_1(T)$	12.67	16.87	13.70	17.75	12.99	17.67
$C_2(I_{0.70})$ and $C_2(T)$	12.59	16.86	20.77	5.20	16.84	4.31
$C_3(I_{0.70})$ and $C_3(T)$	12.55	16.82	3.19	0.31	4.20	4.23
$C_4(I_{0.70})$ and $C_4(T)$	12.52	16.83	2.20	2.24	26.74	5.18
$C_5(I_{0.70})$ and $C_5(T)$	12.51	16.84	31.65	2.23	14.90	6.27
$C_6(I_{0.70})$ and $C_6(T)$	12.49	16.87	30.69	6.16	19.87	4.29
$C_7(I_{0.70})$ and $C_7(T)$	12.48	16.85	33.64	3.16	29.64	3.32
$C_8(I_{0.70})$ and $C_8(T)$	12.53	16.71	28.81	3.22	1.26	1.29
$C_9(I_{0.70})$ and $C_9(T)$	9.26	9.36	2.13	3.13	17.93	3.21
$C_{10}(I_{0.70})$ and $C_{10}(T)$	8.80	8.61	2.12	3.12	32.63	3.14
$C_{11}(I_{0.70})$ and $C_{11}(T)$	6.95	6.34	34.74	2.10	4.13	5.08

TABLE 5. The translation parameters Ψ_X and Ψ_Y obtained by averaging the parameters corresponding to the coarse scale registrations. The actual translation values are 13 in X and 17 in Y.

	Mean Squares	Normalized Correlation	Mutual Information
Ψ_X	12.56	12.99	12.98
Ψ_Y	16.82	17.74	17.37

Before applying our multiscale registration algorithm, we attempt to register $T_{0.40}$ with $I_{0.40}$ using the three registration methods mean squares, normalized correlation, and mutual information. The results shown in Table 6 indicate that registration of the noisy images fails, regardless of the metric used in the optimal linear registration algorithm.

Since ordinary registration of the noisy images fails, we register the images using Algorithm II, the multi-node multiscale registration technique. First, we perform the multiscale decomposition discussed in Section 5 to both noisy images, again using $m = 12$ hierarchical steps, initial scale $\lambda_0 = 0.01$, and $\lambda_j = 2^j \lambda_0$. Let $C_k(I_{0.40})$ and $C_k(T_{0.40})$ denote the k^{th} component in the multiscale decomposition of $I_{0.40}$

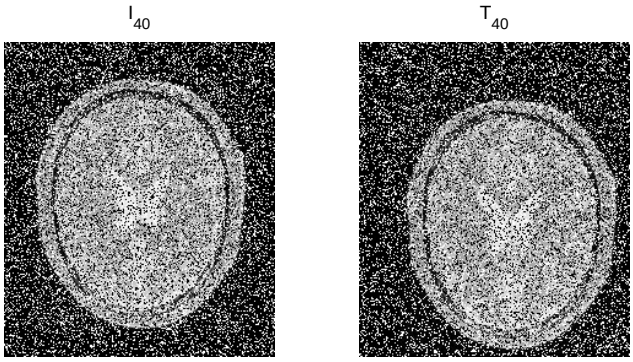


FIGURE 10. The original image I and translated image T with impulse noise of density $\delta = 0.40$

TABLE 6. The results of registering the noisy translated image $T_{0.40}$ with the noisy image $I_{0.40}$, using three different metrics. The actual translation values are 13 in X and 17 in Y.

Registration Method	ϕ_X	ϕ_Y
Mean Squares	11.02	7.04
Normalized Correlation	3.05	0.99
Mutual Information	5.03	2.54

and $T_{0.40}$, respectively. Since both images are noisy, we register the k^{th} component $C_k(T_{0.40})$ with the k^{th} component $C_k(I_{0.40})$. For each registration simulation, denote by ψ the optimal transformation produced by the registration algorithm, and denote by ψ_X and ψ_Y the corresponding X- and Y-translation parameters of the optimal transformation ψ. We present the results of this multiscale registration in Table 7.

To estimate the transformation Ψ, for mutual information we average the parameters corresponding to registration of the first 7 scales. For mean squares, we average the results of the first 2 registrations, and for normalized correlation, we average the registration results from the first 4 registrations. In Table 8, we present the X- and Y-translation values Ψ_X and Ψ_Y.

Note that since the actual translation values are 13 in X and 17 in Y, our multiscale registration technique provides accurate results in the case in which both the fixed and moving images contain significant levels of noise.

REMARKS.

1. *For the sake of brevity, we presented only the multiscale registration results for registration of images that contain levels of noise greater than the level at*

TABLE 7. The results of registering $T_{0.40}$ with $I_{0.40}$ using Algorithm II. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The actual translation values are 13 in X and 17 in Y.

Fixed Image	Moving Image	Mean Squares		Normalized Correlation		Mutual Information	
		ϕ_X	ϕ_Y	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y
$I_{0.40}$	$T_{0.40}$	5.03	2.54	11.02	7.04	3.05	0.99
$C_0(I_{0.40})$	$C_0(T_{0.40})$	13.06	16.92	13.05	16.92	13.05	16.92
$C_1(I_{0.40})$	$C_1(T_{0.40})$	13.05	16.93	13.02	16.22	13.06	16.92
$C_2(I_{0.40})$	$C_2(T_{0.40})$	13.03	16.93	8.11	5.29	13.02	16.27
$C_3(I_{0.40})$	$C_3(T_{0.40})$	13.02	16.94	5.40	12.19	13.02	16.25
$C_4(I_{0.40})$	$C_4(T_{0.40})$	13.02	16.94	2.20	8.00	2.23	5.09
$C_5(I_{0.40})$	$C_5(T_{0.40})$	13.01	16.93	26.76	1.21	1.17	7.00
$C_6(I_{0.40})$	$C_6(T_{0.40})$	12.99	16.81	23.83	4.11	1.22	2.17
$C_7(I_{0.40})$	$C_7(T_{0.40})$	7.05	6.08	0.20	3.15	0.20	4.15
$C_8(I_{0.40})$	$C_8(T_{0.40})$	6.78	5.05	6.04	2.09	6.04	6.05
$C_9(I_{0.40})$	$C_9(T_{0.40})$	3.05	1.02	9.98	1.10	5.06	10.01
$C_{10}(I_{0.40})$	$C_{10}(T_{0.40})$	12.20	14.01	-1.97	0.99	-3.93	3.04
$C_{11}(I_{0.40})$	$C_{11}(T_{0.40})$	4.80	3.19	1.01	5.98	3.91	0.72

TABLE 8. The translation parameters Ψ_X and Ψ_Y for registration of $T_{0.40}$ with $I_{0.40}$ obtained by averaging the translation parameters of the coarse scale registrations. The actual translation values are 13 in X and 17 in Y.

	Mean Squares	Normalized Correlation	Mutual Information
Ψ_X	13.03	13.03	13.04
Ψ_Y	16.92	16.57	16.59

which ordinary registration methods fail. However, we also performed multiscale registration simulations for noise densities lower than those presented here, and in all cases, the multiscale technique was either as accurate as or more accurate than ordinary registration techniques.

2. *The method of estimating the translation parameters in X and Y by averaging the parameters corresponding to the coarse scale registrations is based on determining the scales that should be taken into consideration. For the results presented in this paper, as well as for all of the other simulations that we studied, we found a drastic jump in the translation parameters such as that between the eighth and ninth scales in the mutual information column of Table 4. In such cases, the natural choice is to average the parameters corresponding to the coarse scales before the jump, and to exclude the remaining values; indeed, in all cases considered for this study, we have found that estimating the parameters in this way yields extremely accurate results. More generally, we expect that for most problems of this type, there should be a noticeable jump in the multiscale registration parameters, thus enabling a determination of the*

coarse scales that should be averaged. This jump occurs because once a certain level of detail, and hence noise, appears in the scales, the registration process fails. More specifically, the optimization of the registration metric does not produce meaningful results if enough noise is present, and the registration parameters that result differ significantly from those that result from registration of images in which less noise is present.

8. **Summary.** While there are many existing medical image registration techniques, common approaches are shown to fail to give accurate results when one or more of the images to be registered contains high levels of noise. Further, if the noise level is significantly high, image registration can fail even when a denoising algorithm is applied to the noisy images before registration. We have presented an image registration technique based on the hierarchical multiscale decomposition of [20] of the images to be registered. The multiscale decomposition of an image results in a hierarchical representation that separates the coarse and fine scales of the image. Upon obtaining the decomposition of one or both of the images to be registered, we register the components of the moving image with the components of the fixed image. Since the coarse scale components of an image contain the essential features and shapes of the image, registration of the coarse scale components of the moving image with the coarse scale components of the fixed image provides an accurate estimate of the actual transformation that brings the moving image into spatial alignment with the fixed image. Using images in which the precise transformation that maps one to the other is known, we have shown that the multiscale approach is indeed accurate for levels of noise much higher than the noise levels at which ordinary optimal linear registration and denoising prior to ordinary registration methods fail; moreover, for all levels of noise, the multiscale technique either matches or outperforms ordinary registration techniques. Finally, we hope to extend these techniques to other image registration problems in which ordinary registration techniques are not successful due to degradation or other factors present in the images to be registered.

Acknowledgments. The work of D. Levy was supported in part by the National Science Foundation under Career Grant No. DMS-0133511. The work of L. Xing was supported in part by the Department of Defense under Grant No. PC040282 and the National Cancer Institute under Grant No. 5R01 CA98523-01.

REFERENCES

[1] N.M. Alpert, J.F. Bradshaw, D. Kennedy, and J.A. Correia, *The principal axes transformation–A method for image registration*, Journal of Nuclear Medicine, vol. 31, no.10, pp. 1717-1722, 1990.

[2] M. Bro-Nielsen and C. Gramkow, *Fast fluid registration of medical images*, Lecture Notes in Computer Science, 1131, pp. 267-276, Springer, Berlin and Heidelberg, 1996.

[3] C. Broit, *Optimal Registration of Deformed Images*, Ph.D. thesis, Computer and Information Science, University of Pennsylvania, 1981.

[4] A. Collignon, D. Vadermeulen, P. Suetens, and G. Marchal, *3d multi-modality medical image registration based on information theory*, Computational Imaging and Vision, vol. 3, pp. 263-274, 1995.

[5] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[6] B. Fischer and J. Modersitzki, *Fast inversion of matrices arising in image processing*, Numerical Algorithms, vol. 22, pp. 1-11, 2001.

[7] *Insight Segmentation and Registration Toolkit (ITK) Software Guide*, http://www.itk.org.

[8] W.A. Light, *Variational methods for interpolation, particularly by radial basis functions*, Numerical Analysis. pp. 94-106, 1995.
[9] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, *Multimodality image registration by maximization of mutual information*, IEEE Transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, 1997.
[10] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Paris, 1998.
[11] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series 22, AMS, Providence, RI, 2002.
[12] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford, 2004.
[13] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629-639, 1990.
[14] T. M. Peters, B. L. K. Davey, P. Munger, R. M. Comeau, A. Evans, and A. Oliver, *Three-dimensional multi-modal image-guidance for neurosurgery*, IEEE Transactions on Medical Imaging, vol. 15, no. 2, pp. 121-128, 1996.
[15] J. P. W. Pluim, J .B. A. Maintz, and M. A. Viergever, *Image registration by maximization of combined mutual information and gradient information*, IEEE Transactions on Medical Imaging, vol 19, no. 8, pp. 809-814, 2000.
[16] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, *Mutual information based registration of medical images: A survey*, IEEE Transactions on Medical Imaging, vol. 22, no. 8 , pp. 986-1004, 2003.
[17] K. Rohr, *Landmark-based Image Analysis*, Computational Imaging and Vision, Kluwer Academic, Dordrecht, 2001.
[18] L. Rudin and V. Caselles, *Image recovery via multiscale total variation*, Proceedings of the Second European Conference on Image Processing, Palma, Spain, 1995.
[19] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, vol. 60, pp. 259-268, 1992.
[20] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Modeling and Simulations, vol. 2, no. 4, pp. 554-579, 2004.
[21] P. Thévenaz and M. Unser, *Optimization of mutual information for multiresolution image registration*, IEEE Transactions on Image Processing, vol. 9, no. 12, pp. 2083-2099, 2000.
[22] P. Viola, W. Wells, H. Atsumi, S. Nakajima, and R. Kikinis, *Multi-modal volume registration by maximization of mutual information,* Medical Image Analysis, vol. 1, no. 1, pp. 35-51, 1995.
[23] A. R. Weeks, *Fundamentals of Electronic Image Processing*, SPIE Optical Engineering Press and IEEE Press, 1996.
[24] J. West, et al., *Comparison and evaluation of retrospective intermodality brain image registration techniques*, Journal of Computer Assisted Tomography, vol. 21, no. 4, pp. 554-566, 1997.

Received on December 29, 2005. Accepted on January 26, 2006.

E-mail address: dpaquin@stanford.edu
E-mail address: dlevy@math.stanford.edu
E-mail address: eduards@stanford.edu
E-mail address: lei@reyes.stanford.edu

Image registration with auto-mapped control volumes

Eduard Schreibmann and Lei Xing[a)]
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305-5847

(Received 9 August 2005; revised 5 February 2006; accepted for publication 8 February 2006; published 30 March 2006)

Many image registration algorithms rely on the use of homologous control points on the two input image sets to be registered. In reality, the interactive identification of the control points on both images is tedious, difficult, and often a source of error. We propose a two-step algorithm to automatically identify homologous regions that are used as *a priori* information during the image registration procedure. First, a number of small control volumes having distinct anatomical features are identified on the model image in a somewhat arbitrary fashion. Instead of attempting to find their correspondences in the reference image through user interaction, in the proposed method, each of the control regions is mapped to the corresponding part of the reference image by using an automated image registration algorithm. A normalized cross-correlation (NCC) function or mutual information was used as the auto-mapping metric and a limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) was employed to optimize the function to find the optimal mapping. For rigid registration, the transformation parameters of the system are obtained by averaging that derived from the individual control volumes. In our deformable calculation, the mapped control volumes are treated as the nodes or control points with known positions on the two images. If the number of control volumes is not enough to cover the whole image to be registered, additional nodes are placed on the model image and then located on the reference image in a manner similar to the conventional BSpline deformable calculation. For deformable registration, the established correspondence by the auto-mapped control volumes provides valuable guidance for the registration calculation and greatly reduces the dimensionality of the problem. The performance of the two-step registrations was applied to three rigid registration cases (two PET-CT registrations and a brain MRI-CT registration) and one deformable registration of inhale and exhale phases of a lung 4D CT. Algorithm convergence was confirmed by starting the registration calculations from a large number of initial transformation parameters. An accuracy of $\sim$2 mm was achieved for both deformable and rigid registration. The proposed image registration method greatly reduces the complexity involved in the determination of homologous control points and allows us to minimize the subjectivity and uncertainty associated with the current manual interactive approach. Patient studies have indicated that the two-step registration technique is fast, reliable, and provides a valuable tool to facilitate both rigid and nonrigid image registrations. © *2006 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2184440]

Key words: image registration, image fusion, deformable model, spline, IGRT

I. INTRODUCTION

Imaging is at the foundation of many clinical interventional procedures and development of an effective image registration technique has been one of the most important research areas in medical applications of imaging technology. Generally speaking, registration is to align two imaging data sets in a common coordinate space by transforming the higher resolution data set (the "model") while keeping the other one (the "reference") fixed. Depending on the mathematical nature of the transformation, image registration is divided into rigid and deformable registrations. In rigid transformations, it is assumed that the geometry of the object is identical in the two input images and no distortion occurs in the image acquisition process. When working in the "world coordinate" in which the lengths in both images are measured in the same scale, a rigid transformation consists of six degrees of freedom: three displacement parameters and three rotational parameters. Deformable registration, on the other hand, is more complicated and entails the modeling of voxel-dependent distortion in addition to the translation and rotation.[1–11] Clinically, the need for a robust image registration algorithm to compare/fuse images representing the same structures imaged under different conditions or on different modalities is ever increasing because of the extensive use of multimodality imaging and the emergence of new imaging techniques and methods.

Computer-based rigid image registration has gained widespread popularity in the last decade and is used in routine clinical practice. In this approach, the matching of the two input images is formulated into an optimization problem and the best registration of the two images is obtained by iteratively comparing various possible matches until no better registration can be found. The search for the optimal match of the two input images is usually gauged by a ranking func-

tion constructed based on some physical considerations. Depending on the nature of the input images, the formulation of the problem can be highly complicated and the truly optimal solution may not be readily attainable. To facilitate the computer decision-making process, image preprocessing or user interaction may be required, especially when dealing with a deformable image registration. Along this line, a commonly used strategy is to locate several homologous anatomic landmark pairs on the two input images, as it is generally true that inclusion of prior system knowledge often leads to a better matching. In order to accurately identify the control points on both images, a prerequisite is that the user must have a detailed understanding of the patient anatomy and the characteristics of the two modalities. The point pairs are usually obtained interactively with the user repetitively exploring the input image sets and each time trying to locate a point in both of them. Due to the 3D nature, the process is rather tedious and difficult to perform. Inaccuracy and subjectivity exist in the resulting match due to the user dependence in selecting the control points.

The purpose of this work is to develop a general method to facilitate the selection of control points for both rigid and deformable image registrations. Under an assumption that sufficient common features exist in the two input studies in most practical registration problems, the registration of two images in our approach proceeds in two steps. First, a number of small control regions having distinct anatomical features are identified on the model image in a more or less arbitrary fashion. Instead of attempting to find the correspondences of the regions in the reference image through the judgment of the user, in the proposed method, each of the control regions is mapped to the corresponding part of the reference image by using an automated image registration algorithm. The mapping of a control region is generally efficient and robust provided that sufficient information is included in the volume. After the mapping, a conventional automated image registration algorithm utilizing the predetermined control points can be employed to complete the remaining image registration process. This new way of image registration eliminates the need for the manual placement of the homologous control points and allows us to register the two images accurately and efficiently.

II. METHODS AND MATERIALS

A. Software platform

All calculations are implemented using an open-source software toolkit named the Insight Toolkit (ITK),[12] which consists of template-based codes for a large number of image visualization, segmentation, and registration classes. The programs contained in ITK are easily extendable, making it an ideal platform for the development of image registration methods. Concise and clear descriptions on the use of the available subroutines are provided in the ITK manual.

B. Selection of control regions on the model image

Image registration is a highly nontrivial optimization problem and suboptimal or even nonoptimal solution may result if it is done without *a priori* knowledge of the system. A method for enhancing the success rate of the calculation is to provide partial guidance to the optimization program by introducing a number of homologous control points identified on the model and reference images at different spatial locations. In general, the point pair should represent the same anatomic location in the two data sets. While the method is inherently efficient because it poses strong constraints on the solution search process, a challenge here is that the spatial arrangements of the reference and model points must be brought into optimal agreement. This is often performed by a skilled user and there exists considerable interperson variation in the identification of the control point pairs.

Different from the previous approaches, we replace each control point on the model image by a small volume, coined as control volume. A control volume is placed on a location where deformation is negligible (e.g., in or nearby a bony structure). It is preferable that sufficient imaging feature/information is contained within the volume so that its counterpart in the reference image can be easily and uniquely identified by computer. The underlying assumption here is that the regions on a model image are generally not equivalent and some regions can be more reliably mapped onto the reference image than others. The use of a "rigid" control region that possesses rich internal intensity pattern makes the mapping of the region from model image to reference image simple and unique. The determination of the size of the control region is generally a balance of a few factors. If the volume is too small, there would be not enough structurally unique features in the region to warrant an accurate mapping of the region. On the other hand, it may be difficult to specify a large control volume that meets the general selection criteria of the control volume when there is tissue deformation. Typically, the volume is spherical or cubic shaped with a dimension of $1-2$ cm in the case of intramodality registration, but, depending on the application, other shaped volume or dimension should also be acceptable.

C. Mapping of control regions from the model image to reference image

In Fig. 1 we show the flow charts of the new image registration scheme for rigid and deformable registrations. The input to the software is the images to be registered: a model image and a reference image, described by their intensity distributions $I_\alpha(\mathbf{x})$ and $I_\beta(\mathbf{x})$, respectively. A virtue of the new approach is that the control regions need to be identified only on the model image in a fairly arbitrary fashion. The correspondences of these regions on the reference image are determined through the use of an automatic image registration algorithm. In a sense, the determination of control regions here is a process of registration within a registration. For intramodality registration, we use the normalized cross correlation (NCC) of control regions between the model points and the corresponding reference points,

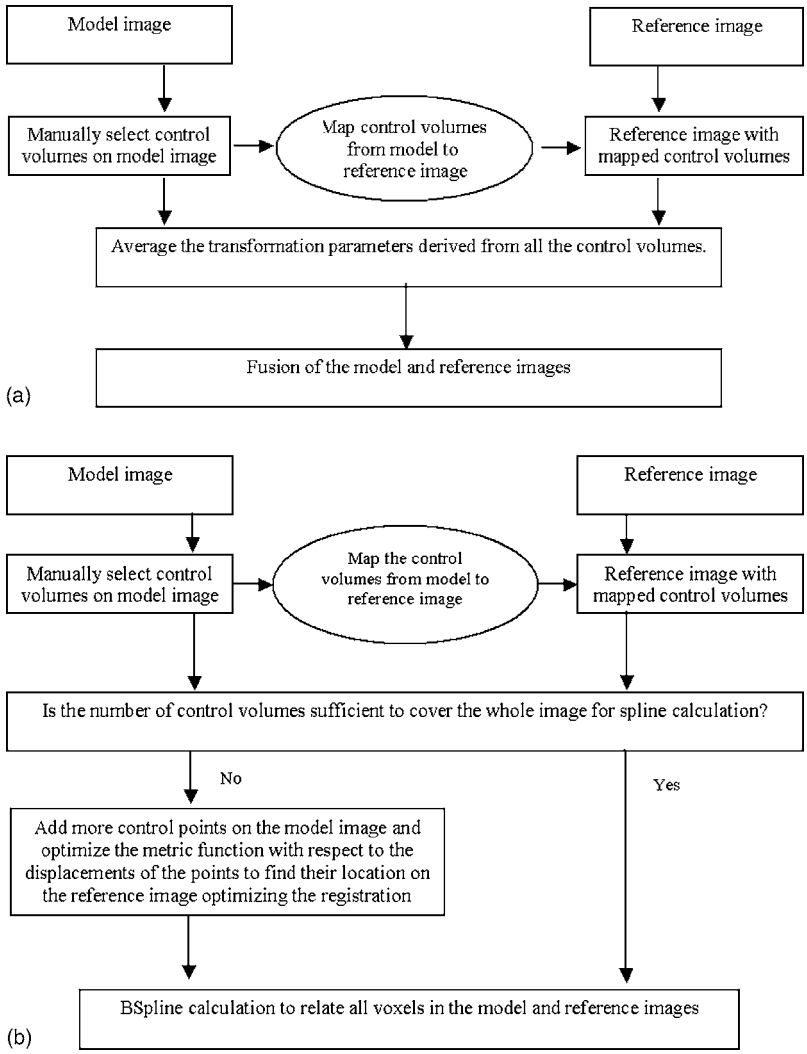


FIG. 1. Flow chart of the proposed calculation procedure for rigid (a) and deformable registration (b). The control volumes are selected by the user only on the model image. A rigid registration algorithm automatically maps the selected control volumes to their corresponding locations on the reference image.

$$f = -\frac{\sum_{i=1} I_\alpha(\mathbf{x}_i) I_\beta(T\mathbf{x}_i)}{\sqrt{\sum_{i=1} I_\alpha^2(\mathbf{x}_i) \sum_{j=1} I_\beta^2(T\mathbf{x}_j)}}, \quad (1)$$

as the criterion for assessing the goodness of the mapping of the control regions. In Eq. (1), i and j are the voxel indices within a control region α in the model image, $I_\alpha(\mathbf{x}_i)$ is the intensity of the voxel at $\mathbf{x}_i$ on the model image α, and $I_\beta(T\mathbf{x}_i)$ is the intensity of the reference image β where the voxel $\mathbf{x}_i$ is mapped. We note that other types of metrics should be equally applicable to deal with the problem here. For intermodality registration such as CT and MRI, the commonly used mutual information metric can be used as the matching criterion of the control regions.

The above mapping directly uses the value of all pixels within a control volume as the information feature with the assumptions that both data sets show the control volume similarly and that the control volume pair is related by a rigid transformation. In intramodality registration, an idealized situation is that the intensity of each voxel in a control region in the model image differs from its corresponding reference pixel by a factor. In practice, the two data sets may be affected by noise, sampling, and clinical or experimental situation. The mapping of a control volume is considered optimal if the above NCC function is minimal or mutual information is maximal. Numerous optimization algorithms can be employed to minimize/maximize the matching criteria. In our calculation, the control volumes are mapped sequentially and independently. For a chosen control region, an iterative optimization algorithm is used to update the six transformation parameters until no further improvement can be achieved. The iterative calculation is outlined in the next section. After one volume pair is finished, the calculation moves to the next control region and the iterative calculation is repeated.

D. Optimization of the NCC function

The NCC given in Eq. (1) or the mutual information metric for a given control volume is a function of its translational and rotational parameters. For each control volume, the optimization of the function with respect to these parameters is done by using the L-BFGS algorithm.[13] The details of this algorithm have been presented[11,14,15] and will not be repeated here. During the mapping of a control volume, the

iterative calculation based on the LBFGS algorithm continues until a preset maximum number of iterations (typically, 100 iterations) is reached.

E. Rigid and BSpline deformable registration with incorporation of the mapped control volume information

For rigid registration, a single control volume is, in principle, sufficient to derive the translational and rotational parameters of the system. To be robust, we typically select two to three control volumes and the final registration parameters are determined by averaging that derived from the individual control volumes. The use of control volume for rigid image registration is especially useful when there are artifacts in one or both input images.

For deformable registration, we usually select five or more control regions for each plane and the translation and rotation of each volume are generally different. The obtained homologous control volume pairs play a similar role as the nodes in the BSpline model. After the nodal set is determined, spline deformable registration can be used to relate the remaining part of the images. The detail of the calculation is described as follows.

Mathematically, the task of a deformable image registration is to find the transformation matrix, $T(\mathbf{x})$, that maps an arbitrary point $\mathbf{x}$ on the model image α to the corresponding point $\mathbf{x}'$ on the reference image β (or vice versa) in such a way that the best possible match, as measured by the registration metric, is achieved. In BSpline deformable registration calculation, the image is generally divided into a grid with N3 cells. The corner of a lattice cell is referred to as a node and is indexed by i $(i=1,2,\ldots,N3)$. The spacing between the nodes are usually 2–5 cm. The displacement of a node i is specified by a vector $\mathbf{x}_i$ and the displacement vectors, $\{\mathbf{x}_i\}$, of a collection of nodes characterize the tissue deformation. The displacement at a location $\mathbf{x}$ on the image is deduced by fitting a polynomial expressed using the basis spline (BSpline)[16] to the grid nodes $\mathbf{x}_i$. In reality, the nodes do not have to form a lattice on the model image. In our BSpline deformable calculation, the mapped control volumes are treated as the spline control points. Upon the completion of the mapping of the control volumes, the positions of these nodes on both input images are known. If the number of control volumes is not enough to cover the whole image volume, additional control points are placed on the model image and the locations of the added nodes are determined by optimizing the registration metric function with respect to their displacements, as is done in conventional spline deformable calculation for the whole nodal set. Otherwise, no additional nodes are required and one can proceed directly to the spline interpolation[1,7,17] to relate the voxels on the two input images, as depicted in Fig. 1(b). In this calculation, the existing homologous control volume pairs provides valuable *a priori* information of the registration and greatly reduces the dimensionality of the problem of finding an adequate set of control points.

To recapitulate, a deformable registration calculation in the proposed approach is constituted of two steps: mapping of the control volumes and registration of the image data. The first step is rigid registration in nature and yields the translational and rotational parameters for each control volume. To complete the second step, a strategy of incorporating the mapped control volumes needs to be in place. In our calculation, the coefficients of the transformation matrix $T(\mathbf{x})$ are the node displacements and need to be calculated only for those added nodal points, because the locations of the auto-mapped control volumes are already known with high confidence.

F. Search space characteristics and convergence analysis

Image registration is to establish a voxel-to-voxel mapping between the two input entities. The change of the metric function with the relative displacement of the two inputs provides a useful evaluation of the search space. The value of the metric function reaches its minimum or maximum when the two images are aligned and increases/decreases as they moved away from each other. A desirable metric function should exhibit a single distinct global minimum/maximum and vary smoothly as the two images are slide away gradually. For the rigid registration, we plot the metric as a function of the relative displacement of the two mirror control volumes. For comparison, we also plot the metric function for conventional registration, in which the final transformation parameters are obtained by iteratively adjusting the relative positions of the two images. For nonrigid registration, the behavior of each control volume is different and the metric functions for some selected control volumes are plotted when the volumes are displaced away from their ideally mapped locations.

In general, the optimization result should be independent of the initial values of the system variables. The convergence behavior of the proposed method is studied for both rigid and nonrigid image registrations by starting the registration calculation with a number of initial transformation parameters chosen randomly. For each calculation, we plot the metric value as a function of the iteration step. The results are compared with that obtained using conventional approach when treating each input image as an entity. In addition, the shifts toward the optimal match in the x, y, and z directions resulting from each of the calculations are plotted for the two-step registration and the conventional registration.

A checkerboard display tool, where images to be compared are merged together in a chesslike pattern, is used to assess the difference between the two images before and after registration. At the transition zone it is possible to measure the mismatch between anatomical structures in the two images. Ideally, when the two images are perfectly aligned, no difference should be visible in the checkerboard of the two images.

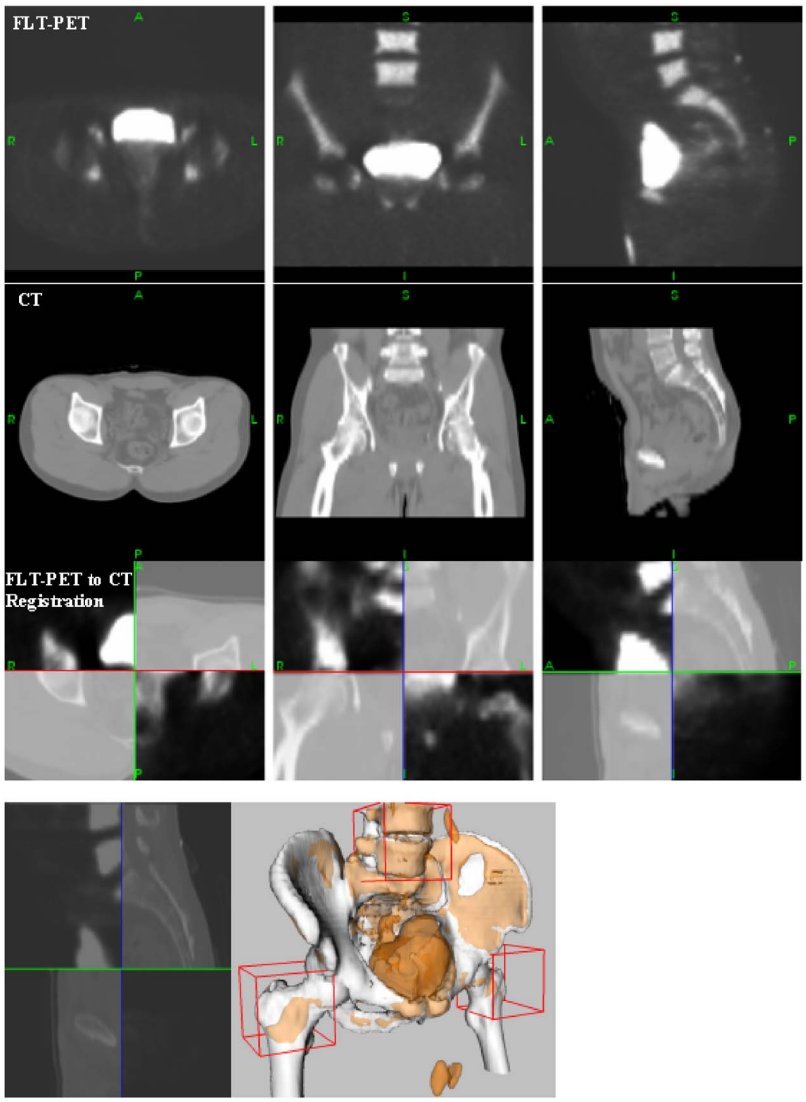


FIG. 2. Sagittal, coronal, and axial views of the FLT-PET images (first row) and CT images (second row). The checkerboard of the CT and FLT-PET images after registration is shown in the third row and the left panel of the fourth row. The right panel of the fourth row shows a stereotactic view of the matched PET and CT images after registration.

G. Case studies

The utility of the proposed two-step registration method is demonstrated by studying a few rigid registrations and a non-rigid registration case. The first two are cancer patients with tumors in the abdomen regions, who have undergone both CT and [^{18}F] fluorodeoxythymidine (FLT)-PET scans. Here FLT is a marker for thymidine kinase activity, representing one of the potentially more selective tracers that have been under intense investigations.[18] A side feature of FLT-PET is that bony structures appear clearly because of the high uptake of the bone matrix, which makes the registration of PET and other anatomic imaging modalities simple. The technique described above is used to fuse the CT and FLT-PET images and the results are compared with that obtained using conventional approach without relying on the use of control volumes. For the first case we use three cubic volumes with the side length of 7 cm, and the second case $5\times10\times10$ cm^3.

The next case consists of the registration of head CT and MRI images. Although the head is commonly believed to have no deformation, in practice, subtle differences may exist between images acquired using different modalities or under different conditions, which may adversely affect the registration. For example, the patient studied here uses an immobilization mask in the CT scan but not in the MRI. Additionally, this patient has a tumor in his left eye, which makes the MRI appear differently from a regular situation. Moreover, CT artifacts exist in the mandible region due to the denture, which does not have much impact on the MRI data. These factors complicate the registration calculation.

The input to the deformable image registration study is the CT images acquired at the expiration and inspiration phases using a 4D CT protocol.[19] A set of control volumes is selected on the exhale image [Fig. 9(c)] and a successive rigid registration is carried out to find their locations in the inhale image. These homologous control volumes pairs then serve as the nodes for the subsequent spline deformable cal-

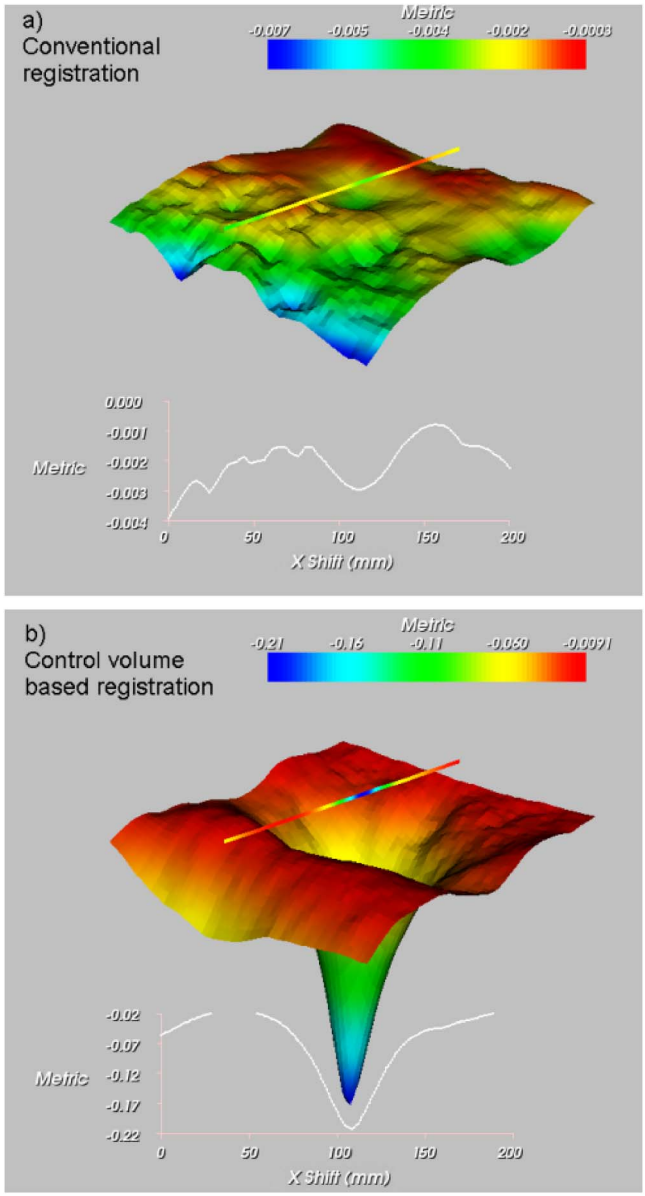


FIG. 3. Metric space of FLT-PET and CT registration when the whole image set is used (a) and when three control volumes on the bony structure are used (b).

culation to match voxels on the exhale image with that of the inhale image for the purpose of 4D treatment planning.[20,21]

III. RESULTS

A. Study 1: Rigid registration of CT and FLT-PET images

The input CT and FLT-PET images used for the first case in this study are shown in the first and second rows of Fig. 2. The control volumes relative to patient geometry are depicted on the right of the fourth row of Fig. 2 as red wire frames. To visually evaluate the registration results, in the third row of Fig. 2 we show a checkerboard comparison of the PET and CT images after registration. The sagittal checkerboard image of the PET and CT is shown in the left of the fourth row of Fig. 2. The images are registered accurately, with no geometric mismatch visible in the transition zone. A 3D view of the registration is presented on the right of the fourth row of Fig. 2, where an excellent coincidence is observed between the bony structures revealed in the CT (white) and the PET images (orange). In both 3D rendering and checkerboard display, the bladder is bright in the PET image, but barely visible in the CT images. The feature inherent to a certain type of images but not others is problematic in multimodality image registration since it "disturbs" the mathematical correspondence between the two images. Another commonly seen example of this is the registration of CT images containing metal artifacts with MRI data (see next section), in which no artifacts present. Control volume-based registration permits us to exclude the undesirable regions based on *a priori* knowledge and provides an effective way to find the truly optimal solution in this situation.

In Fig. 3 we compare the NCC metric space when the whole image is used with that when control volumes are used. The surface plot in Fig. 3 represents the NCC function values when the two input images [Fig. 3(a)] or two corresponding control volumes [Fig. 3(b)] placed in the lumbar region are shifted away from each other. The function values are color-coded from red, representing suboptimal matching, to blue, representing the optimal registration sought after. In the former situation, it is seen that the metric function is not smooth and multiple local minima exist. Aside from the fact that the PET image is generally noisy, the high-intensity voxels in the bladder (caused by the rapid excretion of FLT through the urinary system) that do not have a correspondence in the CT image also influence the behavior of the metric function in the former case. A single minimum in the control volume-based registration is, on the other hand, clearly visible and the search space is very smooth. Moreover, the NCC value at the minimum in this case is 0.22, which is two orders of magnitude deeper than that of the conventional registration. The improvement in the functional behavior of the metric makes the image registration calculation much more efficient. This is also supported by the following analysis.

To examine the algorithm's ability to find the correct shifts to the best match, we assigned a known displacement to the two input images and studied the registration process by starting the iterative optimization from different initial transformation parameters. In Fig. 4(a) we show the NCC as a function of iteration step for 50 randomly assigned initial trial transformation parameters. While the trajectory for each optimization calculation is different, they all converge to the same NCC value in the control volume-based calculations. The convergence of the calculation is also evident from Fig. 4(b), where it is shown that, for a given mismatch of the two input control volumes, all 50 calculations with different initial parameters lead to the same shift values, $(x,y,z) = (-1.8, 3.3, 5.3)$, where the coordinates are in mm. This is, however, not the case for the conventional approach based on the information contained in the whole image entity. As seen from Fig. 4(c), the NCC converges to different values

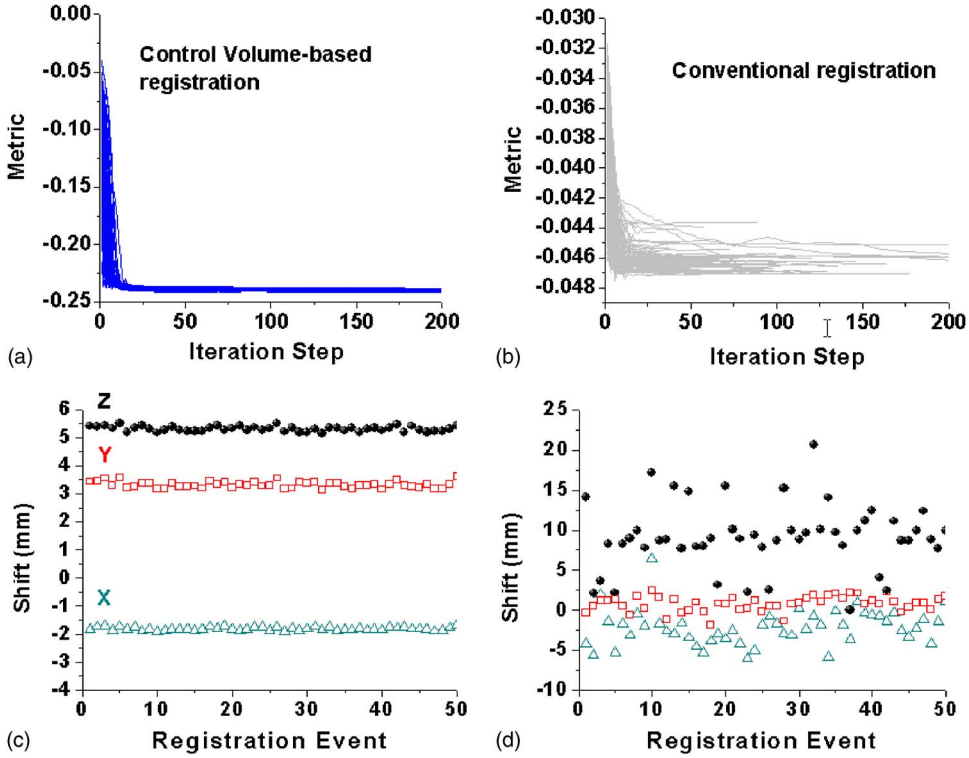


FIG. 4. Convergence analysis for a rigid registration case with (left column) and without (right column) the use of control volumes. The control volume-based registration converges to the same cost function value and leads to reproducible shift values in the x, y, and z directions, labeled by triangles, squares, and circular dots, respectively. In the standard registration, large variations are observed in the shifts along the x, y, and z directions.

when different initial conditions are used. Furthermore, the shifts [Fig. 4(d)] resulted from the approach fluctuate from calculation to calculation. In this situation, the average shift of the 50 runs in the x, y, and z directions are found to be $(-2.1, 0.85, 9.9)$. The variation range of the shifts in the control volume-based calculations are all within 0.2, 0.5, and 0.4 mm in the x, y, and z directions. The fluctuation in the conventional approach is, however, much greater: 15.8, 6.2, and 21.0 mm, respectively. In most cases, the results came out of the conventional approach represent, at best, suboptimal solutions. The misregistration is the z direction is particularly large, presumably due to the high intensity bladder in PET, which has no correspondence in CT.

The input CT and FLT-PET images for the second case are shown in the first and second rows of Fig. 5, along with a checkerboard display of the PET and CT images after registration (the third row). Once again, no geometric mismatch is visible in the transition zone. A 3D view of the registration is also presented in Fig. 5, where an excellent coincidence is observed between the bony structures revealed in the CT (white) and PET images (orange). In both 3D rendering and checkerboard display, the bladder and the liver are bright in the PET image, but not on the CT images. The NCC metric function and the convergence behavior of the registration calculation are similar to the previous case, as illustrated in Figs. 3 and 4. The control volume-based calculation yielded the correct transformation parameters, $(x,y,z)=(0,0,0)$, in all 100 calculations with different initial parameters, whereas the conventional approach based on the information contained in the whole image entity was problematic and failed to give the optimal solution. Indeed, for the 100 calculations starting with different initial conditions, the variation range of the shifts in the control volume-based calculations were within 0.1, 0.5, and 0.7 mm in the x, y, and z directions, consistent with that resulting from matching the DICOM coordinates of the PET and CT images. To illustrate the ability of the algorithm in finding the optimal mapping, in Fig. 6 we show 5 (out of the 100) independent registrations starting from different initial mismatches with inclusion of rotational degrees of freedom. Because the patient was scanned using a hybrid PET/CT scanner, the DICOM coordinate match of the two sets of images can be regarded as a "gold standard" of registration, and the above agreement indicates that the control volume-based method is accurate in dealing with clinical image registration. The fluctuation in the conventional calculation was, however, much greater: 11.0, 75.2, and 65.0 mm, respectively. The results coming out of the conventional approach are clinically unacceptable.

In general, PET images are noisy and have little anatomic information to yield descent registration with CT images. Most, if not all, image registration algorithms are problematic when applied to the registration of PET and CT images because there is simply not enough common information between the two datasets (which makes the software PET-CT registration an extremely ill-posed problem). When dealing with general FDG-PET and CT registration, there is no exception for the proposed method because it relies on the common information between the two images to reliably map the control volumes. However, in a special case when the FLT tracer is used for PET imaging, the patient's bony structures have high FLT uptake and show up on the PET images and it is possible to use image registration algorithm to register the PET and CT images. For pelvic FLT-PET imaging, however, the bladder shows up as a high-intensity

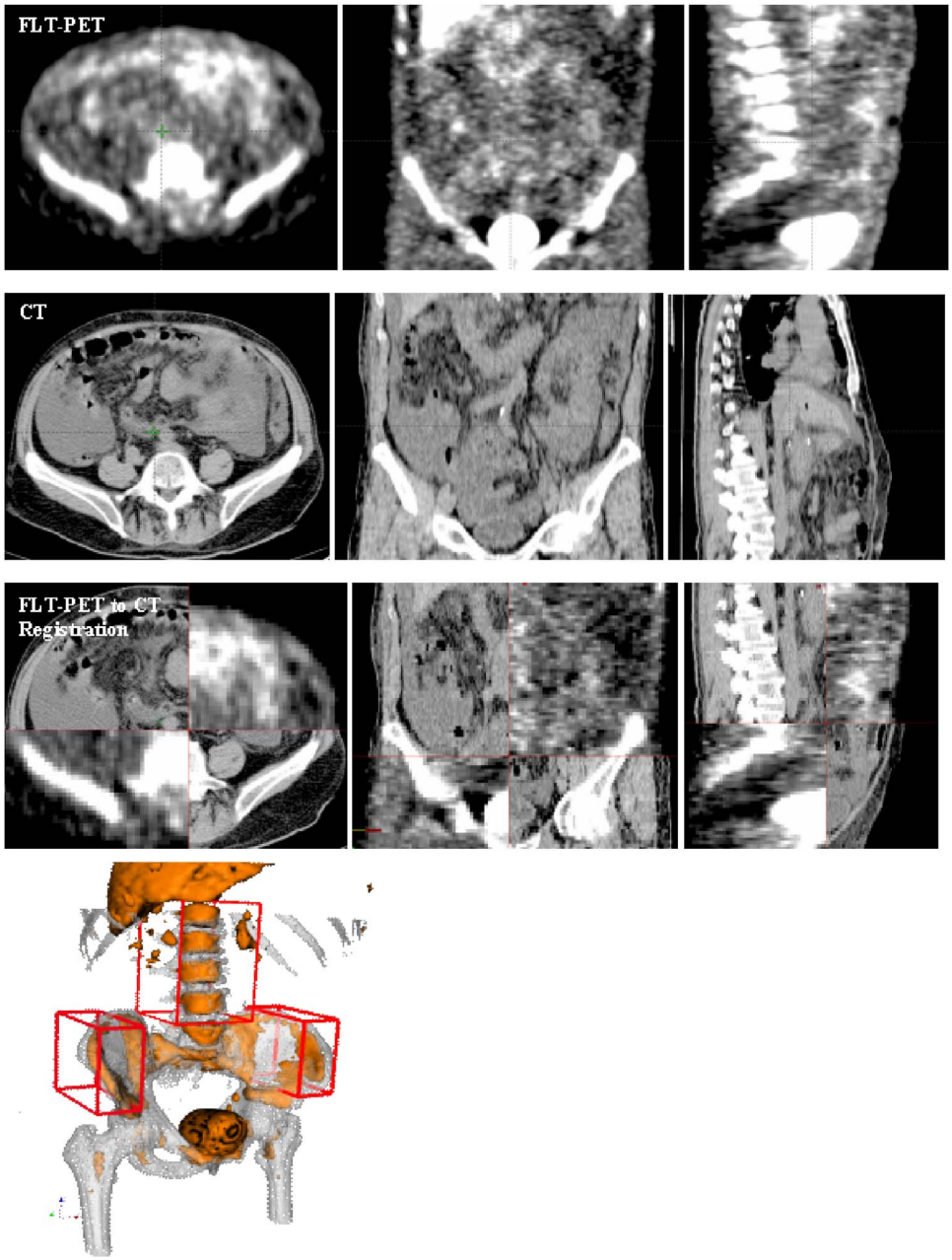


FIG. 5. Screenshots of sagittal, coronal, and axial views of the FLT-PET images (first row) and CT images (second row). The checkerboard of the CT and FLT-PET images after registration is shown in the third row. The fourth row shows a stereotactic view of the matched PET and CT images after registration.

region because of the accumulation of FLT in bladder, which has no correspondence in the CT images. The control volume-based registration affords an effective way for us to deal with the issue.

B. Study 2: Rigid registration of CT and MRI images for a brain case

In Fig. 7 we compare the control volume-based and conventional MI-based registration of CT and MRI data at four different slice locations for a brain case. In general, a MI metric is preferable for multi-modality image registration such as CT-MRI registration. Three cubic control volumes are placed in the model image. The control volume-based registration method outperforms the conventional approach in both computational speed and the quality of the registration. Indeed, the computing time for the two different techniques to find their "optimal" solutions is reduced by almost an order of magnitude in the control volume-based calculation. In the conventional registration, artifacts caused by the mask, tumor in the eye, and the denture modify the metric function and lead to unrealistic solutions. In Fig. 7, it is seen that the whole CT set is dragged down relative to its expected location. The misregistration is evident especially in the second row of Fig. 7 where the tumor visualized in the MRI has no correspondence in the CT image. This mismatch is eliminated in the control-volume based calculation, since only artifact-free volumes are considered in the metric calculation.

Starting from an initial spatial association of the two input images, we have performed 50 independent calculations,

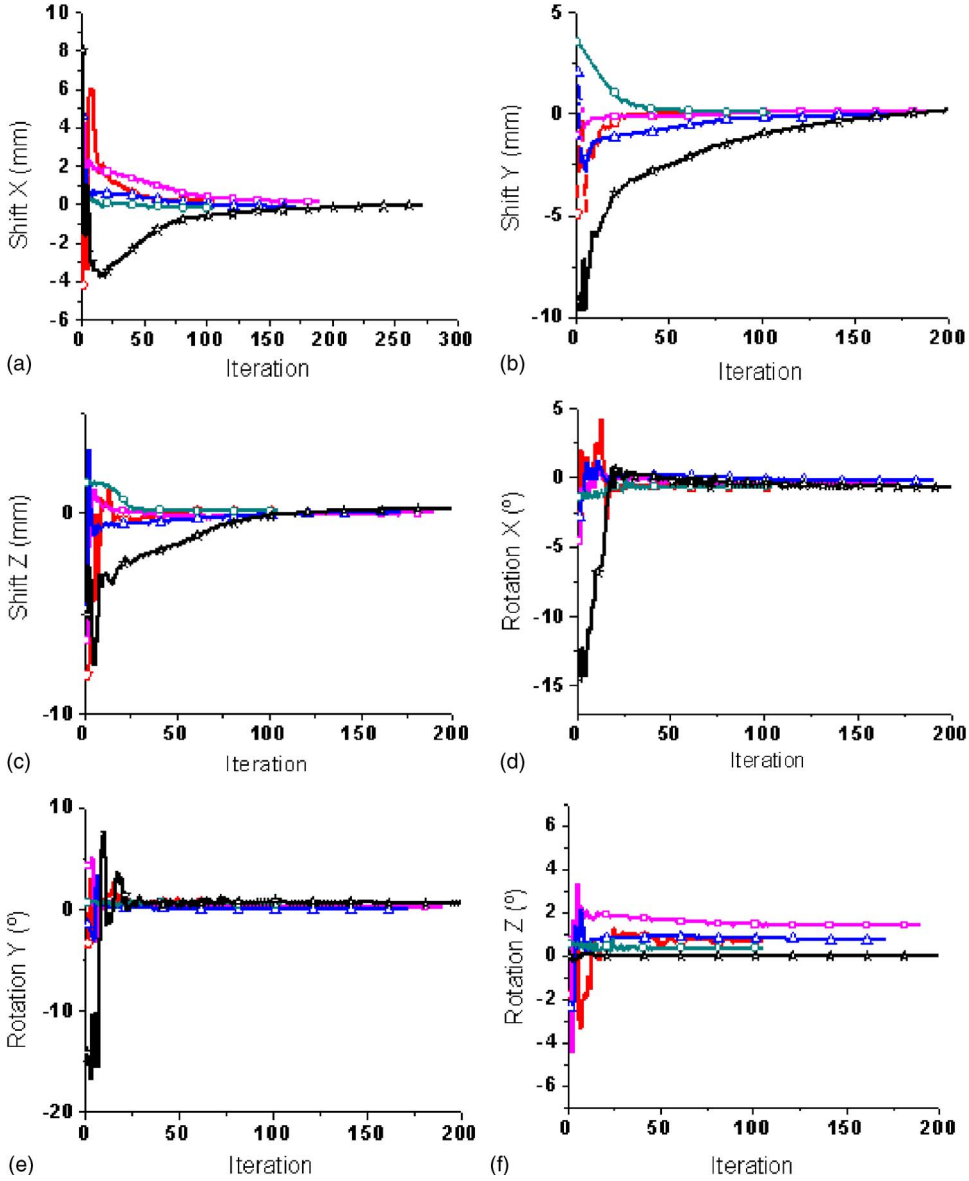


FIG. 6. Iterative calculation processes for five independent registrations starting from different initial matches for the patient shown in Fig. 5. Some of the curves stop earlier because the convergence criteria is met.

each with different initial trial transformation parameters. Similar to the PET-CT registration, the control volume-based algorithm yields almost identical shift values in the x, y, and z directions for all 50 runs. Interestingly, the average shifts for the 50 independent calculations in the conventional registration are fairly close to that of the control volume-based calculation: (6.0 mm, 44.7 mm, 82.7 mm) for the control volume-based versus (6.3 mm, 44.6 mm, 83.7 mm) for the conventional calculation. For the conventional approach, however, the final transformation parameters vary from test to test and do not converge to the same values. The variation range in the control volume-based calculation are all within 1.9, 3.0, and 2.8 mm in the x, y, and z directions, whereas the fluctuations in the conventional approach are much greater: 10.5, 6.7, and 13.6 mm, in the three directions. As seen from Fig. 8(d), the misregistrations for most individual runs in the x and z directions are particularly serious.

In addition, we have carried out quantitative evaluations of how registration results depend on the size and placement of control volumes for the brain CT-MRI registration. Three different sized control volumes were studied. For each size, the volume was placed at 20 randomly selected artifact-free locations and the registration parameters were then recorded. The results are summarized in Fig. 9. It is seen that as the control volume increases the fluctuation from one control volume placement to another decreases dramatically. A millimeter order of accuracy is achievable with a reasonable sized control volume. We wish to emphasize here that the MI metric used for CT-MRI registration is known to be statistically noisy. For intramodality registration such as CT-CT or CT-cone beam CT, our experience indicated that a control volume of $1-2$ cm^3 would result in a similar statistic as the data shown in the third column of Fig. 9.

C. Study 3: Deformable registration of exhale and inhale CT images for a thorax patient

The proposed method is applicable to both rigid and nonrigid image registration problems. The inputs to be registered

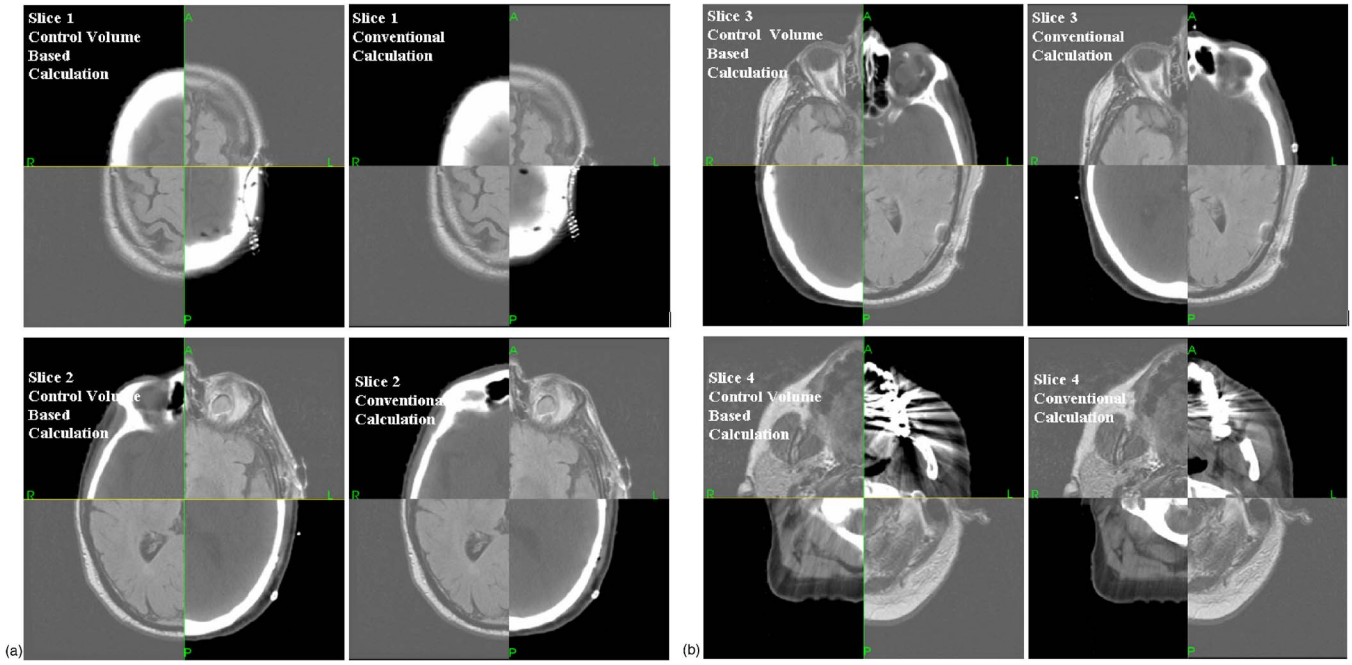


FIG. 7. Comparison of the newly proposed control volume-based (left column) and conventional whole image-based algorithms (right column) for rigid image registration. In conventional registration, the metric tries to accommodate all the voxels, which becomes less adequate in the presence of image artifacts or other noises. The control volume-based calculation eliminates the influence of imaging artifacts and produces better registration.

in this example are CT images acquired at expiration and inspiration [Figs. 10(a) and 10(b)]. Five control volumes are selected on the model image [Fig. 10(c)] and they are then mapped independently to their corresponding locations in the reference image [Fig. 10(d)].

Similar to the convergence analysis presented in the first example, in Fig. 11 we plot the NCC metric for three of the five control volumes when they are shifted away from their ideal match positions. To better understand the system, we have chosen two different sized control volumes: 20 mm (upper row) and 40 mm (lower row). In both situations, the metric space is smooth and has a very pronounced ridge. Obviously, this behavior is resulted from the fact that the intensity variation in the selected control volume is predominantly in the direction perpendicular to the boundary of the lung. When comparing the metric function for different sized control volumes, we note that the search space characteristics do not change significantly. For the size of 20 mm, the peaks and valleys of the search space are more pronounced because, as the number of pixels reduces, a small mismatch would lead to a large variation in the metric function. As the size of the control volume increases, the details of the function are "evened out" by the large number of voxels. In both situations, the positions of the minimum are found to be at $(x,y,z)=(-5.5,-11.3,3.4)$, $(-1.8, -12.0, 3.3)$, and $(-3.8, 2.6, 0.6)$ for the three control volumes, respectively, where the coordinates are in millimeters. In practice, determining the size of a control volume is a matter of tradeoff between a few factors, such as the accuracy, sensitivity against a small mismatch of the control volume, and calculation speed.

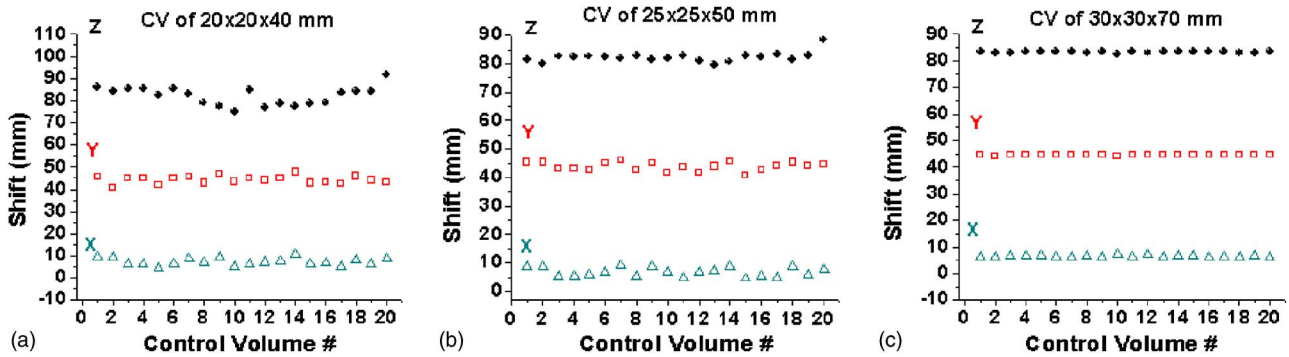


FIG. 8. Convergence analyses of a rigid brain CT-MRI registration using the control volume-based and whole image-based algorithms. In the former case, all 50 calculations with different initial transformation parameters converge to the same shift values in the x, y, and z directions, labeled by triangles, squares, and circular dots, respectively. For the latter case, the fluctuation in the final shift values are much larger.

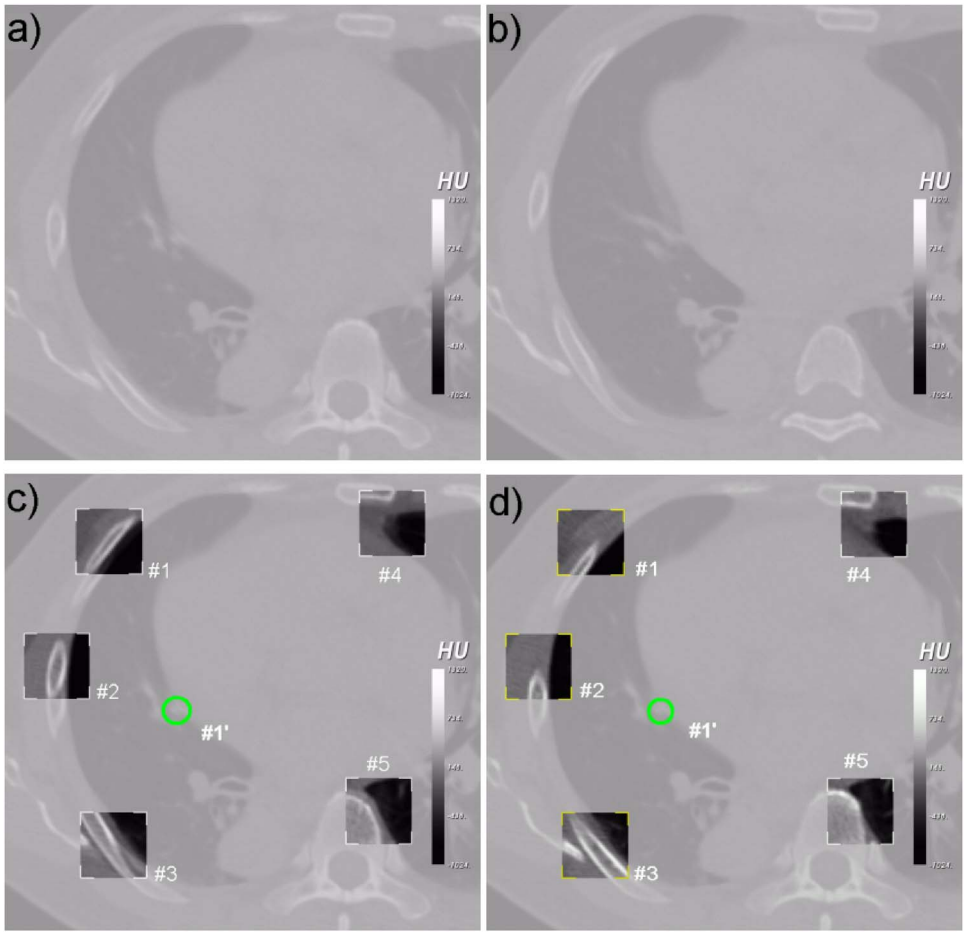


FIG. 9. Dependence of registration on the placement of the control volume for three different sized volumes. For each size, the shifts resulting from 20 random placements of the control volume are shown.

After the mapping of the selected control volumes, a warping using the spline model is used to register the remaining part of the image with the control volumes serving as the control points. To have enough control points to warrant a robust spline calculation, we added one control point inside the region encompassed by the points. Because the added point is in a deformable region, it is less justifiable to use the auto-mapping procedure described above to find its mirror point on the reference image. Following the step outlined in Fig. 1(b), we used a B Spline-like algorithm to optimally locate its mapping with the known nodal points #1–#5 as constraints. The mapped point is shown in Fig. 10(d) by a circle. The final result of the deformable registration is shown in Fig. 10(d). The model image is represented as semi-transparent background in Fig. 10(c). The selected rectangular control volumes are shown in Fig. 10(c) as overlays on the background image, with the image in each control volume cropped from the reference image to assess the

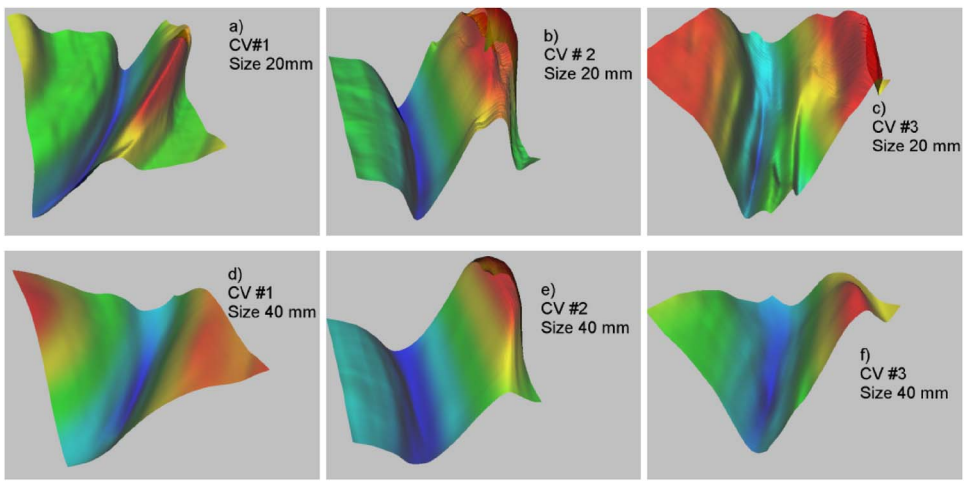


FIG. 10. Axial slices of the model (a) and reference (b) images for the deformable registration study. Rectangular control volumes, #1 to #5, are placed in the model image and their mappings are shown on the reference image. The anatomy in the regions affected by the respiration does not match initially (c). After the control volume-based deformable registration, the anatomy matches very well, as confirmed by the overlay image (d).

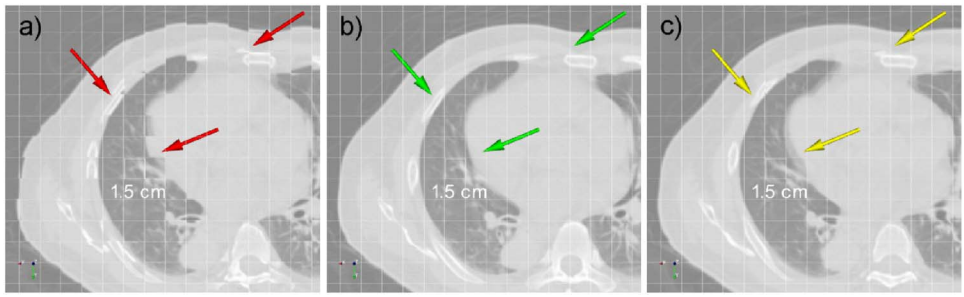


FIG. 11. Metric function for the first three control volumes for two different control volume sizes (20 mm for upper row and 40 mm for the lower row). In both cases, the search spaces are smooth, with a very pronounced ridge. The search space characteristics and the location of the global minima do not depend significantly on the control volume size.

difference between the two input images. After registration, no difference between the mapped model image and the reference image is observable [Fig. 10(d)].

The results obtained with the control volume-based method are compared with that obtained using a "brute-force" BSpline formalism.[22] In Fig. 12(a) we show the checkerboard display of the two input images as given in Figs. 10(a) and 10(b). Before registration, deformations up to 1 cm in the two phases are clearly observable in the checkerboard transition zone, as marked by red arrows. Both control volume-based and the BSpline models [Fig. 12(b)] are able to register the two images. The mapped inhale images from the two different approaches are displayed in Fig. 12(b) using a checkerboard tool. The difference between the two mapped images is found to be less than 2 mm in any region. The control volume-based registration took only a few minutes to complete the calculation process, whereas one–two hours are required by the conventional BSpline technique to find the solution because it was needed for computer to optimize the metric function with respect to a large number of nodal variables.

The convergence behavior of the deformable registration is assessed in a similar fashion as described previously. For a pair of control volumes with intentionally introduced starting mismatch, we repeated the registration calculation 100 times with randomly assigned initial transformation parameters. The resultant transformation parameters of the first three control volume pairs for the 100 independent registration calculations are plotted in Fig. 13. For the first control volume, the shifts in x, y, and z directions are found to be (−5.5, −11.3, 3.4). The transformation parameters in the three directions resulted from the calculations are all within the range of 0.9, 1.3, and 1.5 mm, respectively. Similar convergence is achieved for the other control volumes. The variation ranges for the second and third regions in the directions, for example, are (1.4, 1.9, 1.3) and (1.1, 1.6, 1.5) in millimeters, respectively.

IV. DISCUSSION

With the recent advancements in image-guided radiation therapy (IGRT),[23] multi-modality imaging becomes increasingly important. Full realization of the potential of IGRT and highly conformal IMRT would be impossible without a robust and efficient image registration technique. In general, image registration has two important aspects: formulation of the problem and optimization of the transformation parameters required to match the two input images. In this work we proposed a general two-step registration technique, in which homologous pairs of control volumes are obtained using an auto-mapping algorithm and the pairs are then used as the *a priori* knowledge of the system to facilitate the registration process. In particular, we present our experience with the selection of model parameters, optimization algorithm, and the validation of the technique. The development of the proposed technique was motivated by an intuitive observation that, in image registration, the reliability of information in an image is generally spatially heterogeneous and some regions in the model image can be more reliably mapped to the reference image than others. In other words, there are at least

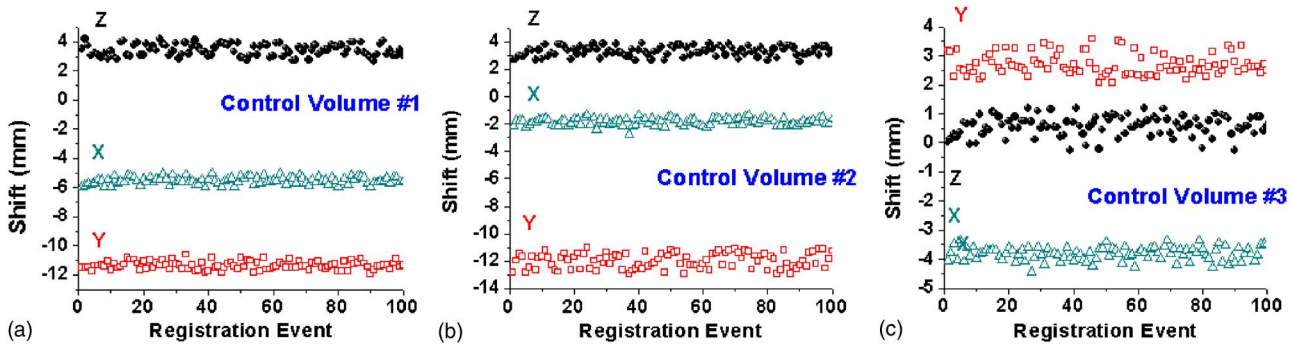


FIG. 12. Checkerboard comparison of the model and reference images before (a) and after (b) control volume-based deformable registration. Displacement up to 1 cm is visible before registration, as marked by the arrows in (a). The control volume-based algorithm yielded the same registration as that of the conventional BSpline calculation, but with a significantly reduced computational time. A checkerboard comparison of the mapped model images from the two algorithms is shown in (c).

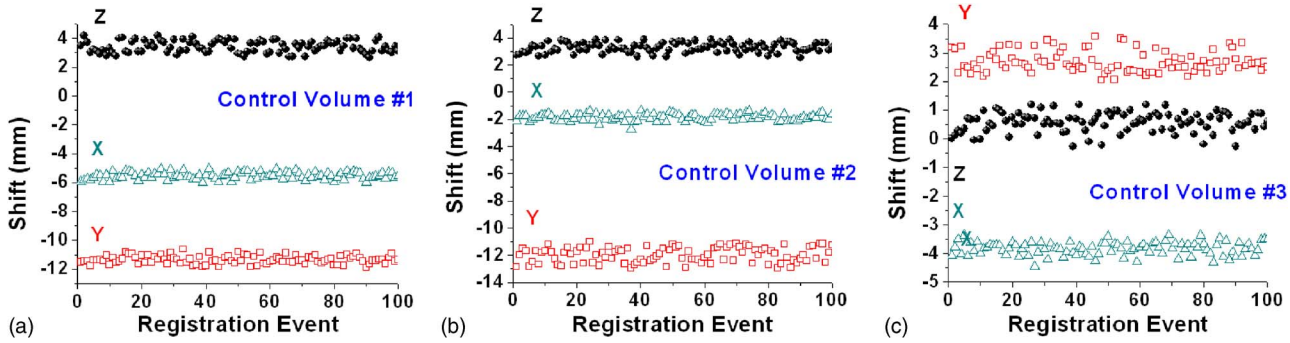


FIG. 13. Convergence analysis for the first three control volumes shown in Fig. 8. Presented are the shifts of the final transformation as obtained in 100 registration tests starting from different initial transformation parameters. All calculations converged to the same solution to within a range of less than 2 mm for the x (triangles), y (squares), and z (circular dots) directions.

two types of regions in a model image: regions whose correspondence in the reference image are easily identifiable and regions whose mapping to the reference image is less obvious for a variety of practical reasons, such as tissue deformation, breathing movement, lack of distinct image features, image artifacts, or inherent difference in the imaging modality. Mapping of the two types of regions should be treated differently, rather than equally as implemented in current image registration algorithms. The use of control volumes placed on the reliable regions permits us to eliminate the uncertainty and/or disturbance arising from the less reliable regions, thus improving the performance and fidelity of the image registration calculation.

A control volume is generally placed in a region where the anatomical correspondence between the model and reference images can be easily realized by a quick visual judgment. In the case of rigid registration, the mapping of K ($K \geqslant 3$) control volumes yields $K+1$ sets of transformation parameters, in which K from individual control volume and one from the coordinates of the center of mass of all the control volumes. Each set of transformation parameters include three translational and three rotational parameters. The first K sets of transformation parameters tend to be regional, whereas the last one global. In our calculation, the final registration parameters are derived by averaging the $K+1$ sets of transformation parameters. Obviously, more sophisticated estimation based on the $K+1$ sets of transformation data is possible. In particular, some consistency check between the $K+1$ sets of data can be imposed to ensure that the parameters derived from the mapping of the individual control volumes are all within a certain tolerance. This should be able to alert the user if an affine transformation is needed to better describe the transformation data. The approach is superior over the existing registration technique based on the auto-matching of the two whole image sets because it is not influenced by the local disturbance from imaging artifacts. In reality, difference in image features due to the use of different acquisition protocols may constitute "noise" to the registration algorithm and lead to inaccurate or unrealistic registration. Using control volumes provides an effective way to avoid the disturbance and allows us to fully utilize the features that are known with high confidence to enhance the fidelity of the registration. The two-step registration technique has the advantage of the landmark-based registration without having to go through the hassle of interactively selecting a homologous control point pairs. The proposed technique also eliminates the subjectivity often associated with the landmark-based registration and offers operator independence and reproducibility. In reality, as noted by West *et al.*,[2] a careful visual inspection of the results obtained from any automatic registrations is crucial in a clinical setting to ensure that the registration found by a computer is clinically reasonable.

For deformable registration, the benefit of the two-step registration is even more significant. In addition to the increased robustness and confidence level, the technique greatly speeds up the calculation process. In conventional BSpline method, for example, the mapping of the nodes or the transformation of the nodes from the model image to the reference image is accomplished by a searching algorithm, which treats the transformation parameters of the nodes as interrelated system variables and hence necessitates an iterative examination of a huge number of trial parameters under the guidance of the registration metric. The amount of computation is proportional to $O(N^3)$, where N is the number of the nodes. In the proposed two-step registration, the selected control volumes can be regarded as spline control points. The determination of most, if not all, nodes is done independently, which essentially reduces the calculation into an $O(N)$ problem. In reality, the number of nodes chosen by independent mapping may not be enough to cover the whole patient volume to yield a reliable deformation field everywhere inside a patient. In this situation, additional nodes may be introduced by using the conventional BSpline method with the positions of the added nodes determined collectively (the control volume method-mapped nodes serve as constraints in the modified BSpline method). The modified or hybrid BSpline registration method, in which one first maps a fraction of nodes located on the reliable regions through the control volume mapping and then proceeds with the standard BSpline calculation for the determination of remaining nodes, also reduces the dimensionality of the deformable image registration problem. This not only significantly speeds up the calculation but also improves the convergence behavior of the registration. The reduction in computational time

may have practical implication as the radiation therapy is moving toward more sophisticated image-guided patient localization and dose delivery and verification.[23]

Our study has also shown that the use of control volumes can greatly simplify the optimization calculation incurred in the image registration problem. Over the years, powerful but computationally intensive algorithms, which are capable of statistically escaping from local minima, such as simulated annealing or genetic algorithms, have often been used to attack the optimization problem. Even with these stochastic algorithms, finding the true solution of the problem can still be challenging because of the wide variety of possible pixel intensities within an organ and the complex behavior of the metric function. This has been clearly demonstrated in the examples presented above. The distinct feature of the control volume smoothens the metric function space and makes it simple for the search algorithm to converge to the global solution.

V. CONCLUSIONS

In conventional techniques the information contained in the model and reference images is employed as input entity without "filtering" or "prioritization." As thus, any artifact is treated as part of the input information and adversely affects the final registration result. We have proposed a novel method for auto-selection of the control volumes and described a two-step technique for both rigid and deformable image registrations. Instead of relying on the interactive selection of homologous control point pairs on both model and reference images, the user needs only to identify some control volumes on the model image in a somewhat arbitrary fashion. The approach has two major advantages: (*i*) It allows us to incorporate *a priori* knowledge into the image registration process to avoid/reduce potential registration errors caused by image artifacts. (2) It improves the computational efficiency because the control volume involves fewer voxels. The method was applied to both rigid and nonrigid image registration problems and our results indicated that the registration is reliable and provides a valuable tool for intra- or intermodality image registration. The increased robustness and confidence in the registration and the improved convergence behavior of the calculation are the important features of the new technique. Compared to the manual rigid registration, this method eliminates the nuisance of the control point pair selection and removes a potential source of error in registration. Compared to the automated method, the approach is more intuitive and robust, especially in the presence of image artifacts. Thus the method should find useful applications in radiation oncology and other image-guided intervention procedures.

ACKNOWLEDGMENTS

We would like to thank Dr. A. Koong, Dr. A. Quon, Dr. B. Thorndyke, Dr. T. Li, Dr. Y. Yang, Dr. D. Levy, and Dr. D. Paquin for useful discussions. This work is supported in part by research grants from the U.S. Department of Defense (DAMD17-03-1-0023), the Susan G. Komen Breast Cancer Foundation (BCTR0504071), and the National Cancer Institute (5 R01 CA98523-01).

[a]Author to whom correspondence should be addressed. Electronic mail: lei@reyes.stanford.edu

[1]H. Wang, L. Dong, M. F. Lii, A. L. Lee, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung, "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **61**, 725–735 (2005).

[2]H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," Phys. Med. Biol. **50**, 2887–2905 (2005).

[3]A. Bharatha, M. Hirose, N. Hata, S. K. Warfield, M. Ferrant, K. H. Zou, E. Suarez-Santana, J. Ruiz-Alzola, A. D'Amico, R. A. Cormack, R. Kikinis, F. A. Jolesz, and C. M. Tempany, "Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging," Med. Phys. **28**, 2551–2560 (2001).

[4]K. K. Brock, S. J. Hollister, L. A. Dawson, and J. M. Balter, "Technical note: creating a four-dimensional model of the liver using finite element analysis," Med. Phys. **29**, 1403–1405 (2002).

[5]D. Yan, D. A. Jaffray, and J. W. Wong, "A model to accumulate fractionated dose in a deforming organ," Int. J. Radiat. Oncol., Biol., Phys. **44**, 665–675 (1999).

[6]X. Wu, S. J. Dibiase, R. Gullapalli, and C. X. Yu, "Deformable image registration for the use of magnetic resonance spectroscopy in prostate treatment planning," Int. J. Radiat. Oncol., Biol., Phys. **58**, 1577–1583 (2004).

[7]J. Lian, L. Xing, S. Hunjan, B. Spielman, and B. Daniel, "Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: a deformable registration and validation study," Med. Phys. **31**, 3087–3094 (2004).

[8]W. Lu, M. L. Chen, G. H. Olivera, K. J. Ruchala, and T. R. Mackie, "Fast free-form deformable registration via calculus of variations," Phys. Med. Biol. **49**, 3067–3087 (2004).

[9]M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," Med. Phys. **31**, 2942–2948 (2004).

[10]E. Schreibmann, G. T. Y. Chen, and L. Xing, "Image interpolation in 4D CT using a BSpline deformable registration model," Int. J. Radiat. Oncol., Biol., Phys. (2006).

[11]E. Schreibmann and L. Xing, "Image registration with auto-mapped control points," Disclosure filed to the Office of Technology Licensing of Stanford University (2005).

[12]L. Ibanez, W. Schroeder, and L. Ng, "ITK Software Guide," Kitware Inc. (2003).

[13]D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," Math. Program. **45**, 503–528 (1989).

[14]R. H. Byrd, P. Lu, J. Nocedal, and C. Y. Zhu, "A limited memory algorithm for bound constrained optimization," SIAM J. Sci. Comput. (USA) **5**, 1190–1208 (1995).

[15]M. Lahanas, E. Schreibmann, and D. Baltas, "Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms," Phys. Med. Biol. **48**, 2843–2871 (2003).

[16]S. Lee, G. Wolberg, K. Y. Chwa, and S. Y. Shin, "Image metamorphosis with scattered feature constraints," IEEE Trans. Vis. Comput. Graph. **2**, 337–354 (1996).

[17]B. Schaly, G. S. Bauman, J. J. Battista, and J. Van Dyk, "Validation of contour-driven thin-plate splines for tracking fraction-to-fraction changes in anatomy and radiation therapy dose mapping," Phys. Med. Biol. **50**, 459–475 (2005).

[18]A. K. Buck, G. Halter, H. Schirrmeister, J. Kotzerke, I. Wurziger, G. Glatting, T. Mattfeldt, B. Neumaier, S. N. Reske, and M. Hetzel, "Imaging proliferation in lung tumors with PET: 18F-FLT versus 18F-FDG," J. Nucl. Med. **44**, 1426–1431 (2003).

[19]T. Li, E. Schreibmann, B. Thorndyke, G. Tillman, A. Boyer, A. Koong, K. Goodman, and L. Xing, "Radiation dose reduction in 4D computed tomography," Med. Phys. **12**, 3650–3660 (2005).
[20]E. Rietzel, G. T. Chen, N. C. Choi, and C. G. Willet, "Four-dimensional image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion," Int. J. Radiat. Oncol., Biol., Phys. **61**, 1535–1550 (2005).
[21]A. Trofimov, E. Rietzel, H. M. Lu, B. Martin, S. Jiang, G. T. Chen, and T. Bortfeld, "Temporo-spatial IMRT optimization: concepts, implementation and initial results," Phys. Med. Biol. **50**, 2779–2798 (2005).
[22]D. Mattes, R. D. Haynor, H. Vesselle, K. T. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," IEEE Trans. Med. Imaging **22**, 120–128 (2003).
[23]L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T. Li, G. Y. Kim, G. Luxton, and A. Koong, "Overview of image guided radiation therapy," Med. Dosim 1 (2006).

Int. J. Radiation Oncology Biol. Phys., Vol. 64, No. 5, pp. 1537–1550, 2006
Copyright © 2006 Elsevier Inc.
Printed in the USA. All rights reserved
0360-3016/06/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2005.11.018

PHYSICS CONTRIBUTION

IMAGE INTERPOLATION IN 4D CT USING A BSPLINE DEFORMABLE REGISTRATION MODEL

EDUARD SCHREIBMANN, PH.D.,* GEORGE T. Y. CHEN, PH.D.,† AND LEI XING, PH.D.*

*Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA; †Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA

Purpose: To develop a method for deriving the phase-binned four-dimensional computed tomography (4D CT) image sets through interpolation of the images acquired at some known phases.

Methods and Materials: Four-dimensional computed tomography data sets for 3 patients were acquired. For each patient, the correlation between inhale and exhale phases was studied and quantified using a BSpline deformable model. Images at an arbitrary phase were deduced by an interpolation of the deformation coefficients. The accuracy of the proposed scheme was assessed by comparing marker trajectories and by checkerboard/difference display of the interpolated and acquired images.

Results: The images at intermediate phases could be derived by an interpolation of the deformation field. An analysis of marker movements indicated that 3 mm accuracy is achievable by the interpolation. The subtraction of image analysis indicated a similar level of success. The proposed technique was useful also for automatically mapping the organ contours in a known phase to other phases, and for designing patient-specific margins in the presence of respiratory motion. Finally, the technique led to a 90% reduction in the acquired data, because in the BSpline model, a lattice of only a few thousand values is sufficient to describe a CT data set of 25 million pixels.

Conclusions: Organ deformation can be well modeled by using a BSpline model. The proposed technique may offer useful means for radiation dose reduction, binning artifacts removal, and disk storage improvement in 4D imaging. © 2006 Elsevier Inc.

Four-dimensional computed tomography, Deformable registration, Image-guided radiotherapy, Delineation, Respiratory motion.

INTRODUCTION

Respiratory motion degrades anatomic position reproducibility during imaging (1–10), distorts the shapes of the tumor target, displaces the organs, and causes errors during radiation delivery (11–15). In the presence of breathing motion, tumors and sensitive structures in the thorax and abdomen can move as much as 2–3 cm, posing a significant challenge for radiation therapy planning and delivery. Until recently, tumor motion has been handled primarily by using respiratory-gating or breath-hold technique (16). A more advanced and potentially more beneficial approach is four-dimensional computed tomography (4D CT), which adopts techniques developed for cardiac CT imaging and allows one to acquire image data at specified phases over several respiratory cycles and then combines the data into phase-binned images. Vedam *et al.* (9), Low *et al.* (17), and Rietzel *et al.* (18) have refined the imaging techniques developed for phase binning based on cardiac motion and applied them to the problem of respiratory motion. The signals used to stamp the time point of the image data are either from a respiration monitor (RPM; Varian Medical Systems, Palo Alto, CA) (9) or spirometry-based tidal volume measurement (17). By these methods, thoracic 4D CT images accounting for respiratory motion have been successfully acquired using single-slice scanners (8, 9) and multislice 4D CT scans (17, 19, 20). Also, 4D cone beam CT scans have been acquired using a benchtop system (21) and on-board cone beam CT (22, 23).

Although the 4D CT provides a powerful tool to study respiratory motion, a hurdle in realizing all the potential gains is the need to acquire images for all breathing phases. In the above-mentioned methods, the data acquisition is

Reprint requests to: Lei Xing, Ph.D., Stanford University School of Medicine, Department of Radiation Oncology, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Tel: (650) 498-7896; Fax: (650) 498-4015; E-mail: lei@reyes.stanford.edu

This work is supported in part by a research grant from the National Cancer Institute (1 R01 CA98523-01) and the U.S. Department of Defense (DAMD17-03-1-0023).

Acknowledgments—We would like to thank Drs. G. Mageras, Y. Erdi, and S. Nehmeh from Memorial Sloan-Kettering Cancer Center for providing some of the 4D CT images used in this study. We would like to express our appreciation to Drs. A. Koong, Q. T. Le, K. Goodman, B. Loo, T. Li, B. Thorndyke, and Y. Yang for the helpful discussions.

Received Feb 28, 2005, and in revised form Nov 11, 2005. Accepted for publication Nov 16, 2005.

"brute force" in nature, and 10–20 sets of phase-resolved 3D CT images are needed. In addition to greatly increasing the workload of the CT scanner, the radiation dose to the patient becomes a major concern (24). The resultant data are typically comprised of 1,500–3,000 images occupying several hundred megabytes (MB) of disk space. Additionally, the approaches may lead to artifacts if the correlation between RPM spirometer and scanner (17) is not accurate.

A 4D CT can be either prospective or retrospective. In the former case, the scanner collects images at only one of the breathing phases of the patient instead of scanning continuously. Thus, the system acquires a series of contiguous images at appropriate predetermined phases and creates a single volumetric image corresponding to a specific phase. The retrospective 4D CT scan results in multiple image sets, corresponding to different breathing phases of the patient. It consists of three relatively orthogonal processes (9, 17, 18, 24): recording of respiratory signal(s), acquisition of time-dependent CT projection data, and construction of a 4D image from these data. The question we ask in this research is that, given a few sets of phased 3D CT (for example, 3D CT images obtained at inhale and exhale points), is it possible to deduce the intermediate phases by warping or "interpolating" the images with a deformable registration model? If successful, the method can greatly reduce the radiation dose to the patient while maintaining the benefits of 4D CT. As a byproduct, the contour information outlined for one phase can be automatically transformed to the other phases. A major task in this endeavor is the determination of patient-specific organ deformations occurring between inhale and exhale phases. Several relevant image registration techniques have been reported in the literature (25–38). A deformable procedure based on the finite element model, where images are described as blocks of elastic materials on which forces apply, was proposed by Bharath *et al.* (25) and Brock *et al*. (26). However, the values of the elasticity and density constant for various tissues are not readily available and have to be found by a trial and error procedure. A fluid flow registration was applied in radiation therapy to automatically warp contours delineated in one phase into next phases (27). Recently, a simpler technique based on spline interpolation was proposed (28). This model uses only a lattice of nodes overlaid on the image, where deformation at any location in the image is deduced by spline interpolation of the closest node coefficients (29). In reality, the spline coefficients can come also from a set of user-defined control points, as was done by Fei *et al.* (28) in a study of warping and rigid registration of magnetic resonance volumes. The simplicity and robustness of the free-form deformation (BSplines) defined on discrete nodes make it useful for clinical applications (30, 31). This method was applied also to directly register prostate CT and MRI/MRSI and was validated by using a series of phantom measurements (32, 33). Rohlfing *et al.* (34) and Kaus *et al.* (35) used a BSpline-based registration to study liver deformation, and an accuracy of 3 mm was achieved. Berlinger *et al*. (36) and Schweikard *et al*. (37) obtained synthetic digitally reconstructed radiographs at different breathing phases using the same approach. Coselmon *et al.* (39) used a similar technique to study the accuracy of mutual information–based CT registration of the lung at exhale and inhale respiratory states.

In the following text, we present a general BSpline model for deriving the images at the intermediate phases by starting from two or three image sets at some distinct phase points such as inhale/exhale phases. The performance of the algorithm is assessed by monitoring the displacements of implanted or surface markers. The general reference drawn from this study is that it may not be necessary to acquire the CT images at all phases to obtain the detailed 4D picture of a patient. The information at a few distinct phases is sufficient, because the behavior of the system at the intermediate phases can be reasonably predicted by using a deformable image registration model. Although the focus of this study is to deduce the images at the intermediate phases, the technique developed here can be generalized also for auto-segmentation of the organs at all phases by starting from the contours at a given phase point. The tool can be employed also to analyze the geometric displacements of various organs for the determination of the minimal target margin in presence of respiratory motion.

METHODS AND MATERIALS

Image acquisition

The 4D CT images of the first patient were acquired by using a GE Light Speed QX/I scanner (GE Medical Systems, Milwaukee, WI). Ten phase bins were set for the 4D study, and 110 images with a 3.75-mm slice thickness were acquired for each phase. The recorded exposure time was 800 s. The 10 breathing phases recorded contained over 402 MB of data in DICOM image format. The patient had three fiducial markers implanted in liver, kidney, and vertebral body.

Two additional cases (a female and male torso) were acquired using a GE Discovery QX/I CT scanner. In each case, 10 phases, each consisting of 90 images with 2.5-mm slice thinness, were acquired. The size of the 4D scans was 402 and 502 MB, respectively. A single marker was placed on the patient's abdomen or torso to track the displacement. The image acquisition time was 500 s, and exposure time was 750 s.

The study was approved by the Institutional Review Board, and all participants gave written informed consent. For each patient, the 10 bins of the 4D CT data set spanned over the whole breathing cycle and were indexed from CT0 to CT90, with CT0 corresponding to the start point of the expiration and CT40 the full inspiration point. Image sets from CT10 to CT30 represent the patient during the inspiration, whereas CT50 to CT80 represent expiration.

We selected CT0 and CT40 as the input to the image interpolation algorithm with the intent of deriving the images CT10, CT20, and CT30. The synthetic images deduced by the registration algorithm are termed "interpolated" images. The effort was focused on the inspiration process, but the same calculation can be similarly done for other phase points. To assess the accuracy of the image warping, the interpolated images were compared quantitatively with the actual CT10, CT20, and CT30 images acquired during the 4D CT scanning.

Software and hardware platform

The image warping software was implemented using two related open-source software toolkits named ITK (40) and VTK (41). ITK consists of template-based code for a large number of image registration algorithms and was used in this work for the image registration study. VTK consists of tools for 3D visualization and contouring and was used to determine the marker positions, extract and interpolate surfaces, and compute distances between corresponding surfaces. For convenience, in the following text we outline the methods used in our calculation with attention paid to the issues specific to CT image registration. All calculations were done on a standard PC computer with a Windows XP operating system and a Pentium 4 Processor at 1.6 GHZ, 256 MB of RAM. The DICOM protocol was used for image data communication and transfer. The VTK and ITK have a DICOM filter configured to read image files in the DICOM format.

Coregistration of images at known phases

There are two problems that need to be solved to derive the images at all intermediate phases from sets of 3D images acquired at a few known phase points. First, the known image sets must be registered using a deformable registration model. An interpolation can then be performed to obtain the images at the phases falling between the known phases (see sections below). The interpolated images find their phase tag in the respiratory cycle through the phase monitoring curve obtained using external means such as an RPM or spirometry device. The overall process of the calculation is depicted in Fig. 1.

Image coregistration is to map a floating image to a fixed or target image by using a mathematical model. The process is shown in the top part of Fig. 1. The input to the registration software is the images to be registered: a fixed image and a floating image, described by their intensity distributions $I_a(x)$ and $I_b(x)$, respectively. In our calculation, the fixed image was CT0, and the floating image was CT40. The resulting transformation describes the 3D deformation field describing the patient anatomy change from CT0 (starting inspiration) to CT40 (ending inspiration).

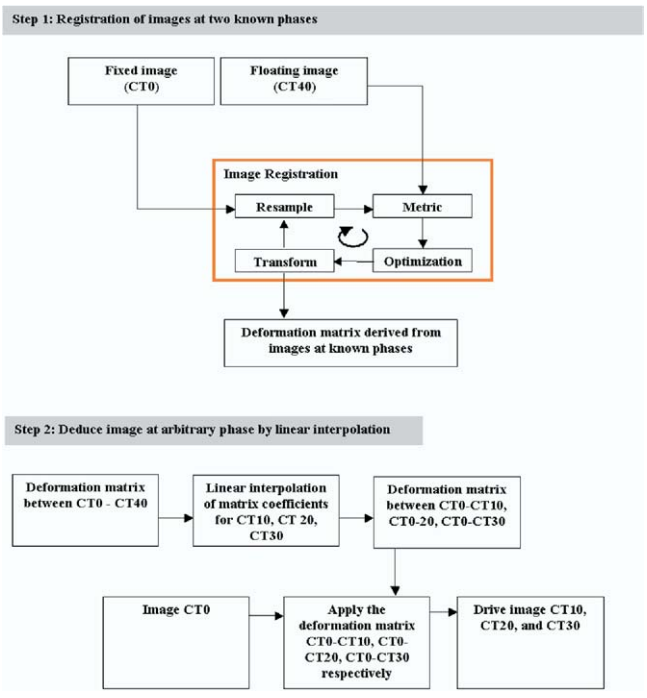


Fig. 1. Flow chart of the image interpolation process. First, a deformable image registration procedure deduces patient-specific displacements between the inhale and exhale phases. The displacements are subsequently used to create interpolated images at intermediary phases.

For convenience, the image is divided into a grid with N3 cells. The corner of a lattice cell is referred to as a node and is indexed by i ($i = 1, 2, \ldots$ N3). The displacement of a node, i, is specified by a vector, x_i, and the displacement vectors, $\{x_i\}$, of a collection of nodes characterize the tissue deformation. The displacement at a location, x, on the image is deduced by fitting a polynomial expressed using the basis spline (BSpline) (42, 43) to the grid nodes x_i. Unlike other spline models, the BSplines are locally controlled. That is, the displacement of an interpolation point is influenced only by that of the closest grid points, and changing a lattice node affects the transformation only regionally, making it efficient in describing local deformations.

Mathematically, the task of image registration is to find the transformation matrix, $T(x)$, that maps an arbitrary point, x, on the fixed image to the corresponding point, x', on the floating image (or vice versa) so that the best possible match, as measured by the registration metric, is achieved. As shown in the top part of Fig. 1, the image registration proceeds in an iterative fashion. The matrix coefficients of $T(x)$ are the node displacements and are adjusted iteratively to minimize the normalized cross correlation (NCC) between the two images, defined as follows (Eq. 1):

$$f = -\frac{\sum_{i=1} I_a(x_i)\, I_b(Tx_i)}{\sqrt{\sum_{i=1} I_a^2(x_i) \sum_{j=1} I_b^2(Tx_j)}}, \qquad (1)$$

where i and j are the node indices on the fixed image, $I_a(x_i)$ is the intensity of the node at x_i on the fixed image a, and $I_b(Tx_i)$ is the intensity of the image b at where the node x_i is mapped.

Optimization of the NCC function with respect to the displacements of the nodes, $\{x_i\}$, yields the transformation coefficients $T(x)$ that map the points of image a to image b. Because the two images do not necessarily have the same size, an interpolation may be needed to compute intensity at a mapped point, $x = T(x)x$. To facilitate the optimization, it is preferable that both the deformable model and the metric are differentiable (30). This condition is satisfied for the system that we are dealing with, as demonstrated in a previous mathematical study (44).

It was found by previous researchers (30, 31, 45) that a lattice with spacing of $<$40 mm is needed to model liver deformations. For all cases, we used a lattice of 15 nodes for each dimension, corresponding to a spacing of approximately 30 mm. Because three variables are associated with each node, this setup leads to a problem of 10,125 variables, requiring the use of efficient optimizers to find the minimum of the NCC metric.

To optimize the system, we used the limited memory BFGS algorithm (L-BFGS) (46), which is known for its superior performance in dealing with high-dimensionality problems. Starting from a positive definitive approximation of the inverse Hessian H_0 at x_0, L-BFGS derives the optimization variables by iteratively searching through the solution space. At an iteration, k, the calculation proceeds as follows:

1. Determine the descent direction $p_k = -H_k \nabla f(x_k)$.
2. Line search with a step size $\alpha_k = \arg\min_{\alpha>0} f(x_k + \alpha p_k)$, where $\alpha \geq 0$ is the step size defined in the L-BFGS software package.
3. Update $x_{k+1} = x_k + \alpha_k p_k$.
4. Compute H_{k+1} with the updated H_k.

At each iteration, a backtracking line search is used in L-BFGS (46) to determine the step size of movement to reach the minimum of f along the ray $x_k + \alpha p_k$. For convergence, α has to be chosen such that a sufficient decrease criterion is satisfied, which depends on the local gradient and function value and is specified in L-BFGS by the Wolfe conditions (46).

During an image registration process, the above iterative calculation based on L-BFGS algorithm continues until the following stopping criterion is fulfilled:

$$\frac{\|\nabla f(x_k)\|_2}{\max(1, \|x_k\|_2)} < \varepsilon \qquad (2)$$

with $\varepsilon = 10^{-6}$ in this study, or a preset maximum number of iterations (typically, 50 iterations) is reached. In reality, tracking the value of registration metric during the iterative calculation also provides useful information and was used to assess the convergence of the optimization algorithm.

Performance of image coregistration

Image coregistration is to relate the two input images: the fixed and the floating images (see Fig. 1). Ideally, the mapped floating image and the fixed image should be identical after the registration. To evaluate the performance of the BSpline algorithm, we used the popular checkerboard display tool, where images to be compared are merged together in a chess-like pattern. At the transition zone of the squares, difference in the two images can be easily visualized and evaluated. If two images are similar, no differences should be observed in the checkerboard tool. An analysis of the subtraction image of the two registered images was also carried out to evaluate the closeness of the two images. Histograms characterizing the fractional voxels for a range of the Hounsfield number (HU) differences were presented before and after deformable image registration.

Image interpolation based on the deformable field derived from the images at known phases

Synthetic CT images were obtained by warping CT0 with the given deformation field defined by the BSpline lattice coefficients. In our calculation, the range between CT0 and CT40 was divided into 30 incremental steps. The deformation field for each step was obtained by interpolating the transformation matrix derived from the image coregistration between CT0 and CT40 (see the previous section for the calculation details). In general, the interpolated images lack the phase information. A phase stamp can be imposed on the interpolated images by correlating the location of the external marker(s) on the images with the RPM signal characterizing the patient's respiratory motion.

Assessment of the interpolated images

The accuracy of the interpolated images at the intermediate phases was evaluated in three ways. First, we compared the positions of the implanted and surface markers in the interpolated images with those in the known 4D images at the corresponding phases. The trajectories of the markers during the respiration process should ideally coincide with their positions in the known 4D CT images. The trajectories of the markers were computed by warping coordinates observed in CT0 with the deformation field obtained at each interpolation step. The locations of the markers in these images were measured using the display tools available in VTK.

The next level of tests involved tracking the surfaces of anatomic structures. The involved organs were delineated in both the interpolated and the known 4D CT images. We term them the complementary surfaces. For each point on one surface, a scalar value representing the distance to its complementary surface was assigned. Color-coded visualization of the distance permits direct assessment of the regions. In addition, the checkerboard display and subtraction image tools were also employed to evaluate the difference between the two sets of images. An analysis of the subtraction image of the interpolated and the actual images was carried out to evaluate the closeness of the two images. Histograms characterizing the fractional voxels for a variety of HU differences were presented for all 3 patients.

Segmentation of organ contours in 4D CT

The model was also used to map the contours delineated in one phase to another phase, as was previously done by other investigators using a fluid flow method (27). Although both algorithms can successfully find the deformation field between the inhale and exhale images, turbulence may occur during evolution of a fluid-based model, and thus a successive application of the fluid flow algorithm is needed between each pair of images. This procedure is simplified by the usage of the BSpline model, because the deformation field for the intermediary phases is deduced by a direct interpolation. Starting from a mesh of the contour at CT0, the warping of the contours was implemented by modifying coordinates of each point according to the deformation field derived by the BSpline method. The vertices were unaltered during the process.

RESULTS

Convergence analysis

The convergence behavior of the registration was analyzed by monitoring the value of the NCC as a function of the iteration step for three different numbers of nodes: 5, 10, and 15. The results are presented in Fig. 2. In all cases, convergence was achieved in less than 50 iterations. The initial values of the metric are relatively good (−0.985), because only voxel intensities near the organ boundaries are affected by the respiration. Indeed, the affected voxels (the affected voxels are the voxels whose HU units are changed during the breathing cycle) represented only a small percent of the total number of voxels. It is remarkable that the algorithm was sensitive enough to detect and correct these small deviations.

The convergence analysis is useful for the determination of optimal algorithm parameters, most notably the number of nodes. In practice, the value of this parameter is a result of tradeoff between the computational speed and the accuracy. A loose grid would not be able to describe small or local deformation, whereas a dense grid may unnecessarily increase the number of variables in the L-BFGS optimization and thus prolong the computation. For five nodes per di-

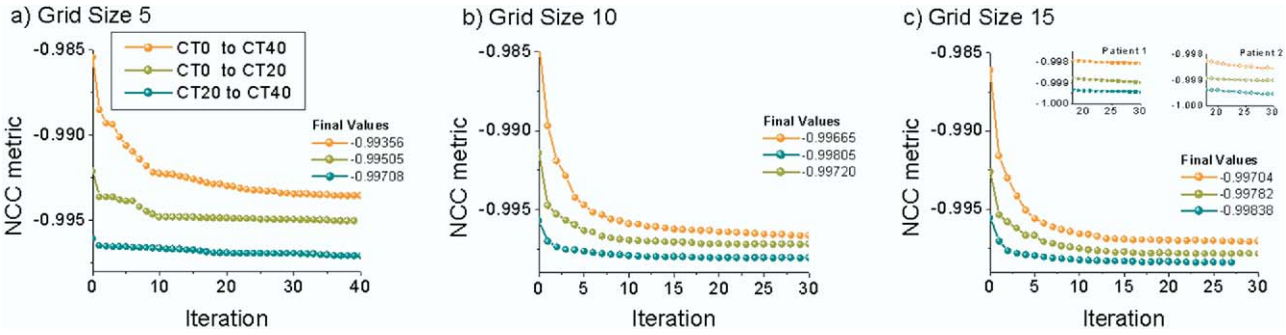


Fig. 2. Assessment of optimal parameters for the deformable registration algorithm. Presented is the normalized correlation coefficient metric as a function of iteration step for three different grid sizes (5, 10, and 15 nodes per dimension). Each panel follows for different grid sizes the three registration problems used in this work, namely between the phases of CT0–CT40, CT0–CT20, and CT20–CT40. The registration quality is assessed by the final values of the metric. Note that a change from 5 to 10 nodes resulted in significant improvements, whereas a further increase from 10 to 15 nodes resulted in only a marginal improvement (2%) in the final value of the metric.

mension, the NCC between the CT0 and CT40 phases was found to be −0.993. The correlation was improved to −0.996 and −0.997, respectively, when 10 and 15 grid nodes per dimension were used. Computation time is in the range of 3 to 5 hours, depending on the number of slices in the study.

For comparison, the above analysis was repeated with the insertion of an intermediate phase image, CT20. For the case of five nodes per dimension, the final metric in the CT0–CT40 registration could not reach the value in the CT20–CT40 registration, indicating that the deformation between CT0 and CT40 was too complex to be described by such a loose BSpline grid. However, for 15 nodes per dimension, the registration calculations converged to −0.9970 for CT0–CT40, −0.9978 for CT0–CT20, and −0.9983 for CT20–CT40 registration. Similar behavior was observed for the remaining two cases. The final metric values for these two cases are presented in the inset of Fig. 2c. A further increase of node number dramatically increases the computational burden but does not significantly improve the accuracy. We therefore selected in this study a grid with 15 nodes per dimension.

Deformation as determined by the deformable registration of the inhale and exhale phases

The range and characteristics of the anatomy deformation occurring during the respiration are of clinical interest. This feature can be directly evaluated by using the coefficients of the BSpline grid. We focus our discussion here on the lung deformation. In Fig. 3a, we present a 3D visualization of the BSpline grid nodes (the nodes are denoted by green spherical dots) in the coregistration of CT0–CT40 images. In Figs. 3b, 3c, and 3d, the deformations at all nodal points as derived from the registration algorithm are displayed by arrows. The orientation and length of an arrow show the deformation direction and magnitude at the corresponding points. The inferior parts of the lungs (and thus the liver) move most. A deformation as large as 43.2 mm was observed in the posterior region of the liver, which is confirmed by the checkerboard comparison of CT0 and CT40 (Fig. 4a), where most discrepancies are observed in the upper liver, stomach, and thoracic region. For the second patient, a similar analysis shows a displacement up to 14.2 mm in the posterior part of the liver and 12.0 mm in the thoracic region. However, the deformation was more uniformly distributed. The deformation of the third patient was predominantly located in the liver, but with displacements oriented differently from the previous 2 patients. Our results seem to suggest that organ deformation and displacement are patient specific (47), which hinders the development of a common mathematical model for the respiratory process. The deformation depends on anatomy and needs to be determined individually.

Performance of image coregistration

The performance of image coregistration is usually evaluated by using the checkerboard display and subtraction image. For the first case, the checkerboard images before and after registration for an axial, sagittal, and coronal slice are presented in Fig. 4a. It is seen that before the image coregistration of CT0–CT40, the movements in these slices are as large as 8 mm. After the deformable registration, the two images become almost indiscernible (Fig. 4b). The subtraction image between the original CT0 and CT40 images shows large HU differences in the liver and thoracic regions (Fig. 4c). The difference in the voxel intensity ranges from 0 to 875. After the deformable registration, the maximum difference is reduced to 250 HU (Fig. 4d). The differences appear mainly in the regions where the motion of small structures such as bronchi cannot be easily described by the BSpline lattice. In Fig. 4e we plot the histograms of the fractional voxel number for a number of HU differences. In addition to the fact that the initial maximum mismatch of 875 HU is reduced to 250 HU, the number of pixels having a large HU difference is dramatically reduced. For example, pixels with an HU difference greater than 20 HU represent only less than 0.1% of the total number

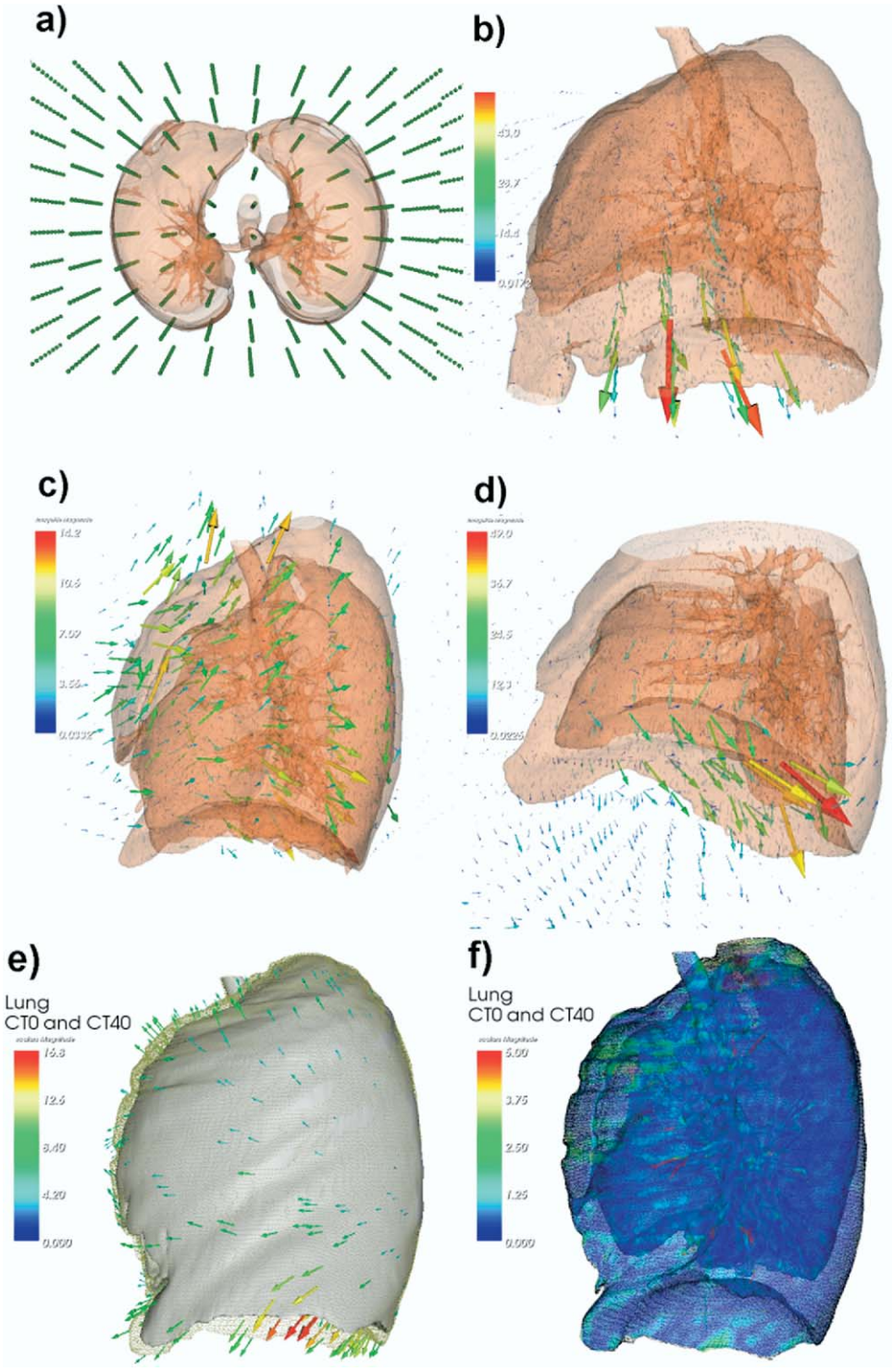


Fig. 3. Results from BSpline deformable calculation. (a) Representation of the grid nodes superimposed on lung contours. (b) Visualization of resulting deformations. On each node, arrow length and orientation are proportional to deformation magnitude and direction. (c, d) Same analysis for Patients 2 and 3 demonstrates that deformations are patient dependent. (e) The node deformation can be interpolated to any location, such as the surface of an organ. In this view, the lung surface extracted from CT0 is represented as surface, whereas homologs surface extracted from CT40 as points. The arrows represent deformations from CT0 to CT40 in random points on the lung surface. (f) The lung surface at CT0 was deformed with the deformations presented in (e) to contour the lung surface at CT40. Color represents contouring error, ranging from less than 1.5 mm on most of the surface to 5 mm observed only for small regions of the bronchial tree, where the registration algorithm performs poorly.

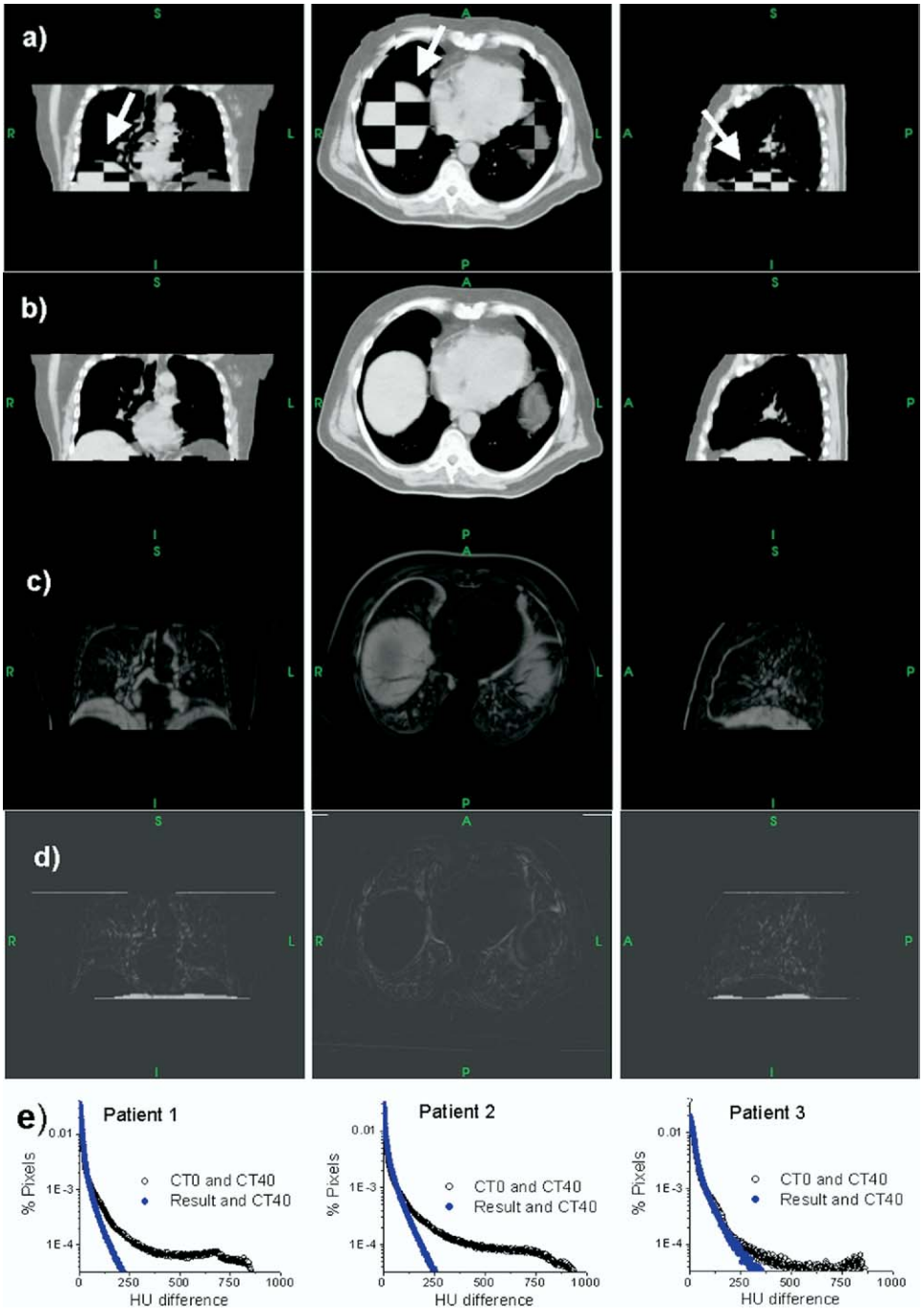


Fig. 4. Checkerboard display of the inhale and exhale phases (a) before and (b) after deformable image registration. The results show that the BSpline algorithm can properly model the lung deformation caused by the respiratory motion. The subtract images of the two phases before and after registration are shown in (c) and (d). (e) In the histogram analysis of the image differences, large differences of up to 750 Hounsfield units are observed initially for voxels located on the lung–liver interface. After registration, the difference is reduced to 250 Hounsfield units.

of voxels after the image coregistration. We anticipate a further improvement if more nodes are used. This gain is, of course, achieved at the expense of an increased computational time.

For the last 2 patients, the registration results as assessed by the checkerboard tool are presented in Fig. 5. Once again, excellent registrations were achieved between CT0 and CT40 images.

Image interpolation

Comparison of the trajectories of implanted and surface markers in the interpolated and actual images affords an effective way to assess the performance of the proposed image interpolation scheme. In Fig. 6, we show the 4D trajectories of all the markers for two different interpolation schemes. The curves in the top row for each fiducial marker

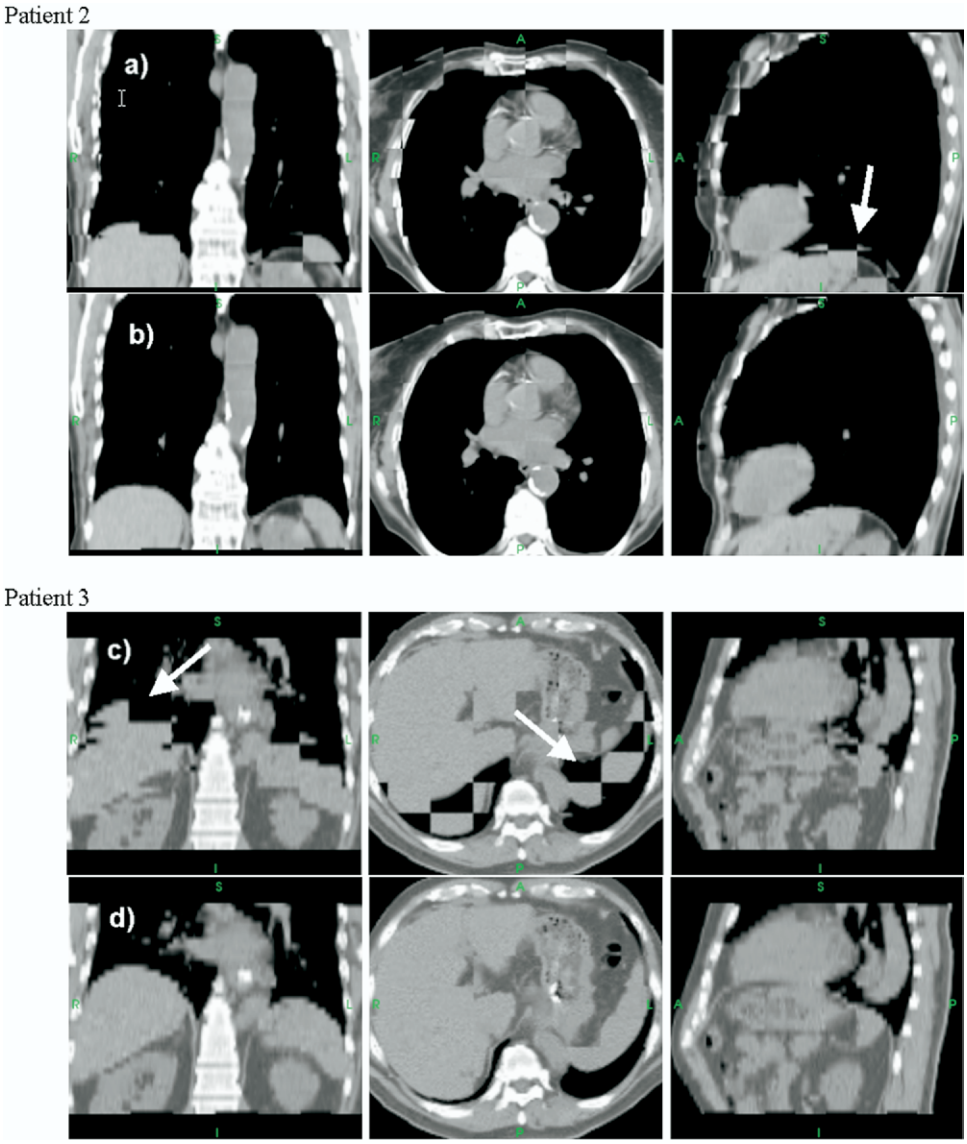


Fig. 5. Checkerboard comparisons of CT0 and CT40 for the second and third cases. (a) Discordance up to 8 mm is observed before registration. (b) After the deformable registration, virtually no difference between the two images is observable. (c, d) Similar results are obtained for the third case.

show the results derived from interpolating the CT0–CT40 registration, along with the marker trajectories obtained from the actual 4D images (dots). The curves in bottom row for each fiducial show the marker trajectories derived by interpolation from CT0–CT20–CT40. In the former case, we found that the discrepancy of the marker positions in the two image sets was less than 3 mm. When the intermediate 3D CT set, CT20, was added, the uncertainty was reduced to less than 2 mm. It is sensible that the usage of an intermediate phase would lead to an improvement in the interpolation accuracy. In general, the determination of the number of image sets for image interpolation is a tradeoff between a few factors, notably the computational accuracy and convenience.

We have also assessed the interpolation scheme by tracking the surface of anatomic structures. In Fig. 7, we show the 3D plot of the distance between some structural surfaces in the interpolated and actual CT10, CT20, and CT30 images for the 3 patients. To illustrate the utility of the structural surface distance plot, we have elected to display different structures for different patients. We noted that, for the first patient (the first row of Fig. 7), a large error appeared consistently at the slices near the top of the liver, where binning artifacts have been identified in the acquired images. Binning is not used to generate the interpolated images, making them artifact free.

Checkerboard display for the first patient, comparing the interpolated and actual images of CT10, CT20, and CT30, is shown in Fig. 8. The two sets of images are highly similar, with virtually no difference visible in the checkerboard. The subtraction images for the same case are displayed in Fig. 9. The histograms describing the differences in the voxel intensities in the subtraction images are shown in the bottom of Fig. 9. Overall, no significant difference was observed

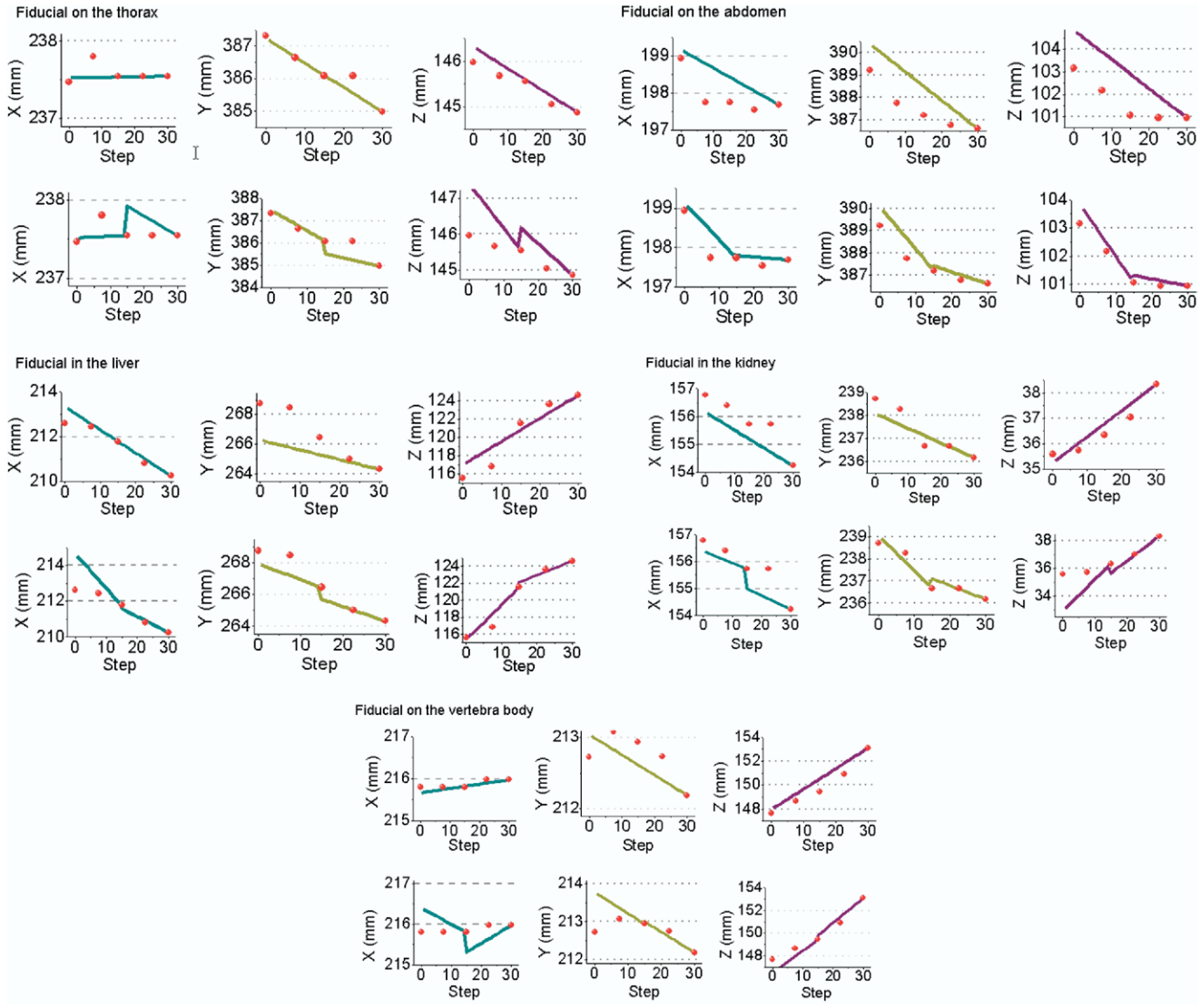


Fig. 6. Assessment of interpolation accuracy by tracking the trajectories of the implanted markers. Presented are coordinates as function of interpolation step. The positions of the markers in the acquired 4D images (red dots) are plotted together, with their correspondence deduced from the interpolation of the deformation field (curves). Deformation field from either direct CT0–CT40 registration (upper row) or use of an intermediary phase CT20 (lower row). For all markers, maximum deviations are 3 and 2 mm, respectively.

between the interpolated and actual images, and the major discrepancies occurred at the boundaries of small structures.

Segmentation in 4D CT

The deformation between CT0 and CT40 can be used to deduce both images and contours at the intermediate phases. Surfaces delineated in CT0 were warped to CT10, CT20, and CT30, and the results are used in Fig. 10. Generally, the difference between the two complementary surfaces is less than 3 mm, as illustrated for Patient 1 in the top row of Fig. 7 for the lung surfaces. Even for small structures such as the trachea, the error was generally less than 5 mm (middle row of Fig. 7). The performance of the registration algorithm is interesting, because it is commonly believed that a large number of nodal points would be required to describe the deformation of such small organs. Large error may also appear in a given set of slices where the binning artifact was present.

To illustrate the application of the algorithm to deduce tumor motion, in Fig. 10 we present the sagittal contours corresponding to CT20 and CT40, warped from the tumor contour delineated on CT0, overlaying on the background image (CT40). This application allows us to monitor the tumor path and shape changes as the patient breathes and helps us to better define the margin for radiation treatment.

DISCUSSION

In radiation therapy, respiratory motion poses significant challenges for treating tumors in the thorax or abdomen. It can distort the shape of an object, degrade the anatomic

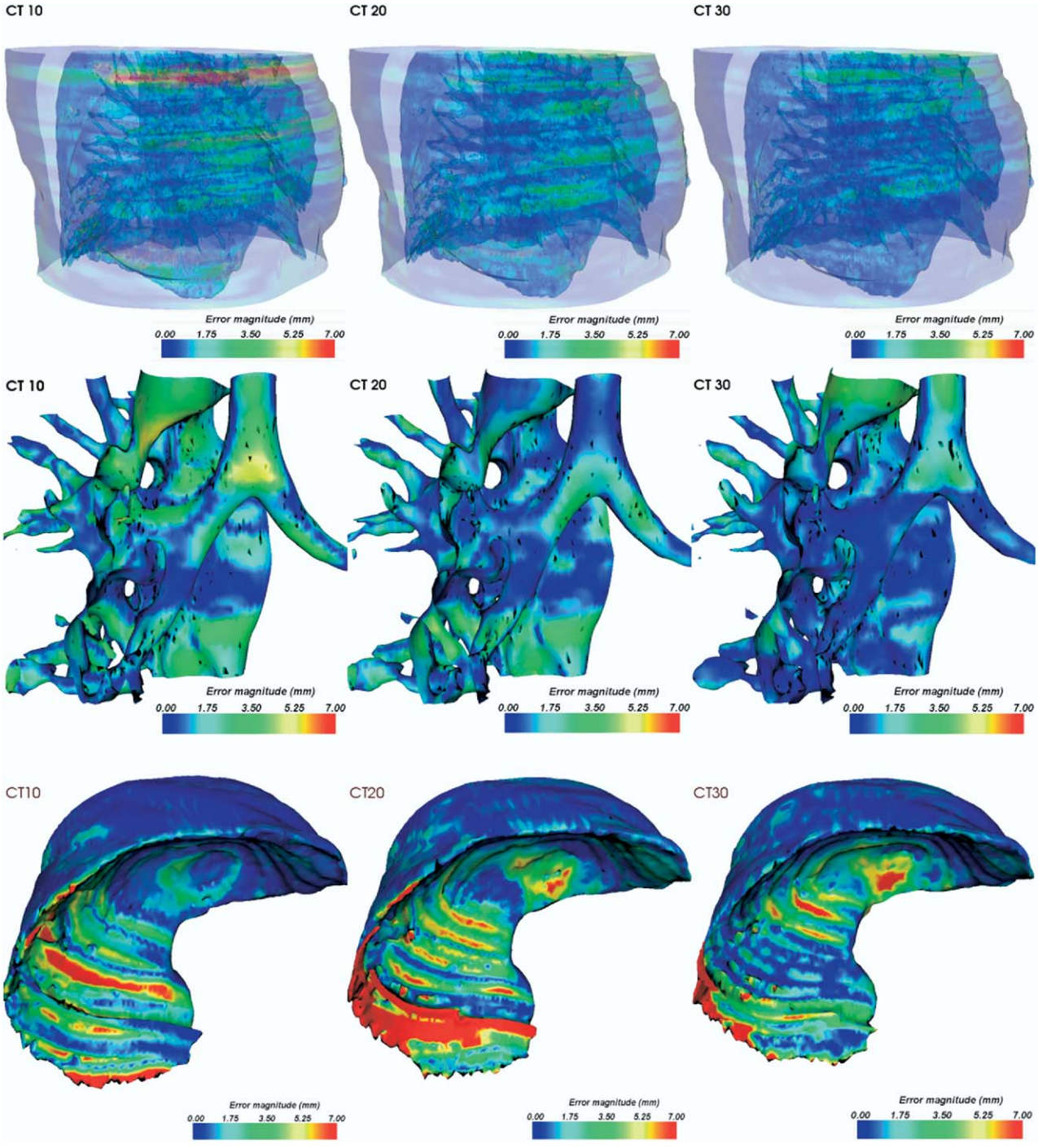


Fig. 7. Assessment of interpolation accuracy by surface comparison. The error distances between interpolated and acquired images are color-coded representation. Presented are the lung and external contour of the first patient, the bronchial tree (second patient), and the diaphragm (third patient). Color coding ranges from 0 mm (blue) to 7 mm (red).

position reproducibility during imaging, and necessitate larger margins during radiotherapy planning. It also causes inaccuracy in estimating the tumor volume, thereby preventing an effective dose escalation for the treatment of a target tumor. How to minimize its adverse effects on radiation therapy represents a significant problem in achieving the goal of conformal radiation therapy. Four-dimensional CT scans, acquired synchronously with a respiratory signal, provide not only the 3D spatial information, but also temporal changes of the anatomy as a function of the respiratory phase during the imaging and can be employed to guide the treatment planning to explicitly account for the respiratory motion. The availability of 4D imaging techniques thus provides a useful tool to better understand the

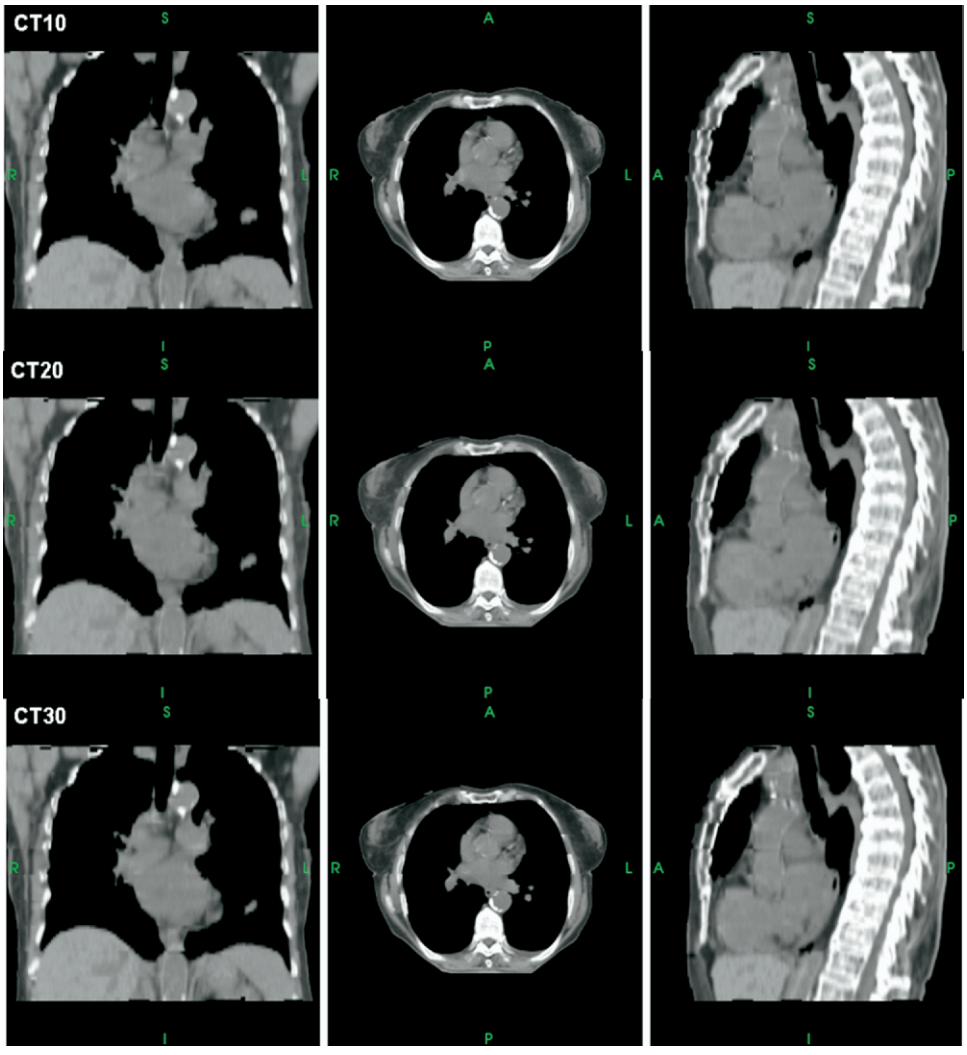


Fig. 8. A checkerboard comparison of interpolated and acquired images for the phases of CT10, CT20, and CT30. The interpolated and actually acquired images are practically identical.

physiologic respiratory process and develop solutions to the problem.

In this study we have extensively studied the interphase correlation of the images during the respiratory process and presented a method to interpolate the 4D images based on a few sets of 3D images at different phases. The underlying assumption of the approach is that the patient's anatomy at different phases can be related by using a deformable image registration model. Using a BSpline technique, we were able to interpolate the images at the intermediate phases to within 3 mm even when only two sets of images at inhale and exhale phases were used.

The proposed technique may have three practical implications. First, by reducing the number of 3D image sets, the radiation dose to the patient can be reduced significantly. When a modern multislice CT is used for a regular clinical examination, the dose received by the patient may approach 10 mSv for head and 20 mSv for the chest or abdomen. With a 4D acquisition, because a patient is scanned multiple times at each couch position during the imaging, the radiation exposure will be considerably higher than the regular CT scan (up to 1 order of magnitude higher). Effective dose reduction is thus highly desirable for clinical application of the cutting-edge 4D CT scanning technology. With the proposed interpolation scheme, the radiation dose to the patient can be reduced by 50% to 80%, which may have significant clinical implication.

Second, the technique reduces the requirement for storage disk space and may afford an effective method to facilitate 4D CT data compression, storage, and retrieval. The BSpline model used in this study provides a concise and artifact-free representation of the deformation field and the 4D images. The "decompression" of the image can be realized by applying the deformation lattice on the initial CT image. Our experience has shown that the disk space usage can be reduced from 402 MB to 40 MB per patient.

Finally, the model can be applied to interpolate contours between phases, providing an alternative to the optical flow methods reported by other groups (27). To incorporate 4D

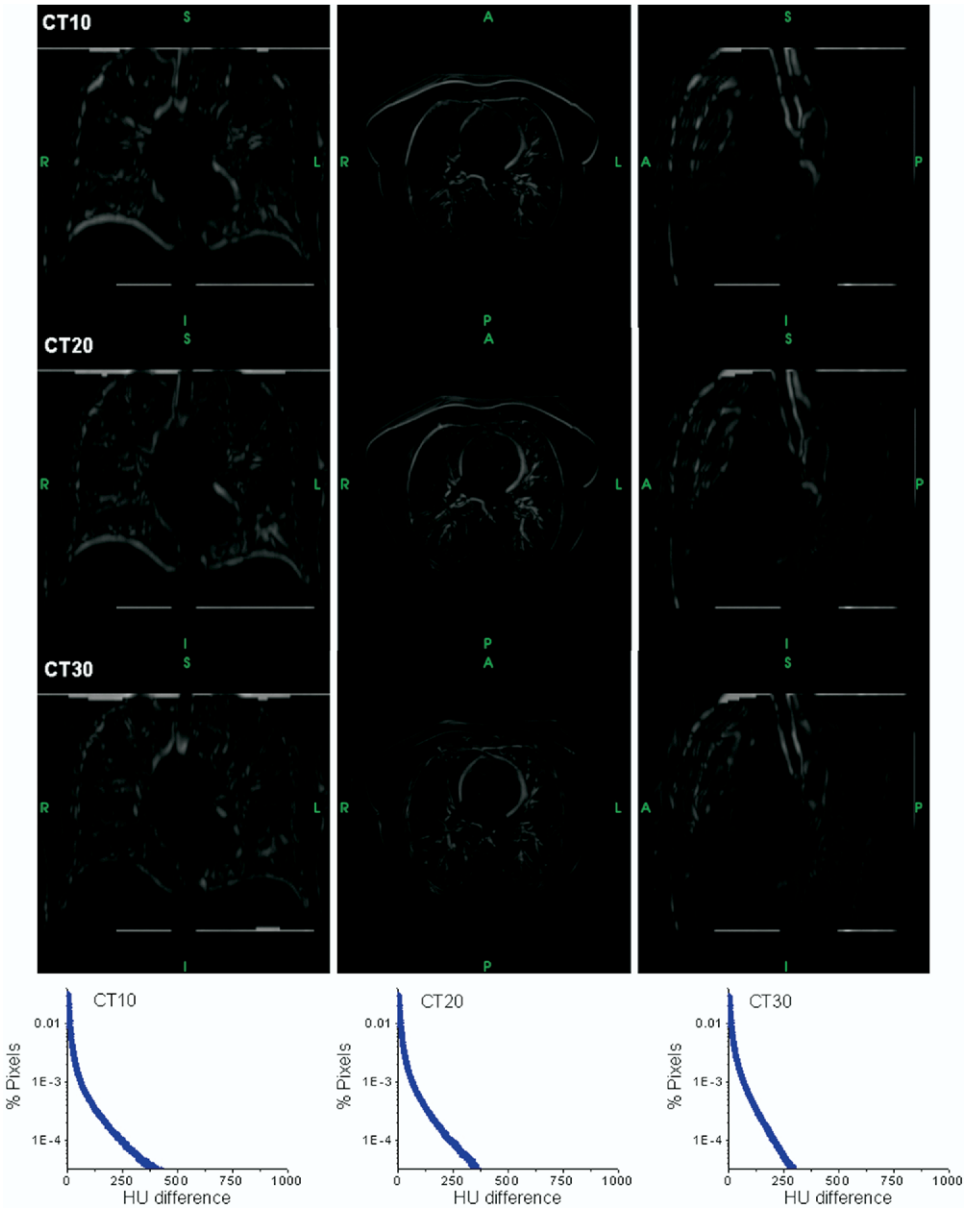


Fig. 9. Subtraction display of the interpolated and acquired images for the phases of CT10, CT20, and CT30. Histogram analysis of the differences in images (last row) documents a maximum difference of 250 Hounsfield units. Only 0.01% of the image pixels have a difference larger than 20 Hounsfield units.

CT data into radiation treatment planning, a necessary step is the delineation of the normal anatomic structures on the data sets. Because the segmentation process in 3D imaging is already a tedious and labor-intensive process, it seems to be impractical for clinicians to outline the structures slice by slice and phase by phase. Because the voxel-to-voxel mapping between different phases is established during the deformable registration, mapping of the segmented structure contours from one phase to another takes essentially no additional time.

Theoretically, there is no limitation in the number of nodes that can be used with the BSpline model, with an extreme case where one node is located in each voxel. Such an extreme case should be able to model any trajectory and structures as small as conceivable but would require optimization of 45 million variables. In reality, a grid of 15 nodes means 10,125 variables to be optimized, and a "small" increase to 20 nodes per dimension leads to 24,000 variables, which pushes the calculation to the limit of the current optimization algorithm. A realistic registration software should balance the number of nodes and the accuracy of registration. Although there are a number of deformable image registration algorithms, the BSpline technique is known for its simplicity and reasonable accuracy. It is a preferred option, because the technique provides a good tradeoff between a number of practical factors. Other formalisms include, to name a few, fluid flow algorithm and finite-element model. The fluid flow algorithm is

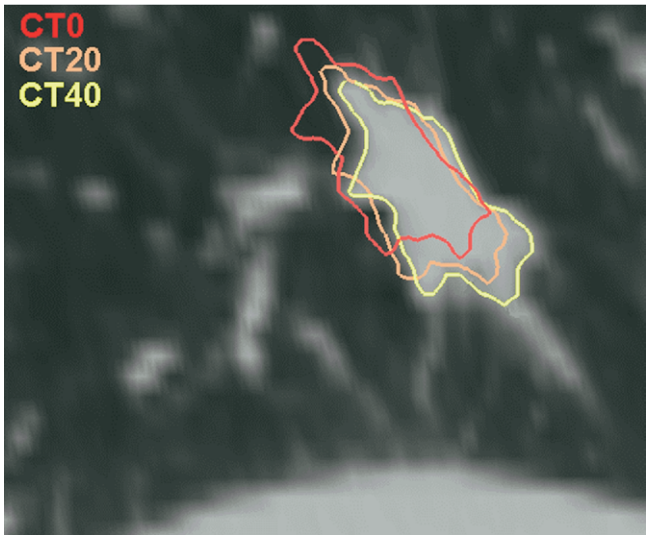


Fig. 10. Automatic mapping of contours based on the BSpline deformable calculation. The tumor trajectory is deduced by applying the deformation field on the contour delineated in CT0. Presented are contours warped to CT20 and CT40 from CT0. The background image of CT40 confirms agreement of warped contour and observed tumor.

better suited for small deformations, which is hardly the case in 4D CT applications. The finite element model is in principle most accurate, because it models the organ motion based on real anatomic forces occurring in nature. A major drawback of the approach is that the elasticity and viscosity parameters that model the tissues/organs are poorly known, hindering its practical application. Because the deformable image registration is a fast developing field, better and faster algorithms may be more available in the future.

CONCLUSIONS

We have demonstrated the feasibility of using a BSpline deformable image registration model to interpolate the intermediate phases by starting from two or three sets of 3D CT images acquired at different phase points. The study suggests that the organ deformation during the breathing process can be well modeled by using a BSpline deformable algorithm. The proposed technique may find useful applications in reducing radiation dose, removing binning artifacts, generating images at a finer-phase resolution for better evaluation of the physiologic respiratory process, and decreasing the disk storage usage in 4D imaging.

REFERENCES

1. Mayo JR, Muller NL, Henkelman RM. The double-fissure sign: a motion artifact on thin-section CT scans. *Radiology* 1987;165:580–581.
2. Tarver RD, Conces DJ Jr, Godwin JD. Motion artifacts on CT simulate bronchiectasis. *AJR Am J Roentgenol* 1988;151: 1117–1119.
3. Ritchie CJ, Hsieh J, Gard MF, Godwin JD, Kim Y, Crawford CR. Predictive respiratory gating: a new method to reduce motion artifacts on CT scans. *Radiology* 1994;190:847–852.
4. Ritchie CJ, Godwin JD, Crawford CR, Stanford W, Anno H, Kim Y. Minimum scan speeds for suppression of motion artifacts in CT. *Radiology* 1992;185:37–42.
5. Balter JM, Ten Haken RK, Lawrence TS, Lam KL, Robertson JM. Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing. *Int J Radiat Oncol Biol Phys* 1996;36:167–174.
6. Shimizu S, Shirato H, Kagei K, *et al.* Impact of respiratory movement on the computed tomographic images of small lung tumors in three-dimensional (3D) radiotherapy. *Int J Radiat Oncol Biol Phys* 2000;46:1127–1133.
7. Shimizu S, Shirato H, Ogura S, *et al.* Detection of lung tumor movement in real-time tumor-tracking radiotherapy. *Int J Radiat Oncol Biol Phys* 2001;51:304–310.
8. Ford EC, Mageras GS, Yorke E, Ling CC. Respiration-correlated spiral CT: a method of measuring respiratory-induced anatomic motion for radiation treatment planning. *Med Phys* 2003;30:88–97.
9. Vedam SS, Keall PJ, Kini VR, Mostafavi H, Shukla HP, Mohan R. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Phys Med Biol* 2003;48:45–62.
10. Allen AM, Siracuse KM, Hayman JA, Balter JM. Evaluation of the influence of breathing on the movement and modeling of lung tumors. *Int J Radiat Oncol Biol Phys* 2004;58:1251–1257.
11. Yu CX, Jaffray DA, Wong JW. The effects of intra-fraction organ motion on the delivery of dynamic intensity modulation. *Phys Med Biol* 1998;43:91–104.
12. Keall PJ, Kini VR, Vedam SS, Mohan R. Motion adaptive x-ray therapy: a feasibility study. *Phys Med Biol* 2001;46:1–10.
13. Chui CS, Yorke E, Hong L. The effects of intra-fraction organ motion on the delivery of intensity-modulated field with a multileaf collimator. *Med Phys* 2003;30:1736–1746.
14. George R, Keall PJ, Kini VR, *et al.* Quantifying the effect of intrafraction motion during breast IMRT planning and dose delivery. *Med Phys* 2003;30:552–562.
15. Jiang SB, Pope C, Al Jarrah KM, Kung JH, Bortfeld T, Chen GT. An experimental investigation on intra-fractional organ motion effects in lung IMRT treatments. *Phys Med Biol* 2003; 48:1773–1784.
16. Xing L, Thorndyke B, Schreibmann E, *et al.* Overview of image guided radiation therapy. *Med Dosim* 2006. In press.
17. Low DA, Nystrom M, Kalinin E, *et al.* A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. *Med Phys* 2003;30:1254–1263.
18. Rietzel E, Pan T, Chen GT. Four-dimensional computed tomography: image formation and clinical protocol. *Med Phys* 2005;32:874–889.
19. Mageras GS, Pevsner A, Yorke ED, *et al.* Measurement of lung tumor motion using respiration-correlated CT. *Int J Radiat Oncol Biol Phys* 2004;60:933–941.
20. Pan T, Lee TY, Rietzel E, Chen GT. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med Phys* 2004;31:333–340.
21. Taguchi K. Temporal resolution and the evaluation of candidate algorithms for four-dimensional CT. *Med Phys* 2003;30: 640–650.
22. Sonke JJ, Zijp L, Remeijer P, van Herk M. Respiratory correlated cone beam CT. *Med Phys* 2005;32:1176–1186.
23. Li T, Schreibmann E E, Yang Y, Xing L. Motion correction

for improved target localization with on-board cone-beam CT. *Phys Med Biol* 2006;51:253–267.
24. Li T, Schreibmann E, Thorndyke B, *et al.* Radiation dose reduction in 4D computed tomography. *Med Phys* 2005;32: 3650–3660.
25. Bharath A, Hirose M, Hata N, *et al.* Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Med Phys* 2001;28: 2551–2560.
26. Brock KK, McShan DL, Ten Haken RK, Hollister SJ, Dawson LA, Balter JM. Inclusion of organ deformation in dose calculations. *Med Phys* 2003;30:290–295.
27. Guerrero T, Zhang G, Huang TC, *et al.* Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method. *Phys Med Biol* 2004;49:4147–4161.
28. Fei B, Kemper C, Wilson DL. A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes. *Comp Med Imag Graph* 2003;4:267–281.
29. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol* 2004;49:3067–3087.
30. Mattes D, Haynor RD, Vesselle H, Lewellen KT, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imag* 2003;22:120–128.
31. Rohlfing T, Maurer CR Jr, O'Dell WG, Zhong J. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Med Phys* 2004;31:427–432.
32. Lian J, Hunjan S, Dumoulin C, *et al.* Integrating deformable MRI/MRSI and CT image registration into the prostate IMRT treatment planning. *Int J Radiat Oncol Biol Phys* 2003;57: S207.
33. Lian J, Xing L, Hunjan S, Spielman B, Daniel B. Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: a deformable registration and validation study. *Med Phys* 2004; 31:3087–3094.
34. Rohlfing T, Mauer CR Jr, O'Dell WG, Zhong J. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Med Phys* 2004;31:427–432.
35. Kaus MR, Netsch T, Kabus S, Pekar V, McNutt T, Fischer B. Estimation of organ motion from 4D CT for 4D radiation therapy planning of lung cancer. *LNCS* 2004;3217:1017–1024.
36. Berlinger K, Roth M, Fisseler J, Sauer O, Schweikard A, Vences L. Volumetric deformation model for motion compensation in radiotherapy. *LNCS* 2004;3217:925–932.
37. Schweikard A, Shiomi H, Fisseler J, *et al.* Fiducial-less respiration tracking in radiosurgery. *LNCS* 2004;3217:992–999.
38. Schreibmann E, Xing L. Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies. *Int J Radiat Oncol Biol Phys* 2005;62:595–605.
39. Coselmon MM, Balter JM, McShan DL, Kessler ML. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med Phys* 2004;31:2942–2948.
40. Ibanez L, Schroeder W, Ng L. ITK software guide. Clifton Park, NY; Kitware Inc., 2003.
41. Schroeder W, Martin K, Lorensen B. The visualization toolkit: An object-oriented approach to 3D graphics.
42. Lee S, Wolberg G, Chwa KY, Shin SY. Image metamorphosis with scattered feature constraints. *IEEE Trans Vis Comp Graph* 1996;2:337–354.
43. Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. *IEEE Trans Vis Comp Graph* 1997;3: 228–244.
44. Thevenaz P, Unser T. Spline piramids for inter-modal image registration using mutual information. *IEEE Trans Med Imag* 1997;16:187–198.
45. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imag* 1999;18:712–721.
46. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program* 1989;45:503–528.
47. Koch N, Liu HH, Starkschall G, *et al.* Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part I–correlating internal lung motion with skin fiducial motion. *Int J Radiat Oncol Biol Phys* 2004;60:1459–1472.

Medical Dosimetry, Vol. 31, No. 2, pp. 91-112, 2006
Copyright © 2006 American Association of Medical Dosimetrists
Printed in the USA. All rights reserved
0958-3947/06/$–see front matter

ELSEVIER

doi:10.1016/j.meddos.2005.12.004

OVERVIEW OF IMAGE-GUIDED RADIATION THERAPY

LEI XING, PH.D., BRIAN THORNDYKE, PH.D., EDUARD SCHREIBMANN, PH.D., YONG YANG, PH.D., TIAN-FANG LI, PH.D., GWE-YA KIM, PH.D., GARY LUXTON, PH.D., and ALBERT KOONG, M.D.

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA

(*Accepted* 21 *December* 2005)

Abstract—Radiation therapy has gone through a series of revolutions in the last few decades and it is now possible to produce highly conformal radiation dose distribution by using techniques such as intensity-modulated radiation therapy (IMRT). The improved dose conformity and steep dose gradients have necessitated enhanced patient localization and beam targeting techniques for radiotherapy treatments. Components affecting the reproducibility of target position during and between subsequent fractions of radiation therapy include the displacement of internal organs between fractions and internal organ motion within a fraction. Image-guided radiation therapy (IGRT) uses advanced imaging technology to better define the tumor target and is the key to reducing and ultimately eliminating the uncertainties. The purpose of this article is to summarize recent advancements in IGRT and discussed various practical issues related to the implementation of the new imaging techniques available to radiation oncology community. We introduce various new IGRT concepts and approaches, and hope to provide the reader with a comprehensive understanding of the emerging clinical IGRT technologies. Some important research topics will also be addressed. © 2006 American Association of Medical Dosimetrists.

Key Words: **IGRT, Organ motion, Image guidance, Dose optimization, 4D imaging.**

INTRODUCTION

Radiotherapy is an image-guided intervention, and imaging is involved in every key step of the process, ranging from patient staging, simulation, treatment planning, and radiation delivery, to patient follow-up. The evolution of radiation therapy has been strongly correlated with the development of imaging techniques. During the early days when Roentgen first discovered x-rays, 2-dimensional (2D) transmission images of the human body provided unprecedented imagery of bony landmarks, which allowed radiologists to deduce the location of internal organs. Using planar radiographs, radiologists planned cancer treatments by collimating rectangular fields that circumscribed the presumed tumor location. Additional blocks placed daily to match marks on the patient's skin, and later using low-temperature-melting dense alloys. The emergence of computed tomography (CT) in the 1970s revolutionized radiation therapy and allowed us to use image data to build a 3-dimensional (3D) patient model and design 3D conformal radiation treatment. In general, 3D conformal radiation therapy (3DCRT) is a method of irradiating a tumor target volume defined in a 3D anatomical image of the patient with a set of x-ray beams individually shaped to conform to the 2D beam's-eye-view (BEV) projection of the target. The reduction in normal tissue irradiation when moving from 2D to 3D should theoretically improve the therapeutic ratio and allow the tumor target volume to be treated to a higher dose, thereby improving the probability of tumor control. Recent technical advances in planning and delivering intensity-modulated radiation therapy (IMRT) provide an unprecedented means for producing exquisitely shaped radiation doses that closely conform to the tumor dimensions while sparing sensitive structures.[1–3] The development of 3DCRT and IMRT places more stringent requirements on the accuracy of beam targeting. In practice, large uncertainties exist in tumor volume delineation and in target localization due to intra- and inter-organ motions. The utility of modern radiation technologies, such as 3DCRT and IMRT, cannot be fully exploited without eliminating or significantly reducing these uncertainties. The need to improve targeting in radiation treatment has recently spurred a flood of research activities in image-guided radiation therapy (IGRT).

While all radiation therapy procedures are image guided *per se*, traditionally, imaging technology has primarily been used in producing 3D scans of the patient's anatomy to identify the location of the tumor prior to treatment. The verification of a treatment plan is typically done at the level of beam portals relative to the patient's bony anatomy before patient treatment. In current literature, the term of IGRT or IG-IMRT is employed loosely to refer to newly emerging radiation planning, patient setup, and delivery procedures that integrate cutting-edge image-based tumor definition methods, pa-

Reprint requests to: L. Xing, Ph.D., Department of Radiation Oncology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. E-mail: lei@reyes.stanford.edu

tient positioning devices, and/or radiation delivery guiding tools. These techniques combine new imaging tools, which interface with the radiation delivery system through hardware or software, and state-of-the-art 3DCRT or IMRT, and allow physicians to optimize the accuracy and precision of the radiotherapy by adjusting the radiation beam based on the true position of the target tumor and critical organs. With IGRT, it is also possible to take tumor motion into account during radiation therapy planning and treatment. Because IGRT improves precision, it raises the possibility of shortening the duration of radiation therapy by reducing the number of treatment sessions for some forms of cancer.

The purpose of this article is to highlight the recent developments of various available imaging techniques and present an overview of IGRT. Stanford experience on various aspects of clinical IGRT will also be presented. We hope that readers will gain an overall picture of IGRT and find it easier to navigate themselves through the subsequent articles in this issue, which focus on providing technical details and/or specific clinical applications of the available IGRT tools.

ISSUES IN IGRT

In current 3DCRT or IMRT, uncertainties exist in many circumstances, such as tumor target definition, patient immobilization, and patient breathing motion, which make it difficult to administer a high radiation dose to the planned location. The exact locations of the boundaries of the tumor target and the adjacent sensitive structures are often not known precisely, and a population- and disease site-based safety margin is used routinely to cope with a problem that is otherwise insoluble. An important task of IGRT is to eliminate or significantly reduce the margins involved in defining the clinical and planning target volume (CTV and PTV, respectively).

Many IGRT solutions have been proposed to resolve the problem of target definition and beam targeting. Briefly, IGRT developments are focused in four major areas: (1) biological imaging tools for better definition of tumor volume; (2) time-resolved (4D) imaging techniques for modeling the intra-fraction organ motion; (3) on-board imaging system or imaging devices registered to the treatment machines for inter-fraction patient localization; and (4) new radiation treatment planning and delivery schemes incorporating the information derived from the new imaging techniques. These are discussed in more detail in the following.

TUMOR TARGET VOLUME DEFINITION

CT, MRI, and ultrasound (US) imaging techniques

To be able to "see" the extent of disease more clearly and define the tumor target volume relative to the patient's anatomy have been among the most important issues in radiation oncology. CT has played a pivotal role in the process. Many radiation oncology departments have acquired dedicated CT scanners. A typical patient's 3D CT data set has more than 100 axial slices, each of which contains 512×512 pixels. With 16 bits per pixel, a CT data set can easily run over 50 megabytes. CT has many advantages, including high spatial integrity, high spatial resolution, excellent bony structure depiction, and the ability to provide relative electron density information used for radiation dose calculation. The recent development of ultra-fast multi-slice CT has opened a new dimension to CT technology and allows time-resolved (4D) CT imaging of patient's cardiac and breathing cycles. Using array detectors, multisection CT scanners can acquire multiple slices or sections simultaneously and thereby greatly increase the speed of CT image acquisition. Currently, all manufactures are moving toward 8-, 16- and even higher slice CT technology. Radiation oncology application of 4D CT will be discussed later.

MRI provides superior soft tissue discrimination, especially for central nervous system (CNS) structures and within the abdomen and pelvis, and has been widely used in the diagnosis and tumor delineation. MRI is also utilized for virtual simulation of radiation treatment for some specific disease sites. Physically, MRI involves the determination of the bulk magnetization of nuclei within a given voxel through use of radio-frequency (RF) radiation and magnetic fields. In a clinical setting, MRI is typically employed together with CT images with the help of image fusion software to delineate the extent of the malignancy. As with other imaging techniques, MR technology has gone through a series of revolutions in the past 3 decades. MRI technology is moving toward higher field strengths to further improve the quality of MR images, as evidenced by the installations of 3T scanners in many institutions (9.4 T MRI scanners have been installed in a few institutions). Fast-cine MRI is also becoming increasingly available and may offer physicians an alternative for imaging the temporal process of patient breathing or even heart beating. Figure 1 shows an example of MRI images acquired at 2 different phases for a liver cancer patient. In addition, the development of some specialized MRI scans has also attracted much attention. These include diffusion and perfusion MRI, dynamic contrast MRI, MR angiography, MR spectroscopic imaging (MRSI), and functional MRI (fMRI). The recent development of diffusion tensor imaging (DTI), for instance, enables diffusion to be measured in multiple directions and the fractional anisotropy in each direction to be calculated for each voxel. fMRI measures signal changes in the brain that are due to changing neural activity. These techniques enable researchers to make axonal and functional maps to examine the structural connectivity of different regions in the brain and may allow better definition of brain tumors and better sparing of sensitive regions.[4]

Ultrasound (US) is another useful imaging modality for radiation therapy. US utilizes high-frequency (1~10

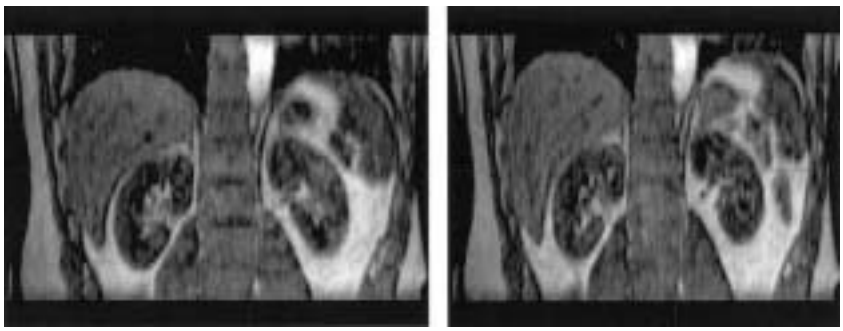


Fig. 1. Cine MR images at inhale and exhale phases for a liver cancer patient.

MHz) sound waves to generate anatomical images that have high spatial resolution and tissue characterization discrimination power through image texture analysis. In radiation therapy, it has been particularly useful in prostate imaging. Transrectal US permits an examination/localization of the prostate gland[5,6] and is the imaging modality of choice in guiding the prostate seed implant procedure.

Biological imaging

Regardless of the course of therapy, current standard imaging modalities such as CT and MRI do not always provide an accurate picture of the tumor extent, especially in the zone of infiltration that may be the limiting factor in an attempt of a radical treatment approach. This has been shown to be the case for gliomas before surgical intervention. It is also true when attempting to determine the volume of residual tumor for additional therapy owing to problems in differentiating post-therapy changes from residual tumor. Indeed, the above-mentioned imaging modalities are anatomic in nature, *i.e.*, they provide snapshot of a patient's anatomy without biological information of various organs or structures. Biological imaging, defined as the *in vivo* characterization and measurement of biological processes at the cellular and molecular level, is an emerging multidisciplinary field resulting from the developments of molecular biology and diagnostic imaging and shows significant promise to revolutionize cancer detection, staging/re-staging, treatment decision-making, and assessment of therapeutic response. MRSI and positron emission tomography (PET) are 2 valuable modalities for radiation therapy planning. ^{1}H MRSI combines the advantages of obtaining biochemical data by water-suppressed ^{1}H MR spectroscopy with the spatial localization of that data. MR spectroscopy is useful in characterization of brain and prostate tumors. In the brain, for example, malignant tumors have an increased rate of membrane turnover (increased level of choline) and a decreased concentration of neurons. Furthermore, spectroscopy allows for the noninvasive monitoring of the response of residual tumor to therapy and for differentiating tumor recurrence from tissue necrosis. Recently, Pirzkall *et al.*[7] have applied multi-voxel MRSI to assess the impact of MRSI on the target volumes used for radiation therapy treatment planning for high-grade gliomas. It was found that, although T_2-weighted MRI estimated the region at risk of microscopic disease as being as much as 50% greater than by MRSI, metabolically active tumor tissue still extended outside the T_2 region in 88% of patients by as much as 28 mm. In addition, T_1-weighted MRI suggested a lesser volume and different location of active disease compared to MRSI. The discordance of high-grade-glioma target volumes resulting from MRI was also observed in other functional imaging modalities such as (PET) and single-photon emission computed tomography (SPECT).

While there is a growing body of evidence now indicating that *in vivo* MRSI provides unique information on metabolism that will ultimately affect clinical diagnosis, choice and monitoring of therapies, and treatment planning, in reality, MRSI has mainly remained a research tool confined to a small number of academic institutions.[8–12] PET, on the other hand, is more widely used and has been harnessed into the planning process in many clinics. In general, PET has lower image resolutions than CT images and, with commonly used fluorine-18-labeled deoxyglucose (FDG) tracer, contains no anatomic information about normal structures. Information derived from PET needs to be fused with the corresponding CT images for treatment planning. The fusion of PET and CT images are simplified with the use of the hybrid PET/CT scanner.[13,14] Figure 2 shows the data flow of a typical PET/CT scanner.

Hybrid PET/CT systems have several positive features that are absent in stand-alone PET and CT units. PET/CT is a hardware-based image-fusion technology that virtually eliminates the uncertainty and inconvenience of currently available software fusion of separate PET and CT images, which are often acquired with patients in different positions. It should be emphasized that the PET/CT unit is not simply a PET and CT combination—not from the perspective of system design, nor the practical utility. Other than the fact that one does

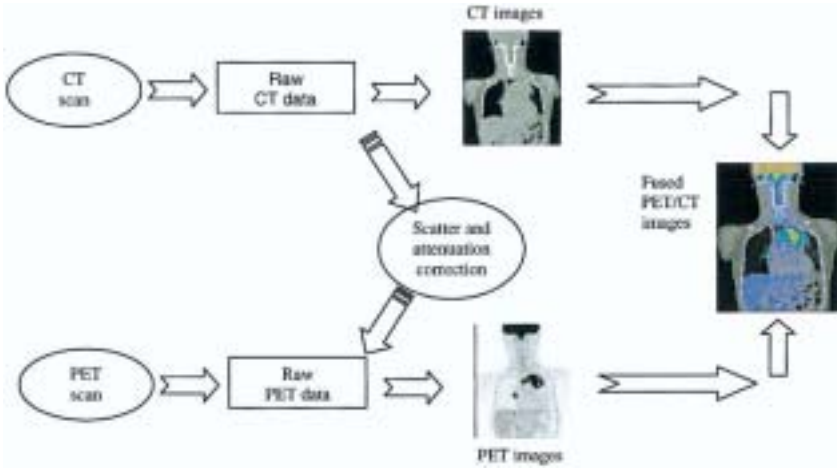


Fig. 2. Schematic drawing of the data flow in a hybrid PET/CT.

not have to go through the cumbersome and time consuming software fusion process, it has the advantages of simultaneous availability of the fused images, convenience to the patient and the physician, increased physician confidence in interpreting the image findings, and ~30% of reduction in PET scanning time due to the use of CT data for PET attenuation correction.

Integration of biological imaging techniques and multimodality image fusion

FDG-PET provides a means to study metabolic activity of tumors *in vivo*. Initial studies incorporating FDG-PET into treatment planning have been reported.[15–17] Bradley *et al*.[17] have carried out a prospective study to determine the impact of functional imaging with FDG-PET on target volumes among non-small cell lung cancer (NSCLC) patients being considered for definitive radiation therapy. They found that radiation targeting with fused FDG-PET and CT images resulted in alterations in radiation therapy planning in over 50% of patients by comparison with CT targeting. The changes included the alterations in the AJCC TNM stage (31% of the patients studied) and modification of target volume (58% of the patients studied). In a separate study, MacManus *et al*.[16] reported that 30% of patients with locally-advanced NSCLC became ineligible for curative radiotherapy because of detection of either distant metastatic disease or intrathoracic disease too extensive for radical radiation. Recently, Howard *et al*.[18] have studied the value of FDG-PET/CT for esophagus cancer and reported similar findings.

Emerging PET tracers for oncologic imaging

While FDG-PET has been shown to be effective for a number of malignancies, imaging of many other neoplasms, such as breast cancer and prostate cancer, with FDG has shown less success.[19,20] Many pitfalls have previously been described with FDG-PET imaging. The FDG tracer can be nonspecifically taken up by several benign conditions such as inflammatory disease, pneumonia, brown fat, muscle, bowel uptake, and granulomatous disease. Also, slow-growing indolent tumors may exhibit only a mild increase in glucose metabolism and therefore be missed by FDG-PET.[21–23] Thus, FDG-PET is only minimally useful for the evaluation of indolent tumors such as organ-confined prostate cancer. The recent development of fluorothymidine (FLT)[24–26] provided a new opportunity to improve the sensitivity and specificity of PET imaging of cancer. Because there is upregulation of thymidine transport into malignant cells due to accelerated deoxyribonucleic acid synthesis, either ^{11}C or ^{18}F-labeled thymidine radiotracers can be used to determine cellular proliferation. Several studies have shown that the accumulation of FLT correlates better with proliferation in comparison with the commonly used FDG tracer.[25,26] Recently, Smyczek-Gargya *et al*.[27] have reported FLT-PET imaging experiments involving 12 patients with 14 primary breast cancer lesions (T2–T4). Thirteen of the 14 primary tumors demonstrated focally increased FLT uptake. The study showed that FLT-PET is suitable for the diagnosis of primary breast cancer and locoregional metastases and the high image contrast of the technique may facilitate the detection of small foci.

Agents, such as antisense molecules, aptamers, antibodies, and antibody fragments, can be aimed at molecular targets for biological imaging. Tumor receptors and certain cellular physiologic activities, including metabolism, hypoxia, proliferation, apoptosis, angiogenesis, and infection, provide such targets. In addition to FLT, there are several other new nuclide imaging tracers under clinical or laboratory investigations,[21,28–35] which include, to name a few, ^{11}C-Acetate,[36–38] ^{18}F-choline,[39,40] ^{11}C-choline,[41,42] ^{64}Cu-DOTA-Bombesin[43], ^{18}F-FMISO,[44,45] ^{18}F-FAZA,[46] ^{64}Cu-ATSM.[47] For example, carcinogenesis is often characterized by enhanced cell proliferation

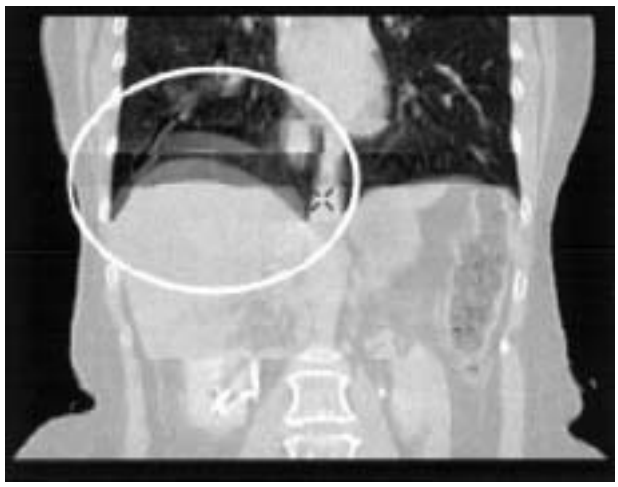


Fig. 3. Example of 4D CT where respiratory cycle irregularities have produced significant interbed mismatches near the base of the lung.

and transformation, and elevated levels of choline and choline kinase activity in certain neoplasmic dieases have motivated the development of positron-labeled choline analogs for noninvasive detection of cancer using PET.[41] Choline acts as a precursor for the biosynthesis of phospholipids, *e.g.*, phosphoatidylcholine, the major components of cell membrane. Several preliminary studies have demonstrated the potential of the new tracer for prostate cancer and many other cancers.[37,41,48]

Biologically conformal radiation therapy (BCRT)

The current 3DCRT or IMRT inverse planning is typically aimed at producing a homogeneous target dose under the assumption of uniform biology within the target volume. In reality, it is well known that the spatial biology distribution (*e.g.*, clonogen density, radiosensitivity, tumor proliferation rate, functional importance) in most tumors and sensitive structures is heterogeneous. Recent progress in biological imaging is making the mapping of this distribution increasingly possible. This new development opens a new avenue of research, coined BCRT.[49–53] The goal of BCRT is to take the inhomogeneous biological information derived from biological imaging into account and to produce customized nonuniform dose distributions on a patient specific basis. The simultaneous integrated boost (SIB) to elective volumes recently appearing in the literature represents a simple example of BCRT.

To establish BCRT, 3 major aspects must be addressed: (1) determination of the distribution of biological properties of the tumor and critical structures; (2) prescription of the desired dose distribution for inverse planning; and (3) inverse planning to generate most faithfully the prescribed nonuniform dose distribution. While the development of molecular imaging techniques is critically important in mapping biology distributions, the successful integration of this information into IMRT planning through steps (2) and (3) is also indispensable to fully exploit the obtained biology information to improve patient care. With the optimistic assumption that spatial biology distributions within a patient can be reliably determined using biological imaging in the future, Yang and Xing[53] have established a theoretical framework to quantitatively incorporate the spatial biology data into IMRT inverse planning. To implement this method, they first derived a general formula for determining the desired dose to each tumor voxel for a known biology distribution of the tumor based on a linear-quadratic (LQ) model. By maximizing the TCP under the constraint of constant integral target dose, they obtained

$$D_0^T(i) = \frac{\alpha'_{ref}}{\alpha'_i} D_{ref} - \frac{1}{\alpha'_i}(\gamma_{ref} - \gamma_i)\Delta T - \frac{1}{\alpha'_i}\ln\left(\frac{\alpha'_{ref}\rho_{ref}}{\alpha'_i\rho_i}\right), \quad (1)$$

where $D_0^T(i)$ is the desirable prescription dose at the voxel i with the tumor cell density, radiosensitivity, and proliferation rate given by (ρ_i, α_i, γ_i), and D_{ref} is the reference dose for the voxel with reference radiobiological parameters (ρ_{ref}, α_{ref}, γ_{ref}). For a given disease site, the radiation dose used in current clinical practice with "intent to cure" can be used as a good starting point in selecting the value of D_{ref}. The relation is quite general and can be used as prescription dose to guide an arbitrary inverse planning objective function aimed at producing a customized dose distribution in accordance with the spatial biology information.

INTRA-FRACTION ORGAN MOTION: MANAGING THE RESPIRATORY MOTION

Components affecting the reproducibility of target position during and between subsequent fractions of radiation therapy include the displacement of internal organs between fractions and internal organ motion within a fraction. Depending on the disease site, these components contribute differently to the margins that are to be added around the CTV to ensure adequate coverage. In the thorax and abdomen, intra-fraction internal anatomy motion due to respiration is a principal cause for large safety margins. Motion can distort target volumes and result in positioning errors as different parts of the tumor move in and out of the image window with the patient's breathing cycle. Several studies, conducted to examine the extent of diaphragm excursion due to normal respiration, reported the range of motion from ~0.5 to 4.0 cm in the superioinferior direction. As a consequence of a significant margin added around the CTV, a large amount of normal tissue surrounding the CTV is irradiated. Accounting for such motion during treatment has the potential to reduce margins, leading to reduced radiation toxicity and risk of treatment-induced complications, and yielding room for dose escalation.

A complete solution compensating for respiratory motion should ideally start at the simulation stage. There have been several studies to characterize the amplitude, phase and periodicity of organ motion[54–56] using fluoroscopic x-rays, ultrasound,[57,58] and magnetic or RF markers.[59,60] The development and deployment of spiral and multi-detector CT scanners have made practical the acquisition of time-resolved or 4D CT images. The reconstructed images acquired with patients in treatment positions provide 4D models upon which geometric as well as dosimetric computations can be performed. 4D PET is also becoming clinically available.[61–63] Treatment-wise, respiratory gating technology and tumor tracking techniques to synchronize delivery of radiation with the patient's own respiratory cycle are under intensive investigations.

4D CT imaging

A 4D CT can be either prospective or retrospective. In the former case, the scanner collects images at only one of the breathing phases of the patient instead of scanning continuously. The retrospective 4D CT scan results in multiple image sets, corresponding to different breathing phases of the patient, and consists of 3 relatively orthogonal processes[64–68]: recording of respiratory signal(s), acquisition of time-dependent CT projection data, and construction of a 4D image from these data. The first objective can be achieved by tracking a surrogate of respiration-related organ and tumor motion, such as tidal volume measured with a spirometer,[66,69] chest expansion monitored by a pneumatic bellows,[70] or a reflecting external marker placed on the abdomen and tracked with a camera.[64] Time-dependent CT data can be acquired by oversampling in either helical or cine mode, and constructing several CT slices over the full respiratory cycle at each axial location.[67,71] Finally, the respiratory signal and CT data must be combined into a 4D series, providing a CT volume at several points throughout the respiratory cycle. In this section, we will focus primarily on the implementation of 4D CT provided by the Varian Real-time Position Management (RPM) camera/software and the GE Discovery ST multislice PET/CT scanner.

4D CT patient setup proceeds along the same lines as a standard 3D CT exam. The patient is immobilized on the scanner bed, and aligned using room and scanner lasers. Sagittal and coronal scout images are used to verify patient positioning, and the setup is adjusted as necessary. At this stage of the setup, the 4D procedure begins to diverge from the 3D exam.

The RPM system consists of an infrared source, CCD camera, and a reflecting block. The block is attached to the patient's abdomen, typically just inferior to the xiphoid process, and the anterioposterior motion of the block is captured by the camera. This motion is analyzed in real-time by Varian software on a computer connected to the RPM camera. The breathing pattern is recorded for the duration of the scan, and is referred to as the "respiratory trace." Once the scan has finished, the software retrospectively computes the phase at each point of the respiratory trace by determining the location of the peaks at end-inspiration, and assigning percentages to interpeak points based on a linear interpolation of the peak-to-peak distance. For example, under this scheme, end-inspiration occurs at 0%, while end-expiration typically appears near 50–60%. The peak-to-peak distance can vary between respiratory cycles, as can the position of end-expiration with respect to end-inspiration.

Irregularities in a patient's respiratory pattern can often be reduced by encouraging the patient to breathe calmly and consistently, and then relying on the patient's compliance during the scan. If this free-breathing approach is insufficient, the RPM software can provide audio coaching in the form of a "breathe in, breathe out" recording, which is manually or automatically timed to the patient's natural rhythm. Some groups have used video feedback either alone or concurrently with audio instructions.[72] While audio and video coaching can help by stabilizing the respiratory period, amplitude and baseline, they can complicate matters for patients with compromised respiratory function, who find it difficult or impossible to maintain a regular rhythm. Another solution is active breath control (ABC)[73–75] which uses modified ventilator equipment to control the airflow, albeit at the (possibly significant) expense of patient comfort.

Once a sufficiently regular breathing pattern has been established, the CT data is acquired in "cine" mode. This is a step-and-shoot technique, whereby the gantry rotates several times at each bed position to acquire data over the full respiratory cycle. The raw data is partitioned into bins corresponding to a user-selected time interval (typically less than 1/10th the average cycle), and CT slices are automatically reconstructed from these bins. Because several respiratory points are sampled at each bed position, a 4D CT scan can take several times as long as a corresponding 3D CT, resulting in typically 1500–3000 CT slices for a 20–40-cm axial FOV.

The respiratory and scan data are combined at a separate computer, the Advantage Workstation (AW) (GE Medical Systems), which uses the respiratory trace to sort the oversampled CT slices according to their phase. The AW does perform the phase calculations, but rather relies on the phase stamp computed by the RPM during the creation of the respiratory trace file. Missing phases for any couch position are replaced with their nearest neighbor, providing a sorted image without any phase gaps. The user can navigate through the data in each axial direction, similar to standard viewing software, but can also scroll through the respiratory phases from end-inspiration to end-expiration. Individual phases can be subsequently extracted, or combined into averaged or minimum/maximum intensity projections, and exported to planning software in the form of standard

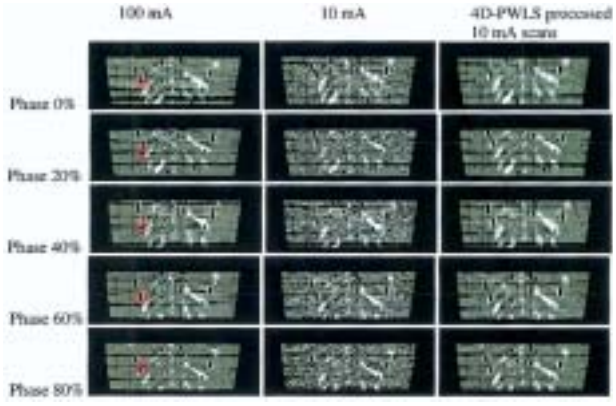


Fig. 4. Motion phantom study for the 4D-PWLS method with the thorax phantom. The left and middle columns are the original phases obtained from the GE Advantage Workstation, for 100 mA and 10 mA, respectively; the right column shows the 10-mA phases after 4D-PWLS processing. The red rectangles represent the selected ROI for calculation of SNRs, each of which contains $5 \times 5 \times 5$ voxels. PWLS smoothed 10-mA scan resulted in more than 2-fold increase in the SNR for every phase of the periodically moving phantom. Similar results were obtained in a patient 4D CT study.

DICOM series. These exported image series form the basis of 4D treatment planning.

Unresolved issues in 4D CT

The AW sorts the data by phase rather than amplitude. If the breathing were perfectly regular from cycle to cycle, then phase- and amplitude-based sorting would give very similar results. The problem arises when there is variation in amplitude, period, or baseline, or when the onset of end-expiration does not occur at the same point each cycle. When these inconsistencies arise, the sorted CT images may contain mismatch artifacts at the interface between bed positions (see Fig. 3). Recent studies have investigated amplitude-based binning as an alternative to the phase-based approach, and it appears that amplitude sorting can improve image quality in many cases.[76–78] Other researchers have matched adjacent CT slices without using a respiratory trace, by maximizing the continuity of CT units integrated over regions of interest.[71] Yet another promising approach involves interpolating the CT data continuously between end-cycle peaks using deformable models.[79]

A second issue arises in the correlation between external fiducial movement and tumor/organ motion. Amplitude ratios between fiducial and tumor displacement may vary from cycle to cycle, and thoracic and abdominal points may involve relative phase shifts.[54,80] These shifts may be especially crucial for tumors near the lung, where hysteresis is prevalent. Finally, larger organs such as the liver can experience substantial deformation during inspiration and expiration, which may not be adequately captured by rigid-body interpolation between points in the respiratory cycle

Finally, even if the 4D CT images have been acquired without problem, there remains the issue of reproducibility at treatment.[81] If treatment planning and delivery are based on 4D CT, there is an implicit assumption that anatomic motion during treatment will match the tumor and organ motion observed during setup. This assumption can be checked to some degree through frequent gated or breath-hold portal imaging.[82] On the other hand, it is reasonable to assume the patient will relax over time, so that their breathing becomes shallower or changes tempo. Indeed, studies have demonstrated that some patients exhibit systematic respiratory changes over a several-week course of radiation therapy, even with visual and audio coaching.[83] These issues strike at the heart of IGRT, and provide a fertile ground for research.

4D CT usually delivers more radiation dose than the standard 3D CT, because multiple scans at each couch position are required to provide the temporal information. We have developed a method to perform 4D CT scans at relatively low current, hence reducing the radiation exposure of patients.[68] To deal with the increased statistical noise caused by the low current, we proposed a novel 4D penalized weighted least square (4D-PWLS) smoothing method, which can incorporate both spatial and phase information. The 4D images at different phases are registered to the same phase via a deformable model, whereby a regularization term combining temporal and spatial neighbors is designed. The proposed method was tested with phantom experiment (see Fig. 4). for an example) and patient study, and superior noise suppression and resolution preservation were observed.

4D PET and related issues

4D PET poses a problem distinct from 4D CT, in that signal is inherently limited by the tolerable patient dose. The result is that any PET scan requires a significant amount of time per bed position (usually a few minutes) to acquire sufficient data to produce a good image. This limitation makes it difficult to partition PET data with the same time resolution possible in 4D CT, but nonetheless, acquisition methods are clinically available to obtain PET images at end-inspiration or end-expiration. The most common solution is to gate the PET scan at the desired respiratory end-point, and reconstruct a single bin of gated data.[84–86]

Patient setup proceeds in the same manner as an ungated PET scan, and a CT image is acquired for attenuation correction just prior to the PET. At this point, the RPM system monitors patient breathing by tracking the reflecting block, and the acquisition trigger is set by the user to occur at some given point (say, end-inspiration) in the cycle. Each time the RPM camera determines that the reflecting block (and, by extension, the patient's respiration) reaches this point in the respiratory cycle, a trigger is sent to the scanner, and data accumulation is initiated. Gated PET differs fundamen-

tally from the 4D CT protocol, by elevating the RPM system to this active role in data acquisition.

In gated mode, the user is able to select both the width of the acquisition window and the number of sequential bins to record each respiratory cycle. The bin width directly affects image quality, because the signal-to-noise ratio within an image asymptotically approaches the square root of the signal level.[87] Multiple bin acquisition allows the capture of the full respiratory cycle in several bins, offering the possibility of retrospectively sorting into 2 or more respiratory phases. Each time the RPM trigger is received, data is directed to the initial bin, and then to the remaining bins sequentially until the next trigger. This process continues for the duration of the scan. Ideally, the scan duration would be chosen such that the first bin (the respiratory point of most interest) would accumulate as many data points as a comparable ungated scan (*i.e.*, divide the bin width by the duty cycle). In reality, because this would lengthen the typical PET scan by a factor of 4 or 5, practical clinical considerations may require the gated scan to be shortened, with corresponding image degredation.

Once the scan has finished, it is possible to associate each bin (beyond the first bin) with a corresponding point in the respiratory cycle. Because the respiratory trace is recorded by the RPM, it is a relatively simple matter to analyze the respiratory motion offline and make this correspondence. It is also possible to retrospectively combine multiple bins into a single bin, merging all the data to create an effectively ungated scan. However, these methods are not yet available from the vendor as a clinical tool, and must be performed by the user in the context of research efforts. Once the desired bin has been selected, its data can be reconstructed using the vendor-supplied filtered backprojection or OS-EM algorithms. The image results can subsequently be exported to treatment planning systems for review, similar to ungated PET series.

A salient point in the PET reconstruction process is the specification of the attenuation correction map. The current clinical design uses the CT scan acquired just prior to the PET specifically for this purpose. This attenuation correction CT can be an acquired during either free breathing or breath-hold. Some research has indicated that PET reconstructions can be quite sensitive to distortions in the attenuation correction map,[88–90] and investigations are ongoing into the use of 4D CT or other models to accurately account for attenuation.[61,91] On the Varian/GE system, this requires selecting the appropriate images from the 4D CT on the AW, sending these series back to the scanner, generating the attenuation correction maps for each 4D PET bin, and then reconstructing each bin separately. Once again, this is a research solution, and not yet available from the vendor for clinical use.

Combining 4D PET with 4D CT and enhancement of the performance of 4D PET with post-acquisition data processing

Once the 4D PET has been acquired (either a single phase, or perhaps several), it is possible to create a 4D PET/CT.[61] This involves manually selecting the PET and CT images with corresponding respiratory phases (or amplitudes), and fusing them on viewing/planning software. We have recently developed a 4D-4D image registration algorithm, which allows us to automate the process. If the CT and PET scans are acquired with the same patient position on the same exam, then the process is a particularly simple hardware-based registration. On the Eclipse treatment planning system, for example, 2 images (not just PET/CT, but other modalities as well) can be automatically fused if they share the same DICOM coordinates. If the DICOM coordinates are not identical, the registration is more difficult, requiring manual or automated shifts and rotations to match anatomical landmarks or fiducials. Fusion may be additionally complicated by organ deformation[92,93](see Rigid and Deformable Image Registration Section below). At the present time, PET/CT hardware fusion for ungated scans is well established and readily available within the clinical setting.[13,14] 4D PET/CT registration, however, remains primarily within the research domain.

The major issue in 4D PET is the lack of statistics. Because the collected photons are divided into several frames, the quality of each reconstructed frame is decreased with increasing number of frames. The increased noise in each frame heavily degrades the quantitative accuracy of the PET imaging. We have recently developed 2 corrective methods to enhance the performance of 4D PET. The first method, coined "retrospective" stacking (RS),[62,63,94] combines retrospective amplitude-based binning of data acquired in small time intervals, with rigid or deformable image registration methods. Unlike gating techniques, RS uses data along the entire respiratory cycle, thereby minimizing the need for lengthened scans while providing a 4-dimensional view of the region of interest.[62,63] In the second approach,[94] we reconstruct each frame with all acquired 4D data by incorporating an organ motion model derived from 4D-CT images by modifying the well-known maximum-likelihood expectation-maximization (ML-EM) algorithm. During the processes of forward- and backward-projection in the ML-EM iterations, all projection data acquired at different phases are combined together to update the emission map with the aid of the deformable model, the statistics are therefore greatly improved. Both phantom and patient studies have indicated promising potential of the 2 methods.

Radiation treatment planning based on 4D information

How to maximally utilize the time-resolved image information derived from 4D CT or PET/CT represents one of the challenges in IGRT. In reality, the information

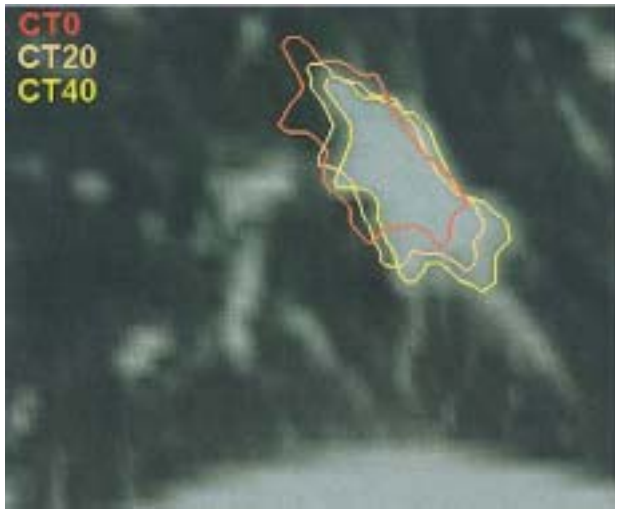


Fig. 5. Tumor contours for 3 breathing phases. The contours labeled as CT20 and CT 40 were produced by applying the deformation field on the tumor contours delineated in CT0. The "trajectories" of the tumor boundary pictorially show the extent of tumor movement and allow us to specify patient specific margin in accounting for the intra-fraction organ motion.

can be integrated into radiation treatment planning and delivery at different levels. At the lowest level, the 4D images can be employed to determine the extent of tumor movement on a patient specific basis and the information can then used to design the CTV margin and the radiation portals to accommodate the motion. Figure 5 shows an example of lung patient, in which tumor boundaries at 3 distinct respiratory phases are plotted. We have referred to this type of treatment as "3.5-dimensional" radiation therapy. The 4D information can also be used for guiding breath-hold or gated radiation therapy. There is also strong interest in using the 4D data to establish a 4D patient model and then to carry out a 4D radiation therapy plan. These are the subjects of the following 2 sub-sections.

Breath-hold and respiratory gating

Various methods have been worked out to counteract respiratory motion artifacts in radiotherapy imaging. Among them are breath-hold, respiration gating, and 4D or tumor-tracking techniques.[55,56,73,75,95] Breath-hold techniques either actively or passively suspend the patient's respiration and treat the patient during this interval. Deep inspiration breath-hold, active breathing control (ABC) (which forces shallow breathing and thereby "freezes" the tumor motion for a small part of the treatment time[73]), and self-held breath-hold are suitable for different types of therapy targeting different cancers. Different types of equipment, such as stereotactic frames, fiducial tracking systems, timers, respirometer, RPM, or interlocks, may be needed depending on the method of breath-hold.

Respiration-gating methods involve tracking the patient's natural breathing cycle and periodically turning the beam on when the patient's respiration signal is in a certain phase of the breathing cycle (generally end-inhale or end-exhale). The patient's respiration is continuously monitored and the beam switches off as the tumor moves out of the target range. Gated radiation therapy can offset some of the motion but requires specific patient participation and active compliance. In gated treatment, it is required that the CT images used for treatment planning faithfully represent the actual treatment situation. While gated CT acquisition at the treatment respiratory phase is possible, our gating protocol proceeds by picking up the CT data at an appropriate phase from the patient's 4D CT acquired using the method described above. The gating interval is typically centered at end-expiration because of the increased reproducibility at this point, and spans 20–30% of the breathing period to provide a reasonable duty cycle. Treatment plans are optimized for this phase range by planning on an averaged composite of the scans within the interval, and using maximum- and minimum-intensity pixel views to incorporate intra-gate margins. The averaged, maximum-intensity and minimum-intensity composites for a lung patient are displayed in Fig. 6.

Tumor tracking

Similar to the establishment of a 3D geometric modeling based on traditional CT data, the availability of 4D imaging information makes it possible to build a patient specific 4D model. Figure 7 shows the the 4D model for a lung patient.[79] In obtaining the models, a BSpline deformable registration technique (see Rigid and Deformable Image Registration section below) was used to register different phases of the 4D CT. Ideally, organ motion represented by the 4D model can be incorporated into the radiation treatment plan optimization to overcome the adverse effect of respiratory motion on IMRT delivery.[96] A few groups[95–100] have explored the feasibility of MLC-based tumor tracking. However, the interplay between different phases has been ignored during the plan optimization in most of these studies. Webb has presented a technique to model the dosimetric effect of elastic tissue movement when modulated beams are delivered.[101] In general, the quadratic inverse planning objective function becomes

$$F = \sum_{t}\sum_{k} \cdot \sum_{i} w_k \left[d_p^k - \sum_{t} d_i^k(\vec{\mathrm{r}}, t) \right]^2 \qquad (2)$$

where d_p^k is prescribed dose for kth structure, w_k is the importance factor, and d_i^k $(\vec{r}, t)$ is the calculated dose in voxel i at time t, and the summation over t represents the integral dose to ith voxel. For 4D planning, it is necessary to know the path of each material coordinate during the treatment, which involves registering the voxels in different respiratory phases. This can be achieved by using a deformable registration algorithm. The optimization of the above objective function or alike[95,102–106]

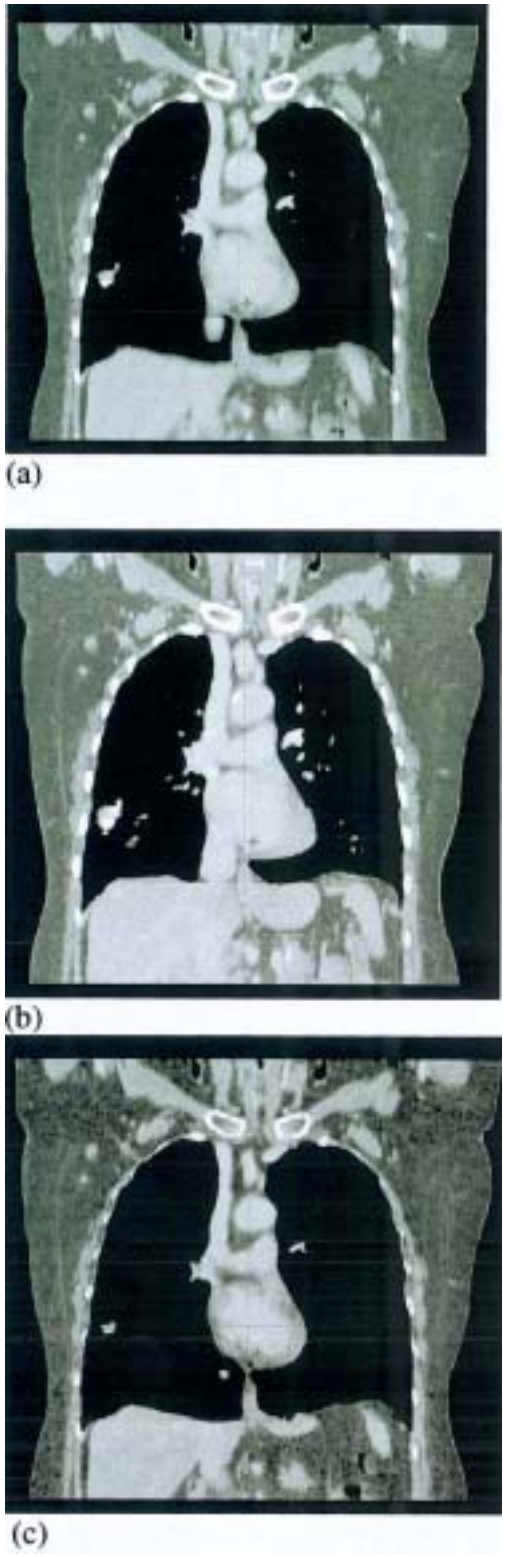


Fig. 6. Composite scans of a 4D CT lung patient. (a) Average pixel; (b) maximum-intensity pixel; (c) minimum-intensity pixel. The maximum-intensity pixel composite reveals the motioextent of hyperdense tissue (*e.g.*, lung tumor), while the minimum-intensity pixel view provides the motion range of hypodense regions (*e.g.*, lung air volume).

can proceed in a similar fashion as conventional 3D inverse planning to derive the optimal trajectories of the movements of the MLC leaves. An aperture-based optimization[107–109] seems to be more adequate for dealing with the organ motion.[96]

4D methods propose to track the tumor with the radiation beam as the tumor moves during the respiration cycle. These techniques require acquisition of some form of respiration signal (infrared reflective markers, spirometry, strain gauges, video tracking of chest outlines, and fluoroscopic tracking of implanted markers are some of the techniques employed to date), which is assumed to be correlated with internal anatomy motion. Fluoroscopy and the cine model electronic-portal-imaging device (EPID) have been proposed as a means for real-time guidance.[110,113] While tumor tracking seems to be the ultimate goal of 4D radiation therapy, the real challenge is clarifying whether the 4D model is repeatable at the time of fractionated treatments, and determining how to correctly synchronize the MLC movements with the patient breathing. Real-time imaging and/or adaptive approaches will likely play a role in this aspect and the issue will surely need more research for many years to come.

INTER-FRACTION ORGAN MOVEMENT

Current techniques in dealing with inter-fraction organ movement

Uncertainty in patient setup has long been known as a limiting factor to conformal radiation therapy. Currently, the accuracy of patient setup is verified by megavoltage (MV) radiograph acquired with either radiographic film or EPID.[114,115] The patient's bony landmarks are used to guide patient alignment. Poor soft tissue contrast and often unclear projection of the bony anatomy are major problems of the approach. To improve the situation, planar kV x-ray imaging has been implemented in a variety forms.[110–113] While these systems show significantly increased contrast for bony structure differentiation, observing soft-tissue detail remains problematic and correction of daily organ motion is still challenging. Attempts have been made to use CT imaging to facilitate the patient setup process. Along this line, the offline adaptive-radiation-therapy (ART) strategy[116] and in-room CT approach[117] have been studied. The former method aims to partially compensate for organ motion by carrying out multiple CT scans in consecutive days in the first week of treatment. The image data are then employed to construct a patient specific PTV model from the composite of the CTVs with inclusion of statistical variations of the observed motions. While beneficial, the approach is hardly an ideal solution for dealing with the inter-fraction organ motion. It relies on establishing a statistical ensemble of all possible setup scenarios under a strong assumption that a limited number of off-line CT scans can ade-

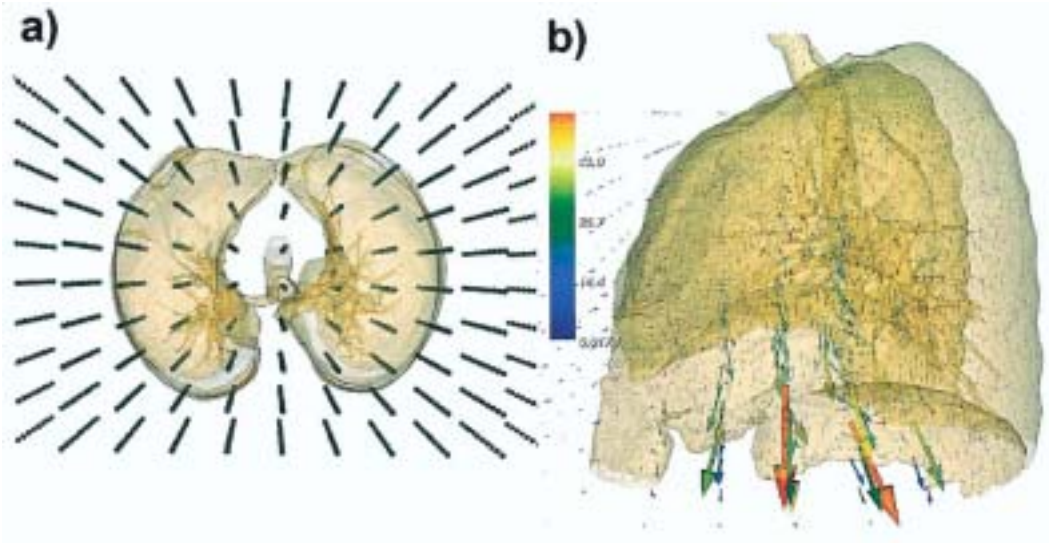


Fig. 7. (a) The BSpline grid superimposed on lung contours. (b) On each node, deformation is represented by arrows, where arrow length is proportional to the deformation.

quately describe the inherently complex, often unpredictable inter-fraction organ motion. Even when it is achievable, the ART margin is not optimized on a daily basis and there is still room for further improvement. An integrated CT/LINAC combination, in which the CT scanner is located inside the radiation therapy treatment room and the same patient couch is used for CT scanning and treatment (after a 180° couch rotation), should allow more accurate correction of interfractional setup errors. Some major radiotherapy vendors provide options to install a CT scanner in the treatment room. The overall precision of EXaCT Targeting™ from Varian has been evaluated by Court *et al.*[117] However, the approach assumes a fixed relationship between the LINAC isocenter and the CT images and relies heavily on the mechanical integrity of the 2 otherwise independent systems. Increased capital cost and prolonged imaging and treatment are other concerns.

Other patient localization techniques available include ultrasound-based methods, video-based surface tracking, on-board cone-beam CT or kV x-ray imaging, CyberKnife and Tomotherapy, etc. For prostate radiation therapy, on-line ultrasound imaging has gained substantial interest[118–120] but in practice has been found susceptible to subtle sources of error and inter-user variability. On-board CBCT holds promise to become a robust integrated on-line imaging technology that can yield unambiguous soft-tissue detail at the time of treatment. Furthermore, CT numbers correlate directly with electron density, thereby providing the potential for reconstruction of the actual dose delivered on a daily basis, in addition to simple anatomic structure alignment. The details of emerging CBCT will be presented in the next section. The robotic CyberKnife™TM from Accuray Inc. (Sunnyvale, CA) represents another promising technology. The system has a feedback mechanism in which motion of the CTV, determined through the Accutrak infrared-x-ray-correlated imaging system, can be fed back to the robot to track the CTV.[121] However, while this improves the duty cycle, there is a finite time between measuring tumor position and arranging the compensation for motion. Helical tomotherapy is an alternative means of delivering IMRT using a device that combines features of a linear accelerator and a helical CT scanner.[122] The commercial version, the HI-ART IITM, can generate CT images using the same MV radiation beam that is used for treatment. Because the unit uses the actual treatment beam as the x-ray source for image acquisition, no surrogate telemetry systems are required to register image space to treatment space. Objective measures of noise, uniformity, contrast, linearity, and spatial resolution, and comparison with that of a commonly utilized CT simulator, have recently performed by Meeks *et al.*[123]

CBCT for patient localization

CBCT based upon flat-panel technology integrated with a medical linear accelerator has recently become available from linac vendors for therapy guidance. The volumetric images may be used to verify and correct the planning patient setup in the linac coordinates by comparing with the patient position defined in treatment plan. Both kV and MV beams have been utilized for the application. The former typically consists of a kV-source and flat-panel combination mounted on the drum of a medical accelerator,[124] with the kV imaging axis orthogonal to that of MV therapy beam. The fan-beam and cone-beam MV CT in clinical applications have been reported by Meeks *et al.*[123] and Poulliot *et al.*[125], respectively. It appears that the MV images contain sufficient

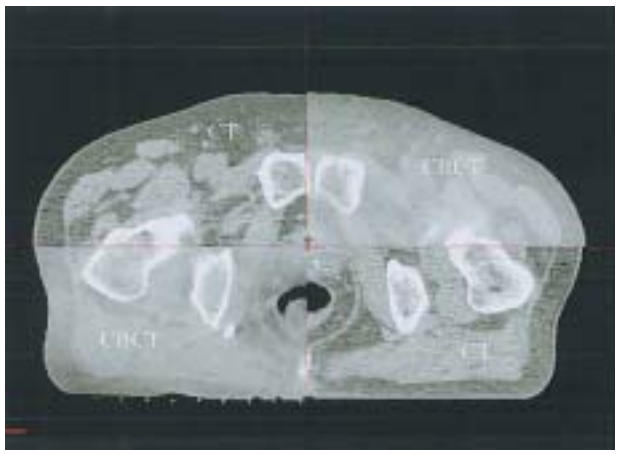


Fig. 8. The fusion of the 2 types of CT and CBCT images for a prostate case.

resolution of bone and air cavities to register them to structures in the planning CT with millimeter precision.[124,125]

Currently, CBCT is primarily used for guiding the patient setup.[126,127] The procedure is not much different from the current patient treatment, other than the fact that the AP/LAT portal images are replaced by volumetric data. In Fig. 8, we show 3D CBCT images of a prostate case in one of the fractional treatments fused with the patient's planning CT image. It is seen that soft-tissue structures and boundaries are visible to varying degrees in the CBCT images. The patient has implanted fiducials, which show up on both CBCT and planning CT. Our experience indicates that the cone beam data can clearly reveal setup error, as well as the anatomical deformations and other physiological changes. During the patient setup process, the 3D CBCT images are registered with the planning CT data through the use of either manual or automated 3D image registration software that calculates shifts in x-, y-, and z-directions (depending on the manufacturer, rotations can also be included). The movements determined during the registration represent the required setup corrections that should be applied to the patient. Both phantom and patient studies from our group have shown that the volumetric imaging is superior to the conventional MV or kV AP/LAT patient setup procedure. We note that, if only translational shifts are permissible, the level of improvement is generally within 2 mm as compared with kV AP/LAT setup procedure (2D/2D match). However, CBCT can readily detect rotational errors that may otherwise be missed by the 2D/2D match. In Fig. 9, we show the localization image for a head phantom with kV/kV 2D/2D match and 3D/3D match (CBCT/planning CT). The latter approach was found to be sensitive enough to identify a rotational error as small as 2°.

In practice, much effort is needed to improve the robustness and efficiency of the volumetric image registration process. Furthermore, to fully utilize the volumetric data, a new paradigm with seamlessly integrated simulation, planning, verification, and delivery procedure is urgently needed. Until this is realized clinically, the volumetric imaging is nothing but an expensive extension of the already functional planar verification approach. The capital cost and other related overheads do not seem to justify the marginal benefit if the volumetric data is simply used for determining the patient shift in the space. However, one should not underestimate the potential of the volumetric imaging for the future of radiation therapy, as it opens a new avenue (perhaps the only avenue), for us to realize the planned dose distribution with high confidence in clinical settings.

A few groups are working on deformable model based segmentation and patient setup procedures.[93,126–128] When deformable registration is used, there are a few options to achieve the registration depending on whether the primary aim is to match soft-tissue, or to align 3D bony structures. In Fig. 10, we show a patient's CBCT and planning CT registration results using different registration schemes. The multiple choices result from the fact that the dimensionality of the patient data is much greater than that in the patient setup procedure, and suggest that deformable registration is not the ultimate solution to volumetric image-guided radiation therapy. Nonetheless, the technique improves the current method,[127] because it partially takes into account organ deformation by achieving the closest overlay match possible between the planning and CBCT data sets according to our clinical objective, and serves as a useful interim solution before a better integrated approach becomes available.

CBCT-based dose verification

Another important application of on-board volumetric imaging is verification of dose delivery. We have recently evaluated the accuracy of kV CBCT-based dose calculation and examined if current CBCT is suitable for the daily dose verification of patient treatment.[129,130] A CT-calibration phantom was first used to calibrate both conventional CT and CBCT. CT and CBCT images of the calibration phantom, an anthropomorphic phantom and 2 patients (a lung and a prostate case) were then acquired for this study. Our results indicated that the imperfect quality of CBCT images has minimal impact ($< 3\%$) on the dosimetric accuracy when the intra-fractional organ motion is small. When intra-fractional organ motion is large and motion artifacts is severe (*e.g.*, in the case of lung cancer), the dosimetric discrepancy due to the poor image quality of current CBCT was found to be clinically significant. Furthermore, in the latter case, we found that it is possible to use a deformable registration algorithm to map the corresponding electron density information from planning CT to CBCT and then to proceed with conventional dose calculation.

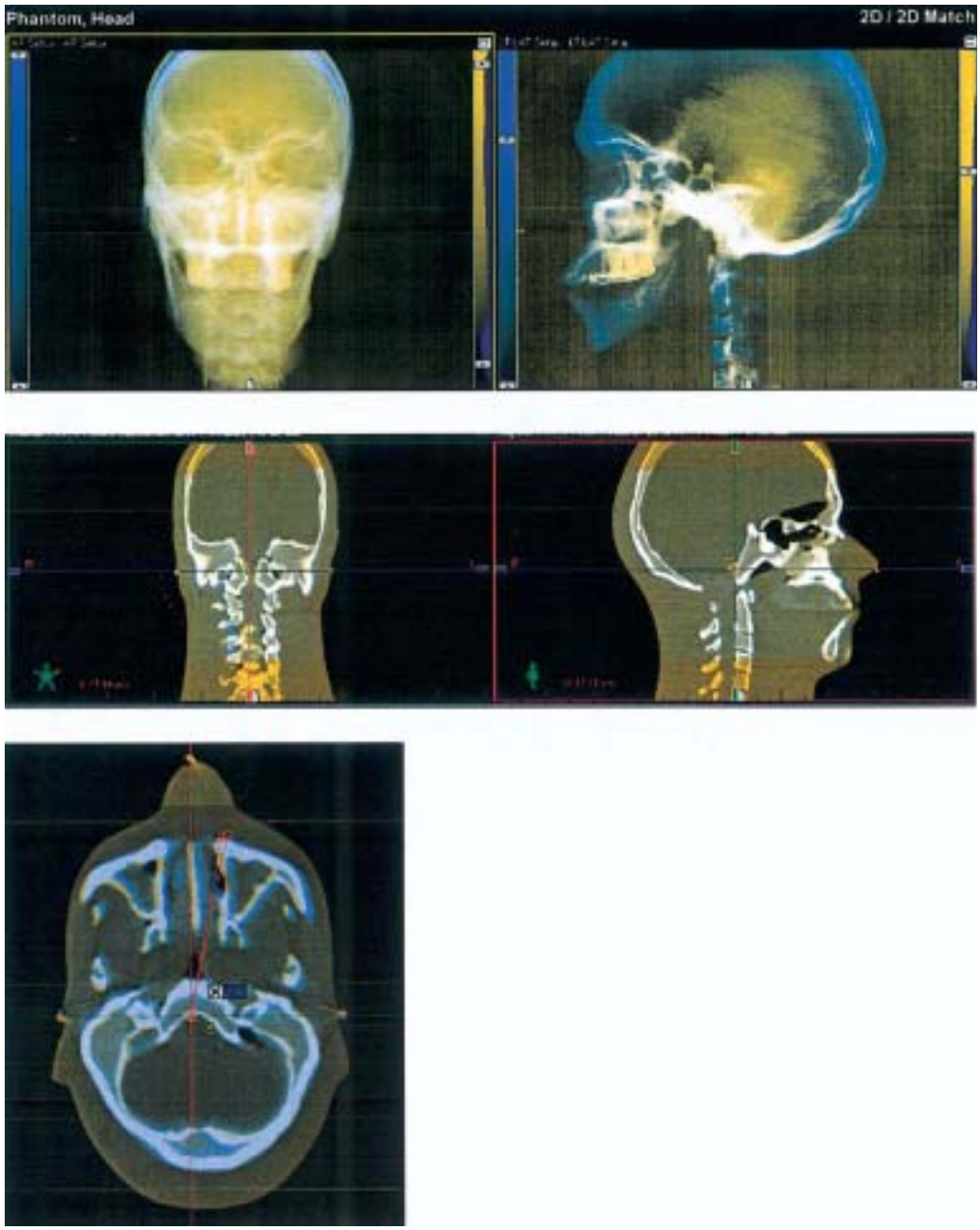


Fig. 9. Setup localization image for head phantom with kV/kV 2D/2D match (top) and 3D/3D CBCT match (middle). The image shown in the bottom panel illustrates that the CBCT is a sensitive technique capable of picking up a 2° rotational miss-match between the planning CT and CBCT.

Respiratory motion artifacts in CBCT

Superior to the common approaches based on 2 orthogonal images, CBCT can provide high-resolution 3D information of the patient in the treatment position, and thus has great potential for improved target localization and irradiation dose verification. In reality, however, scatter and organ motion are 2 major factors limiting the quality of current CBCT. When CBCT is used in imaging thorax or upper abdomen of a patient, respiration induced artifacts such as blurring, doubling, streaking, and distortion are observed, which heavily degrade the image quality, and affect the target localization ability, as well as the accuracy of dose verification. These artifacts are much more severe than those found in conventional CT exams, in which each rotation of the scan can be completed within one second. On the contrary, in CBCT scan, the gantry rotation speed is much slower, typically 40 seconds to 1 minute for a full 360° scan in acquiring the projection data, which is more than 10 breathing cycles for most patients. In Fig. 11, we show the influence of the same motion on a regular "fast" CT scanner and CBCT for a motion phantom, where it is clearly seen

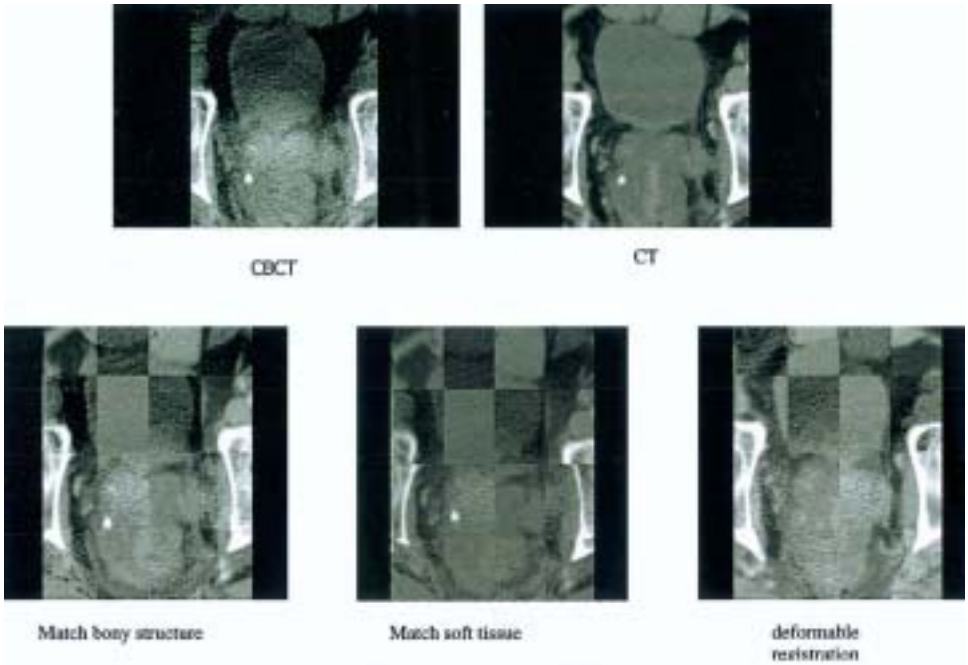


Fig. 10. Image registration of CBCT and planning CT based on bony structure matching, soft tissue matching, and deformable registration. Different matching techniques emphasize on different aspect of the multidimensional problem.

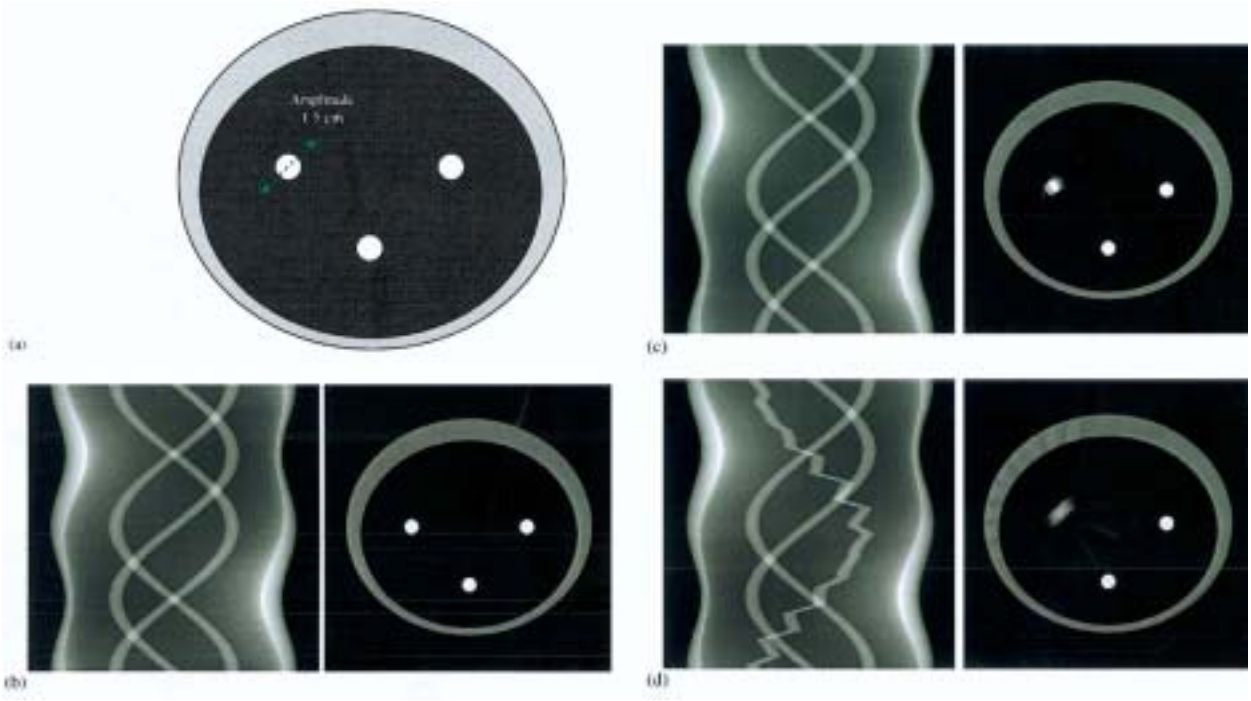


Fig. 11. (a) Motion phantom for CT and CBCT simulation study. The left circle moves diagonally with an amplitude of 1.5 cm and a period of 3.52 seconds. (b) Simulated sinograms and their corresponding reconstructed images with standard FBP algorithm when the circles are stationary. (c) and (d) show the sinograms and their corresponding reconstructed images for 1 s/rotation acquisition (conventional CT scan speed) and 40 s/rotation acquisition (on-board CBCT scan speed), respectively.

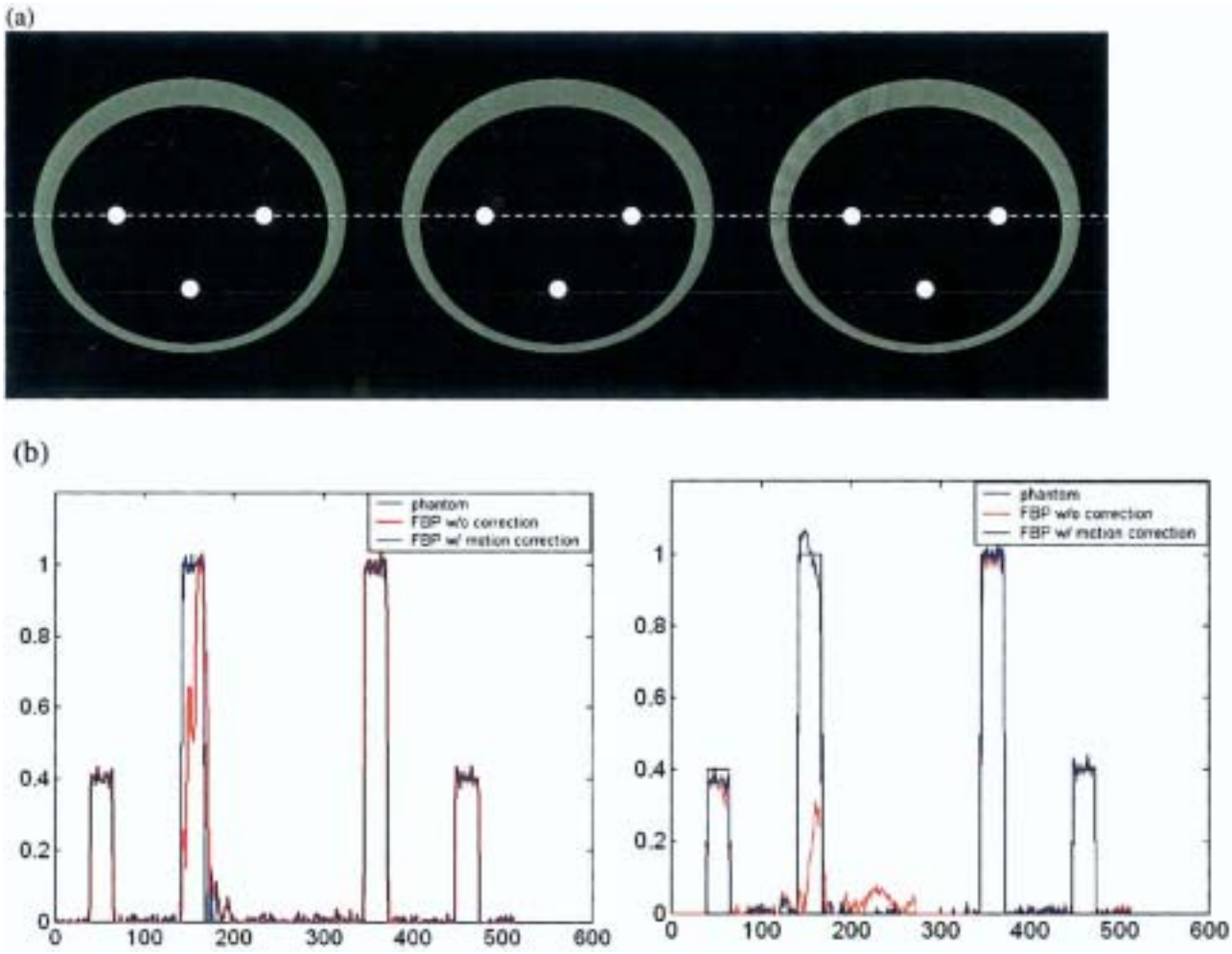


Fig. 12. (a) Phantom and images reconstructed with motion correction for CT and CBCT settings. The 3 images represent the reconstructed image of stationary phantom (left), the conventional "fast" CT (middle), and the CBCT (right). (b) Horizontal profiles through the moving circle for the reconstructed CT (left panel) and CBCT (right panel) images shown in the middle and right of (a). The profiles are in blue. For comparison, the profiles for the stationary phantom (left panel) and images reconstructed without motion artifacts removal mechanism are also plotted in each case (black and red curves, respectively).

that the motion artifacts are much greater than that in a fast scanner.

In the last decade, considerable effort has been devoted to finding solutions to remove motion artifacts and to obtain time-resolved medical images. Wang and Vannier[131] presented a patient-motion estimation and compensation technique for helical CT systems. Willis and Bresler[132] cast the motion artifact problem as a time-varying tomography problem and required special-purpose hardware to optimally sample the spatially and temporally band-limited CT signal space. A parametric model for the respiratory motion was used in MRI, and the motion artifacts were successfully reduced by modifying the reconstruction algorithm.[133] Crawford *et al.*[134] brought the concept into CT imaging, and derived an exact reconstruction formula for motion compensation for CT scans. Generally, motion correction algorithms that assume a motion model work well when the motion conforms to the model, but have limited success when it does not. As described above, 4D CT has been developed in radiation oncology application to explicitly account for the respiratory motion. The 4D CT can be used to derive a patient-specific deformation field and then incorporated into the CBCT filtered-backprojection (FBP) image reconstruction process.[135] The algorithm was tested with simulations at different settings corresponding to conventional CT and CBCT scan protocols, with translational motion and more complex motion, and with and without Gaussian noise. In Fig. 12, we show the result for the motion phantom depicted in Fig. 11.[135] Because the motion model is directly derived from the patient images, it should be more accurate than other artificial modeling, and therefore more efficient motion correction is expected. In addition to this approach, Sonke *et al.*[136] developed a CBCT procedure consisting of retrospective sorting in projection space, similar to

that used in 4D CT. The subsets of projection data are then reconstructed into 4D CBCT dataset. To achieve a sufficient temporal resolution, however, this will require slowing down the gantry rotation. The assumption of periodicity of the respiratory motion is also necessary. Li *et al.*[137] have recently established a novel 4D CBCT reconstruction formalism, in which the reconstruction of a phase will consider not only the projections corresponding to that phase but also those of other phases. By incorporating information from other phases, the efficacy and quality of 4D CBCT images are substantially improved. Zeng *et al.*[138] proposed a method to estimate the parameters of a non-rigid, free-breathing motion model from a set of projections of thorax that are acquired using a slow rotating CBCT scanner.

RIGID AND DEFORMABLE IMAGE REGISTRATION

Development of an effective image registration technique has been one of the most important research areas. Depending on the mathematical nature of the transformation, image registration is divided into rigid and deformable registrations. In rigid transformations, it is assumed that the geometry of the object is identical in the 2 input images and no distortion occurs in the image acquisition process. A rigid transformation consists of 6 degrees of freedom: 3 displacement parameters and 3 rotational parameters. Deformable registration, on the other hand, is more complicated and entails the modeling of voxel dependent distortion. Clinically, the need for a robust image registration algorithm to compare/fuse images representing the same structures imaged under different conditions or on different modalities is ever increasing because of the extensive use of multi-modality imaging and the emergence of new imaging techniques and methods.

Computer-based rigid image registration has gained widespread popularity in the last decade and is used in routine clinical practice. In this approach, the matching of the 2 input images is formulated into an optimization problem and the best registration of the 2 images is obtained by iteratively comparing various possible matches until no better registration can be found. The search for the optimal match of the 2 input images is usually gauged by a ranking function constructed based on some physical considerations. Depending on the nature of the input images, the formulation of the problem can be highly complicated. Court and Dong[139] used a rigid transformation for the correction of tissue displacement. A deformable procedure based on the finite element model (FEM), in which images are described as blocks of elastic materials on which forces apply, was proposed by Bharath *et al.*[140] and Brock *et al.*[141] In this approach, the parameters that control the behavior of the elastic material and are responsible for the conversion of forces into local deformations of the elastic material are Young's elastic modulus and Poisson's ratio. Although powerful, the model has the drawback that values of the elasticity and density constant for various tissues are not readily available and have to be found by a trial and error procedure. The method also relies on using complicated software to generate a FEM mesh and masks of the involved structures. Schreibmann and Xing have proposed a general narrow-band approach for deformable registration.[93] Depending on the problem, modeling of individual voxel movement can also be made using either B-splines,[79] thinplate splines[142,143,], optical flow algorithms,[144] or fluid flow algorithms.[145] Spline interpolation is a relatively simple approach and our experience with the algorithm suggested that the free-form registration is stable and accurate for dealing with IGRT image registration problems.[146] An improvement to this method can be achieved by using a spline model with the smoothness of the deformation field assured by the interpolation between a grid of fixed control points. A simple method along this line is to deduce the spline coefficients from a set of user-defined control points, as was done by Fei *et al.*[147] and Lian *et al.*[143] in warping and registration of MR volumes. Coselmon *et al.*[148] used a similar technique to study the accuracy of mutual-information-based CT registration of the lung at exhale and inhale respiratory states.

To facilitate the computer decision-making process, image pre-processing or user interaction may be required, especially when dealing with a deformable image registration. The use of homologous anatomic landmark pairs on the 2 input images or the control points is an example of this. In reality, the user must have a detailed understanding of the patient anatomy and the characteristics of the 2 modalities in order to accurately identify the control points on both images. The point pairs are usually obtained interactively with the user repetitively exploring the input image sets and each time trying to locate a point in both of them. Due to the 3D nature, the process is rather tedious and difficult to perform. Schreibmann and Xing[149] have developed a general method to facilitate the selection of control points for both rigid and deformable image registrations. Instead of relying on the interactive selection of homologous control point pairs on both model and reference images, in the proposed approach the user needs only to identify some small control volumes on the model image in a somewhat arbitrary fashion. This new way of image registration eliminates the need for the manual placement of the homologous control points and allows us to register the 2 images accurately and efficiently. The method was applied to both rigid and non-rigid image registration problems and our results indicated that the registration is reliable and provides a valuable tool for intra- or inter-modality image registration. In Fig. 13, we show the registration result of a rectal cancer patient who has undergone both CT and FLT-PET scans. The increased robustness and confidence in the registration and the

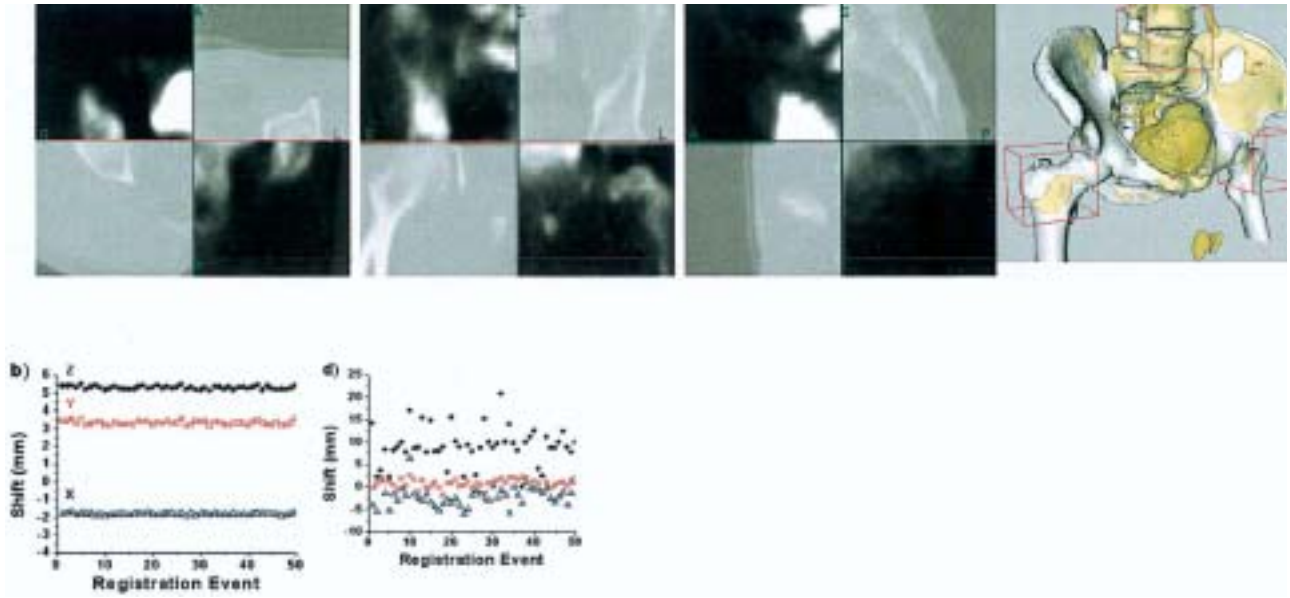


Fig. 13. Sagittal, coronal, and axial views of the CT and FLT-PET registration. In addition to the checkerboard display, a 3D view of the registration is also presented, where an excellent coincidence is observed between the bony structures revealed in CT (white) and PET images (orange). The right 2 panels of the 2nd row show the convergence behaviors of our method and the conventional method for 50 independent calculations. Our method leads to reproducible shifts in x-, y-, and z-directions, and the conventional approach based on the information contained in the whole image entity leads to large variations in the shifts.

increased speed of calculation, especially in the case of the deformable registration, are important features of the new technique. Compared to the manual rigid registration, this method eliminates the nuisance of the control point pair selection and removes a potential source of error in registration. Compared to the automated method, the technique is more intuitive and robust, especially in the presence of image artifacts.

CLINICAL EXPERIENCE WITH IGRT

Clinically implemented IGRT techniques at Stanford include 4D CT, 4D PET, Varian OBI (both planar and CBCT), gating, and Accuray CyberKnife. Several image-guided clinical protocols are under investigation. 4D CT/PET information are used in about 40% of the thorax and upper abdomen cases for patient specific tumor margin definition in 3.5D radiation therapy or for treatment planning of gated radiation therapy. CBCT is mainly applied for patient setup in the treatment of head-and-neck, and prostate and other pelvic diseases. For these sites, the CBCT image quality is reasonable to visualize soft tissues, but the quality is generally notably inferior to that of the state-of-the-art multi-slice fan beam CT scanner. Scan truncation artifacts because the patient shadow does not fit on the detector and/or organ motion often cause Hounsfield unit calibration problems. While this does not seem to influence the image registration, the use of CBCT for dose calculation should proceed with caution. Our initial experience indicates that, when compared with traditional CT-based calculation, the dosimetric error is typically less than 3% for prostate or head-and-neck cases but could be significantly greater in the thoracic region. Comparison between cone beam data and portal image derived setup errors show only slight differences (< 2 mm). However, we should note that the differences are derived purely based on the use of manufacturer-provided image-fusion software, which often emphasizes the high-intensity voxels in bony structures. The next step is to implement soft-tissue based setup corrections clinically. In reality, volumetric data contain much more information compared to planar images, and CBCT promises to be more useful in the future when it is better integrated with treatment planning and delivery systems. An ideal integration would be to use volumetric image-derived information to "tweak" or re-optimize the treatment plan. This work is still in progress at Stanford.

As another example of IGRT treatment, we describe our phase I and II pancreatic tumor dose escalation protocol. The aim is to use CyberKnife to target pancreatic tumors more precisely and to limit the toxicities associated with treatment. In a phase I study, we treated patients with a single fraction of 15, 20, and 25 Gy to unresectable pancreatic tumors using the Cyberknife stereotactic radiotherapy (SRT) system (Accuray).[150] To track tumor movement, we implant fiducial seeds percutaneously into the pancreatic tumor. Using the Accuray Synchrony platform, a model in which the position of the internal fiducials is correlated with the patient's respira-

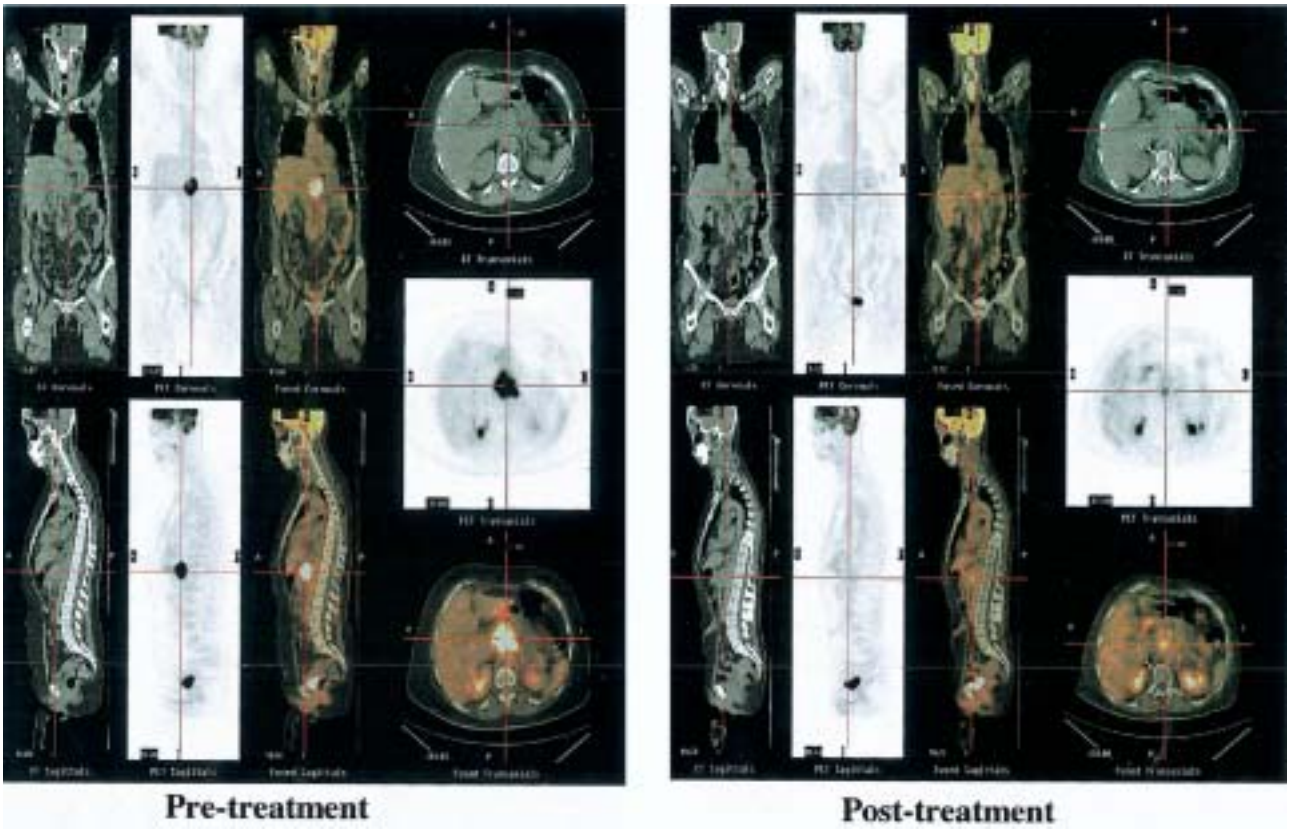


Fig. 14. FDG-PET images of a pancreatic patient before and after radiation therapy.

tory motion is developed. The Cyberknife is able to make real-time corrections to compensate for tumor movement during respiration. Prior to treatment, patients underwent 4D planning CT scans. Using this data set, we are able to visualize how the pancreatic tumor moves/deforms through respiration and compensate for these dynamic changes.[151] Minimal acute gastrointestinal toxicity was observed even at the highest dose. All patients who received 25 Gy had no further local progression of their tumor until death. In a follow up phase II study, a cohort of 19 patients were treated with 45-Gy conventionally fractionated radiation therapy using IMRT to the pancreas and regional lymph nodes followed by a 25-Gy Cyberknife stereotactic radiotherapy boost to the primary tumor.[152] An excellent rate of local control with this therapy was confirmed. Because of the rapid progression of systemic disease, we did not observe a significant improvement in overall survival as compared to historic controls. However, most patients had a clinical benefit (decreased pain, increased activity) and decreased serum tumor marker for pancreatic cancer (CA-19-9) following therapy. To document that SRS truly resulted in an anti-tumor effect, we routinely obtain FDG-PET/CT scans before and after treatment. Figure 14 is an example of one such study. There was intense metabolic activity of the pancreatic tumor prior to therapy with a near complete resolution of FDG uptake in this patient 4 weeks following therapy. The technological challenge for IGRT to minimize toxicity in this clinical scenario is the precision delivery of high-dose radiotherapy. This cannot be accomplished without taking into account the respiratory associated motion of pancreatic tumors. This movement takes place in multiple planes and can result in tumor displacement of up to 1–2 cm. Furthermore, tumor and organ deformation during respiration must also be compensated for during therapy.

SUMMARY

With the development of IMRT during the 1990s, radiation therapy entered a new era. This new process of treatment planning and delivery shows significant potential for improving the therapeutic ratio and offers a valuable tool for dose escalation and/or radiation toxicity reduction. The improved dose conformity and steep dose gradients necessitate enhanced precision and accuracy in patient localization and spawn the development of IGRT, in which various metabolic and anatomical imaging techniques are integrated into the radiation therapy process. The overall goal of IGRT is to target tumors more accurately while better sparing the normal tissues. Much recent effort is focused on removing the uncertainty in the definition of the target volume and in the determination of the position of mobile and often deformable

organs. Biological imaging described in this article will allow us not only to delineate the boundary of the tumor volume based on the tumors' biological characteristics but also to map out the biology distribution of the cancer cells, affording a significant opportunity for BCRT treatment in the future. Developments of effective 4D CT/PET techniques will provide effective means for us to understand the temporal dependence of the involved structures and design the best possible strategy for targeting the moving tumor. Integration of various imaging tools for off-line and on-line application is also of paramount importance, enabling us to ensure the planned dose distributions can be realized in the clinical setting. With the newly available IGRT tools, physicians will be able to optimize radiotherapy accuracy and precision by adjusting the radiation beam based on the actual positions of the target tumor and critical organs during radiation therapy planning and treatment. We should mentioned that IGRT is still in its infancy and many technical issues remain to be resolved, such as the establishment of a robust deformable registration method, auto-mapping of the contours outlined on the planning CT to CBCT or to different phases of 4D CT, and management of the sheer volume of acquired image sets (both 4D CT/PET and CBCT). However, it is believed that much of these technical hurdles will be resolved with time, and that IGRT will become the standard of practice in the future through the effort of researchers around the world.

Acknowledgments—The authors thank the following individuals for their input and help: Drs. S. Hancock, C. King, K. Goodman, Q. Le, B. Loo, K. Horse, A. Boyer, D., P. Maxim, A. Hsu, T. Pawlicki, B. Ambruster, Spielman, S. Gambhir, A. Quon, D. Levy from Stanford University; S. Johnson, R. Morse, P. Munro, C. Zankowski, R. Wicha, M. Svatos, K. Kennedy, and R. Stark from Varian Medical Systems; G. Chen from Massachusetts General Hospital; and G. Mageras from Memorial Sloan Kettering Cancer Center. This work is supported in part by a research grants from the National Cancer Institute (5R01 CA98523-01), Department of Defense (PC040282), the American Cancer Society (RSG-01-022-01-CCE), and Varian Medical Systems. Permission for using copyrighted figures from *Medical Physics* and *International Journal of Radiation Oncology, Biology, Physics* is gratefully acknowledged.

REFERENCES

1. IMRT Collaborative Working Group: Intensity-modulated radiotherapy: Current status and issues of interest. *Int. J. Radiat. Oncol. Biol. Phys*. **51**:880–914; 2001.
2. AAPM IMRT Sub-Committee: Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. *Med. Phys*. **30**:2089–115; 2003.
3. Galvin, J.M.; Ezzell, G.; Eisbrauch, A.; *et al.* Implementing IMRT in clinical practice: A joint document of the American Society for Therapeutic Radiology and Oncology and the American Association of Physicists in Medicine. *Int. J. Radiat. Oncol. Biol. Phys.* **58:**1616–34; 2004.
4. Hamilton, R.J.; Sweeney, P.J.; Pelizzari, C.A.; *et al.* Functional imaging in treatment planning of brain lesions. *Int. J. Radiat. Oncol. Biol. Phys.* **37:**181–8; 1997.
5. Sawada, A.; Yoda, K.; Kokubo, M.; *et al.* A technique for noninvasive respiratory gated radiation treatment system based on a real time 3D ultrasound image correlation: A phantom study. *Med. Phys.* **31:**245–50; 2004.
6. Yu, Y.; Anderson, L.L.; Li, Z.; *et al.* Permanent prostate seed implant brachytherapy: Report of the American Association of Physicists in Medicine Task Group No. 64. *Med. Phys.* **26:**2054–76; 1999.
7. Pirzkall, A.; McKnight, T.R.; Graves, E.E.; *et al.* MR-spectroscopy guided target delineation for high-grade gliomas. *Int. J. Radiat. Oncol. Biol. Phys.* **50:**915–28; 2001.
8. Kurhanewicz, J.; Hricak, H.; Sokolov, D.L.; *et al.* Prostate cancer: Localization with three-dimensional proton MR spectroscopic imaging—clinicopathologic study. *Radiology* **206:**785–90; 1998.
9. DiBiase, S.J.; Hosseinzadeh, K.; Gullapalli, R.P.; *et al.* Magnetic resonance spectroscopic imaging-guided brachytherapy for localized prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **52:**429–38; 2002.
10. Kim, D.H.; Margolis, D.; Xing, L.; *et al.* In vivo prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 T. *Magn. Reson. Med.* **53:**1177–82; 2005.
11. Zaider, M.; Zelefsky, M.J.; Lee, E.K.; *et al.* Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging. *Int. J. Radiat. Oncol. Biol. Phys.* **47:**1085–96; 2000.
12. Hunjan, S.; Gibbs, I.; Spielman, D.M.; *et al.* Validating magnetic resonance spectroscopic imaging for radiation therapy guidance. *Int. J. Radiat. Oncol. Biol. Phys.* **57:**1159–73; 2003.
13. Townsend, D.W.; Carney, J.P.; Yap, J.T.; *et al.* PET/CT doday and tomorrow. *J. Nucl. Med.* **45:**4S–14S; 2004.
14. Czernin, J.; Schelbert, H. PET/CT imaging, facts, options, hopes, and questions. *J. Nucl. Med.* **45:**1S–3S; 2004.
15. Hicks, R.J.; Kalff, V.; MacManus, M.P.; *et al.* (18)F-FDG PET provides high-impact and powerful prognostic stratification in staging newly diagnosed non-small cell lung cancer. *J. Nucl. Med.* **42:**1596–604; 2001.
16. Mac Manus, M.P.; Hicks, R.J.; Ball, D.L.; *et al.* F-18 fluorodeoxyglucose positron emission tomography staging in radical radiotherapy candidates with nonsmall cell lung carcinoma: Powerful correlation with survival and high impact on treatment. *Cancer* **92**:886–95; 2001.
17. Bradley, J.D.; Dehdashti, F.; Mintun, M.A.; *et al.* Positron emission tomography in limited-stage small-cell lung cancer: A prospective study. *J. Clin. Oncol.* **22:**3248–54; 2004.
18. Howard, A.; Mehta, M.P.; Ritter, M.A.; *et al.* The value of PET/CT in gross tumor volume delineation in lung and esophagus cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **60:**S536–537; 2004.
19. Hofer, C.; Laubenbacher, C.; Block, T.; *et al.* Fluorine-18-fluorodeoxyglucose PET is useless for the detection of local recurrence after radical prostatectomy. *Eur. Urol.* **36:**31–5; 1999.
20. Liu, I.J.; Zafar, M.B.; Lai, Y.H.; *et al.* Fluorodeoxyglucose PET studies in diagnosis and staging of clinically organ-confined prostate cancer. *Urology* **57:**108–11; 2001.
21. Gambhir, S.S.; Srinivasan, A.; Banerjee, P.K.; *et al.* PET in oncology: Will it replace the other modalities? *J. Nucl. Med.* **39:**729–34; 1998.
22. Gambhir, S.S. Molecular imaging of cancer with PET. *Nat. Rev. Cancer* **2:**683–93; 2002.
23. Schèoder, H.; Larson, S.M. Positron emission tomography for prostate, bladder, and renal cancer. *Semin. Nucl. Med.* **34:**274–92; 2004.
24. Buck, A.K.; Halter, G.; Schirrmeister, H.; *et al.* Imaging proliferation in lung tumors with PET: 18F-FLT versus 18F-FDG. *J. Nucl. Med.* **44:**1426–31; 2003.
25. Shields, A.F.; Grierson, J.R.; Dohmen, B.M.; *et al.* Imaging proliferation in vivo with [F-18]FLT and positron emission tomography. *Nat. Med.* **4:**1334–6; 1998.
26. Buck, A.K.; Schirrmeister, H.; Hetzel, M.; *et al.* 3-deoxy-3-[(18)F]fluorothymidine-PET for noninvasive assessment of proliferation in pulmonary nodules. *Cancer Res.* **62:**3331–4; 2002.
27. Smyczek-Gargya, B.; Fersis, N.; Dittmann, H.; *et al*. PET with [18F]fluorothymidine for imaging of primary breast cancer: A pilot study. *Eur. J. Nucl. Med. Mol. Imag*. **31**:720–724.
28. Shreve, P.D.; Gross, M.D. Imaging of the pancreas and related diseases with PET carbon-11-acetate. *J. Nucl. Med.* **38:**1305–10; 1997.

29. Seltzer, M.A.; Jahan, S.A.; Sparks, R.; *et al.* Radiation dose estimates in humans for (11)C-acetate whole-body PET. *J. Nucl. Med.* **45:**1233–6; 2004.
30. Jager, P.L.; Que, T.H.; Vaalburg, W.; *et al.* Carbon-11 choline or FDG-PET for staging of oesophageal cancer? *Eur. J. Nucl. Med.* **28:**1845–9; 2001.
31. Ishiwata, K.; Kasahara, C.; Hatano, K.; *et al.* Carbon-11 labeled ethionine and propionine as tumor detecting agents. *Ann. Nucl. Med.* **1:**115–22; 1997.
32. Iozzo, P.; Osman, S.; Glaser, M.; *et al.* In vivo imaging of insulin receptors by PET: Preclinical evaluation of iodine-125 and iodine-124 labelled human insulin. *Nucl. Med. Biol.* **29:**73–82; 2002.
33. Herlin, G.; Persson, B.; Bergstrèom, M.; *et al.* 11C-harmine as a potential PET tracer for ductal pancreas cancer: In vitro studies. *Eur Radiol.* **13:**729–3; 2003.
34. Brown, W.D.; Oakes, T.R.; DeJesus, O.T.; *et al.* Fluorine-18-fluoro-L-DOPA dosimetry with carbidopa pretreatment. *J. Nucl. Med.* **39:**1884–91; 1998.
35. Blankenberg, F.G.; Strauss, H.W. Nuclear medicine applications in molecular imaging. *J. Magn. Res. Imaging* **16:**352–61; 2002.
36. Oyama, N.; Akino, H.; Kanamaru, H.; *et al.* 11C-acetate PET imaging of prostate cancer. *J. Nucl. Med.* **43:**181–6; 2002.
37. Kotzerke, J.; Volkmer, B.G.; Glatting, G.; *et al.* Intraindividual comparison of [11C]acetate and [11C]choline PET for detection of metastases of prostate cancer. *Nuklearmedizin* **42:**25–30; 2003.
38. Mathews, D.; Oz, O.K. PET in prostate and renal cell carcinoma. *Curr. Opin. Urol.* **12:**381–5; 2002.
39. DeGrado, T.R.; Baldwin, S.W.; Wang, S.; *et al.* Synthesis and evaluation of (18)F-labeled choline analogs as oncologic PET tracers. *J. Nucl. Med.* **42:**1805–14; 2001.
40. Hara, T.; Kosaka, N.; Kishi, H. Development of (18)F-fluoroethylcholine for cancer imaging with PET: Synthesis, biochemistry, and prostate cancer imaging. *J. Nucl. Med.* **43:**187–99; 2002.
41. Hara, T.; Kosaka, N.; Kishi, H. PET imaging of prostate cancer using carbon-11-choline. *J. Nucl. Med.* **39:**990–5; 1998.
42. Sutinen, E.; Nurmi M.; Roivainen A.; *et al.* Kinetics of [(11)C]choline uptake in prostate cancer: A PET study. *Eur. J. Nucl. Med. Mol. Imaging* **31:**317–24; 2004.
43. Chen, X.; Park, R.; Hou, Y.; *et al.* MicroPET and autoradiographic imaging of GRP receptor expression with 64Cu-DOTA-[Lys3]bombesin in human prostate adenocarcinoma xenografts. *J. Nucl. Med.* **45:**1390–7; 2004.
44. Rasey, J.S.; Koh, W.J.; Evans, M.L.; *et al.* Quantifying regional hypoxia in human tumors with positron emission tomography of [18F]fluoromisonidazole: A pretherapy study of 37 patients. *Int. J. Radiat. Oncol. Biol. Phys.* **36:**417–28; 1996.
45. O'Donoghue, J.A.; Zanzonico, P.; Pugachev, A.; *et al.* Assessment of regional tumor hypoxia using 18F-fluoromisonidazole and 64Cu(II)-diacetyl-bis(N4-methylthiosemicarbazone) positron emission tomography: Comparative study featuring microPET imaging, Po2 probe measurement, autoradiography, and fluorescent microscopy in the R3327-AT and FaDu rat tumor models. *Int. J. Radiat. Oncol. Biol. Phys.* **61:**1493–502; 2005.
46. Piert, M.; Machulla, H.J.; Picchio, M.; *et al.* Hypoxia-specific tumor imaging with 18F-fluoroazomycin arabinoside. *J. Nucl. Med.* **46:**106–13; 2005.
47. Chao, K.S.; Mutic, S.; Gerber, R.L.; *et al.* A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guided intensity-modulated radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **49:**1171–82; 2001.
48. de Jong I.J.; Pruim, J.; Elsinga, P.H.*; et al.* 11C-choline positron emission tomography for the evaluation after treatment of localized prostate cancer. *Eur. Urol.* **44**:327–8; discussion 38–9; 2003.
49. Ling, C.C.; Humm, J.; Larson, S.; *et al.* Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality. *Int. J. Radiat. Oncol. Biol. Phys.* **47:**551–60; 2000.
50. Brahme, A. Individualizing cancer treatment: Biological optimization models in treatment planning and delivery. *Int. J. Radiat. Oncol. Biol. Phys.* **49:**327–37; 2001.
51. Xing, L.; Cotrutz, C.; Hunjan, S.; *et al.* Inverse planning for functional image-guided IMRT. *Phys. Med. Biol.* **47:**3567–78; 2002.
52. Alber, M.; Paulsen, F.; Eschman, S.M.; *et al.* On biologically conformal boost dose optimization. *Phys. Med. Biol.* **48:**N31–5; 2003.
53. Yang, Y.; Xing, L. Towards biologically conformal radiation therapy (BCRT): Selective IMRT dose escalation under the guidance of spatial biology distribution. *Med. Phys.* **32:**1473–84; 2005.
54. Seppenwoolde, Y.; Shirato, H.; Kitamura, K.; *et al.* Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **53:**822; 2002.
55. Shirato, H.; Seppenwoolde, Y.; Kitamura, K.; *et al.* Intrafractional tumor motion: Lung and liver. *Semin. Radiat. Oncol.* **14:**10–8; 2004.
56. Shimizu, S.; Shirato, H.; Ogura, S.; *et al.* Detection of lung tumor movement in real-time tumor-tracking radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **51:**304–10; 2001.
57. Xu, Q.; Hamilton, R. Novel respiratory gating method based on automated analysis of ultrasonic diaphragm motion. *Med. Phys.* **32:**2124; 2005.
58. Huang, M.H.; Lin, Y.S.; Lee, C.L.; *et al.* Use of ultrasound to increase effectiveness of isokinetic exercise for knee osteoarthritis. *Arch. Phys. Med. Rehabil.* **86:**1545–51; 2005.
59. Seiler, P.G.; Blattmann, H.; Kirsch S.; *et al.* A novel tracking technique for the continuous precise measurement of tumour positions in conformal radiotherapy. *Phys. Med. Biol.* **45:**N103–10; 2000.
60. Balter, J.M.; Wright, J.N.; Newell, L.J.; *et al.* Accuracy of a wireless localization system for radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **61:**933–7; 2005.
61. Nehmeh, S.A.; Erdi, Y.E.; Pan, T.; *et al.* Four-dimensional (4D) PET/CT imaging of the thorax. *Med. Phys.* **31:**3179; 2004.
62. Thorndyke, B.; Schreibmann, E.; Maxim, P.; et al. Enhancing 4D PET through retrospective stacking. Presented at 2005 Annual Meeting of American Association of Physicists in Medicine, Seattle, WA, 2005.
63. Thorndyke, B.; Schreibmann, E.; Xing, L. Reducing motion artifacts in PET through retrospective stacking. Med. Phys. **33:** 2006; In press.
64. Vedam, S.S.; Keall, P.J.; Kini, V.R.; *et al.* Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Phys. Med. Biol.* **48:**45; 2003.
65. Keall, P.J.; Starkschall, G.; Shukla, H.; *et al.* Acquiring 4D thoracic CT scans using a multislice helical method. *Phys. Med. Biol.* **49:**2053–67; 2004.
66. Low. D.A.; Nystrom, M.; Kalinin, E.; *et al*. A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. *Med. Phys.*
67. Rietzel, E.; Pan, T.; Chen, G.T.Y. Four-dimensional computed tomography: Image formation and clinical protocol. *Med. Phys.* **32:**874; 2005.
68. Li, T.; schreibmann, E.; Thorndyke, B.; *et al*. Radiation dose reduction in 4D CT. Med. Phys. In press.
69. Kubo, H.D.; Len, P.M.; Minohara, S.; *et al.* Breathing-synchronized radiotherapy program at the University of California Davis Cancer Center. *Med. Phys.* **27:**346; 2000.
70. Rueckert, D.; Sonoda, L.I.; Hayes, C.; *et al.* Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging* **18:**712–21; 1999.
71. Pan, T.; Lee, T.-Y.; Rietzel, E.; *et al.* 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med. Phys.* **31:**333; 2004.
72. Kini, V.R.; Vedam, S.S.; Keall, P.J.; *et al.* Patient training in respiratory-gated radiotherapy. *Med. Dosim.* **28:**7; 2003.
73. Wong, J.W.. Sharpe, M.B.; Jaffray, D.A.*; et al.* The use of active breathing control (ABC) to reduce margin for breathing motion. *Int. J. Radiat. Oncol. Biol. Phys.* **44**:911; 1999.
74. Frazier, R.C.; Vicini, F.A.; Sharpe, M.B.; *et al.* Impact of breathing motion on whole breast radiotherapy: A dosimetric analysis using active breathing control. *Int. J. Radiat. Oncol. Biol. Phys.* **58:**1041–7; 2004.
75. Dawson, L.A.; Brock, K.K.; Kazanjian, S.; *et al.* The reproducibility of organ position using active breathing control (ABC) during liver radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **51:** 1410–21; 2001.

76. Starkschall, G.; Himanshu, S.; Keall, P.J.; et al. Displacement-based binning of 4-D CT image data sets. XIVth International Conference on the Use of Computers in Radiation Therapy, 2004, pp 53.
77. Thorndyke, B.; Schreibmann, E.; Li, T.; *et al.* A comparison of amplitude- and phase-based 4D CT. *Med. Phys.* **32:**1919; 2005.
78. Lu, W.; Parikh, P.; Bradley, J.; *et al.* A comparison between amplitude sorting and phase sorting using external respiratory measurements for 4D CT. *Med. Phys.* **32:**2094; 2005.
79. Schreibmann, E.; Chen, G.T.Y.; Xing, L. Image interpolation in 4D CT using a BSpline deformable registration model. *Int. J. Radiat. Oncol. Biol. Phys*. 2006; In press.
80. Schweikard, A.; Shiomi, H.; Fisseler, J.; et al. Fiducial-less respiration tracking in radiosurgery. LNCS **3217**:992–99; 2004.
81. Vedam, S.S.; Kini, V.R.; Keall, P.J.; *et al.* Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. *Med. Phys.* **30:**505; 2003.
82. Ford, E.C.; Mageras, G.S.; Yorke, E.; *et al.* Evaluation of respiratory movement during gated radiotherapy using film and electronic portal imaging. *Int. J. Radiat. Oncol. Biol. Phys.* **52:**522; 2002.
83. Yorke, E.; Rosenzweig, K.E.; Wagman, R.; *et al.* Interfractional anatomic variation in patients treated with respiration-gated radiotherapy. *J. Appl. Clin. Med. Phys.* **6:**19; 2005.
84. Klein, G.J.; Reutter, B.W.; Ho, M.H.; *et al.* Real-time system for respiratory-cardiac gating in positron tomography. *IEEE Trans. Nucl. Sci.* **45:**2139; 1998.
85. Nehmeh, S.A.; Erdi, Y.E.; Ling, C.C.; *et al.* Effect of respiratory gating on quantifying PET images of lung cancer. *J. Nucl. Med.* **43:**876–81; 2002.
86. Nehmeh, S.A.; Erdi, Y.E.; Rosenzweig, K.E.; *et al.* Reduction of respiratory motion artifacts in PET imaging of lung cancer by respiratory correlated dynamic PET: Methodology and comparison with respiratory gated PET. *J. Nucl. Med.* **44:**1644; 2003.
87. Cherry, S.R.; Dahlborn, M. Physics, Instrumentation and Scanners, PET: Molecular Imaging and its Biological Applications. New York: Springer-Verlag; 2004.
88. Sarikaya, I.; Yeung, H.W.; Erdi, Y.; *et al.* Respiratory artefact causing malpositioning of liver dome lesion in right lower lung. *Clin. Nucl. Med.* **28:**943; 2003.
89. Erdi, Y.E.; Nehmeh, S.A.; Pan, T.; *et al.* The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. *J. Nucl. Med.* **45:**1287; 2004.
90. Pevsner, A.; Nehmeh, S.A.; Humm, J.L.; *et al.* Effect of motion on tracer activity determination in CT attenuation corrected PET images: A lung phantom study. *Med. Phys.* **32:**2358; 2005.
91. Nehmeh, S.A.; Erdi, Y.E.; Pan, T.; *et al.* Quantitation of respiratory motion during 4D-PET/CT acquisition. *Med. Phys.* **31:** 1333–8; 2004.
92. Mattes, D.; Haynor, R.D.; Vesselle, H.; *et al.* PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging* **22:**120–28; 2003.
93. Schreibmann, E.; Xing, L. Narrow band deformable registration of prostate MRI/MRSI and CT studies. *Int. J. Radiat. Oncol. Biol. Phys.* **62:**595–605; 2005.
94. Li, T.; Thorndyke, B.; Schreibmann, E.; et al. Model-based image reconstruction for four-dimensional PET. Med. Phys. Submitted.
95. Trofimov, A.; Rietzel, E.; Lu, H.M.; *et al.* Temporo-spatial IMRT optimization: Concepts, implementation and initial results. *Phys. Med. Biol.* **50:**2779–98; 2005.
96. Armbruster, B.; Yang, Y.; Schreibmann, E.; et al. Inverse planning for time-resolved aperture modulated radiation therapy. Annual ASTRO Meeting, Denver, CO, 2005.
97. Bortfeld, T.; Jiang, S.B.; Rietzel, E. Effects of motion on the total dose distribution. *Semin. Radiat. Oncol.* **14:**41–51; 2004.
98. Jiang, S.B.; Pope, C.; Al Jarrah, K.M.*; et al.* An experimental investigation on intra-fractional organ motion effects in lung IMRT treatments. *Phys. Med. Biol.* **48**:1773–84; 2003.
99. Rietzel, E.; Chen G.T.; Choi, N.C.; *et al.* 4D image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion. *Int. J. Radiat. Oncol. Biol. Phys.* **61:**1535; 2005.
100. Keall, P.J.; Joshi, S.; Vedam, S.S.; *et al.* Four-dimensional radiotherapy planning for DMLC-based respiratory motion tracking. *Med. Phys.* **32:**942–51; 2005.
101. Webb, S. The effect on IMRT conformality of elastic tissue movement and a practical suggestion for movement compensation via the modified dynamic multileaf collimator technique. *Phys. Med. Biol.* **50:**1163–90; 2005.
102. Yang, Y.; Xing, L. Clinical knowledge-based inverse treatment planning. *Phys. Med. Biol.* **49:**5101–17; 2004.
103. Yang, Y.; Xing, L. Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme. *Med. Phys.* **31:**2839–44; 2004.
104. Wu, Q.; Mohan, R.; Niemierko, A.; *et al.* Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int. J. Radiat. Oncol. Biol. Phys.* **52:**224–35; 2002.
105. Olivera, G.H.; Shepard, D.M.; Reckwerdt, P.J.; *et al.* Maximum likelihood as a common computational framework in tomotherapy. *Phys. Med. Biol.* **43:**3277–94; 1998.
106. Lee, E.K.; Gallagher, R.J.; Silvern, D.; *et al.* Treatment planning for brachytherapy: An integer programming model, two computational approaches and experiments with permanent prostate implant planning. *Phys. Med. Biol.* **44:**145–65; 1999.
107. Shepard, D.M.; Earl, M.A.; Li, X.A.; *et al.* Direct aperture optimization: A turnkey solution for step-and-shoot IMRT. *Med. Phys.* **29:**1007–18; 2002.
108. Bednarz, G.; Michalski, D.; Houser, C.; *et al.* The use of mixed-integer programming for inverse treatment planning with predefined field segments. *Phys. Med. Biol.* **47:**2235–45; 2002.
109. Cotrutz, C.; Xing, L. Segment-based dose optimization using a genetic algorithm. *Phys. Med. Biol.* **48:**2987–98; 2003.
110. Murphy, M.J.; Adler, Jr.; J.R Bodduluri, M.*; et al.* Image-guided radiosurgery for the spine and pancreas. *Comput. Aided Surg.* **5**:278–88; 2000.
111. Murphy, M.J. Tracking moving organs in real time. *Semin. Radiat. Oncol.* **14:**91–100; 2004.
112. Sharp, G.C.; Jiang, S.B.; Shimizu, S.; *et al.* Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys. Med. Biol.* **49:**425–40; 2004.
113. Berbeco, R.I.; Neicu, T.; Rietzel, E.; *et al.* A technique for respiratory-gated radiotherapy treatment verification with an EPID in cine mode. *Phys. Med. Biol.* **50:**3669–79; 2005.
114. Herman, M.G.; Balter, J.M.; Jaffray, D.A.*;et al.* Clinical use of electronic portal imaging: Report of AAPM Radiation Therapy Committee Task Group 58. *Med. Phys.* **28:**712–37; 2001.
115. Yang, Y.; Xing, L. Quantitative measurement of MLC leaf displacements using an electronic portal image device. *Phys. Med. Biol.* **49:**1521–33; 2004.
116. Yan, D.; Wong, J.; Vicini, F.; *et al.* Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors. *Int. J. Radiat. Oncol. Biol. Phys.* **38:**197–206; 1997.
117. Court, L.E.; Dong, L.; Taylor, N.; *et al.* Evaluation of a contour-alignment technique for CT-guided prostate radiotherapy: An intra- and interobserver study. *Int. J. Radiat. Oncol. Biol. Phys.* **59:**412–8; 2004.
118. Orton, N.P.; Tome, W.A. The impact of daily shifts on prostate IMRT dose distributions. *Med. Phys.* **31:**2845–8; 2004.
119. Fuss, M.; Salter, B.J.; Cavanaugh, S.X.; *et al.* Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies. *Int. J. Radiat. Oncol. Biol. Phys.* **59:**1245–56; 2004.
120. Saw, C.B.; Ayyangar, K.M.; Zhen, W.; *et al.* Clinical implementation of intensity-modulated radiation therapy. *Med. Dosim.* **27:** 161–9; 2002.
121. Schweikard, A.; Glosser, G.; Bodduluri, M.; *et al.* Robotic motion compensation for respiratory movement during radiosurgery. *Comput. Aided Surg.* **5:**263–77; 2000.
122. Mackie, T.R.; Balog, J.; Ruchala, K.; *et al.* Tomotherapy. *Semin. Radiat. Oncol.* **9:**108–17; 1999.
123. Meeks, S.; Harmon, J.J.; Langen, K.; *et al.* Performance characterization of megavoltage computed tomography imaging on a helical tomotherapy unit. *Med. Phys.* **32:**2673–81; 2005.
124. Jaffray, D.A.; Siewerdsen, J.H.; Wong, J.W.; *et al.* Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **53:**1337–49; 2002.
125. Pouliot, J.; Bani-Hashemi, A.; Chen, J.; *et al.* Low-dose megavoltage cone-beam CT for radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **61:**552–60; 2005.

126. Oldham, M.; Letourneau, D.; Watt, L.; et al. Cone-beam-CT guided radiation therapy: A model for on-line application. Radiother. Oncol. **75:**271–8; 2005.
127. Li, T.; Yang, Y.; Schreibmann, E.; et al. A new cone-beam CT reposition technique through deformable registration. Annual Meeting of American Association of Physicists in Medicine, Seattle, WA, 2005.
128. Mohan, R.; Zhang, X.; Wang, H.; *et al.* Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes. *Int. J. Radiat. Oncol. Biol. Phys.* **61:**1258–66; 2005.
129. Yang, Y.; Li, T.; Schreibmann, E.; et al. Is cone beam CT suitable for dose verification? Annual Meeting of American Association of Physicists in Medicine, Seattle, WA, 2005.
130. Yang, Y.; Schreibmann, E.; Xing, L.; et al. Dosimetric evaluation of kV cone-beam CT (CBCT)-based dose calculation. *Phys. Med. Biol.* 2006; Submitted.
131. Wang, G.; Vannier, M.W. Preliminary study on helical CT algorithms for patient motion estimation and compensation. *IEEE Trans. Med. Imaging* **14:**205–11; 1995.
132. Willis, N.P.; Bresler, Y. Optimal scan for time-varying tomography. I: Theoretical analysis and fundamental limitations*IEEE Trans. Med. Imaging* **14:**642–53; 1995.
133. Atalar, E.; Onural, L. A respiratory motion artifact reduction method in magnetic resonance imaging of the chest. *IEEE Trans. Med. Imaging* **10:**11–24; 1991.
134. Crawford, C.; King, K.F.; Ritchie, C.J.; *et al.* Respiratory compensation in projection imaging using a magnification and displacement model. *IEEE Trans. Med. Imaging* **15:**327–32; 1996.
135. Li, T.; Schreibmann, E.; Yang, Y.; Xing, L. Respiratory motion correction with deformation field for improved on-board imaging with cone-beam CT. *Phys. Med. Biol.* **51:**253–67; 2006.
136. Sonke, J.J.; Zijp, L.; Remeijer, P.; *et al.* Respiratory correlated cone beam CT. *Med. Phys.* **32:**1176–86; 2005.
137. Li, T.; Xing, L.; Munro, P.; et al. 4D cone beam CT using an on-board imager. *Med. Phys.* 2006; Submitted.
138. Zeng, R.; Fessler, J.A.; Balter, J.M. Respiratory motion estimation from slowly rotating x-ray projections: Theory and simulation. *Med.Phys.* **32:**984–91; 2005.
139. Court, L.E.; Dong, L. Automatic registration of the prostate for computed-tomography-guided radiotherapy. *Med. Phys.* **30:** 2750–7; 2003.
140. Bharath, A.; Hirose, M.; Hata, N.; *et al.* Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Med. Phys.* **28:**2551–60; 2001.
141. Brock, K.K.; McShan, D.L.; Balter, J.M. A comparison of computer-controlled versus manual on-line patient setup adjustment. *J. Appl. Clin. Med. Phys.* **3:**241–7; 2002.
142. Bookstein, F.L. Thin plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **11:**567–85; 1989.
143. Lian, J.; Xing, L.; Hunjan, S.; *et al.* Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study. *Med. Phys.* **31:**3087–94; 2004.
144. Guerrero, T.; Zhang, G.; Huang, T.C.; *et al.* Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method. *Phys. Med. Biol.* **49:**4147–61; 2004.
145. Mageras, G.S.; Pevsner, A.; Yorke, E.D.; *et al.* Measurement of lung tumor motion using respiration-correlated CT. *Int. J. Radiat. Oncol. Biol. Phys.* **60:**933–41; 2004.
146. Schwartz, D.L.; Ford, E.C.; Rajendran, J.; *et al.* FDG-PET/CT-guided intensity modulated head and neck radiotherapy: A pilot investigation. *Head Neck* **27:**478–87; 2005.
147. Fei, B.; Wheaton, A.; Lee, Z.; *et al.* Automatic MR volume registration and its evaluation for the pelvis and prostate. *Phys. Med. Biol.* **2002:**823–38; 2002.
148. Coselmon, M.M.; Balter, J.M.; McShan, D.L.; *et al.* Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med. Phys.* **31:**2942–8; 2004.
149. Schreibmann, E.; Xing, L. Image registration with auto-mapped control volumes. *Med. Phys.* Submitted.
150. Koong, A.C.; Le, Q.T.; Ho, A.; *et al.* Phase I study of stereotactic radiosurgery in patients with locally advanced pancreatic cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **58**:1017–21; 2004.
151. Loo, B.; Thorndyke, B.; Maxim, P.; et al. Determining margin for target deformation and rotation in respiratory motion-tracked stereotactic radiosurgery of pancreatic cancer, Annual ASTRO Meeting, Denver, CO, 2005.
152. Koong, A.C.; Le, Q.T.; Ho, A.; et al. Phase II study of stereotactic radiosurgery in patients with locally advanced pancreatic cancer. *Int. J. Radiat. Oncol. Biol. Phys.* In press.

Dosimetric Evaluation of kV Cone-Beam CT (CBCT)-Based Dose Calculation[†]

Yong Yang, Eduard Schreibmann, Tianfang Li, Chuang Wang,
and Lei Xing[a)]

Department of Radiation Oncology, Stanford University School of Medicine,

Stanford, CA 94305-5847

Short title: Mapping electron density distribution from planning CT to CBCT

[a)] Author to whom correspondence should be addressed:
Department of Radiation Oncology
Stanford University School of Medicine,
Clinical Cancer Center
875 Blake Wilbur Drive, Rm G-204
Stanford, CA 94305-5847
Telephone: (650) 498-7896
Fax: (650) 498-4015
Email: lei@reyes.stanford.edu

Physics in Medicine and Biology, in press.

[†] Part of this work was presented in 2005 Annual Meeting of American Association of Physicists in Medicine, Seattle, WA.

Abstract

Kilovolt (kV) CBCT based on flat-panel technology integrated with linear accelerator has recently become available from linac vendors for therapy guidance. Currently, the system is primarily utilized to guide the patient alignment. As an advanced tool of obtaining a patient's 3D representation, CBCT also affords an effective means for us to examine the actual dose distribution to be delivered or already delivered to the patient on a daily basis. Before this can be implemented clinically, the accuracy of kV CBCT-based dose calculation must be evaluated and some logistic issues related to the application need to be addressed. Indeed, image quality of current CBCT is not as good as conventional diagnostic CT due to the scatter and organ motion artifacts, which may lead to dosimetric inaccuracy. This work is aimed to investigate the feasibility and accuracy of CBCT-based dose calculation and to propose a deformable electron density mapping (DEDM) method that is potentially useful to facilitate CBCT-based dose calculation. In the proposed DEDM technique, the CBCT and planning CT are first registered by using a deformable image registration model. The electron density distribution is then mapped from the planning CT to the CBCT. The CBCT with the mapped electron density information is useful for more accurate CBCT-based dose calculation. For disease sites where intra-fractional organ motion is not an issue, this study indicates that CBCT can be employed directly for dose calculation and the results agree with the planned dose distributions to within 1~2%. The use of DEDM further reduces the dosimetric inconsistency and provides a sanity check of the CBCT-based dose calculation. While the true solution for using CBCT to calculate dose lies in the improvement of image quality, the DEDM approach seems to afford a useful interim technique for better CBCT-based dose calculation.

Key word: CBCT, IGRT, Dose verification, Deformable registration, IMRT

I. Introduction

Modern radiation therapy techniques, such as 3D conformal radiotherapy (3DCRT) and intensity-modulated radiation therapy (IMRT), provide unprecedented means for producing exquisitely shaped radiation doses that closely conform to the tumor dimensions while sparing sensitive structures. As a result of greatly enhanced dose conformality, more accurate beam targeting becomes an urgent issue in radiation therapy. In current practice, large uncertainties exist in tumor target localization due to intra- and inter-organ motions during the course of radiation treatment. As thus, large safety margins around the tumor targets and sensitive structures are introduced to cope with the otherwise insoluble patient localization problem. The use of non-optimal margins compromises the patient care and adversely affects the treatment outcome (1-7). The need to improve targeting in high precision radiation therapy has recently spurred a flood of research activities in image-guided radiation therapy (IGRT) (7-11).

CBCT based upon flat-panel technology integrated with a medical linear accelerator has recently become available from Linac vendors for therapy guidance. The volumetric images may be used to verify and correct the planning patient setup in the linac coordinates by comparing with the patient position defined in treatment plan. Both kV and MV beams(12-14) have been utilized for this application. The former typically consists of a kV-source and flat-panel combination mounted on the drum of a medical accelerator(8-11, 15), with the kV imaging axis orthogonal to that of MV therapy beam. In addition to guide the patient setup process, CBCT data acquired prior to the treatment can, in principle, be used to recalculate or verify the treatment plan based on the patient anatomy of the treatment day. The recalculation starts with the intended fluence maps from the patient's treatment plan, whereas the verification is done by using the fluence maps measured at the exiting sides of the incident beams. If CBCT-based dose calculation is accurate enough (say, with an accuracy within 1~2%), this will provide a valuable option for us to predict/assess the patient dose routinely. In reality, because of the presence of organ movement/deformation, it is conceivable that the dose distributions delivered to the patient are usually different from fraction to fraction. It is paramount to be able to monitor the actual patient dose for each fraction as well as the accumulated

doses to the target and sensitive structures while the fractional treatment proceeding. This will not only give physician more confidence about the treatment but may, in future, afford us an effective means to adaptively modify the patient's treatment plan during the course of a radiation therapy based on the dose that has already been delivered.

The accuracy of MV fan-beam and cone-beam CT has recently been assessed by Langen et al (16) and Poulliot et al (14). The potential of its counterpart, the kV CBCT, for dosimetric calculation has, on the other hand, not been examined systematically. Different from conventional fan beam CT, CBCT covers a much larger field of view (FOV) in the longitudinal direction, and scatter poses a much severe problem in the resultant image. In addition, the gantry rotation speed is limited to ~1 min. by IEC regulation, which makes the CBCT less prone of motion artifacts. The deteriorated image quality raises serious concern about the dosimetric reliability of CBCT-based dose calculation. The purpose of this work is two-fold: to evaluate the dosimetric accuracy of kV CBCT-based dose calculation and to explore a strategy, coined as the deformable electron density mapping technique, for improving the CBCT-based dose calculation.

II. Method and Materials

A. Data acquisition

The onboard imager (OBI) integrated in a Trilogy™ medical linear accelerator (Varian Medical Systems, Palo Alto, CA) is used in this work to acquire CBCT images. The kV OBI system is capable of obtaining low-dose, high-resolution radiography, fluoroscopy and CBCT. The system is mounted on the treatment machine via robotically controlled arms, which operate along three axes of motion. A 150 kV X-ray tube with maximum 32 ms pulse length for continuous irradiation and maximum 320 ms pulse length for single pulse is designed for generating high-resolution images from a moving gantry. The spot of the tube is located at 90º to the MV source and 100cm from the radiation axis of the accelerator. A 39.7cm X 29.8 cm amorphous silicon flat-panel X-ray image detector (Varian PortalVision™ aS1000) mounted opposite the kV tube is used to acquire digital images with a pixel matrix of 2048 X 1536. Using the OBI system, the CBCT data can be

acquired in two modes: full fan mode and half fan mode. In the full fan mode, the beam central axis passes through the detector center and a full projection of the scanned patient is acquired for each acquirement angle. Total 675 projections are taken during the whole 364º gantry rotation with a maximum field of view (FOV) about 25 cm in diameter and 17cm in length. The data acquisition time is about 60 second and the reconstruction time for 340 slices of 512X512 CBCT images with a voxel size of 0.5mm^3 is also about 1 minute on a PC. The half fan mode is designed to obtain larger FOV. In this mode, the detector is shifted laterally to take only half of the projection of the scanned patient for each acquirement angle. Total about 965 projections are taken during the 364º gantry rotation and a FOV of 50 cm in the axial plane and 15cm in the longitudinal direction can be achieved. The data acquisition and reconstruction time for 512X512 CBCT images with a voxel size of 0.95mm^3 using this mode is about double compared with the full fan mode. The averaged dose for a head and neck CBCT scan is about 2 cGy, and 3cGy for an abdominal scan.

B. Calibration of relative electron density

To use CT or CBCT for radiation dose calculation, it is required to relate the Hounsfield Unit (HU) of the scanner with the actual electron density. A CT-phantom, Catphan-600 module CTP404 (Phantom Laboratory, NY), was used for the calibration of planning CT (GE Discovery-ST PET/CT scanner, Milwaukee, WI) and CBCT. The gantry rotation speed of the 16-slice Discovery-ST scanner is 0.5sec/rotation. The CTP404 has a diameter of 150 mm and contains 17 different sizes of inserts with seven different tissue substitute materials, air, PMP, LDPE, Polystryrene, Acrylic, Delrin and Teflon, respectively. Their relative electron densities ranged from 0 to 1.866. A cross section of the phantom is shown in figure 2. The calibration of a CT scanner involves acquiring CT images, obtaining average HUs for each inserting materials, and plot the HU as a function of the relative electron density. For CBCT calibration, the only difference from the conventional CT is that it is necessary to calibrate separately for full and half fan modes because the beam geometry and characteristics of the two types of scanning modes are different.

In order to test the stability of the CBCT calibration curve with time, the phantom was repeatedly scanned every week for two months. The obtained HU vs relative electron density curves were compared to assess the HU fluctuations with time.

C. Phantom study

CT and CBCT images of the Catphan-600 phantom were acquired using the procedure outlined in Sec. II.A. The phantom was placed on a platform that can be set to one-dimensional cyclic motion with a number of speeds. The axis of the cylindrically shaped phantom, along which the phantom moves cyclically, was angled from the central axis of the CBCT gantry rotation by about 15º in order to study the motion influence on CT/CBCT. The movement of the phantom produces motion artifacts in the images and allows us to evaluate the performance of CBCT-based dose calculation in the presence of organ motion. The full fan mode was used to scan the phantom. CT and CBCT images of the phantom were acquired with and without motion. In the former case, the peak-to-peak amplitude of the motion was 0.5 cm in the left-right direction and the period was 4s. In addition, different sizes of homogeneity cylindrical phantoms, with a diameter of 10.8, 16, and 26.6 cm, respectively, were scanned to evaluate the scatter influence on image quality.

To quantify the difference in the image quality of the CT and CBCT images, we first analyzed the HU distribution for all the acquired images. The influence of phantom motion and scatter radiation on the HU distribution was investigated. The CT and CBCT images were imported to a Varian Eclipse treatment planning system for dosimetric comparison study. For planning and evaluation purpose, a hypothetical spherical target with a diameter of 5cm was created at the center of the phantom and a single 5 X 5cm^2 6MV photon beam was used to irradiate the target. A simple beam configuration was used here because, in this way, the results are more intuitively interpretable. Seven plans, corresponding to the different sets of CT images, were generated using the same target and beam configuration. The pencil beam convolution dose calculation algorithm implemented in Varian Eclipse treatment planning system was adopted for dose calculation. The resultant isodose curves, dose profiles and DVHs were compared.

D. Patient study

Three prostate cancer patients were selected for the evaluation study of CBCT-based dose calculation and to demonstrate the feasibility of the proposed deformable electron density mapping (DEDM) technique (see next sub-section) for improved dose calculation accuracy. For all three cases, the targets included the PTV, consisting of the prostate gland with a margin of 6mm, and the seminal vesicles. The critical structures were rectum, bladder and femoral heads. IMRT plans using five 15MV photon beams with gantry angles of 35^{o}, 110^{o}, 180^{o}, 250^{o}, and 325^{o} (in IEC convention) were adopted for the three cases. All the plans were normalized to deliver a prescription dose of 78Gy to 99% the prostate PTV and no less than 50Gy to the 98% of seminal vesicles in 39 fractions.

After the patients were setup using the current clinical procedure, CBCT images of the patients were acquired using the half fan mode. The CBCT images were transferred to an Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA). For each case, the IMRT planning parameters generated for the patient's treatment, including beam configuration, MU settings, and MLC files, were employed to recalculate the dose based on the CBCT data. The CT and CBCT-based treatment plans were then compared.

E. Deformable electron density mapping

The dosimetric inaccuracy of CBCT-based dose calculation primarily arises from the inability of the CBCT technique to provide accurate HU or relative electron density distribution. The genuine solution to the problem lies in the improvement of the CBCT acquisition technology so that high quality images can be acquired. While this endeavor is still on-going, here we propose an interim solution for dealing with the problem. Under the assumptions that the HU or relative electron density distribution is known from planning CT and an acceptable geometric registration between CT and CBCT is achievable by a deformable registration, we propose to map the electron density in the planning CT onto the daily setup CBCT and then carry out the dose calculation. The CBCT with mapped electron density, referred to as modified CBCT, possesses the anatomical information of CBCT and yet the electron density information of the planning CT. Dose calculation based on the modified CBCT allows us to compute more

accurately the delivered dose with the patient in his/her setup position. The mapping process is described as follows.

A free form spline (BSpline) deformable model (17-21) was employed to register the planning CT and CBCT and map the deformed electron density from planning CT to CBCT. The method was used for several IGRT related projects in our group and others and its simplicity and accuracy have been demonstrated (18, 21-24). Briefly, in the BSpline model, a lattice of user-defined nodes is overlaid on the image. Each node contains a deformation vector, whose components are determined by optimizing a metric function that characterizes the goodness of the registration. The metric is a function of the BSpline nodal parameters. In this work, a voxel-based normal cross correlation (NCC) metric was used. A suitable set of node deformations was determined using the gradient-based algorithm L-BFGS(18-20), which is known for its superior performance in large-scale optimization problems. The optimizer iteratively varies the nodal displacements to optimize the metric. The deformation at any point of the image is calculated by spline interpolation of closest nodes values. Unlike other spline models, the BSplines are locally controlled, such that the displacement of an interpolation point is influenced only by the closest grid points and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations. After the deformable registration, the HU in each voxel in planning CT was mapped to the corresponding point in the reference CBCT to produce the modified CBCT images.

The feasibility of DEDM technique was evaluated by using the three patients mentioned above. For this purpose, the CT and CBCT images were registered using the BSpline model. The targets and sensitive structures contoured on the planning CT were copied to the CBCT using the deformable model. For each patient, the treatment plan parameters were employed to recompute the dose distribution based on the patient's modified CBCT. The resultant isodose curves and DVHs were evaluated and the level of improvement in dosimetry due to the use of DEDM was assessed.

III. Results

A. Calibration of CT and CBCT

The relation between kV HU distribution of CBCT and relative electron dosimetry was established by using a Catphan-600 CT phantom following the procedure described in Sec. II.B. The calibration curves for planning CT, half fan and full fan CBCT modes are shown in figure 1a. Figure 1b compares the calibration curves obtained with an interval of 1 week during a period of two months for full fan CBCT. No significant variations were found in the calibration data, which is similar to what have been observed for MV (16). The stability of the kV CBCT and electron density calibration is a good indicator of the HU number integrity and the overall performance of the CBCT system.

B. Phantom study

Figures 2a to 2d show the same transverse slices of the CT and CBCT images of the Catphan-600 phantom with and without motion. The first two panels are the CT and CBCT images of the phantom in the absence of motion, and the second two show the same with the phantom motion "switched on". It is seen that the quality of CBCT images is worse than that of the conventional planning CT, especially in the presence of motion. The HU profiles of the four images along the two orthogonal lines (lines A-A and B-B as marked in figure 2) are plotted in figure 3. It is found that the HU profiles of the planning CT and CBCT normally agree to within 10% in the static situation. On the other hand, when the motion is "switched on", CBCT shows a much greater level of artifacts (figure 2d) and the HU difference between the conventional CT and CBCT is aggravated, with the maximum difference reaching several hundred HUs.

Because of the cone beam geometry, the influence of scatter radiation in CBCT is much severe as compared to that of a fan beam geometry. In general, X-ray scatter reduces image contrast, increases image noise and may introduce reconstruction error into CBCT. Figure 4 plots the HU profiles along a central axis of three different sized homogeneous cylindrical water phantoms. As expected, the fluctuation range of HU value increases with the phantom size, indicating the increased influence of scatter radiation.

Figures 5, 6 and 7 present the dosimetric results calculated using a single 6 MV 5 X 5cm^2 photon beam. Figures 5a to 5d depict the dose distributions in a transverse slice

calculated based on the four sets of images given in figure 2. Figures 6a and 6b compare the dose profiles along the two orthogonal lines (lines A-A and B-B in figure 2), and figure 7 compares the DVHs of the target for the four different situations. From these results we find that the dose calculated using planning CT agrees with that of CBCT-based calculation to within 1.0%, indicating that it may be acceptable to use kV CBCT for dose calculation if no organ motion presents. However, when phantom motion is involved, the motion-induced artifacts significantly influence the HU distribution and thus the accuracy of CBCT-based dose calculation. For this simple phantom case, we find that the discrepancy between the planning CT- and CBCT-based calculations is about 3%, which is clinically significant. The motion artifacts existing in current CBCT limit the direct use of CBCT for dose calculation when intra-fractional organ motion is not negligible.

C. Patient study

Figures 8a to 8c show the same transverse slices of the planning CT, CBCT, and checkerboard image resulting from the deformable registration of the two sets of images for one of the prostate cases. The modified CBCT obtained by mapping the HUs from the planning CT to CBCT is shown in figure 8d. Our previous studies have indicated that a registration accuracy better than 2mm is achievable by using the BSpline deformable model (18,19). As can be seen from the checkerboard overlay, the registration between CT and CBCT is excellent. Figure 9 shows the isodose distributions for the three calculations based on planning CT, CBCT, and modified CBCT for the same case. A comparison of DVHs of PTV, prostate, seminal vesicles, bladder and rectum for this case is presented in figure 10. Figures 11 and 12 present the DVHs for the targets and sensitive structures for the other two cases. While there is significant dosimetric discrepancy between the planning CT- and modified CBCT-based plans, the results obtained using the CBCT or modified CBCT is similar. For all three cases, we found that the modified CBCT-reconstructed prostate dose agrees with the planned one to within 3~4%. However, the dosimetric differences in the PTV and seminal vesicles are quite significant, which could be as large as ~10% for the PTV and ~50% for the seminal

vesicles. Similar observation was also made by the MD Anderson group in a study using daily CT on-rail (9, 25). For the rectum and bladder, the discrepancies between planned and reconstructed doses could be greater than 8%. We note that for all the structures (except the seminal vesicle in the first case), the differences between the CBCT- and modified CBCT-based calculations are all less than 2%. For the seminal vesicle in the first case, the DVH difference is somewhat large. We attribute this to the relatively small volume of the seminal vesicle and a possible structural mismatch between the CT and CBCT.

In general, the difference between the planned and CBCT-reconstructed dose distributions arises from two factors: (i) patient positioning error and organ deformation/displacement; and (ii) relative electron density difference between the CT and CBCT images. The small discrepancy between the doses computed using CBCT and modified CBCT suggests that, in the prostate cases, the second factor is small and it may be acceptable to directly use CBCT for dose calculation. The dosimetry is predominantly determined by the accuracy of patient setup and the level of interfractional deformation /displacement of the involved target and sensitive structures.

IV. Discussion

The feasibility and accuracy of using kV CBCT to calculate dose have been investigated with phantoms and three clinical prostate cases. In the absence of motion artifacts, it seems to be acceptable to directly use CBCT for dose verification calculation. Otherwise, extra caution is required to avoid significant dosimetric inaccuracy. To cope with the problem of deteriorated imaging quality of CBCT, a DEDM method has been proposed to map the electron density information from the patient's planning CT to the setup CBCT with a deformable image registration. In IGRT, since the registration has to be done for the purpose of patient setup, the computational overhead of introducing DEDM is minimal. Before an effective CBCT image quality improving technique is in place, DEDM provides a useful interim solution to the problem. In the presence of organ motion, our phantom study indicated that significant dosimetric errors could be resulted. Recent developments of 4D CT and 4D CBCT (21, 26, 27), in conjunction with the

proposed DEDM method, may provide a valuable solution to the problem of CBCT-based dose calculation in the regions of thorax and upper abdomen.

Dose distributions computed based on CBCT or modified CBCT represent the dose to be delivered to the patient because the CBCT was acquired prior to the patients' treatments after the patients were repositioned/shifted using the patient setup procedure in current practice. In the prostate IMRT plans, the inherent dosimetric error resulted from the use of CBCT images is found to be small. However, the dosimetric error caused by the inter-fractional organ motion/deformation is not insignificant, as revealed by the dose recalculation results given in the last section. A few groups are working on deformable model based segmentation and patient setup procedures (9, 25, 28, 29). When deformable model is used, one can go beyond simply aligning the 3D bony structures to achieve a registration based on matching soft-tissue organ(s). The problem here becomes multi-dimensional depending on which structure to align during the patient setup process. The multiple choices resulting from the fact that the dimensionality of the patient data is much greater than that available in the patient setup procedure and suggest that deformable registration is not the ultimate solution to volumetric image-guided radiation therapy. However, patient positioning based on deformable model improves the current body-structure-based patient alignment method since it partially takes into account organ deformation by achieving the closest overlay match possible between the planning and CBCT data sets according to our clinical objective, and provides an improved positioning technique. We should emphasize that, even when 3D volumetric based deformable registration is available in the future, the problem of patient positioning will not disappear as relative organ deformations may well persist. A possible solution to accommodate various factors mentioned above is for us to re-optimize or tweak the IMRT plan based on the patient's setup CBCT. Indeed, in order to fully utilize the CBCT volumetric data, a new paradigm with seamlessly integrated simulation, planning, verification, and delivery procedure is urgently needed. Until this is realized clinically, the volumetric imaging is nothing but an expensive extension of the existing planar verification approach.

Finally, we emphasize that a pre-assumption of the proposed DEDM approach is that the image registration between CT and CBCT is sufficiently accurate to avoid wrong assignment of electron density information. Generally, the accuracy of

deformable registration may depend on the quality of CBCT. One could naturally expect that, at a certain level of intensity mismatch between the CT and CBCT, the deformable model may break down. In a parallel study, we have recently studied the influence of noises on the accuracy of rigid and deformable image registration and developed a multiscale image registration technique for the registration of medical images that contain significant levels of noise(30). As usually, after image registration, it is helpful to have an experienced physician or physicist to double check the registration results.

V. Conclusion

Onboard CBCT provides useful volumetric anatomy information for patient positioning verification. When used for dose verification calculation, it is required to have a reliable HU to electron density curve. Our phantom and patient studies have indicated that, in the absence of motion artifacts, the dosimetric accuracy seems to be acceptable for the purpose of dosimetric sanity check. Our motion phantom study indicated that the dosimetric errors may be more pronounced when intra-fractional organ motion is present. In this situation, a direct use of CBCT for dose calculation is not recommended. The use of a reliable deformable registration would allow us to incorporate the electron density distribution from the planning CT and to calculate the dose more accurately. The proposed DEDM approach affords a practical means to estimate the dose to be delivered or already delivered to the patient based on the setup CBCT.

Acknowledgement

We wish to acknowledge Drs. C. King, S. Hancock, T. Pawlicki, P. Maxim, and G. Luxton for useful discussions. The supports from the National Cancer Institute (1 R01 CA98523-01), Department of Defense (PC040282), and Varian Medical Systems are also gratefully acknowledged.

Referrences

1. Langen KM, Jones DT: Organ motion and its management. Int J Radiat Oncol Biol Phys 2001; 50(1):265-78
2. Bortfeld T, Jiang SB, Rietzel E: Effects of motion on the total dose distribution. Semin Radiat Oncol 2004; 14(1):41-51
3. Yan D, Lockman D: Organ/patient geometric variation in external beam radiotherapy and its effects. Med Phys 2001; 28(4):593-602
4. Hugo GD, Agazaryan N, Solberg TD: The effects of tumor motion on planning and delivery of respiratory-gated IMRT. Med Phys 2003; 30(6):1052-66
5. Mohan DS, Kupelian PA, Willoughby TR: Short-course intensity-modulated radiotherapy for localized prostate cancer with daily transabdominal ultrasound localization of the prostate gland. Int J Radiat Oncol Biol Phys 2000; 46(3):575-80
6. Yan D, Xu B, Lockman D, Kota K, Brabbins DS, Wong J, Martinez AA: The influence of interpatient and intrapatient rectum variation on external beam treatment of prostate cancer. Int J Radiat Oncol Biol Phys 2001; 51(4):1111-9
7. Xing L, Thorndyke B, Schreibmann E, Yang Y, Li TF, Kim GY, Luxton G, Koong A: Overview of image-guided radiation therapy. Med Dosim 2006; 31(2):91-112
8. Berbeco RI, Jiang SB, Sharp GC, Chen GT, Mostafavi H, Shirato H: Integrated radiotherapy imaging system (IRIS): design considerations of tumour tracking with linac gantry-mounted diagnostic x-ray systems with flat-panel detectors. Phys Med Biol 2004; 49(2):243-55
9. Mohan R, Zhang X, Wang H, Kang Y, Wang X, Liu H, Ang KK, Kuban D, Dong L: Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes. Int J Radiat Oncol Biol Phys 2005; 61(4):1258-66
10. Letourneau D, Martinez AA, Lockman D, Yan D, Vargas C, Ivaldi G, Wong J: Assessment of residual error for online cone-beam CT-guided treatment of prostate cancer patients. Int J Radiat Oncol Biol Phys 2005; 62(4):1239-46

11. Mackie TR, Kapatoes J, Ruchala K, Lu W, Wu C, Olivera G, Forrest L, Tome W, Welsh J, Jeraj R, Harari P, Reckwerdt P, Paliwal B, Ritter M, Keller H, Fowler J, Mehta M: Image guidance for precise conformal radiotherapy. Int J Radiat Oncol Biol Phys 2003; 56(1):89-105

12. Meeks SL, Harmon JF, Jr., Langen KM, Willoughby TR, Wagner TH, Kupelian PA: Performance characterization of megavoltage computed tomography imaging on a helical tomotherapy unit. Med Phys 2005; 32(8):2673-81

13. Pouliot J, Aubin M, Langen KM, Liu YM, Pickett B, Shinohara K, Roach M, 3rd: (Non)-migration of radiopaque markers used for on-line localization of the prostate with an electronic portal imaging device. Int J Radiat Oncol Biol Phys 2003; 56(3):862-6

14. Pouliot J, Bani-Hashemi A, Chen J, Svatos M, Ghelmansarai F, Mitschke M, Aubin M, Xia P, Morin O, Bucci K, Roach M, 3rd, Hernandez P, Zheng Z, Hristov D, Verhey L: Low-dose megavoltage cone-beam CT for radiation therapy. Int J Radiat Oncol Biol Phys 2005; 61(2):552-60

15. Ghilezan M, Yan D, Liang J, Jaffray D, Wong J, Martinez A: Online image-guided intensity-modulated radiotherapy for prostate cancer: How much improvement can we expect? A theoretical assessment of clinical benefits and potential dose escalation by improving precision and accuracy of radiation delivery. Int J Radiat Oncol Biol Phys 2004; 60(5):1602-10

16. Langen KM, Meeks SL, Poole DO, Wagner TH, Willoughby TR, Kupelian PA, Ruchala KJ, Haimerl J, Olivera GH: The use of megavoltage CT (MVCT) images for dose recomputations. Phys Med Biol 2005; 50(18):4259-76

17. Lian J, Xing L, Hunjan S, Dumoulin C, Levin J, Lo A, Watkins R, Rohling K, Giaquinto R, Kim D, Spielman D, Daniel B: Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: a deformable registration and validation study. Med Phys 2004; 31(11):3087-94

18. Schreibmann E, Chen GT, Xing L: Image interpolation in 4D CT using a BSpline deformable registration model. Int J Radiat Oncol Biol Phys 2006; 64(5):1537-50

19. Schreibmann E, Xing L: Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies. Int J Radiat Oncol Biol Phys 2005; 62(2):595-605
20. Schreibmann E, Xing L: Image registration with auto-mapped control volumes. Med Phys 2006; 33(4):1165-79
21. Li T, Schreibmann E, Yang Y, Xing L: Motion correction for improved target localization with on-board cone-beam computed tomography. Phys Med Biol 2006; 51(2):253-67
22. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W: PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging 2003; 22(1):120-8
23. Loeckx D, Maes F, Vandermeulen D, Suetens P: Non-rigid image registration using a statistical spline deformation model. Inf Process Med Imaging 2003; 18:463-74
24. Li T, Schreibmann E, Thorndyke B, Xing L: Radiation Dose Reduction in 4D Computed Tomography. Medical Physics 2005; 32(6):2094
25. Court LE, Dong L, Lee AK, Cheung R, Bonnen MD, O'Daniel J, Wang H, Mohan R, Kuban D: An automatic CT-guided adaptive radiation therapy technique by online modification of multileaf collimator leaf positions for prostate cancer. Int J Radiat Oncol Biol Phys 2005; 62(1):154-63
26. Sonke JJ, Zijp L, Remeijer P, van Herk M: Respiratory correlated cone beam CT. Med Phys 2005; 32(4):1176-86
27. Li T, McGuinness C, Loo B, Munro P, Koong A, Xing L: Four-dimensional cone-beam CT using an on-board imager. Medical Physics 2006:submitted
28. Oldham M, Letourneau D, Watt L, Hugo G, Yan D, Lockman D, Kim LH, Chen PY, Martinez A, Wong JW: Cone-beam-CT guided radiation therapy: A model for on-line application. Radiother Oncol 2005; 75(3):271 E1-8
29. Li T, Yang Y, Schreibmann E, Xing L: A New Cone-Beam CT Repositioning Technique Through Deformable Registration. Medical Physics 2005; 32(6):2160
30. Paquin D, Levy D, Schreibmann E, Xing L: Multistage image registration. Mathematical Biosciences and Engineering 2006; 3:389-418.

Figure Captions

Fig. 1. (a)The calibration curves (Hounsfield number vs relative electron density) for planning CT, half fan and full fan mode CBCT; (b) the variation of calibration curves with time for the full fan CBCT.

Fig. 2. The CT and CBCT images with and without motion for the Catphan-600: (a) planning CT in the absence of phantom motion; (b) CBCT in the absence of phantom motion; (c) planning CT with moving phantom; and (d) CBCT with moving phantom.

Fig. 3. HU profiles of planning CT and CBCT images (see figure 2) along the A-A line (panel a) and B-B line (panel b).

Fig. 4. HU profiles for three different sized homogeneous cylindrical water phantoms. The diameters for large, medium and small phantoms are 10.8, 16.0 and 26.6 cm, respectively.

Fig. 5. Dose distributions in a transverse slice calculated based on the four sets of CT data shown in figure 2: (a) planning CT; (b) CBCT; (c) planning CT with a motion; and (d) CBCT with a motion. In all four situations, a 5 X 5cm^2 single field plan was used to irradiate a spherical hypothetical target with a diameter of 5cm located at the phantom center.

Fig. 6. Comparison of the dose profiles along the two orthogonal lines shown in figure 2 for the Catphan-600 phantom: (a) profile along the A-A line; (a) profile along the B-B line.

Fig. 7. Comparison of the target DVHs calculated based on the four sets of CT data shown in figure 2 for the phantom case.

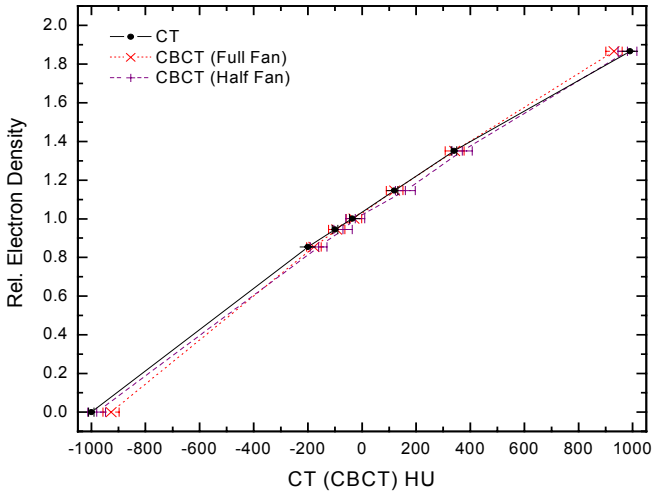
Fig. 8. CT, CBCT and modified CBCT images for the first prostate case: (a) planning CT; (b) daily CBCT; (c) checkerboard overlay of CT and CBCT after the deformation registration; and (d) modified CBCT.

Fig. 9. Dose distributions in a transverse slice calculated based on the: (a) planning CT; (b) CBCT; and (c) modified CBCT for the prostate case.

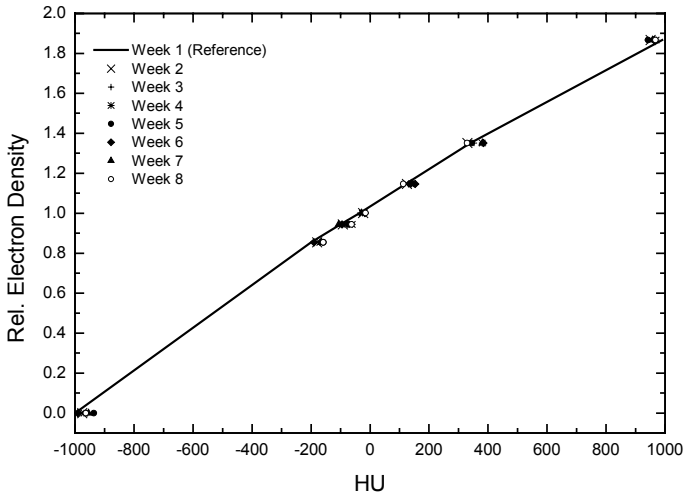
Fig. 10. Comparison of DVHs of the prostate, PTV, rectum and bladder obtained based on the planning CT, CBCT and modified CBCT images for the first prostate case.

Fig. 11. Comparison of DVHs of the prostate, PTV, rectum and bladder obtained based on the planning CT, CBCT and modified CBCT images for the second prostate case.

Fig. 12. Comparison of DVHs of the prostate, PTV, rectum and bladder obtained based on the planning CT, CBCT and modified CBCT images for the third prostate case.


CT
CBCT (Full Fan)
CBCT (Half Fan)
Rel. Electron Density
CT (CBCT) HU
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-1000 -800 -600 -400 -200 0 200 400 600 800 1000

(a)


Week 1 (Reference)
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Rel. Electron Density
HU
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-1000 -800 -600 -400 -200 0 200 400 600 800 1000

(b)

Figure 1

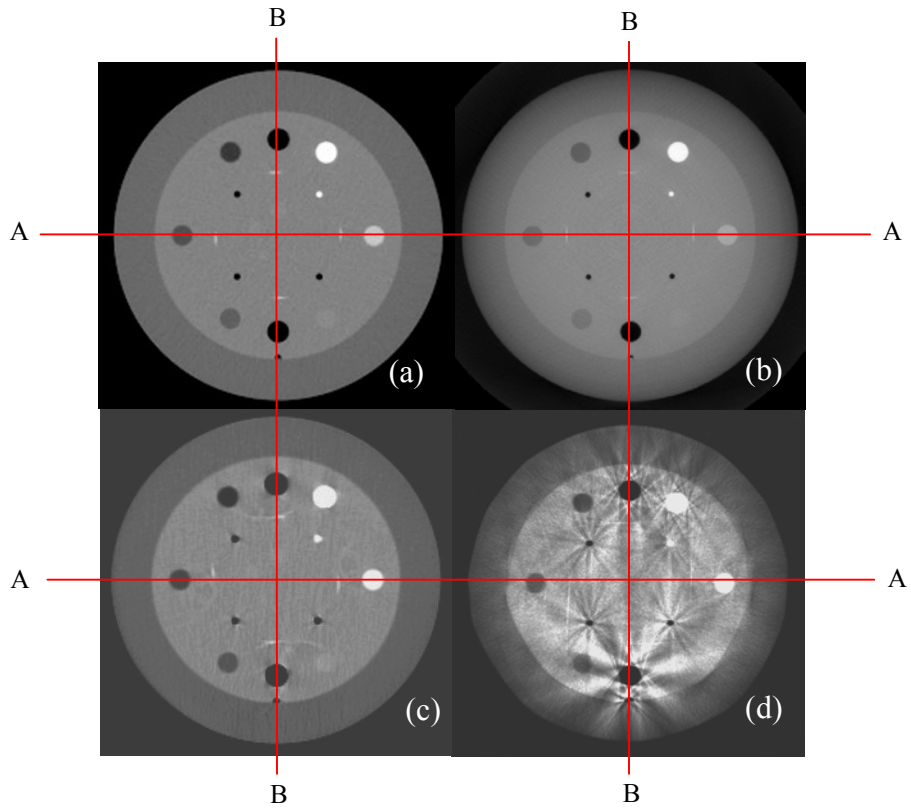
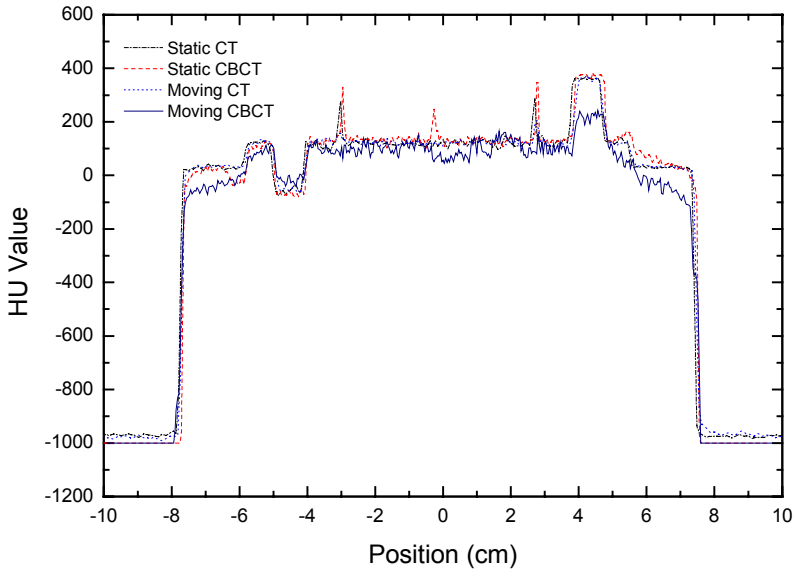
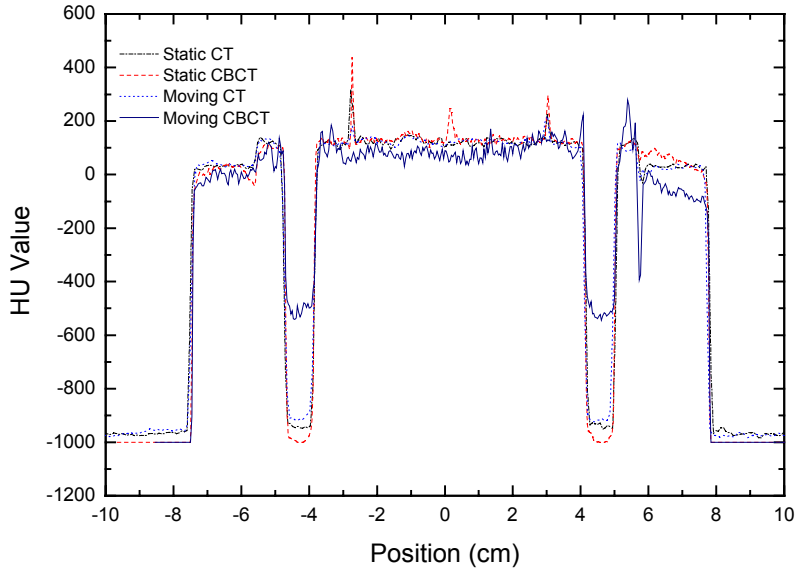


Figure 2


Static CT
Static CBCT
Moving CT
Moving CBCT
HU Value
Position (cm)
600
400
200
0
-200
-400
-600
-800
-1000
-1200
-10
-8
-6
-4
-2
0
2
4
6
8
10

(a)


Static CT
Static CBCT
Moving CT
Moving CBCT
HU Value
Position (cm)
600
400
200
0
-200
-400
-600
-800
-1000
-1200
-10
-8
-6
-4
-2
0
2
4
6
8
10

(b)

Figure 3

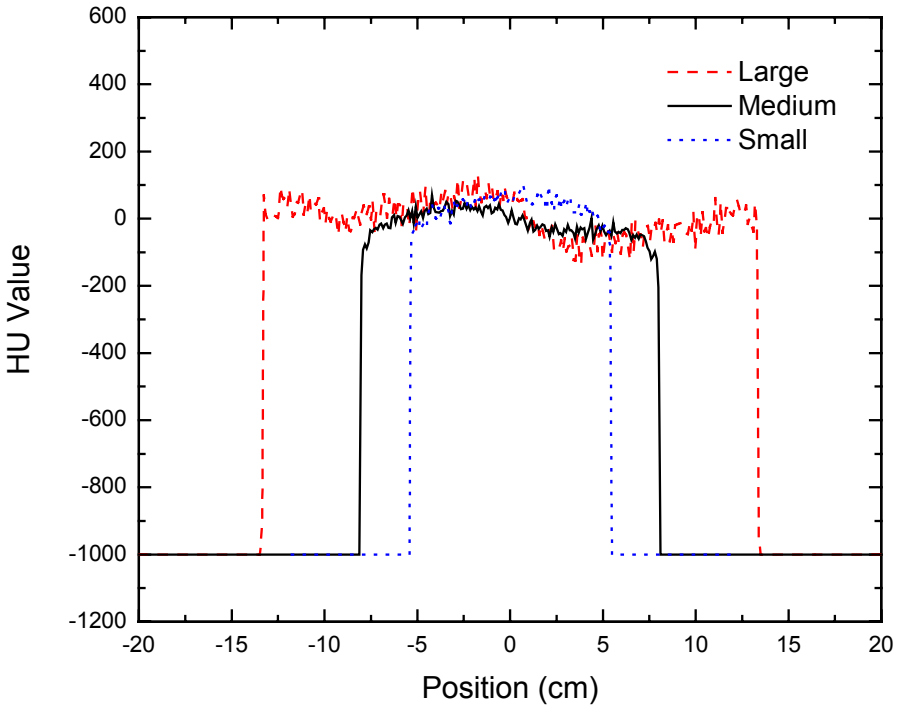

600
400
200
0
-200
-400
-600
-800
-1000
-1200
HU Value
-20
-15
-10
-5
0
5
10
15
20
Position (cm)
Large
Medium
Small

Figure 4

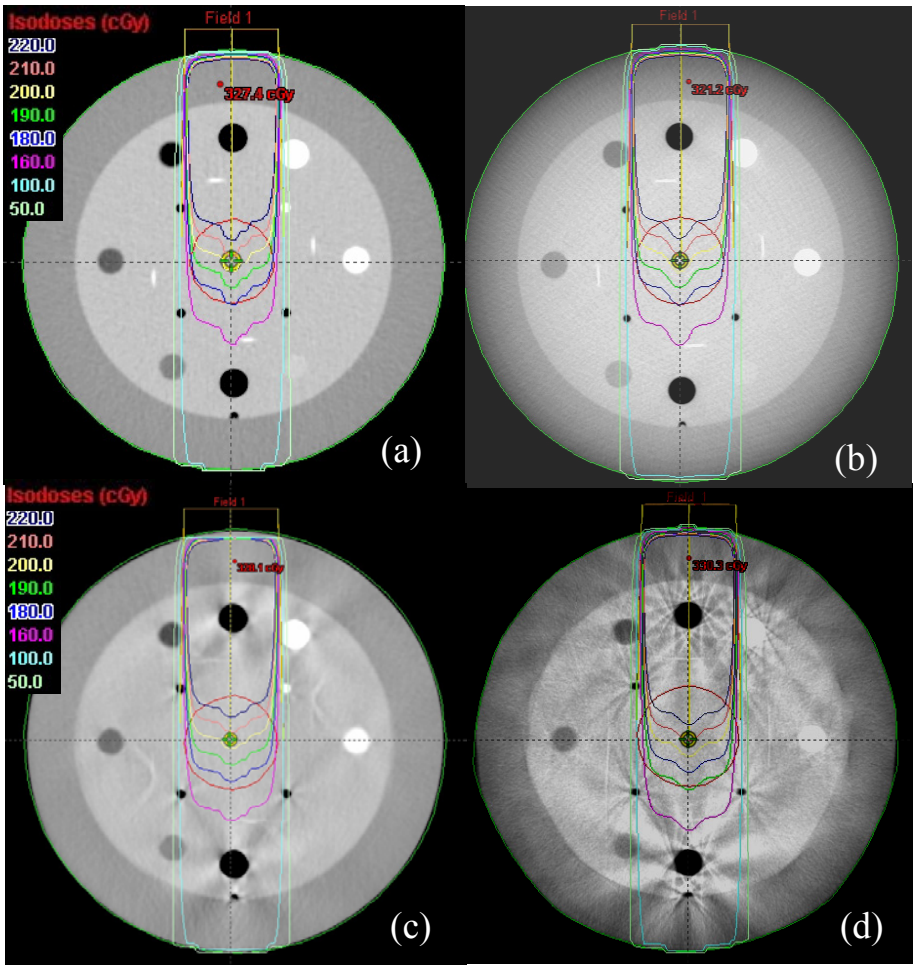
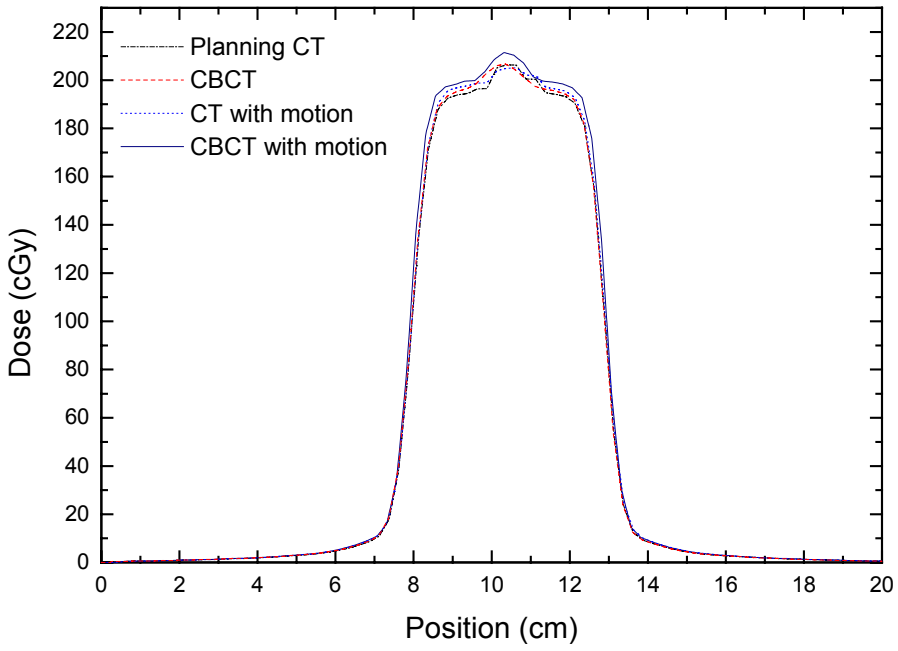
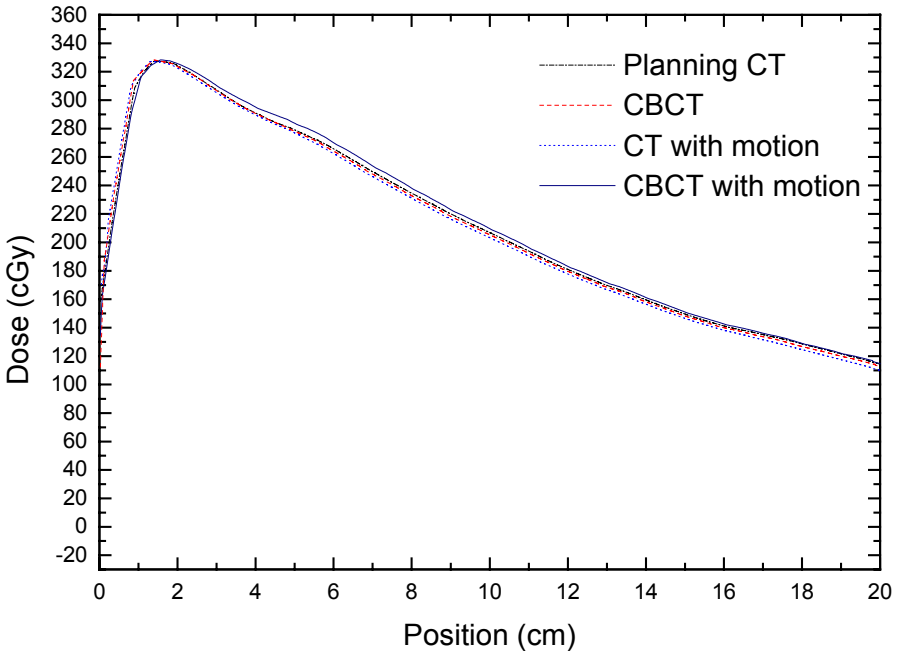

Isodoses (cGy)
220.0
210.0
200.0
190.0
180.0
160.0
100.0
50.0
Field 1
327.4 cGy
(a)
Field 1
321.2 cGy
(b)
Isodoses (cGy)
220.0
210.0
200.0
190.0
180.0
160.0
100.0
50.0
Field 1
(c)
Field 1
(d)

Figure 5


Planning CT
CBCT
CT with motion
CBCT with motion
Dose (cGy)
Position (cm)
0
20
40
60
80
100
120
140
160
180
200
220
0
2
4
6
8
10
12
14
16
18
20

(a)


Planning CT
CBCT
CT with motion
CBCT with motion
Dose (cGy)
Position (cm)
-20
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
0
2
4
6
8
10
12
14
16
18
20

(b)

Figure 6

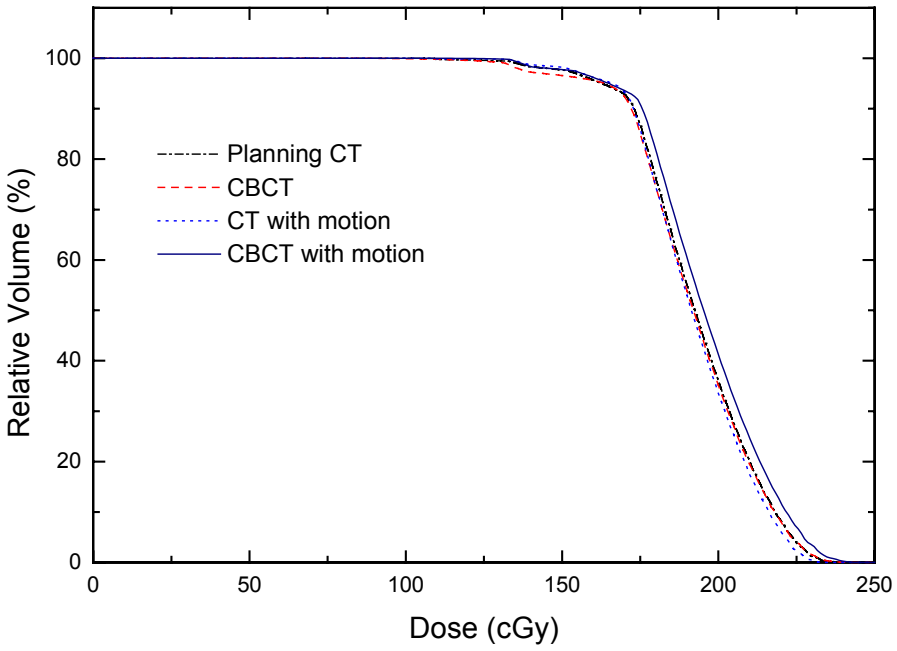

Relative Volume (%)
100
80
60
40
20
0
0
50
100
150
200
250
Dose (cGy)
Planning CT
CBCT
CT with motion
CBCT with motion

Figure 7

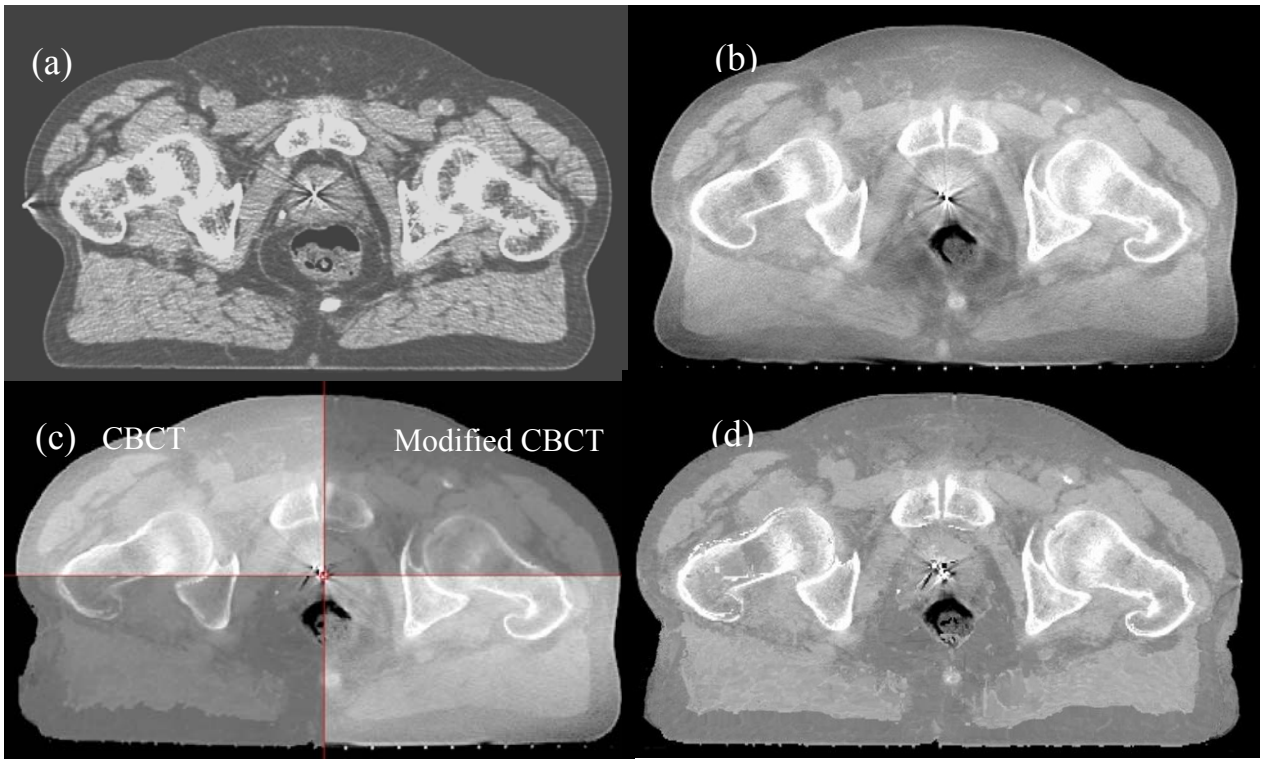


Figure 8

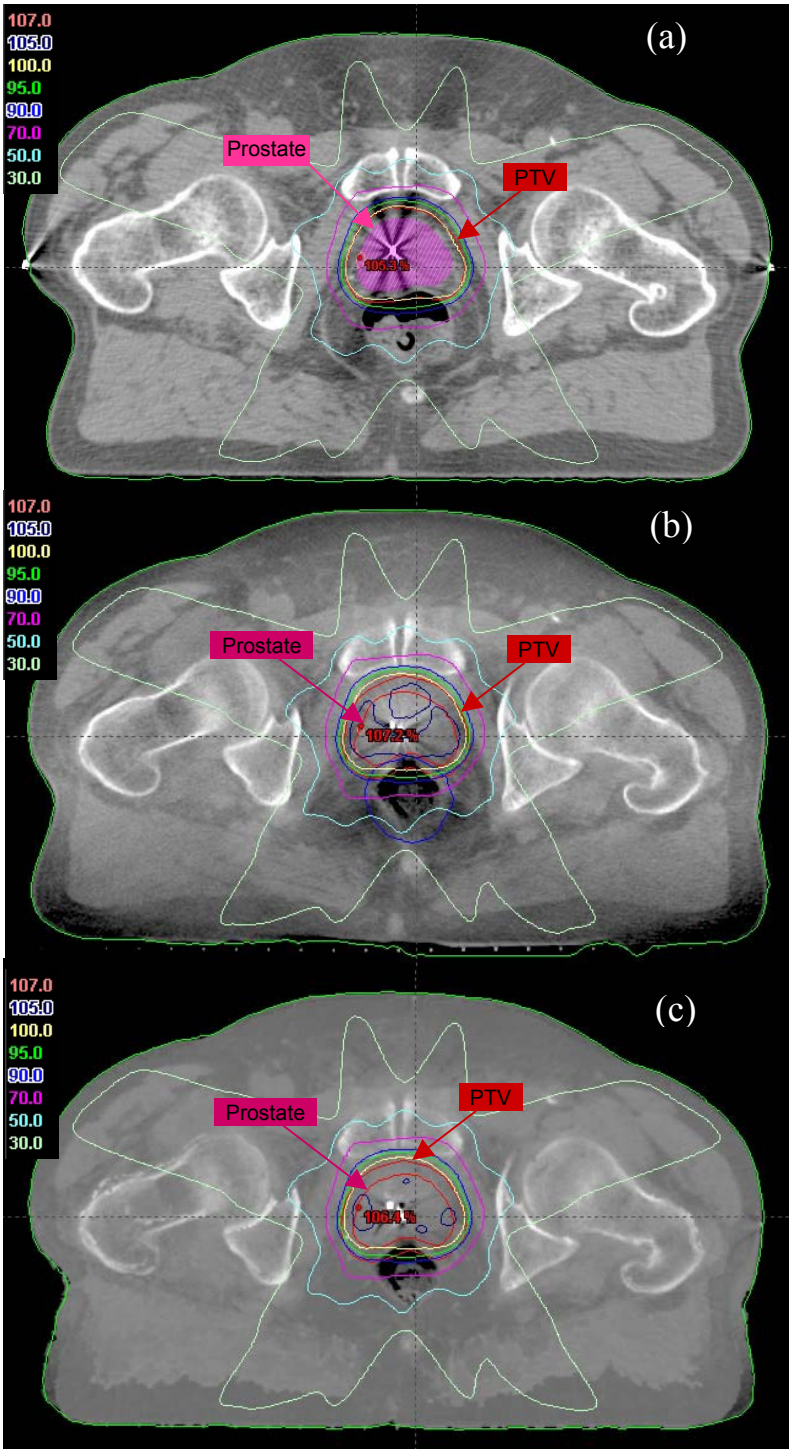

107.0
105.0
100.0
95.0
90.0
70.0
50.0
30.0
(a)
Prostate
PTV
(b)
Prostate
PTV
(c)
Prostate
PTV

Figure 9

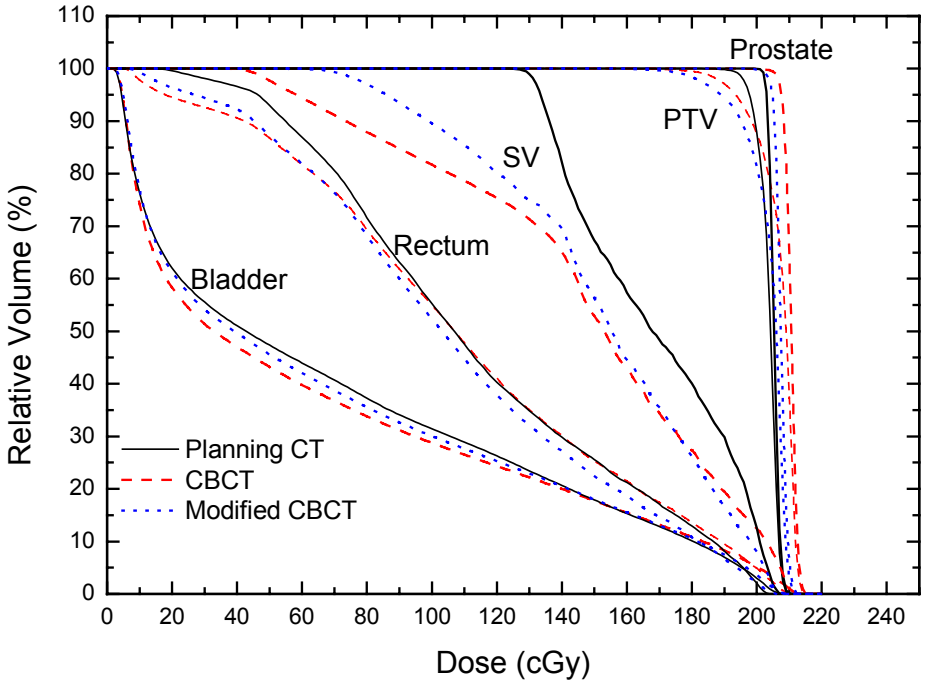

Relative Volume (%)
Dose (cGy)
Prostate
PTV
SV
Rectum
Bladder
Planning CT
CBCT
Modified CBCT
0
10
20
30
40
50
60
70
80
90
100
110
0
20
40
60
80
100
120
140
160
180
200
220
240

Figure 10

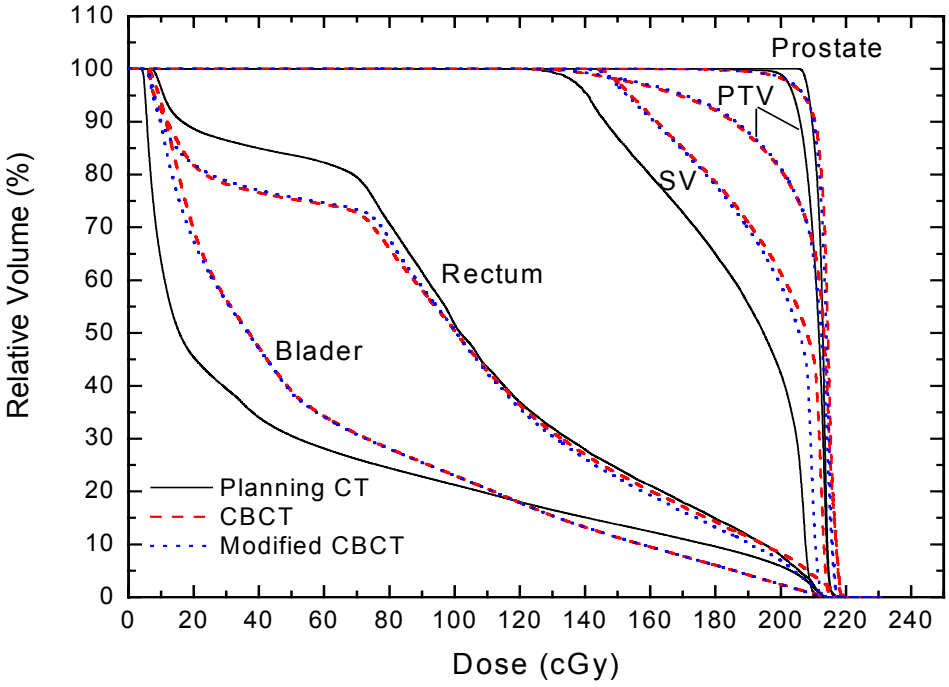

Relative Volume (%)
Dose (cGy)
Prostate
PTV
SV
Rectum
Blader
Planning CT
CBCT
Modified CBCT
0
10
20
30
40
50
60
70
80
90
100
110
0
20
40
60
80
100
120
140
160
180
200
220
240

Figure 11

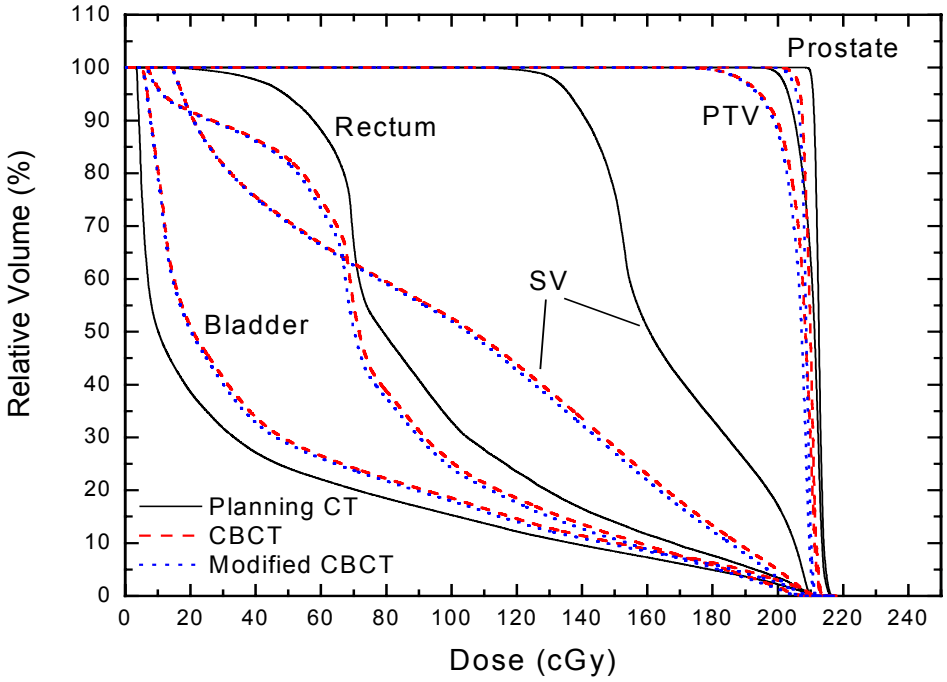

Relative Volume (%)
Dose (cGy)
Prostate
PTV
Rectum
SV
Bladder
Planning CT
CBCT
Modified CBCT
0
10
20
30
40
50
60
70
80
90
100
110
0
20
40
60
80
100
120
140
160
180
200
220
240

Figure 12

Multiscale Deformable Registration

Dana Paquin*, Doron Levy, and Lei Xing

***Abstract*— Multiscale image registration techniques are presented for the deformable registration of medical images. These techniques are shown to be particularly effective for registration problems in which one or both of the images to be registered contains significant levels of noise. Experiments using B-splines deformable registration models demonstrate that ordinary deformable registration techniques fail to produce accurate results in the presence of significant levels of noise. Further experiments in which classical denoising algorithms are applied prior to registration show that registration fails in this case for high levels of noise, as well. The hierarchical multiscale image decomposition of E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Modeling and Simulations, vol. 2, no. 4, pp. 554–579, 2004, is presented, and multiscale image registration algorithms are developed based on the multiscale decomposition. An accurate registration of noisy images is achieved by obtaining a hierarchical multiscale decomposition of the images and registering the resulting components. This approach enables a successful registration of images that contain noise well beyond the level at which ordinary deformable registration fails. Further, an iterated multiscale registration algorithm is presented which improves the accuracy and computational efficiency of other registration methods. The accuracy and efficiency of the multiscale registration techniques are demonstrated using both digital phantom and clinical case studies in two and three dimensions.**

***Index Terms*— Image registration, deformable registration, noise, multiscale decomposition.**

I. Introduction

IMAGE registration is the process of determining the optimal spatial transformation that maps one image to another. Image registration is necessary, for example, when images of the same object are taken at different times, from different imaging devices, or from different perspectives. The two images to be registered, called the fixed and moving images, are the input to the registration algorithm, and the output is the optimal transformation that maps the moving image to the fixed image. Ideally, the transformed moving image should be identical to the fixed image after registration. Applications of image registration include image-guided radiation therapy (IGRT), image-guided surgery, functional MRI analysis, and tumor detection, as well as many non-medical applications, such as computer vision, pattern recognition, and remotely sensed data processing (see [4] and the references therein).

Image registration models are classified into two main categories according to the transformation type: rigid and deformable. Rigid image registration models assume that the transformation that maps the moving image to the fixed image consists only of translations and rotations, while deformable models allow localized stretching of images. While rigid models are sufficient in certain circumstances, many registration problems, particularly in medical applications, are non-rigid. For example, respiratory motion causes non-rigid, or deformable, distortion of the lungs and other organs. As another example, image-guided neurosurgery procedures require deformable registration of pre- and intra-operative images of the brain [14], [19]. For additional applications of deformable registration, see [11], [16], [17] and the references therein.

This paper extends our previous work [13], a multiscale approach to rigid registration in the presence of noise, to deformable registration problems. Our approach is to decompose the images that

Asterisk indicates corresponding author.

*D. Paquin is with the Department of Mathematics, Stanford University, Stanford, CA 94305-2125 (e-mail: paquin@math.stanford.edu).

D. Levy is with the Department of Mathematics, Stanford University, Stanford, CA 94305-2125 (e-mail: dlevy@math.stanford.edu).

L. Xing is with the Department of Radiation Oncology, Stanford University, Stanford, CA 94305-5847 (e-mail: lei@reyes.stanford.edu).

are to be registered into a hierarchical multiscale decomposition. Registration is then obtained with a hierarchical approach in which the various scales are registered consecutively, with each additional scale registration providing a fine tuning to the previous registration.

The structure of this paper is as follows. In Section II, we provide a brief overview of the image registration problem and discuss deformable registration techniques. In Section III, we describe the problem of deformable image registration in the presence of noise, and briefly discuss standard denoising techniques. In Section IV, we review the hierarchical multiscale image decomposition of [18], and we present three multiscale image registration algorithms based on the decomposition. In Section V, we demonstrate the failure of ordinary deformable registration techniques when one or more of the images to be registered contains noise, as well as the failure of ordinary techniques even when the noisy images are denoised using classical filtering methods prior to registration. The accuracy and efficiency of our multiscale registration techniques are studied in Section V with several image registration experiments in both two and three dimensions. Concluding remarks are given in Section VI.

II. Mathematical formulation of the registration problem

Given a *fixed* and a *moving* image, the registration problem is the process of finding an *optimal transformation* that brings the moving image into spatial alignment with the fixed image. While this problem is easy to state, it is difficult to solve. The main source of difficulty is that the problem is ill-posed, which means, for example, that the problem may not have a unique solution. For a detailed overview of the image registration problem and various image registration techniques, see [12].

To formulate the registration problem mathematically, a two-dimensional gray-scale image f is a mapping which assigns to every point $x \in \Omega \subset \mathbb{R}^2$ a gray value $f(x)$ (called the intensity value of the image at the point x). We will consider images as elements of the space $L^2(\mathbb{R}^2)$. Color images can be defined, for example, in terms of vector-valued functions $\mathbf{f} = (f_1, f_2, f_3)$ representing the RGB-color scales. For the medical imaging applications that we are interested in, images are in fact given in terms of discrete data, and the function f must be obtained via interpolation. We will not discuss this construction here, but assume that an interpolation method has been chosen.

Any registration algorithm has three main components:

1) The *transformation model* which specifies the way in which the moving image can be transformed to correspond to the fixed image.
2) The *distance measure*, or metric, used to compare the fixed and moving images.
3) The *optimization process* that varies the parameters of the transformation model in such a way that the transformation produced by the registration process is optimal.

Given a distance measure $D : (L^2(\mathbb{R}^2))^2 \to \mathbb{R}$ and two images $f(x), m(x) \in L^2(\mathbb{R}^2)$, the solution ϕ of the registration problem is given by the following minimization problem:

$$\phi = \underset{\psi:\mathbb{R}^2\to\mathbb{R}^2}{\operatorname{argmin}} \; D(f(x), m(\psi(x))), \tag{1}$$

where ψ is in the specified space of transformation models. Examples of commonly used distance measures are mean squares, normalized correlation, and mutual information. Examples of typical transformation models are rigid, affine, polynomial, and spline transformations [12]. To minimize $D(f, m(\psi))$, we must choose an optimizer which controls the minimization. The most commonly used optimization techniques in image registration are gradient descent and regular step gradient descent methods. The implementation of the registration algorithm works in the following way: at each iteration, the distance D between the two images is computed. The specified transformation is then applied to the moving image, and the distance between the images is recomputed. In theory, this process continues until the distance is minimized (or maximized in certain cases), though in practice a stopping criterion is often applied.

Historically, image registration problems have been classified as either *rigid* or *non-rigid*. As most of the organs in the human body are not confined to rigid motion only, much of the current work in medical image registration is focused on the deformable case. Although deformable image registration clearly allows for more flexibility in the types of images and applications in which it can be used, deformable registration techniques require significantly more computation time than rigid registration techniques, and involve the determination of a very large number of parameters. In this paper, we shall focus on the problem of *deformable registration* in the presence of noise.

A. Deformable registration techniques

Spline-based transformation models are among the most common and important transformation models used in non-rigid registration problems [6]. Spline-based registration algorithms use *control* points in the fixed and moving images and a spline function to define transformations away from these points. The two main spline models used in registration are *thin-plate splines* and *B-splines*. Thin-plate splines have the property that each control point has a global influence on the transformation. That is, if the position of one control point is perturbed, then all other points in the image are perturbed as well. This can be a disadvantage because it limits the ability of the transformation model to model localized deformations. In addition, the computation time required for a thin-plate spline-based registration algorithm increases significantly as the number of control points increases. See [3] for an overview of thin-plate splines.

In contrast, B-splines are only defined in the neighborhood of each control point. Thus perturbing the position of one control point affects the transformation only in a neighborhood of that point. As a result, B-spline-based registration techniques are more computationally efficient than thin-plate splines, especially for a large number of control points. See [9] and [10] for a detailed description of B-splines transformation models. In this paper, we shall use deformable registration algorithms based on B-spline deformation models. To define the spline-based deformation model, let $\Omega = \{(x, y) \mid 0 \leq x \leq X,\ 0 \leq y \leq Y\}$ denote the domain of the image volume. Let α denote a $n_x \times n_y$ mesh of control points $\alpha_{i,j}$ with uniform spacing δ. Then the B-spline deformation model can be written as the 2-D tensor product of 1-D cubic B-splines:

$$\phi(x, y) = \sum_{l=0}^{2} \sum_{m=0}^{2} B_l(x) B_m(y) \alpha_{i+l, j+m}, \quad (2)$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, and B_l represents the l-th basis of the B-spline:

$$\begin{aligned} B_0(u) &= \frac{1}{6}(1-u)^3 \ , \\ B_1(u) &= \frac{1}{6}(3u^3 - 6u^2 + 4) \ , \\ B_2(u) &= \frac{1}{6}(-3u^3 + 3u^2 + 3u + 1) \ . \end{aligned}$$

Changing the control point $\alpha_{i,j}$ affects the transformation only in a local neighborhood of $\alpha_{i,j}$. The control points α act as parameters of the B-spline deformation model, and the degree of non-rigid deformation which can be modeled depends on the resolution of the mesh of control points α. A large spacing of control points allows modeling of global non-rigid deformation, while a small spacing of control points allows modeling of local non-rigid deformations. Additionally, the number of control points determines the number of degrees of freedom of the transformation model, and hence, the computational complexity. For example, a B-spline deformation model defined by a 10×10 grid of control points yields a transformation with $2 \times 10 \times 10 = 200$ degrees of freedom. Thus there is a tradeoff between the model flexibility and its computational complexity.

In Section V, we demonstrate the accuracy of B-splines deformable registration techniques for the registration of non-noisy images. Finally, we note in passing that there are additional deformable registration techniques such as elastic models [2], viscous fluid models [5], and finite element models [7].

III. Registration in the presence of noise

A. Speckle noise model

In imaging, the term noise refers to random fluctuations in intensity values that occur during image capture, transmission, or processing, and that may distort the information given by the image. Image noise is not part of the ideal signal and may be caused by a wide range of sources, such as detector sensitivity, environmental radiation, transmission errors, discretization effects, etc. In this paper, we will study the problem of image registration in the presence of high levels of *speckle* noise (though we have conducted experiments demonstrating that we obtain similar results for other types of noise). See, for example, our results on rigid registration [13].

Speckle noise, or multiplicative noise, is a type of noise that occurs commonly in medical imaging. In particular, speckle noise is often found in ultrasound images [1]. It is defined by the following model. We let $s(x)$ denote the actual image, and $f(x)$ the observed image. Then

$$f(x) = s(x) + \eta(0, \delta) \cdot s(x), \tag{3}$$

where $\eta(0, \delta)$ is uniformly distributed random noise of mean 0 and variance δ. We add speckle noise of increasing variance to the image S, as illustrated in Figure 4. For a given noise variance δ, we denote the noisy image S_δ. In Section V, we present the registration results obtained upon registering the noisy images S_δ with the original image I for increasing values of δ.

B. Classical denoising techniques

Since we are considering the problem of image registration in the presence of noise, it is natural to consider whether or not the application of standard denoising algorithms prior to registration enables a successful deformable registration of noisy images. Image denoising is a fundamental problem in image processing, and there has been much research and progress on the subject. As our primary interest is the problem of image registration of noisy images, and not denoising, we do not focus on the general problem of image denoising, but instead consider a few of the most common and computationally simple denoising techniques.

Spatial filtering is the traditional approach to removing noise from images. Spatial filters use the assumption that noise occupies the higher regions of the frequency spectrum, and thus attenuate high spatial frequencies. Local spatial filtering is a process in which the value of a given pixel in the filtered image is computed by applying some algorithm to the pixel values in a neighborhood of the given pixel. Typical implementations of spatial filters include mean filtering, median filtering, and Gaussian smoothing. See [20] for an overview of classical spatial filtering techniques.

Mean filtering computes the value of each output pixel by computing the statistical mean of the neighborhood of the corresponding input pixel. That is, the intensity value of each pixel is replaced with the mean, or average, value of its neighbors (including itself). The neighborhood size is typically taken to be a 3×3 or a 5×5 square grid (or kernel); larger neighborhoods result in more severe smoothing. In this paper, we use a 3×3 square kernel to perform mean filtering. Applying a mean filter to a noisy image reduces the amount of variation in gray-level intensity between pixels. Although this filter is computationally easy to implement, it is sensitive to the presence of outliers.

Median filtering computes the value of each output pixel by computing the statistical median of the neighborhood of the corresponding input pixel. The median filter has two main advantages over the mean filter. First, the median is a more robust average than the mean, and thus is less sensitive to a single very unrepresentative pixel in a neighborhood. Second, since the median value must actually be the value of one of the pixels in the neighborhood, the median filter does not create unrealistic pixel values when the filter straddles an edge. For this reason, the median filter is much more effective at preserving sharp edges than the mean filter. As with the mean filter, we use a 3×3 square kernel to perform median filtering.

Convolution with a Gaussian kernel is another commonly used spatial filtering technique. The 2-dimensional Gaussian kernel has the form

$$G(x,y) = \frac{1}{2\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{4}$$

where σ is the standard deviation of the distribution, which we assume has mean zero. The Gaussian filter computes a weighted average of each pixel's neighborhood, with the average weighted towards the value of the central pixels, in contrast to the mean filter's uniformly weighted average. In Section V, we present the registration results obtained upon applying classical denoising techniques to noisy images prior to registration.

IV. Multiscale registration algorithms

A. The multiscale decomposition

The multiscale registration techniques to be discussed in this paper are based on the multiscale image representation using the hierarchical (BV, L^2) decompositions of [18]. This multiscale decomposition will provide a hierarchical expansion of an image that separates the essential features of the image (such as large shapes and edges) from the fine scales of the image (such as details and noise). The decomposition is hierarchical in the sense that it will produce a series of expansions of the image that resolve increasingly finer scales, and hence include increasing levels of detail. We will eventually apply the multiscale decomposition algorithm to the problem of image registration in the presence of noise, and will demonstrate the accuracy of the multiscale registration technique for noisy images.

We will use the following mathematical spaces in the decomposition algorithm. The space of functions of bounded variation, BV, is defined by:

$$BV = \big\{ f \quad \big| \quad ||f||_{BV} := \sup_{h \neq 0} |h|^{-1} ||f(\cdot + h) - f(\cdot)||_{L^1} < \infty \big\}.$$

We will also use the Sobolev space $W^{-1,\infty}$ with norm given by:

$$||f||_{W^{-1,\infty}} := \sup_g \left[\frac{\int f(x) g(x)\, dx}{||g||_{W^{1,1}}} \right],$$

where $||g||_{W^{1,1}} := ||\nabla g||_{L^1}$.

Define the J-functional $J(f, \lambda)$ as follows:

$$J(f,\lambda) := \inf_{u+v=f} \lambda ||v||_{L^2}^2 + ||u||_{BV}, \tag{5}$$

where $\lambda > 0$ is a scaling parameter that separates the L^2 and BV terms. This functional $J(f, \lambda)$ was introduced in the context of image processing by Rudin, Osher, and Fatemi [15]. Let $[u_\lambda, v_\lambda]$ denote the minimizer of $J(f, \lambda)$. The BV component, u_λ, captures the coarse features of the image f, while the L^2 component, v_λ, captures the finer features of f such as noise. This model is effective in denoising images while preserving edges, though it requires prior knowledge on the noise scaling λ.

Tadmor, *et al.* proposed in [18] an alternative point of view in which the minimization of $J(f, \lambda)$ is interpreted as a decomposition $f = u_\lambda + v_\lambda$, where u_λ extracts the edges of f and v_λ extracts the textures of f. This interpretation depends on the scale λ, since texture at scale λ consists of edges when viewed under a refined scale. We refer to $v_\lambda = f - u_\lambda$ as the residual of the decomposition. Upon decomposing $f = u_\lambda + v_\lambda$, we proceed to decompose v_λ as follows:

$$v_\lambda = u_{2\lambda} + v_{2\lambda},$$

where

$$[u_{2\lambda}, v_{2\lambda}] = \operatorname*{arginf}_{u+v=v_\lambda} J(v_\lambda, 2\lambda).$$

Thus we obtain a two-scale representation of f given by $f \cong u_\lambda + u_{2\lambda}$, where now $v_{2\lambda} = f - (u_\lambda + u_{2\lambda})$ is the residual. Repeating this process results in the following hierarchical multiscale decomposition of f. Starting with an initial scale $\lambda = \lambda_0$, we obtain an initial decomposition of the image f:

$$f = u_0 + v_0,$$

$$[u_0, v_0] = \operatorname*{arginf}_{u+v=f} J(f, \lambda_0).$$

We then refine this decomposition to obtain

$$v_j = u_{j+1} + v_{j+1},$$

$$[u_{j+1}, v_{j+1}] = \underset{u+v=v_j}{\text{arginf}}\ J(v_j, \lambda_0 2^{j+1}), \quad j = 0, 1, \ldots$$

After k steps of this process, we have:

$$\begin{aligned} f &= u_0 + v_0 \\ &= u_0 + u_1 + v_1 \\ &= u_0 + u_1 + u_2 + v_2 \\ &= \ldots \\ &= u_0 + u_1 + \ldots + u_k + v_k, \end{aligned} \tag{6}$$

which is a multiscale image decomposition $f \sim u_0 + u_1 + \ldots + u_k$, with a residual v_k. As k increases, the u_k components resolve edges with increasing scales $\lambda_k = \lambda_0 2^k$.

Implementation of the multi-scale decomposition:

As described in [18], the initial scale λ_0 should capture the smallest oscillatory scale in f, given by

$$\frac{1}{2\lambda_0} \leq ||f||_{W^{-1,\infty}} \leq \frac{1}{\lambda_0}. \tag{7}$$

However, in practice, we may not be able to determine the size of $||f||_{W^{-1,\infty}}$, so we determine the initial choice of λ_0 experimentally. Following [18], for the applications presented in this paper, we will use $\lambda_0 = 0.01$ and $\lambda_j = \lambda_0 2^j$.

We follow the numerical algorithm of [18] for the construction of our hierarchical decomposition. In each step, we use finite-difference discretization of the Euler-Lagrange equations associated with the $J(v_j, \lambda_{j+1})$ to obtain the next term, u_{j+1}, in the decomposition of the image f. Due to the singularity when $|\nabla u_\lambda| = 0$, we replace $J(f, \lambda)$ by the regularized functional $J^\epsilon(f, \lambda) :=$

$$\inf_{u+v=f} \{\lambda ||v||^2_{L^2} + \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2}\, dx\, dy\}, \tag{8}$$

and at each step, we find the minimizer u_λ of J^ϵ. The Euler-Lagrange equation for $J^\epsilon(f, \lambda)$ is

$$u_\lambda - \frac{1}{2\lambda} \text{div}\left(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2 + |\nabla u_\lambda|^2}}\right) = f \ \text{ in } \Omega\ ,$$

with the Neumann boundary conditions:

$$\left.\frac{\partial u_\lambda}{\partial n}\right|_{\partial\Omega} = 0, \tag{9}$$

where $\partial\Omega$ is the boundary of the domain Ω and n is the unit outward normal. We thus obtain an expansion $f \sim \sum_{j=0}^{k} u_j$, where the u_j are constructed as approximate solutions of the recursive relation given by the following elliptic PDE:

$$\begin{aligned} u_{j+1} - \frac{1}{2\lambda_{j+1}} \text{div}\left(\frac{\nabla u_{j+1}}{\sqrt{\epsilon^2 + |\nabla u_{j+1}|^2}}\right) \\ = -\frac{1}{2\lambda_j} \text{div}\left(\frac{\nabla u_j}{\sqrt{\epsilon^2 + |\nabla u_j|^2}}\right). \end{aligned} \tag{10}$$

To numerically implement the method, we cover the domain Ω with a grid $(x_i := ih, y_j := jh)$, and discretize the elliptic PDE of Eq. (10) as follows:.

$$\begin{aligned} u_{i,j} = f_{i,j} \\ + \frac{1}{2h^2}\left[\frac{u_{i+1,j} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j})^2 + (D_{0y}u_{i,j})^2}}\right] \\ - \frac{1}{2h^2}\left[\frac{u_{i,j} - u_{i-1,j}}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j})^2 + (D_{0y}u_{i-1,j})^2}}\right] \\ + \frac{1}{2h^2}\left[\frac{u_{i,j+1} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j})^2 + (D_{+y}u_{i,j})^2}}\right] \\ - \frac{1}{2h^2}\left[\frac{u_{i,j} - u_{i,j-1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1})^2 + (D_{-y}u_{i,j})^2}}\right], \end{aligned} \tag{11}$$

where D_+, D_-, and D_0 denote the forward, backward, and centered divided differences, respectively. To solve the discrete regularized Euler-Lagrange equations (11), we use the Gauss-Siedel iterative method to obtain:

$$
\begin{aligned}
u_{i,j}^{n+1} = f_{i,j} \qquad & (12)\\
&+ \frac{1}{2h^2}\left[\frac{u_{i+1,j}^{n} - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j}^{n})^2 + (D_{0y}u_{i,j}^{n})^2}}\right]\\
&- \frac{1}{2h^2}\left[\frac{u_{i,j}^{n+1} - u_{i-1,j}^{n}}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j}^{n})^2 + (D_{0y}u_{i-1,j}^{n})^2}}\right]\\
&+ \frac{1}{2h^2}\left[\frac{u_{i,j+1}^{n} - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j}^{n})^2 + (D_{+y}u_{i,j}^{n})^2}}\right]\\
&- \frac{1}{2h^2}\left[\frac{u_{i,j}^{n+1} - u_{i,j-1}^{n}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1}^{n})^2 + (D_{-y}u_{i,j}^{n})^2}}\right].
\end{aligned}
$$

To satisfy the Neumann boundary conditions (9), we first reflect f outside Ω by adding grid lines on all sides of Ω. As the initial condition, we set $u_{i,j}^0 = f_{i,j}$. We iterate this numerical scheme for $n = 0, 1, \ldots N$ until $||u^{n_\infty} - u^{n_\infty - 1}||$ is less than some preassigned value so that $u_{i,j}^{n_\infty}$ is an accurate approximation of the fixed point steady solution u_λ.

Finally, we denote the final solution $u_\lambda := \{u_{i,j}^{n_\infty}\}_{i,j}$. To obtain the hierarchical multiscale decomposition, we reiterate this process, each time updating f and λ in the following way:

$$
\begin{aligned}
f_{new} &\leftarrow f_{current} - u_\lambda,\\
\lambda_{new} &\leftarrow 2\lambda_{current}.
\end{aligned} \qquad (13)
$$

That is, at each step, we apply the $J(f_{current} - u_\lambda, 2\lambda)$ minimization to the residual $f_{current} - u_\lambda$ of the previous step. Taking $\lambda_j = \lambda_0 2^j$, we obtain after k steps a hierarchical multiscale decomposition $f = u_{\lambda_0} + u_{\lambda_1} + \ldots + u_{\lambda_k} + v_{\lambda_k}$, where we write $u_{\lambda_j} = u_j$. We call the u_j, $j = 1, 2, \ldots, k$ *the components of* f, and the v_k *the residuals*. For ease of notation, given an image f, we let $C_k(f)$ denote *the* k^{th} *scale of the image* f, $k = 1, \ldots, m$:

$$
C_k(f) = \sum_{i=0}^{k-1} u_k(f). \qquad (14)
$$

Thus $C_k(A)$ will denote the k^{th} scale of the image A, and $C_k(B)$ will denote the k^{th} scale of image B.

B. Multiscale registration

In this section, we present multiscale image registration techniques that are based on the hierarchical multiscale decomposition of [18] reviewed in Section IV-A. For the general setup, consider two images A (the fixed image) and B (the moving image), and suppose that we want to register image B with image A. Suppose that one or both of the images contains a significant amount of noise. If only one of the images is noisy, we assume that it is image B. For each of the three algorithms presented below, we first apply the multiscale hierarchical decomposition to both images. Let m denote the number of hierarchical steps used for the multiscale decompositions. We presented Algorithms I and II in the context of rigid registration in [13].

1) Algorithm I: Single-node registration: In our single-node multiscale registration algorithm, Algorithm I, we register the k^{th} scale $C_k(B)$ of image B with the image A, for $k = 1, \ldots, m$. We refer to this algorithm as a *one-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, the moving image is always the image A. That is, in the single-node algorithm, we use only the multiscale components of the fixed image B. Since this algorithm considers scales only of the noisy (moving) image, we expect that it will be particularly successful when only one of the images to be registered is noisy.

Recall that $C_1(B)$ contains only the coarsest scales of the image B, and as k increases, $C_k(B)$ contains increasing levels of detail (and hence, noise) of the image B. Thus, we expect that the registration of $C_k(B)$ with A should be more accurate than ordinary registration for the first few values of k. As k increases, however, we expect that eventually the scale $C_k(B)$ will become too noisy to register successfully.

Upon registering $C_k(B)$ with A for $k = 1, 2, \ldots, m$, we can estimate the actual transformation that maps the images to one another either

by computing a weighted average of the transformations produced by these m registrations, or simply by taking as the estimate of the actual transformation the transformation corresponding to one of the coarse scale registrations. In [13], we showed that in the case of rigid registration, using one of the coarse scale registration transformations to estimate the actual transformation is sufficient for a high degree of accuracy.

2) Algorithm II: Multi-node registration: In our multi-node multiscale registration algorithm, Algorithm II, we register the k^{th} scale of image B with the k^{th} scale of image A, for $k = 1, 2, \ldots m$. We refer to this algorithm as a *multi-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, we consider both the scales of the fixed image A and the scales of the moving image B. As before, we expect that registration of $C_k(B)$ with $C_k(A)$ should be more accurate than ordinary registration for the first few values of k. As k increases, however, we expect that eventually the scales $C_k(A)$ and $C_k(B)$ will become too noisy to register successfully. Since this algorithm considers scales of both images, we expect that it will be particularly successful when both images are noisy.

Upon registering $C_k(B)$ with $C_k(A)$ for $k = 1, 2, \ldots, m$, we estimate the actual transformation that maps the images to one another either by computing a weighted average of the transformations produced by these m registrations, or simply by using one of the coarse scale registrations $C_k(B)$ with $C_k(A)$ (for some small k), as in Algorithm I.

3) Algorithm III: Iterated multiscale registration: The accuracy and speed of convergence of Algorithms I and II can be improved by implementing an iterated multiscale registration algorithm, as follows. To iterate Algorithm I, we first register the first coarse scale $C_1(B)$ of the moving image with the fixed image A. The output of this registration process is a set of deformation parameters that represent the deformation transformation produced by the registration algorithm. We then register the second scale $C_2(B)$ of the moving image with the fixed image A, using the output deformation parameters from the first registration as the starting parameters for the second registration. We repeat this procedure until the last scale (or desired stopping scale) is reached. That is, at each stage, we use the output deformation parameters from the previous registration as the initial parameters for the current registration. The iterated form of Algorithm II is analogous, but at each stage we register the k^{th} scale $C_k(B)$ of the moving image with the k^{th} component $C_k(A)$ of the fixed image. We expect the single-node iterated multiscale registration algorithm to be particularly effective when only one of the images to be registered contains significant levels of noise, and we expect the multi-node iterated multiscale registration algorithm to be particularly effective when both of the images contain noise.

V. Results and discussion

A. Ordinary registration of a deformed image

Consider the mid-sagittal brain slice I and the deformed image S, shown in Figure 1. The mid-sagittal brain slice I is taken from the Insight Segmentation and Registration Toolkit (ITK) data repository [8]. The deformed image S is obtained by applying a known B-spline deformation to the original image I. Since the deformation transformation that maps the deformed image S to the original image I and corresponding deformation field are known, we can effectively evaluate the accuracy of various deformable registration methods by comparing the output deformation fields with the known deformation field. For all registration simulations presented in this paper, we use a B-spline deformable registration technique with a mean squares image metric and a conjugate gradient descent algorithm. However, the multiscale registration algorithms developed in this paper are independent of the registration technique used to register the images.

Using a B-spline deformable registration model, the image S is successfully registered with the image I. In Figure 2, we compare the result of the registration process, namely the image obtained upon applying the optimal deformable transformation determined by the algorithm to the deformed image, with the original image I. Ideally, both figures should be identical. Indeed, the images in Fig-

ure 2 demonstrate that the deformable registration algorithm recovers the deformation transformation.

In Figure 3, we display the exact deformation field corresponding to the deformation transformation between the images I and S (on the left) and the deformation field determined by the deformable registration algorithm, and note that visually the two deformation fields are almost identical. To quantitatively compare the exact and computed deformation fields, and for future reference, we compute the mean square difference (MSD) between the fields. For the deformation fields shown in Figure 2, the MSD is $7.2 \cdot 10^{-2}$. The minimum possible MSD between two fields is 0 (corresponding to no difference between the images), and the maximum possible MSD is 1 (corresponding to a large difference).

B. Ordinary registration of a noisy deformed image

In this section, we present the registration results obtained upon registering the noisy deformed images S_δ with the original image I for increasing noise variances δ. See Eq. (3). Figure 4 shows the noisy deformed images S_δ for $\delta = 0.1, \ldots, 0.8$. In Figure 5, we illustrate the deformation fields produced by the B-spline deformable registration algorithm upon registering the noisy deformed images S_δ with the original image I. Recall that the actual deformation is shown in Figure 3.

A visual comparison of the deformation fields presented in Figure 5 with the exact deformation field in Figure 3 indicates that the deformation registration technique fails to produce physically meaningful results for noise variance δ greater than 0.2. To quantitatively compare the deformation fields determined by the deformable registration algorithm in Figure 5 with the exact deformation field in Figure 3, we compute the mean square differences (MSD) between the computed and exact deformation fields for each speckle noise variance δ. In Table I, we present the mean square values for each noise variance δ. For reference, we also include in the first line of Table I the MSD between the deformation field produced by the registration algorithm when the deformed image contains no noise and the exact deformation field (from Section V-A).

The results presented in Figure 5 and Table I indicate that ordinary deformable registration techniques fail to produce an accurate registration result when one of the images to be registered contains significant levels of noise. As expected, the level of failure increases as the speckle noise variance δ increases. For variances greater than or equal to 0.2 the algorithm fails to produce any meaningful results.

C. Denoising results

Since ordinary deformable registration of the the noisy images fails, we study next the effect of denoising on the noisy image registration problem. We apply a mean, median, and Gaussian convolution filter, as discussed in Section III-B, to the noisy image $S_{0.6}$ and register the denoised images with the original image I. In Figure 6, we illustrate the denoised images and the deformation fields produced by registering the denoised images with the original image.

To quantitatively compare the computed deformation fields with the exact deformation, we compute the mean square differences between the deformation fields in Figure 6 and the exact deformation field in Figure 3. The mean square differences are $3.0 \cdot 10^{-1}$, $3.2 \cdot 10^{-1}$, and $4.6 \cdot 10^{-1}$, for mean, median, and Gaussian denoising techniques, respectively. These mean square differences, combined with a visual comparison of the deformation fields in Figure 6 with the exact deformation field in Figure 3, demonstrate that the application of classical denoising techniques prior to registration does not enable successful deformable registration of the noisy image $S_{0.6}$ with the original image I.

D. Multiscale registration of a noisy deformed image

Next, we demonstrate with several image registration experiments that the multiscale techniques presented in Section IV-B accurately and efficiently register the noisy deformed images considered in Sections V-B and V-C. Consider again the original image I and the noisy deformed image $S_{0.6}$, and recall that the deformation that maps the deformed

TABLE I

THE MSDS BETWEEN THE COMPUTED (USING ORDINARY REGISTRATION TECHNIQUES) AND EXACT DEFORMATION FIELDS FOR INCREASING NOISE VARIANCES δ FOR THE DEFORMATION FIELDS SHOWN IN FIGURE 5.

δ	0	0.1	0.2	0.3	0.4
MSD	$7.2 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$3.3 \cdot 10^{-1}$	$3.6 \cdot 10^{-1}$
δ	0.5	0.6	0.7	0.8	0.9
MSD	$4.3 \cdot 10^{-1}$	$4.4 \cdot 10^{-1}$	$4.5 \cdot 10^{-1}$	$4.8 \cdot 10^{-1}$	$5.7 \cdot 10^{-1}$

(moving) image to the original (fixed) image is given by the deformation field in Figure 3.

1) Single-node multiscale registration: In this section, we use Algorithm I to register the hierarchical scales of the noisy image $S_{0.6}$ with the original image I. In Figure 7, we illustrate the hierarchical multiscale decomposition of the noisy image $S_{0.6}$, and in Figure 8, we illustrate the deformation fields produced by the single-node multiscale registration algorithm.

In Table II, we calculate the MSDs between the deformation fields obtained upon registering the hierarchical scales of the noisy deformed image with the original image and the exact deformation field.

The deformation fields shown in Figure 8 and the mean square differences in Table II demonstrate that the multiscale registration method is successful for registration of the noisy deformed image $S_{0.6}$ with the original image I. Visually comparing the deformation fields in Figure 8 with the exact deformation in Figure 3, we see that the deformation fields obtained upon registering the coarse scales of the noisy deformed image with the original image are a good approximation for the actual deformation field. The MSDs indicate that the first few coarse scales produce the most accurate registration, as expected. As more detail, and hence noise, is added in increasing scales, the registration becomes less accurate for higher scales. However, it is clear both from the deformation fields in Figure 8 and the MSDs in Table II that all scale registrations of the multiscale registration technique produce very accurate results. Since both ordinary deformable registration of the noisy deformed image and the application of classical denoising algorithms prior to registration failed to produce accurate results for noise variances δ greater than 0.2, and we are able to accurately register the deformed noisy image with variance 0.6, we conclude that the multiscale single-node registration technique is a significant improvement over ordinary registration.

2) Iterated multiscale registration: In this section, we register the noisy deformed image $S_{0.6}$ with the original image I using (the single-node version of) Algorithm III. That is, we first decompose the noisy deformed image $S_{0.6}$ into 8 hierarchical scales using the multiscale hierarchical decomposition presented in Section IV-A. We then register the first scale of $S_{0.6}$ with I, and use the resulting transform parameters to register the second scale of $S_{0.6}$ with I. This process is then iterated, at the last stage using the transformation parameters obtained by registering the seventh scale of $S_{0.6}$ with I as the starting parameters for registering the eighth scale of $S_{0.6}$ with I. In Table III, we compute the MSDs between the computed deformation fields using the iterated multiscale registration algorithm with the exact deformation field, and in Figure 9, we illustrate the deformation field produced by the final iteration.

The results presented in Table III and Figure 9 demonstrate that the iterated multi-scale registration algorithm is a significant improvement over ordinary deformable registration techniques. Additionally, the results of the iterated multiscale algorithm are more accurate than those obtained via Algorithm I (single-node multiscale registration), as indicated by the MSDs in Table III. However, the main improvement of the iterated algorithm over the single-node multiscale algorithm is the improvement in computation time. Working on a

TABLE II

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELDS OBTAINED UPON REGISTERING THE HIERARCHICAL SCALES OF THE NOISY IMAGE WITH THE ORIGINAL IMAGE (USING ALGORITHM I).

Scale	1	2	3	4
MSD	$1.7 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
Scale	5	6	7	8
MSD	$1.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$1.0 \cdot 10^{-1}$

TABLE III

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELDS OBTAINED UPON REGISTERING THE NOISY DEFORMED IMAGE $S_{0.6}$ WITH THE ORIGINAL IMAGE I USING THE SINGLE-NODE ITERATED MULTISCALE REGISTRATION METHOD (ALGORITHM III).

Iteration	1	2	3	4
MSD	$1.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$6.9 \cdot 10^{-3}$	$6.4 \cdot 10^{-3}$
Iteration	5	6	7	8
MSD	$6.3 \cdot 10^{-3}$	$6.2 \cdot 10^{-3}$	$6.2 \cdot 10^{-3}$	$6.2 \cdot 10^{-3}$

Dell Dimension 8400 Intel Pentium 4 CPU (3.40 GHz, 2.00 GB of RAM), registering a single scale of the noisy deformed image $S_{0.6}$ with the original image I takes an average of 84.5 seconds. The time required for the registration process increases with each scale, as each scale contains more detail and noise than the previous scale. Thus the total time required to register each scale of the noisy deformed image $S_{0.6}$ with the original image I is approximately 676 seconds. With the iterated multiscale registration method, the initial registration of the first scale of the noisy deformed image with the first scale of the original image takes approximately 29.5 seconds, and each successive iteration takes approximately 8 seconds. Thus the total time required to complete all 8 iterations of the iterated multiscale registration algorithm is approximately 85.5 seconds. Hence we conclude that the iterated multiscale method is significantly more efficient than the single-node multiscale method. Moreover, the mean square differences in Table III indicate that accurate results can be achieved by iterating the iterated multiscale registration algorithm until the fourth or fifth iteration, which further reduces the computation time.

3) Increasing the noise variance: Finally, we demonstrate that the iterated multiscale registration algorithm produces accurate results for noise variances δ significantly greater than those at which ordinary deformable registration fails. In Figure 10, we illustrate the noisy deformed images S_δ for very large values of the noise variance δ ($\delta = 1, \ldots, 6$), and in Figure 11, we illustrate the deformation fields computed using the single-node iterated multiscale registration algorithm (Algorithm III) to register the noisy deformed images S_δ with the original image I for each δ illustrated in Figure 10. In Table IV, we illustrate the MSDs between the computed and exact deformation fields. These results demonstrate that the iterated multiscale registration algorithm accurately registers the noisy deformed image with the original image for noise variances that are significantly greater than those at which ordinary registration fails. Recall from Section V-B that ordinary deformable registration of a noisy deformed image with a non-noisy fixed image fails for noise variances δ greater than 0.2. In Figure 11 and Table IV, we demonstrate that the iterated multiscale registration algorithm produces accurate results for noise variances δ as large as 6. Moreover, the iterated multiscale registration algorithm is more

accurate than ordinary deformable registration when the images contain no added speckle noise. Recall that the MSD between the exact and computed deformation field in Section V-A obtained upon registering the non-noisy deformed image S with the non-noisy original image I using B-splines deformable registration is $7.2 \cdot 10^{-3}$. Using the iterated multiscale registration algorithm, the MSD between the computed and exact deformations fields is $1.8 \cdot 10^{-3}$. Thus the iterated multiscale algorithm improves the accuracy of deformable registration even when the images do not contain added noise.

E. Registration of a noisy deformed image with a noisy fixed image

In this section, we consider the case in which both images to be registered contain significant levels of noise. We add speckle noise of variance 0.6 to the original image I, and denote this noisy image $I_{0.6}$. Our goal is to register the noisy deformed image $S_{0.6}$ with the noisy fixed image $I_{0.6}$. In Figure 12, we illustrate both of the noisy images, as well as the deformation field produced upon registering the noisy deformed image $S_{0.6}$ with the noisy original image $I_{0.6}$ using ordinary B-splines deformable registration techniques.

A visual comparison of the computed deformation field in Figure 12 with the exact deformation field in Figure 3 indicates that ordinary deformable registration of the noisy images fails. The MSD in this case is $4.4 \cdot 10^{-1}$.

1) Multi-node registration: Since ordinary deformable registration of the noisy images fails, we register the images using our multiscale multi-node algorithm (Algorithm II). That is, we apply the multiscale decomposition to both of the images $I_{0.6}$ and $S_{0.6}$ and register the k-th scale of $S_{0.6}$ with the k-th scale of $I_{0.6}$ for $k = 1, 2, \ldots, 8$ (we use $m = 8$ hierarchical steps in the multiscale decomposition of each image). The deformation fields obtained upon registering the scales of $S_{0.6}$ with the scales of $I_{0.6}$ are illustrated in Figure 13. A visual comparison of the deformation fields in Figure 13 with the exact deformation field in Figure 3 indicates that the multi-node registration algorithm is successful. The corresponding MSDs are shown in Table V.

2) Iterated multiscale registration: In this section, we register the noisy deformed image $S_{0.6}$ with the noisy original image $I_{0.6}$ using (the multi-node version of) Algorithm III. That is, we first decompose the noisy deformed image $S_{0.6}$ and the noisy original image $I_{0.6}$ into 8 hierarchical scales each, using the multiscale hierarchical decomposition presented in Section IV-A. We then register the first scale of $S_{0.6}$ with the first scale of $I_{0.6}$, and use the resulting transform parameters to register the second scale of $S_{0.6}$ with the second scale of $I_{0.6}$. This process is then iterated, at the last stage using the transformation parameters obtained by registering the seventh scale of $S_{0.6}$ with the seventh scale of $I_{0.6}$ as the starting parameters for registering the eighth scale of $S_{0.6}$ with the eighth scale of $I_{0.6}$. In Table VI, we compute the MSDs between the computed deformation fields using the iterated multiscale registration algorithm with the exact deformation field, and in Figure 14, we illustrate the deformation field produced by the final iteration.

The results presented in Table VI and Figure 14 demonstrate that the iterated multi-scale registration algorithm is a significant improvement over ordinary deformable registration techniques. Additionally, the results of the iterated multiscale algorithm are more accurate than those obtained via Algorithm II (multi-node multiscale registration), as indicated by the MSDs in Table VI. Additionally, the iterated algorithm is computationally more efficient than Algorithm II.

3) Increasing the noise variance: Finally, we demonstrate as in Section V-D.3 that the iterated multiscale registration algorithm produces accurate results when both of the images contain speckle noise of variance significantly greater than the level at which ordinary deformable registration fails. In Figure 15, we illustrate the deformation fields computed using the iterated multiscale registration algorithm to register the noisy deformed image S_δ with the noisy original image I_δ for increasing noise variances δ. In Table VII, we illustrate the mean square differences between the computed and

TABLE IV

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELDS OBTAINED UPON REGISTERING THE NOISY DEFORMED IMAGES S_δ WITH THE ORIGINAL IMAGE I FOR INCREASING NOISE VARIANCES δ USING THE SINGLE-NODE ITERATED MULTISCALE REGISTRATION METHOD (ALGORITHM III).

δ	0	0.4	0.8	1
MSD	$1.8 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$9.9 \cdot 10^{-3}$	$7.2 \cdot 10^{-2}$
δ	2	3	4	6
MSD	$1.4 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$

TABLE V

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELD OBTAINED UPON REGISTERING THE HIERARCHICAL SCALES OF THE NOISY DEFORMED IMAGE $S_{0.6}$ WITH THE HIERARCHICAL SCALES OF THE NOISY ORIGINAL IMAGE $I_{0.6}$ (USING ALGORITHM II).

Scale	1	2	3	4
MSD	$1.2 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
Scale	5	6	7	8
MSD	$1.2 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$

TABLE VI

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELD OBTAINED UPON REGISTERING THE NOISY DEFORMED IMAGE $S_{0.6}$ WITH THE NOISY ORIGINAL IMAGE $I_{0.6}$ USING THE MULTI-NODE ITERATED MULTISCALE REGISTRATION METHOD (ALGORITHM III).

Iteration	1	2	3	4
MSD	$1.3 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$
Iteration	5	6	7	8
MSD	$1.0 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-3}$

exact deformation fields. These results demonstrate that the iterated multiscale registration algorithm accurately registers the noisy deformed image with the noisy original image for noise variances significantly greater than those at which ordinary techniques fail; recall that ordinary deformable registration failed when *only one* of the images to be registered contain noise of variance 0.2. In Figure 16, we illustrate the noisy original and deformed images I_2 and S_2. These images contain speckle noise with variance $\delta = 2$. As demonstrated by the deformation field in Figure 15, the multi-node iterated multiscale registration algorithm (Algorithm III) accurately registers these very noisy images.

F. Extension to three-dimensional images

Finally, we demonstrate that the the multiscale registration technique accurately registers *three-dimensional* images. To study the three-dimensional (3D) registration problem, we use four-dimensional computed tomography (4D CT) images acquired with a GE Discovery-ST Scanner (GE Medical Systems, Miluakee, WI) at the Stanford University Medical Center. Four-dimensional computed tomography techniques allows one to acquire image data at specified phases over several respiratory cycles, and then combines the data into three-dimensional phase-binned images. We obtained eight phase bins (i.e. eight 3D CT images) corresponding to eight different breathing phases of

TABLE VII

THE MSDS BETWEEN THE COMPUTED AND EXACT DEFORMATION FIELDS OBTAINED UPON REGISTERING THE NOISY DEFORMED IMAGES S_δ WITH THE NOISY IMAGE I_δ FOR INCREASING NOISE VARIANCES δ USING THE MULTI-NODE ITERATED MULTISCALE REGISTRATION METHOD (ALGORITHM III).

δ	0	0.4	0.8	1	1.5	2
MSD	$1.8 \cdot 10^{-3}$	$9.1 \cdot 10^{-3}$	$3.7 \cdot 10^{-2}$	$4.4 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$9.1 \cdot 10^{-2}$

the respiratory cycle. Each phase consists of 80 two-dimensional images, or *slices*, which are combined to obtain the 3D images. The slice thickness for each phase is 2.5-mm, and the eight breathing phases recorded contain approximately 400 MB of data in DICOM image format.

In Figure 17, we illustrate two corresponding sample slices (slice 25) from the first and eighth phases of the 4D CT data set.

To register the 3D CT images with one another (i.e. to register each phase of the respiratory cycle with, for example, the inhale phase), we first extend the hierarchical multiscale decomposition of [18] to 3D images. Although the multiscale decomposition presented in [18] was done in two dimensions only, the hierarchical multiscale expansion in Eq. (6) is independent of the image dimensionality. To implement the iterated multiscale decomposition in 3 dimensions, we cover the image domain Ω with a grid $(x_i := ih, y_j := jh, z_k := kh)$, and let D_+, D_-, and D_0 denote the forward, backward, and centered divided differences, respectively. Then the 3D extension of the iterated multiscale decomposition given by Eq. (12) in Section IV-A is:

$$
\begin{aligned}
u_{i,j,k}^{n+1} = f_{i,k,j} & \\
&+ \frac{1}{2h^2}\left[\frac{u_{i+1,j,k}^{n} - u_{i,j,k}^{n+1}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j,k}^{n})^2 + (D_{0y}u_{i,j,k}^{n})^2 + (D_{0z}u_{i,j,k}^{n})^2}}\right] \\
&- \frac{1}{2h^2}\left[\frac{u_{i,j,k}^{n+1} - u_{i-1,j,k}^{n}}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j,k}^{n})^2 + (D_{0y}u_{i-1,j,k}^{n})^2 + (D_{0z}u_{i-1,j,k}^{n})^2}}\right] \\
&+ \frac{1}{2h^2}\left[\frac{u_{i,j+1,k}^{n} - u_{i,j,k}^{n+1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j,k}^{n})^2 + (D_{+y}u_{i,j,k}^{n})^2 + (D_{0z}u_{i,j,k}^{n})^2}}\right] \\
&- \frac{1}{2h^2}\left[\frac{u_{i,j,k}^{n+1} - u_{i,j-1,k}^{n}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1,k}^{n})^2 + (D_{-y}u_{i,j,k}^{n})^2 + (D_{0z}u_{i,j-1,k}^{n})^2}}\right] \\
&+ \frac{1}{2h^2}\left[\frac{u_{i,j,k+1}^{n} - u_{i,j,k}^{n+1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j,k}^{n})^2 + (D_{0y}u_{i,j,k}^{n})^2 + (D_{+z}u_{i,j,k}^{n})^2}}\right] \\
&- \frac{1}{2h^2}\left[\frac{u_{i,j,k}^{n+1} - u_{i,j,k-1}^{n}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j,k-1}^{n})^2 + (D_{0y}u_{i,j,k-1}^{n})^2 + (D_{-z}u_{i,j,k}^{n})^2}}\right].
\end{aligned}
\tag{15}
$$

Upon decomposing each 3D image, we use the iterated multi-node multiscale registration algorithm (Algorithm III) to register each scale of the moving 3D image with each scale of the fixed 3D image. In Figure 18, we compare the voxel-wise intensity difference between the two sample slices shown in Figure 17 before and after iterated multiscale deformable registration. In these images, black represents exact intensity agreement and brighter regions indicate intensity disagreement. A comparison of the intensity differences before and after registration demonstrates that the iterated multiscale registration method indeed recovers the difference between the two images. Similar results are obtained with all other slices and phases. Working on a Dell Dimension 8400 Intel Pentium 4 CPU (3.40 GHz, 2.00 GB of RAM), the total required computation time for both the 3D multiscale decomposition and iterated multiscale multi-node registration algorithm is on the order of approximately 15-30 minutes, depending on the data set.

VI. Conclusions

While there are many existing deformable registration techniques, common approaches are shown to fail when one or more of the images to be registered contains even moderate levels of noise. Further, for high levels of noise, image registration fails even when classical denoising algorithms are applied to the images before registration. We have presented deformable image registration techniques based on the hierarchical multiscale image decomposition of [18]. The multiscale decomposition of an image results in a hierarchical representation that separates the coarse and fine scales of the image. We presented three separate multiscale registration algorithms based on this decomposition. In the first, we follow a single-node multiscale registration strategy in which we register the scales of the moving image with the fixed image, and use a weighted average to estimate the actual deformation between the images. In the second, we use a multi-node multiscale registration method in which we register the scales of the moving image with the scales of the fixed image, and use a weighted average to estimate the actual deformation between the images. Finally, in the third algorithm, we follow an iterated multiscale registration strategy. Using images in which the precise deformation between the fixed and moving images is known, we have shown that the multiscale registration algorithms are indeed accurate for levels of noise much higher than the noise levels at which ordinary deformable registration techniques fail. Moreover, the iterated multiscale registration technique significantly reduces the computation time necessary to obtain accurate registration of noisy images. We have also demonstrated that the multiscale techniques improve the accuracy of deformable registration even when the images to be registered do not contain additional noise. Additionally, we have extended the hierarchical multiscale decomposition of [18] to three-dimensional images, and have demonstrated the accuracy and efficiency of our multiscale registration techniques for the registration of three-dimensional images. Finally, we would like to emphasize that using the multiscale decomposition is independent of the registration method and of the noise model.

APPENDIX I
FIGURES

ACKNOWLEDGMENT

The work of D. Levy was supported in part by the National Science Foundation under Career Grant No. DMS-0133511. The work of L. Xing was supported in part by the Department of Defense under Grant No. PC040282 and the National Cancer Institute under Grant No. 5R01 CA98523-01.

REFERENCES

[1] A. ACHIM, A. BEZERIANOS, AND P. TSAKALIDES, *Novel Bayesian Multiscale Method for Speckle Removal in Medical Ultrasound Images*, IEEE Transactions on Medical Imaging, **20**, no. 8, pp. 772–783, 2001.

[2] R. BAJCSY, ET AL., *A computerized system for the elastic matching of deformed radiographic images to idealized atlas images*, Journal of Computer Assisted Tomography, **7**, no. 4, pp. 618–625, 1983.

[3] F. L. BOOKSTEIN, *Principal Warps: Thin-Plate Splines and the Decomposition of Deformations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **11**, no. 6, pp. 567–585, 1989.

[4] L. BROWN, *A Survey of Image Registration Techniques*, ACM Computing Surveys, **24**, no. 4, pp. 325–376, 1992.

[5] G. E. CHRISTENSEN, R. D. RABBITT, AND M. I. MILLER, *Deformable templates using large deformation kinematics*, IEEE Transactions on Image Processing, **5**, pp. 1435–1447, 1996.

[6] W. R. CRUM, T. HARTKENS, AND D. L. G. HILL, *Non-rigid image registration: theory and practice*, The British Journal of Radiology, **77**, pp. 140–153, 2004.

[7] M. FERRANT, A. NAVAVI, B. MACQ, P. M. BLACK, F. A. JOLESZ, R. KIKINIS, ET AL., *Serial registration of intraoperative MR images of the brain*, Medical Image Analysis, **6**, pp. 337–359, 2002.

[8] *Insight Segmentation and Registration Toolkit (ITK)*, http://www.itk.org.

[9] S. LEE, G. WOLBERG, K.-Y. CHWA, AND S. Y. SHIN, *Image metamorphosis with scattered feature constraints*, IEEE Transactions on Visualization and Computer Graphics, **2**, pp. 337–354, 1996.

[10] S. LEE, G. WOLBERG, AND S. Y. SHIN, *Scattered data interpolation with multilevel B-splines*, IEEE Transactions on Visualization and Computer Graphics, **3**, pp. 228–244, 1997.

[11] J. LIAN, L. XING, S. HUNJAN, *et al.*, *Integrating deformable MRI/MSRI and CT image registration into the prostate IMRT treatment planning*, Medical Physics, **31**, no. 11, pp.3087-3094, 2004.

[12] J. MODERSITZKI, *Numerical Methods for Image Registration*, Oxford, 2004.

[13] D. PAQUIN, D. LEVY, E. SCHREIBMANN, AND L. XING, *Multiscale image registration*, Mathematical Biosciences and Engineering, **3**, no. 2, pp. 389–418, 2006.

[14] D. W. ROBERTS, A. HARTOV, F.E. KENNEDY, M.I. MIGA, AND K.D. PAULSEN, *Intraoperative brain shift and deformation: A quantitative analysis of cortical displacement in 28 cases*, Neurosurgery, **43**, no. 4, pp. 749–760, 1998.

[15] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, **60**, pp. 259–268, 1992.

[16] E. SCHREIBMANN, G. T.Y. CHEN, AND L. XING, *Image interpolation in 4D CT using a bspline deformable registration model*, International Journal of Radiation Oncology, Biology, Physics, **64**, no. 5, pp. 1537–1550.

[17] E. SCHREIBMANN, AND L. XING, *Narrow band deformable registration of prostate MRI/MRSI and CT studies*, International Journal of Radiation Oncology, Biology, Physics, **62**, no. 2, pp. 595–605, 2005.

[18] E. TADMOR, S. NEZZAR, AND L. VESE, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Modeling and Simulations, **2**, no. 4, pp. 554–579, 2004.

[19] A. TEI, A. BHARATHA, M. FERRANT, P. MCL. BLACK, F.A. JOLESZ, R. KIKINIS, AND S. WARFIELD, *Tracking volumetric brain deformation during image guided neurosurgery*, VISIM: Information Retrieval and Exploration in Large Medical Image Collections, Springer Verlag, 2001.

[20] A. R. WEEKS, *Fundamentals of Electronic Image Processing*, SPIE Optical Engineering Press and IEEE Press, 1996.

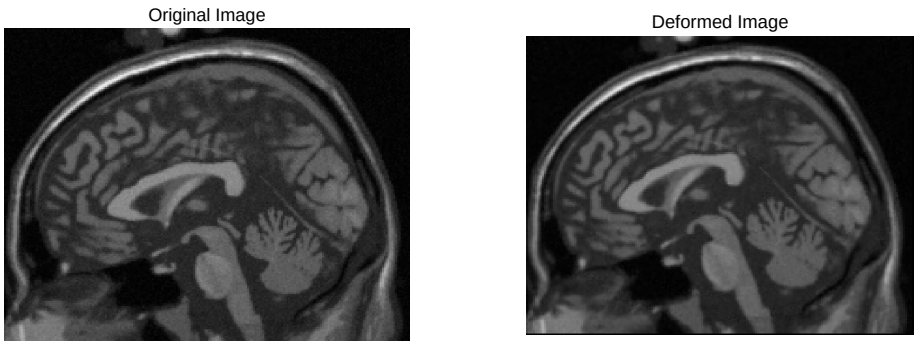


Fig. 1. The mid-sagittal brain slice I (shown on the left) and the deformed image S (shown on the right).

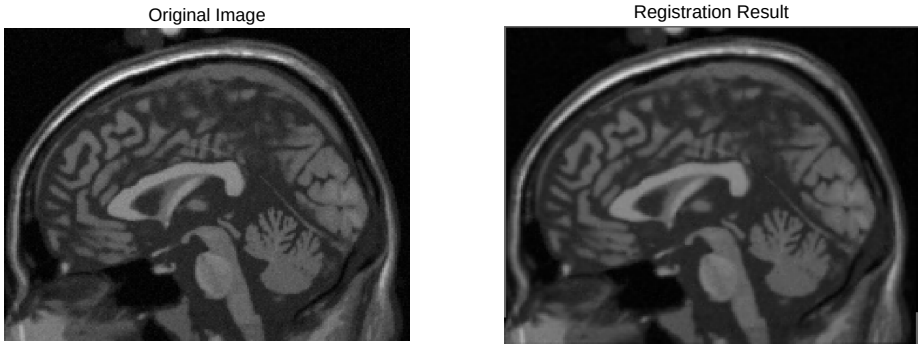


Fig. 2. The result (shown on the right) upon registering the deformed image S with the original image I (shown on the left).

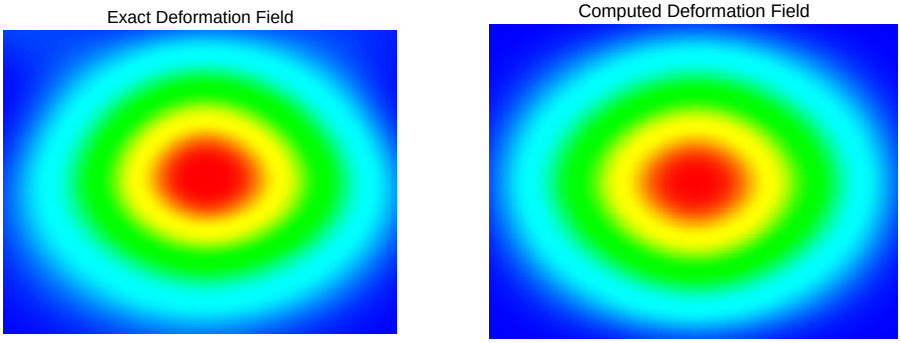


Fig. 3. The exact deformation field corresponding the the deformation transformation between I and S (shown on the left) and the deformation field produced by the registration algorithm upon registering the deformed image S with the original image I (shown on the right).

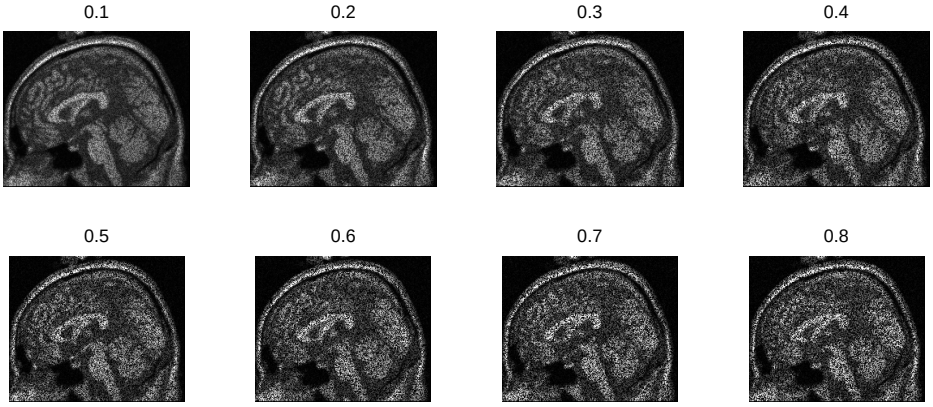


Fig. 4. The noisy images S_δ, for increasing values of δ.

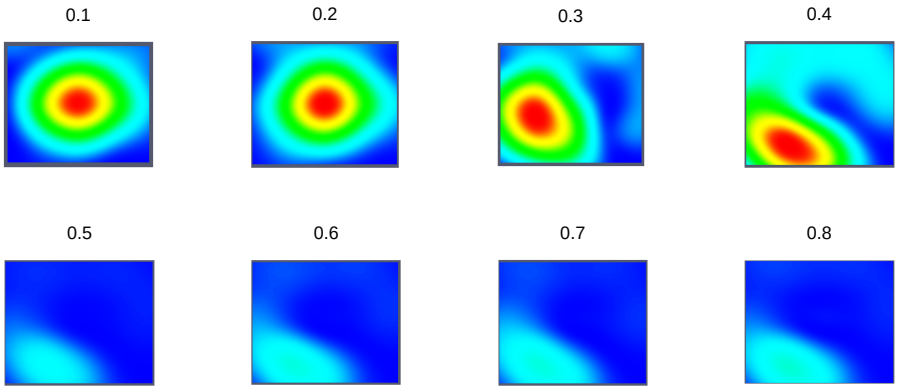


Fig. 5. The deformation fields produced by the registration algorithm upon registering the noisy deformed images S_δ with the original image I, for increasing values of δ.

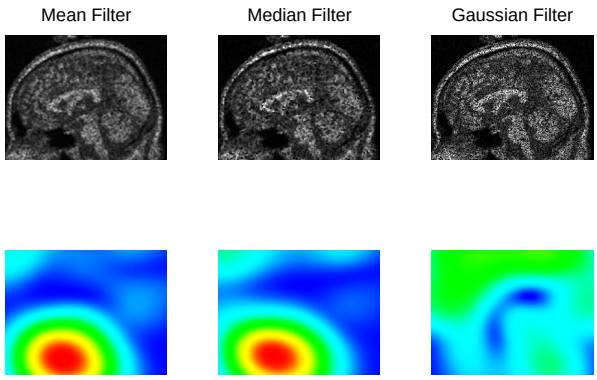


Fig. 6. The denoised images and registration results obtained upon applying mean, median, and Gaussian filters prior to registering the noisy deformed image $S_{0.6}$ with the original image I.

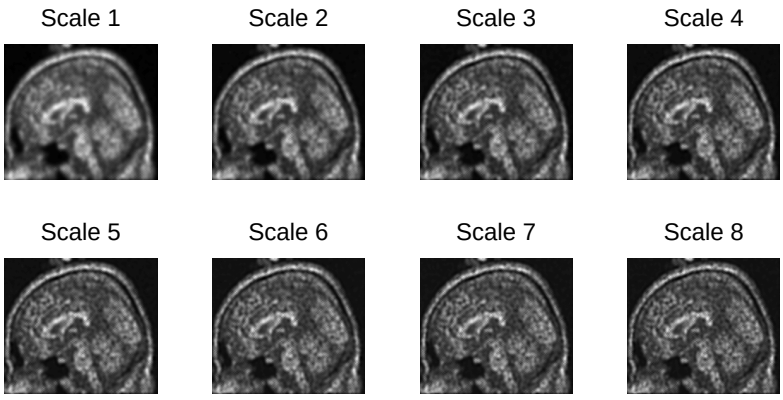


Fig. 7. The multiscale decomposition of the noisy deformed image $S_{0.6}$.

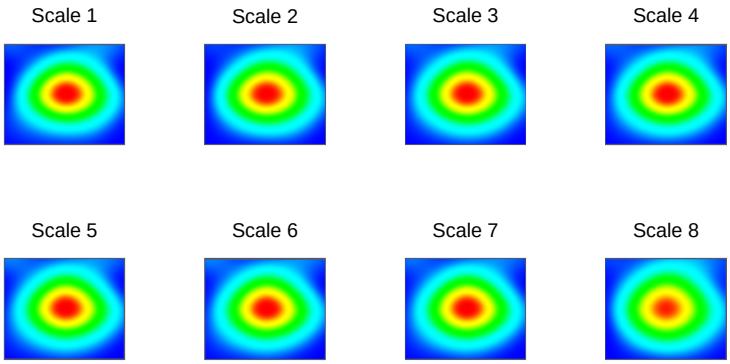


Fig. 8. The deformation fields obtained upon registering the hierarchical scales of the noisy image $S_{0.6}$ with the original image using Algorithm I.

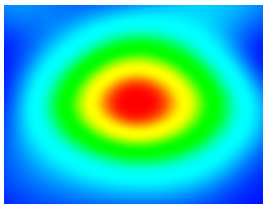


Fig. 9. The deformation field obtained upon registering the noisy deformed image $S_{0.6}$ with the original image using the single-node iterated multiscale registration method (Algorithm III).

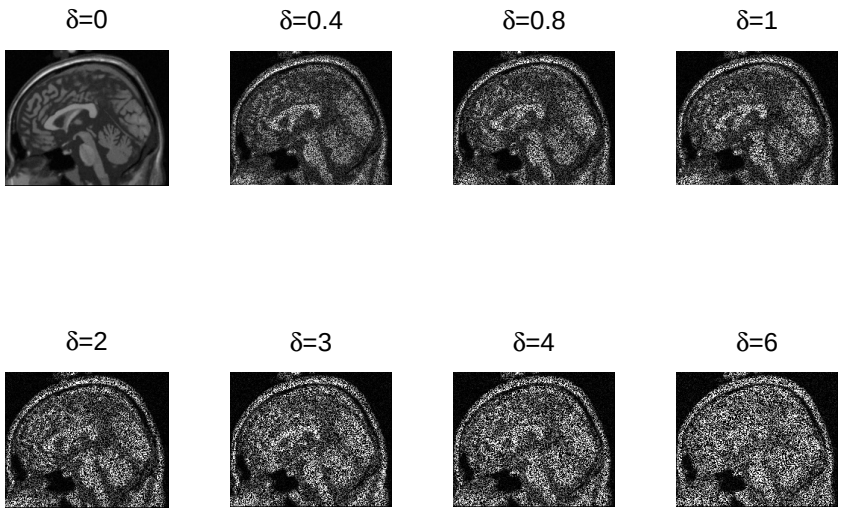


Fig. 10. The noisy deformed images S_δ for increasing noise variances δ.

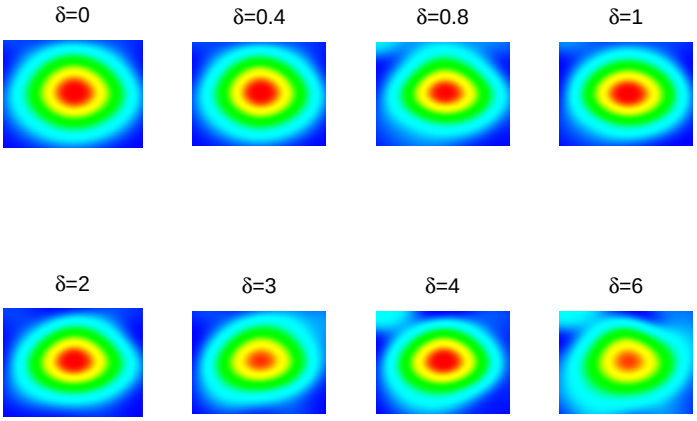


Fig. 11. The deformation fields obtained upon registering the noisy deformed image S_δ with the original image I using Algorithm III for increasing noise variances δ.

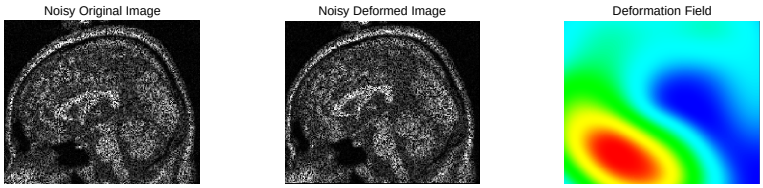


Fig. 12. The noisy mid-sagittal brain slice $I_{0.6}$ (shown on the left), the noisy deformed image $S_{0.6}$ (shown in the center), and the deformation field (shown on the right) produced upon registering $S_{0.6}$ with $I_{0.6}$ using ordinary deformable registration techniques.

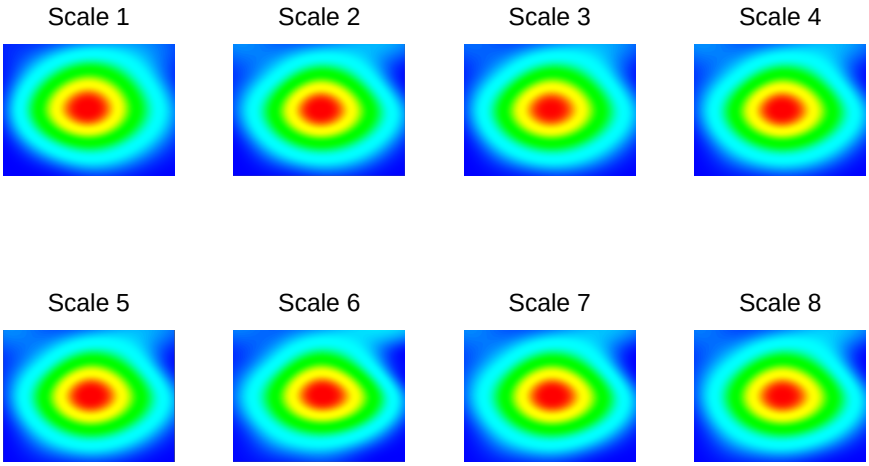


Fig. 13. The deformation fields obtained upon registering the noisy deformed image with the noisy fixed image using the multi-node registration algorithm (Algorithm II).

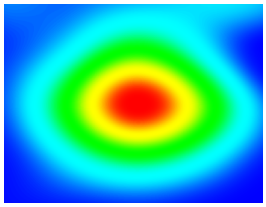


Fig. 14. The deformation field obtained upon registering the noisy deformed image $S_{0.6}$ with the noisy original image $I_{0.6}$ using the multi-node iterated multiscale registration method (Algorithm III).

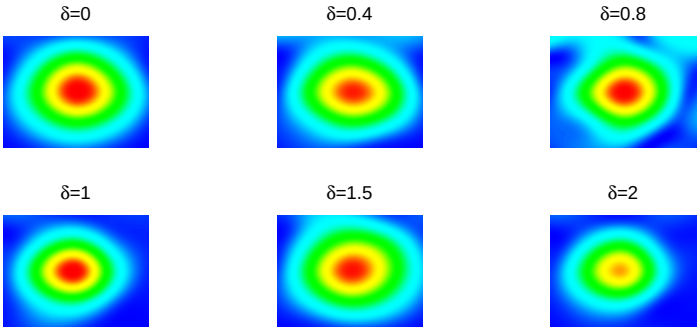


Fig. 15. The deformation fields obtained upon registering the noisy deformed image S_δ with the noisy original image I_δ using the multi-node iterated multiscale algorithm (Algorithm III) for increasing noise variances δ.

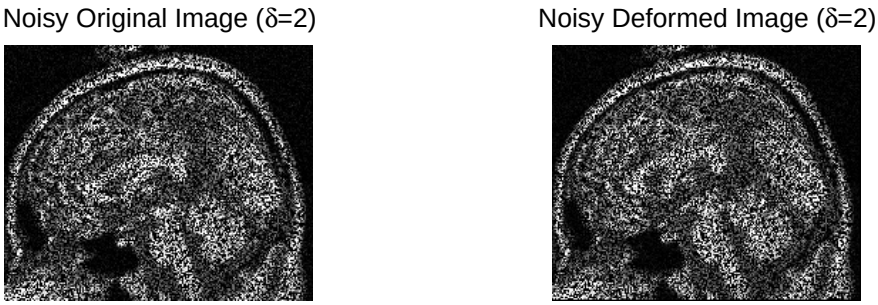


Fig. 16. The noisy original and deformed images I_2 and S_2.

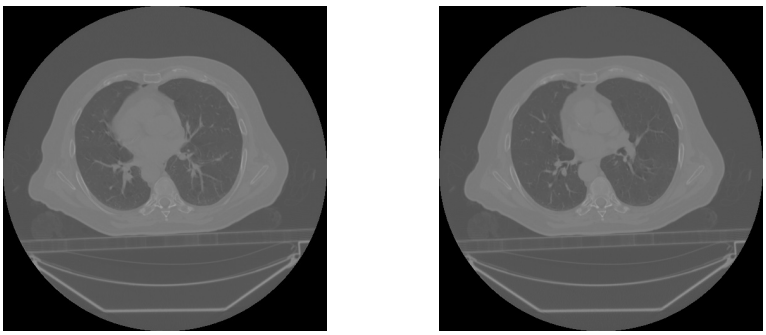


Fig. 17. Two corresponding sample slices from two breathing phases of the same patient.

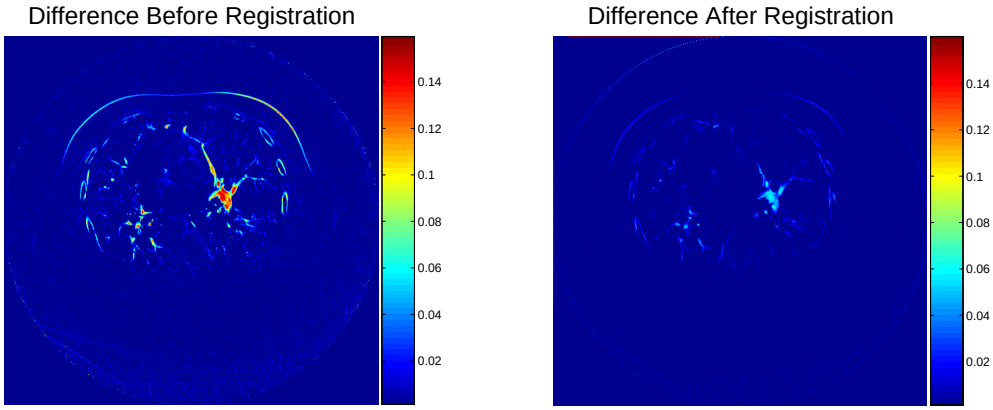


Fig. 18. The voxel-wise intensity difference between two corresponding slices before and after iterated multiscale deformable registration.

Second Pages

Chapter 2

Physics of IMRT

Lei Xing, PhD, Qiuwen Wu, PhD, Yong Yang, PhD, Arthur Boyer, PhD

Radiation therapy (RT) as a means of managing cancer has its roots in the discipline of radiology. From the time Roentgen first discovered x-rays, two-dimensional transmission images of the human body provided unprecedented imagery of bony landmarks, allowing radiologists to deduce the location of internal organs. Using planar radiographs, radiologists planned cancer treatments by collimating rectangular fields encompassing the presumed tumor location. Additional blocks placed daily to match marks on the patient's skin and later the use of low-temperature melting dense alloys provided a cookie-cutter approach to treating the two-dimensional projections of the tumor volumes.

Human anatomy and tumor shapes, however, are inherently three-dimensional. By treating a large amount of nearby normal tissue, physicians were limited by the tolerance of the normal tissue they were treating. Additionally, it was not possible to take the three-dimensional structures into consideration because of the limitations of early dose calculations. The advantage of being able to treat a tumor target conformally can be appreciated by a simple example. Assume that the tumor is a sphere of 5 cm in diameter; it would have a volume of 65.4 cc. If one irradiates it with square fields, directed at the six faces of the cube containing the sphere (an anatomic impossibility that we will allow for the sake of making a theoretical point), a high-dose volume would be created within the sphere containing 125 cc. This represents the three-dimensional nonconformal situation. If one were to treat the volume with circular fields, directed toward the sphere from all directions (which, again, is anatomically impossible), the high dose would be limited to the sphere itself. Approximately 60 cc of normal tissue would be spared. The reduction of tissue irradiated is a factor of $6/\pi$ or about half. This reduction in normal tissue irradiation should theoretically improve the therapeutic ratio and allow the tumor target volume to be treated to a higher dose, thereby improving the probability of tumor control. Other factors play critical roles as well. Tumor biology has a great deal to do with the actual tumor control achieved, but the basic idea of reducing normal tissue irradiation is a valid strategy and the goal for managing local tumor control with a minimum of normal tissue complications. The details on radiobiology are discussed in Chapter 3 "Radiobiology of IMRT."

Three-dimensional conformal radiation therapy (3DCRT) is a method of irradiating target volume defined in a three-dimensional anatomic image of the patient with a set of x-ray beams individually shaped to conform the two-dimensional beam's eye view projection of the target. 3DCRT became feasible with the development of computed tomography (CT). The development of spiral and multislice CT scanners has made the acquisition of large data sets practical. The reconstructed images, acquired with patients in the treatment position, provide a model on which geometric and dosimetric computations can be applied. These data sets can be acquired with spiral scanners capable of recording the transmission data needed to reconstruct 50 to 100 transverse image planes spaced 2 to 5 mm apart. Given adequate immobilization devices to help patients achieve and hold their treatment position for the duration of the image acquisition, these fast scanners provide excellent data sets that can be used for treatment planning. The transmission data are used to reconstruct a three-dimensional data set consisting of Hounsfield numbers associated with voxels. The development of the Digital Imaging and Communication in Medicine (DICOM) standard and its various extensions for data exchange has made possible the use of CT data sets acquired with the equipment from one vendor with treatment planning systems from another vendor and the ability to treat patients with equipment from yet another vendor. The transfer of these data over computer networks has improved the efficiency and accuracy of the entire treatment planning and delivery process.

Evolution from 3DCRT to IMRT

Intensity-modulated radiation therapy (IMRT) emerged in clinical practice as a result of the development of 3DCRT

in the 1980s. Although the exact beginning of the modality depends on one's definition of IMRT, it is generally agreed on that the widespread implementation and realization of the technique occurred in the United States in the early 1990s with the commercially available Peacock IMRT planning system and MIMiC fan beam delivery device (North American Scientific, NOMOS Radiation Oncology Division, Cranberry Township, PA).[1,2] This was then followed by the cone beam multileaf collimator (MLC)-based IMRT in the mid-1990s. MLC allows the rapid and controllable adjustment of field aperture and is thus ideally suited for dynamic radiation beam modulation. In Figure 2-1, different IMRT modalities currently available or under intense investigation are summarized. Physically, a common feature of these IMRT techniques is that they all attempt to enhance control over the three-dimensional dose distribution through the superposition of a large number of independent segmented fields from either a number of fixed directions or from directions distributed on one or multiple arcs.

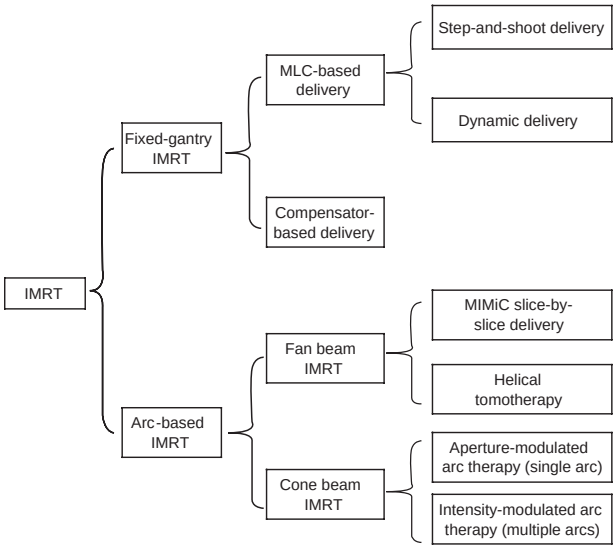


FIGURE 2-1. Currently available intensity-modulated radiation therapy (IMRT) techniques. MLC = multileaf collimator.

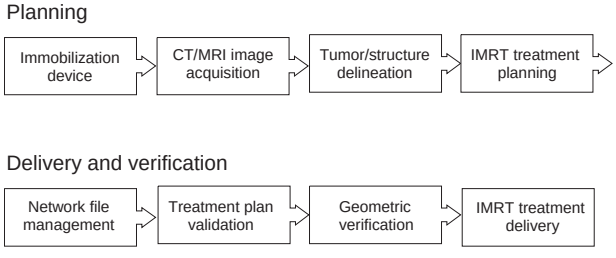


FIGURE 2-2. The intensity-modulated radiation therapy (IMRT) treatment process. CT = computed tomography; MRI = magnetic resonance imaging.

Intensity modulation adds a new degree of freedom to RT planning and provides a more effective means to produce tightly conformal dose distributions in complex treatment situations. The objective of this chapter is to provide an overall comprehension of IMRT and to review the physics aspect of this technology. In Figure 2-2, the overall treatment process of IMRT is illustrated. The key steps involved in the process are discussed in separate sections. In the remainder of this introductory section, we briefly describe the IMRT delivery modes listed in Figure 2-1. Given that fixed-gantry IMRT is by far the most popularly implemented technique, emphasis is given to this mode first.

Fixed-Gantry IMRT

Fixed-gantry IMRT is similar to 3DCRT in that a number of fixed beam directions are used (Figure 2-3A). In this mode, treatment planning is generally done in two steps. First, the dose optimization engine generates a set of intensity profiles, one for each incident beam. Depending on the treatment planning system, the optimized beam profile can be continuous or in a form that is discretized in space and intensity. Without loss of generality, an incident beam is assumed to be already divided into a grid of beamlets, and each beamlet can take a fixed number of intensity levels. The beamlet width (dimension perpendicular to the leaf travel direction) is limited to the MLC leaf width. The beamlet length, or the step size of MLC leaf movement defined as the smallest step in the leaf travel direction, is a parameter specified by the user. A smaller beamlet size or a larger number of intensity levels offers better spatial or intensity

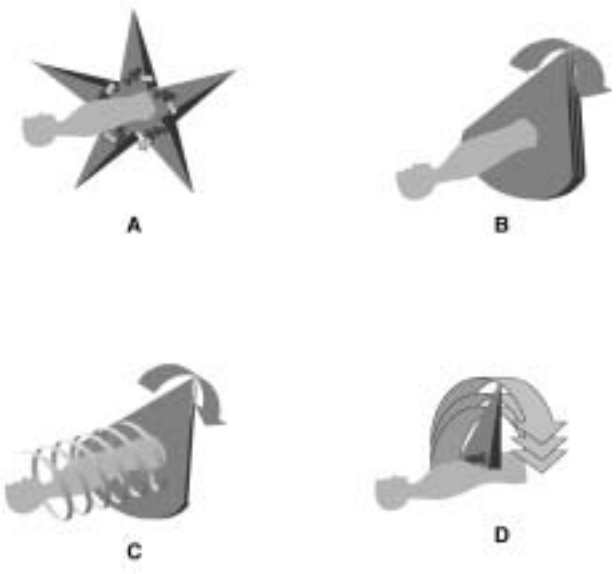


FIGURE 2-3. A schematic drawing of (*A*) fixed-gantry intensity-modulated radiation therapy (IMRT); (*B*) slice-by-slice fan beam delivery; (*C*) tomotherapy delivery; and (*D*) cone beam–based IMRT.

resolution but requires more MLC segments for delivery. Typically, the size of the beamlet and the number of intensity levels in current IMRT treatment are set to 1 × 1 cm and 10, respectively. Figure 2-4 shows an example of an intensity map for a head and neck IMRT treatment, obtained using the *CORVUS* (North American Scientific)) inverse planning system. Occasionally, the beamlet size or the number of intensity levels is varied to meet a specific clinical requirement.

There are many ways to produce a desired fluence map. Conceptually, physical compensators are the most straightforward. The most popular delivery technique is, however, based on computer-controlled MLC. In this approach, an intensity map is decomposed into a set of MLC-formed apertures by using a leaf sequencing algorithm. The MLC sequences are recorded in a computer file, which is then used to control the MLC movement for plan delivery. It is important to note that an intensity map, regardless of its shape, can always be expressed as a superposition of a number of segmented fields (for a given intensity map, generally, a number of ways exist for this decomposition, leading to numerous leaf sequencing algorithms). Depending on the relationship between MLC leaf movements and radiation dose delivery, the delivery can generally be divided into step-and-shoot delivery and dynamic modes. The former is the simplest computer-controlled delivery scheme of the fixed-gantry IMRT, in which MLC leaf movements and dose deliveries are done at different instances. A leaf sequence file consists of alternatives of dose-only and motion-only instances. Dynamic delivery differs from a step-and-shoot mode in that leaf movement and dose delivery are realized simultaneously.

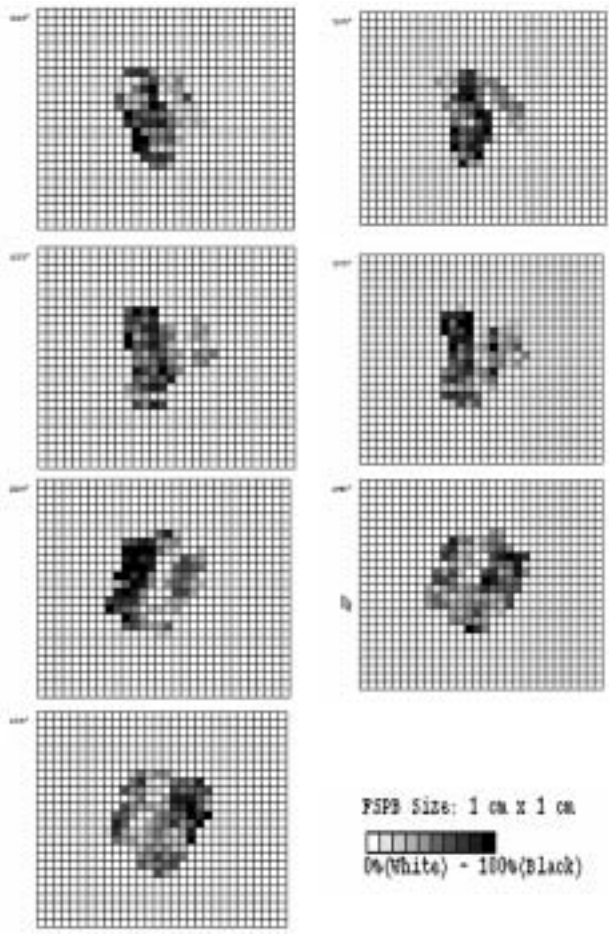


FIGURE 2-4. Intensity patterns of a seven-field intensity-modulated radiation therapy (IMRT) head and neck treatment obtained using the *CORVUS* IMRT planning system.

Arc-Based IMRT

Arc-based treatment delivery has a long history in RT. An early implementation of this method was the so-called Takahashi arc, in which the beam aperture dynamically follows the beam's eye view projection of the target. Stereotactic radiosurgery based on cylindric cones or micro-MLC often uses the arc delivery technique to "spread" the radiation dose to different regions of the brain to avoid overdosing the normal brain tissue. Conformal arc therapy can produce excellent dose conformation to a simple target. However, the target volumes often exhibit significant deviation from the ideal spherical or ellipsoidal shape. In this case, arc-based IMRT treatment, which was first proposed by D[illegible][4], provides a viable option to improve the dose distr[illegible]ions through intensity modulation. The three different forms of arc-based IMRT deliveries are schematically shown in Figure 2-3, B to D, and their features are summarized below.

Fan Beam IMRT

A schematic drawing of the fan beam IMRT is shown in Figure 2-3B. The delivery is realized on a slice-by-slice manner, in which each slice covers 2 to 4 cm in the longitudinal direction and 20 cm in diameter. North American Scientific's Peacock system, which includes the PEACOCK inverse treatment planning system and the MIMiC collimator, is used for this type of treatment. The planning system uses 54 equally spaced beams and optimizes the beamlet maps of each beam. The nominal beamlet sizes on the isocenter plane are 1 × 0.4 cm, 1 × 1 cm, and 1 × 2 cm. An advantage of this modality is that the MIMiC collimator can be retrofitted to an existing linear accelerator without an MLC, allowing IMRT treatment without a substantial hardware upgrade. Use of the arc delivery mode often results in a superior dose distribution in comparison with fixed-gantry IMRT with five to nine beams for deep-seated tumors because of the involvement of a large number of beams in an arc-based treatment.[3]

Tomotherapy

The tomotherapy machine has recently become commercially available (TomoTherapy Inc., Madison, WI). The delivery is also achieved slice by slice but in a helical (or spiral) fashion in which the couch moves at a constant speed during the gantry rotation (see Figure 2-3C). Radiation from the linear accelerator first passes through a single

set of primary collimator jaws, which shape the beam into a rectangular slit that is 40 cm long and up to 5 cm wide at the isocenter. The MLC that is used to modulate the beam intensity consists of 64 tungsten leaves that move across a narrow opening to control the radiation passing through to the target. The computer-controlled MLC has two sets of interlaced leaves that move in and out very rapidly to constantly modulate the beam.

Cone Beam–Based IMRT

To date, the majority of work on arc-based IMRT has been focused on modulated fan beams, and little development has been done using cone beams. The concept of intensity-modulated arc therapy (IMAT) was first proposed in 1995 (see Figure 2-3D), and manufacturers have provided the technical capability for dynamic arc delivery.[4] However, IMAT has not been widely implemented. The lack of enthusiasm for IMAT stems in part from the shortage of effective planning tools and reliable quality assurance (QA) procedures. Reports from several institutions, however, support the notion that a cone beam–based arc technique can generate superior dose distributions, at least for some deep-seated tumors.[5–8]

Cone beam arcs use the arc feature of fan beam IMRT yet take advantage of the cone beam modulation of the fixed-gantry IMRT. To compute dose distributions, an arc is approximated by many fixed fields at small intervals of gantry rotation. Physically, however, the achievement of intensity modulation for cone beam delivery is less straightforward in comparison with its fan beam counterpart. Unlike a slice-by-slice delivery, in which the radiation across the slice can be segmentally blocked from the side by multiple independent vanes, the MLC-shaped aperture cannot change from one shape to another fast enough as the gantry rotates. This problem can be solved, at least in principle, by lowering the gantry rotation speed because, in reality, it is the relative speed between the gantry rotation and MLC leaf movement that determines the level of achievable intensity modulation.

An alternative approach is to use multiple cone beam arcs, as proposed by Yu.[4] At each gantry angle, the beam is considered to be a superposition of a series of subfields, each with uniform intensity from these arcs. When a single arc is used for treatment, the technique is sometimes called aperture-modulated arc therapy. At this time, there are no studies defining how many arcs are sufficient for any disease sites. In Figure 2-5, a comparison of average dose-volume histograms (DVHs) of 3DCRT, IMRT, and IMAT prostate plans for ten patients with prostate cancer is shown.[9] The solid line is IMAT, the dotted line is IMRT, and the dot-dash line is 3DCRT. It is evident that IMAT yields better target coverage and improved bladder and rectum sparing in comparison with fixed-gantry IMRT. Finally, being able to modulate the dose rate while the gantry rotates is a desirable feature, further enhancing the performance of the cone beam arc–based IMRT. To date, however, no linear accelerator manufacturers have provided such technical capability in the clinical mode.

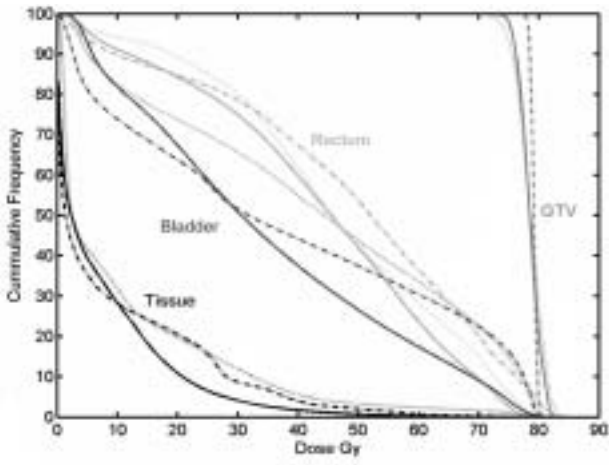


FIGURE 2-5. Average dose-volume histograms for 10 patients with prostate cancer planned using three-dimensional conformal radiation therapy (*dot-dash line*), intensity-modulated radiation therapy (*dotted line*), and intensity-modulated arc therapy (*solid line*). GTV = gross tumor volume. (To view a color version of this image, please refer to the CD-ROM.)

Treatment Planning

RT planning requires the calculation of a set of parameters for the delivery of a radiation dose to the patient. Although manual forward planning may be possible in some simple cases (see the examples below), computer optimization of the beam parameters is almost always used for IMRT treatment planning because of the vast size of search space involved in the problem. In general, this is realized using an inverse treatment planning technique, which derives the optimal beam parameters by starting from a prescribed or desired dose distribution. Although the details of the inverse planning calculation depend on the delivery method, the principle behind the algorithms is essentially the same. Inverse treatment planning is, in fact, a special case of general inverse problem encountered in the sciences and engineering, which attempt to derive the optimal input parameters that will produce the desired output. Before discussing the inverse planning algorithms in detail, it is illustrative to briefly summarize the features of the forward planning approach.

Forward Planning for Segment-Based Treatment

There are two aspects in RT planning: dose conformity and dose uniformity inside the target. What it takes to accomplish the two goals may be different. When the shape of the target is regular and/or when only two or three incident beams are employed, the isodose shaping can often be achieved by beam shaping with an MLC. To achieve a uniform dose distribution within the target volume, one only needs to accommodate the geometric variation of the external contour. Physical

or dynamic wedges are usually used if the patient contour changes monotonically or in some simple hinge field arrangements. In a more general situation, additional MLC-shaped field segments can be introduced to boost a "cold" region or reduce a "hot" region. Examples of this type of case include but are not limited to opposed tangent field breast treatment and anterior-posterior treatment of Hodgkin's disease. For illustration purpose, a forward multisegment breast treatment plan is considered.

The multisegment breast plan starts with the standard opposed tangent fields. In many cases of breast cancer, obtaining a uniform dose within the target volume could be problematic when this approach used. To improve on this, one may proceed to sequentially introduce additional MLC field segments to one or both beam directions to boost the cold region(s) under the guidance of dose distribution in the plane perpendicular to the incident beam direction. Figure 2-6 illustrates the three segments of the lateral and



FIGURE 2-6. Opposed tangential fields for the treatment of a patient with left-sided breast cancer (*top row*). The middle and bottom rows are the multileaf collimator shapes of the three segments of the medial and lateral fields chosen to improve dose uniformity within the treatment volume. (To view a color version of this image, please refer to the CD-ROM.)

medial fields. In this plan, the first segments in the lateral and medial fields and their relative weights are determined using conventional techniques. A physical wedge of 30° is placed on the lateral beam. The two additional segments in each beam direction are then introduced sequentially, and their weights and apertures are adjusted using trial and error to achieve a more uniform dose distribution. The isodose distributions for both plans are shown in Figure 2-7. The maximum dose and the volume receiving a high dose in a multisegment plan are significantly reduced.

Multisegment-based forward planning techniques can be applied only to some relatively simple cases in which the high-dose region is primarily defined by the conventional treatment fields. When isodose conformity to an irregularly shaped target is needed, multiple beams (typically more than five) with a higher level of intensity modulation are needed. In this situation, it becomes tedious to use forward planning–based approaches, and more sophisticated inverse planning techniques become necessary.

Inverse Planning

Inverse planning uses a computer optimization algorithm to determine the optimal beam parameters that lead to a solution as close as possible to the desired output. Mathematically, the problems of image reconstruction, image restoration, signal process, and investment portfolio management can all be formulated as an inverse problem. Roughly speaking, inverse problems can be described as problems in which the output or consequences are known but not the cause. The difference between various treatment planning systems lies in the specifications of the input and output parameters and the criteria used to select the final solution. Specific to RT, the output is generally specified by a desired dose distribution, a set of desired DVHs, or even the tumor control probability (TCP) and normal tissue complication probability (NTCP) for the involved structures. The input parameters to be optimized depend on the delivery scheme. Typically, the number of beams and their incident directions are determined empirically before dose optimization. Each incident beam is discretized into a bixel map (the bixel or beamlet size is typically 1×1 cm^2). The task of inverse planning is then to determine the optimal bixel map or the relative weights of all of the beamlets.



FIGURE 2-7. Conventional opposed tangential breast (*left*) and multisegment (*right*) plans. The isodose lines are (from inside to outside) 105, 100, 95, 90, 80, 50, 20, and 10%. A dose of 50.4 Gy was prescribed to the 90% line. Note that the use of multiple segments improves the dose uniformity to the breast. (To view a color version of this image, please refer to the CD-ROM.)

To better appreciate the problem, assume that six incident beams are used for an IMRT treatment. If each beam is divided into 100 beamlets and each beamlet has 10 permissible intensity levels, there would be $10^6 \times 100$ physically realizable plans. It can be shown that the number of physically realizable solutions for a six-field 3DCRT plan is much less than this number. When wedges are not used, there are 10^6 physically feasible solutions (many of these can be immediately eliminated from being a candidate treatment plan because they do not produce clinically acceptable dose distributions). For a given desired dose distribution D_0, the task is to find a solution D in the physically feasible solution pool {D} that is the same as D_0 or, more appropriately, differs the least from D_0. There are many ways to pick a D that is a "good" representation of the prescribed dose D_0. A commonly used approach for plan optimization is to minimize the distance between D and D_0 in the L^2 norm. For therapeutic applications, it is common to introduce an importance factor $r\sigma$ to control the relative importance of the structure σ. This leads to the following quadratic objective function:

$$F = \frac{1}{N}\sum_{\sigma} r_{\sigma}[D_{\sigma}(n) - D_0(n)]^2, \quad (1)$$

where $r\sigma$ is the importance factor that weights the importance of the structure σ and parameterizes our clinical trade-off strategy and D_0 and D_n are prescribed and calculated doses, respectively. Optimization of this function is essentially a least squares type of estimation in statistical analysis. In addition to equation 1, many other types of objective functions have been proposed for plan optimization. The construct of the objective function plays a crucial role for the success of IMRT treatment and is worthy of detailed discussion.

Models and Model Parameters of Inverse Planning

A common feature of all inverse problems is that they are generally underdetermined and ill-posed. The selection of the final solution depends on the underlying assumption of the model. The objective function quantitatively ranks a candidate treatment plan, and the optimization of the function yields the optimal parameters. In conventional treatment planning, the objective function depends on beam weights, wedge angles, and orientations, whereas in IMRT, it is a function of the beamlet weights. Ideally, an objective function would mimic the decision-making of experienced oncologists and planners. It would rank a given solution (corresponding to a set of parameters) in a way consistent with clinical judgment. In practice, however, a gap exists between mathematical modeling and clinical decision-making, and much effort is being devoted to derive clinically meaningful objective functions for inverse planning. Because the optimization results depend strongly on the objective function, there is inevitably subjectivity associated with the various dose optimization schemes. Therefore, it is essential for physicians to carefully evaluate a treatment plan after optimization to ensure that the "optimal" solution makes clinical sense. Otherwise, the success of an optimization is, at best, mathematical.

If an optimization algorithm is to have a genuine impact on clinical practice, it should incorporate all of the dosimetric and radiobiologic knowledge plus an algorithm for modeling the way in which radiation oncologists and patients balance the risks and benefits. Despite the availability of high-speed computers, state-of-the-art inverse planning algorithms, and improved imaging modalities, we are still a long way from generating truly optimized IMRT treatment plans. For convenience, it is appropriate to classify the currently available dose optimization methods into four categories: (1) dose based, (2) clinical knowledge based, (3) equivalent uniform dose (EUD) based, and (4) TCP or NTCP based. The underlying difference between these models lies in which end points are used to evaluate the treatment plan or which fundamental quantities are used to define the optimal plan. In reality, each type of inverse planning formalism has its own pros and cons in coping with the clinical decision-making process and in practical implementation. These are briefly summarized below.

Dose-Based Formalism

The dose and/or dose volume–based optimization is concerned with accurate dose distributions or DVHs of the involved structures. The quadratic objective function given in equation 1 represents an example of this type. Frequently, DVHs and other physical constraints are imposed to describe certain clinical requirements. The dose or dose volume prescriptions are used implicitly as surrogates of the desired clinical outcome. At this point, the dose-based approach is the most widely employed method, as is evidenced by the fact that all commercial IMRT planning systems have chosen dose-based ranking as the starting point. There are several reasons for this. First, the physical dose objectives reflect the majority of the clinical practice. Although biologic models are available in both research and clinical systems, the uncertainty associated with the predictions often outweighs their guidance. Dose-based objectives will remain the dominant modality of optimization and evaluation for some time. Second, the physical dose is closely related to the optimization parameters, and simple mathematical models, such as the quadratic dose difference expressed in equation 1, can be effectively used.

Clinical Knowledge–Based Formalism

It is highly desirable to incorporate clinical end points in guiding the treatment plan optimization process. The currently available dose-based objective functions do not truly reflect the nonlinear relationship between dose and the response of tumors and normal tissues. In reality, the dose dependence of the clinical end point of a structure may be degenerate in the sense that a given clinical end point may be caused by a variety of dose distributions or DVHs. For the parotid glands, for instance, it is known that the clinical end point is the same if 15 Gy is delivered to 67% of the volume, if 30 Gy is delivered to 45% of the volume, or if 45 Gy is delivered to 24% of the volume. If the dose-based objective function, equation 1, is used, the rankings for the three different scenarios would be different. Even with the use of dose-volume constraints, it is difficult, if not impossible, to incorporate this type of knowledge to correctly model the behavior of the organ in response to radiation. Indeed, a constraint in optimization acts as a "boundary condition" during the optimization (there are methods of treating constrained optimization problem into an equivalent unconstrained one, with a different objective function) and does not change the rankings of dosimetrically different plans.

To overcome these dilemmas, a clinical knowledge–based optimization scheme has recently been developed by Yang and Xing.[10] The central theme of the approach is that clinical outcome data should be used to direct the plan optimization process. In this approach, the quality of a treatment plan is measured by a heuristically constructed objective function which depends not only on the dosimetric properties but also the dose-volume status, which makes it possible to take advantage of the existing outcome data of the involved organs. For the parotid glands, for instance, the three different DVHs mentioned above will be scored equally by the objective function. The final dose distribution or DVHs of the glands will be determined by the optimization algorithm with the consideration of the requirements of other structures. If one of the three possibilities needs to be selected, the one that yields better scores in other involved structures will be favored by the algorithm. The specifics of the plan selection process will, of course, depend on the geometric and dosimetric details of the particular patient.

It is important to emphasize that, at this point, clinical outcome data are sparse and underdetermined and may have large uncertainties. By "underdetermined," we mean that there are not enough clinical data points available to objectively rank all realizable plans. Thus, it is necessary to produce an interpolation/extrapolation scheme for plan ranking. A sensible approach has also been provided in Yang and Xing's work based on the well-known dose response model.[10] The clinical knowledge–based model allows one to more objectively rank treatment plans according to their clinical merits without relying on biological index-based or EUD-based prescriptions.

EUD-Based Formalism

Optimization of the dose distributions can also be cast into the realm of EUD, which is one level higher in terms of the use of biologic information.[11,12] The EUD is defined as the biologically equivalent dose, which, if given uniformly, leads to the same cell kill as the actual nonuniform dose distribution. It can be expressed as follows:

$$EUD = \left(\frac{1}{N}\sum_i D_i^a\right)^{\frac{1}{a}} \qquad (2)$$

In this expression, N is the number of voxels in the anatomic structure of interest, D_i is the dose in the i'th voxel, and a is the tumor- or normal tissue–specific parameter that describes the dose-volume effect. This formulation of EUD is based on the power law dependence of the response of a complex biologic system to a stimulus.

EUD exhibits a dose-response relationship similar to that of the traditional biologic indices. Therefore, it can be a surrogate for them and, in the meantime, is closely related to the physical dose. The objective function based on EUD can be expressed in the following:

$$F = \prod_j f_j \qquad (3)$$

where the component subscore f_j may be either

$$f_T = \frac{1}{1+\left(\frac{EUD_0}{EUD}\right)^a}$$

for tumors or

$$f_{OAR} = \frac{1}{1+\left(\frac{EUD}{EUD_0}\right)^a}$$

for normal tissues. There are several advantages of EUD-based optimization approaches: (1) the formulae are simple, (2) the formulae can be applied to both tumors and organs at risk (OAR) using different parameters, and (3) there are fewer planning parameters than dose volume–based or other biologic indices–based optimization. It has been shown that EUD-based optimization can provide the same or better coverage of targets as dose volume–based optimization and that it offers significantly better protection of OAR. These improvements in the dose distributions to OAR may be due to the fact that there is a larger search space available in EUD optimization because the constraint, or the objective, is determined on the basis of the whole organ rather than the partial volume of the structure. Thus, EUD optimization can be used to search for and evaluate multiple plans that may have different DVHs but the same EUDs. Figure 2-8 shows the dose distributions for IMRT plans optimized using dose volume–based and EUD-based objective functions for a patient with prostate cancer. The OAR are the rectum and bladder. All plans used identical configurations of five coplanar 18 MV photon beams placed at equally spaced gantry angles. The plans were normalized to deliver the prescription dose of 70 Gy to 99% of the target volume. It is clear that, for the same minimal target dose, sparing of the OAR is greatly improved in the EUD-based plan. Furthermore, a sharp dose gradient at the interface between the target and OAR is realized.

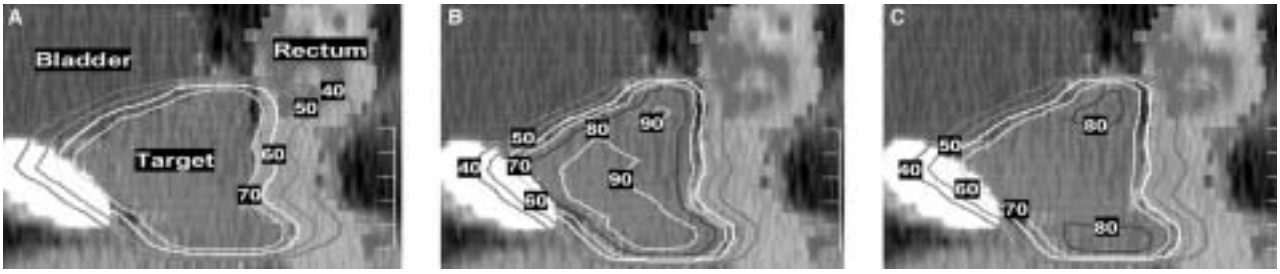


FIGURE 2-8. Sagittal isodose distributions for prostate intensity-modulated radiation therapy plans designed using (*A*) dose volume–based criteria; (*B*) equivalent uniform dose (EUD)-based criteria; and (*C*) EUD-based criteria with target inhomogeneity constraints. Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

The EUD concept can also be incorporated into the framework of physical dose optimization, such as the method of projection over the convex set.[13] In this method, EUD is implemented as an optimization constraint. At each iteration of the optimization, if an organ violates an EUD constraint, a new dose distribution is calculated by projecting the current one onto the convex set of all dose distributions fulfilling the EUD constraint. The cost is slightly more iterations than pure physically constrained optimization. This algorithm is easy to implement and provides better dose sparing of parallel structure organs for which physical constraints may be difficult to define.

Biologic Model–Based Formalism

Biologic model–based optimization proponents argue that plan optimization should be guided by estimates of biologic effects. The biologic effect and the radiation parameters are linked by the radiation dose through the use of a dose-response function. The relationship between the two is not, however, a one-to-one correspondence. A given biologic end point may be produced by many possible dose distributions, which would generally not be equally scored if a dose-based model was used. In principle, biologically based models are most relevant for RT plan ranking.[11,14–21] However, the dose-response function of various structures is not sufficiently understood, and at this point, there is considerable controversy about the models for computing dose-response indices and their use in optimization.

The treatment objective in biologic model–based inverse planning is usually stated as the maximization of the TCP while maintaining the NTCP to within acceptable levels.[22,23] Physical constraints on dose and dose volume are often introduced to ensure that the results are consistent with the clinical judgment of the physicians. Brahme and Kallman and colleagues used the probability of uncomplicated control, P+, in their formalism.[14,24] Practically, the use of dose-response indices for optimization might also pose some problems. For instance, dose response–based optimization may lead to very inhomogeneous target dose distributions. Furthermore, it is difficult for clinicians to specify the optimization criteria in terms of certain dose-response indices (eg, TCP, NTCP, and P_+). This difficulty becomes even more significant when two or more independently optimized plans are to be combined because it is impractical to specify the desired TCP and NTCP of the component plans. Because of these problems, the use of biologic model–based dose optimization has mainly been limited to the research setting and little effort has been made to implement these into commercial IMRT planning systems.

Model Parameters

Any dose optimization framework must deal with trade-offs between the target and OAR.[25] Generally, the objectives of different structures are multifaceted and incommensurable. A combination of the objectives is usually done to form a single objective function. In this process, a set of importance factors is often incorporated into the objective function to parameterize trade-off strategies and prioritize the dose conformity in different anatomic structures. Whereas the general formalism remains the same, different sets of importance factors characterize plans of obviously different flavor and thus determine the final plan. One of the major difficulties is that the influence of these weighting factors on the final solutions is not known until the dose optimization is done, necessitating a trial-and-error determination of the parameters. In most (if not all) of the currently available planning systems, the values of the weighting factors are presented to the user as optimization parameters. A good understanding of the role of these parameters and suitable training on how to empirically determine the parameters are required.

It is possible to use an iterative algorithm to estimate the weighting factors numerically.[25] Plan selection is done in two steps. First, a set of importance factors is chosen, and the beam profiles are optimized under the guidance of a quadratic objective function using an iterative algorithm. The "optimal" plan is then evaluated by a decision function, in which the corresponding trade-off parameters are more easily determinable based on some simple considerations.[25] The importance factors in the objective function are adjusted iteratively toward the direction of improving the ranking of the plan. For every change in the importance factors, the beam parameters are reoptimized. Even though further refinement of the plan may still be needed in selected cases, the technique provides a good starting point for planning.

Dose Optimization Algorithms

Although the modeling of RT treatment is of paramount importance, the optimization of the selected multidimensional objective function provides a vehicle to obtain the optimal solutions. The task of an optimization algorithm is to find the combination of beam parameters that optimize the chosen objective function, possibly subject to some constraints. Numerous algorithms have been developed for the optimization of a multidimensional function in the sciences and engineering over the years, and there is a vast literature on the subject. Generally speaking, the selection of an optimization technique depends on the specific form of the objective function and the imposed constraints. In practice, even for the same class of problem, more than one algorithm may exist for achieving the same goal, and the detailed implementation of different algorithms can be quite different. Many optimization techniques have been used for RT inverse planning. Here we briefly describe a few approaches to illustrate how a multiobjective objective function is optimized and the pros and cons of these common techniques (see Chapter 10 "Treatment Planning").

Iterative Algorithms

The iterative method is perhaps the most widely implemented technique in RT optimization. Starting with an initial approximate solution, it generates a sequence of solutions that converge on the optimal one. For large systems, especially large linear systems, iterative methods prove to be efficient in terms of computer storage and computational time. The available iterative techniques can generally be grouped as non–derivative-based and derivative-based methods. The former incorporates only an objective function value calculation with some systematic method to search the solution space. This technique is generally intuitive, easy to implement, and particularly suitable for simple systems and educational illustration. For a complex system, the convergence behavior may not be as good as more sophisticated gradient-based search techniques. The computational cost and poor convergence in this situation may outweigh the benefit of avoiding derivative calculations.

As an example, Figure 2-9 illustrates the flowchart of an algebraic iterative inverse planning technique (AIIPT) described by Xing and colleagues.[26,27] A schematic drawing of calculation pixels and bins in the AIIPT calculation is shown in Figure 2-10. The algorithm was generalized from the algebraic reconstruction technique (ART) based on the analogy between rotational RT optimization and tomographic image reconstruction. In the AIIPT algorithm, voxels are examined in sequence, and corrections are made immediately after a pixel is addressed. The successive treatment of the system eventually leads to an optimized solution. A geometric interpretation of ART has been published.[28]

The iterative process is described by the following operations: (1) assume an initial set of beam profiles; (2) compute the dose at a voxel; (3) compare the calculated and prescribed doses; (4) obtain correction factors to the beamlets that irradiate the voxel; and (5) apply the corrections to the contributing beamlets and then repeat from step 2 for the next voxel (go back to the first voxel and increase the iteration index by one after all voxels are addressed). This process is repeated until the desired accuracy is achieved. The simultaneous iterative inverse planning and least squares inverse treatment planning algorithms also fall into the same category of the nonderivative method.[26,27] A similar algorithm with a multiplicative beamlet updating scheme was described by Jones and Hoban.[29]

Various gradient-based methods have been successfully applied to RT plan optimization and implemented in commercial IMRT planning systems. A general class of iterative algorithms can be written as

$$\mathbf{I}^{new} = \mathbf{I}^{old} - \lambda \mathbf{M}^{old} \nabla F(\mathbf{I}^{old}) \tag{4}$$

where $\mathbf{I}$ is the fluence vector, $\mathbf{M}^{old}$ is a matrix, and λ is a positive parameter. When $\mathbf{M}^{old}$ is a unit matrix, equation 5 is the well-known steepest descent algorithm, whereas when $\mathbf{M}^{old} = \mathbf{H}^{old}$, it describes Newton's method, in which $\mathbf{H}^{old}$ is the inverse of the Hessian matrix.[30] In the steepest descent algorithm, from a set of fluence functions, $\mathbf{I}^{old}$, we search along the direction of the negative gradient, $-\nabla F(\mathbf{I}^{old})$, to a minimum on this line; this minimum is taken to be $\mathbf{I}^{new}$.

It is fair to say that the choice of a specific algorithm

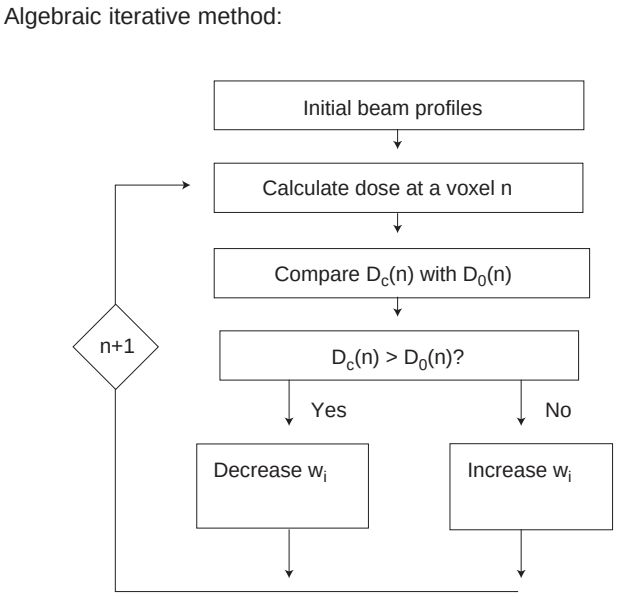


FIGURE 2-9. Flowchart of the algebraic iterative inverse planning technique.

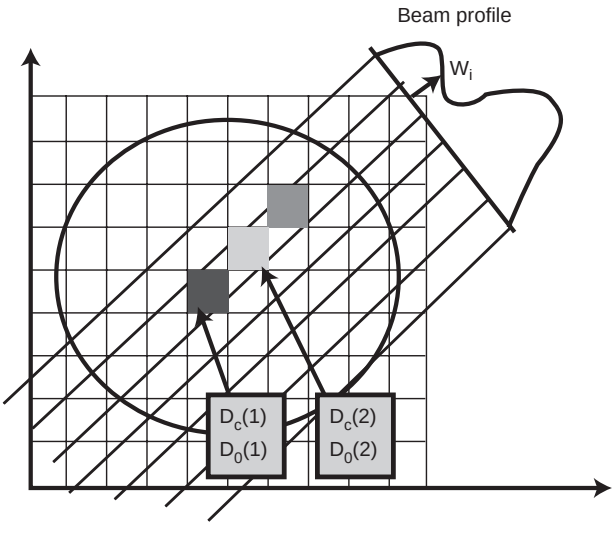


FIGURE 2-10. Pixel and bin configurations used in the algebraic iterative inverse planning technique. In this figure, w_i is the beamlet weight and D_c and D_0 represent the calculated and prescribed doses, respectively. Reproduced with permission.

to solve the inverse planning problem is not unique and is determined by the problem at hand and, to a certain extent, by personal preference. Generally, the iterative approach works well for a nonconvex objective function, and the solution can be trapped in local minima for systems with a complicated form of the objective function. Several commercial systems provide both iterative and stochastic optimizers so that users have the tool to compare the functionality of different approaches and, more importantly, to independently check an optimization calculation.

Computer-Simulated Annealing

The simulated annealing method[31] is an extension of the original Monte Carlo simulation algorithm introduced by Metropolis and colleagues.[32] It attempts to find the optimal solution by mimicking the behavior of a system of interacting particles that are progressively cooled and allowed to maintain thermal equilibrium while reaching the ground state. In physical annealing, the system is heated, thereby conferring randomness to each component. As a result, each variable can temporarily assume a value that is energetically unfavorable, and the system explores configurations that have a higher energy. The fundamental principle here is that even at moderately high temperatures, the system slightly favors regions in the configuration space that are overall lower in energy and hence more likely to contain the global minimum. The algorithm employs a random search that not only accepts changes that decrease objective function but also some changes that increase it. The probability for accepting a trial configuration is controlled by the temperature and is given by

$$P = \begin{cases} 1 & \text{If } \Delta F < 0 \\ \exp(-\dfrac{\Delta F}{T}) & \text{Otherwise} \end{cases} \quad (5)$$

where ΔF is the increase of the objective function and T is the system temperature. The temperature is gradually lowered according to an empirically chosen cooling schedule.[33,34] As the temperature is slowly reduced, the probability of accepting a trial configuration with a higher objective function value is reduced. The starting temperature is chosen to be higher than the largest value of objective function calculated for a random set of variable configurations. In principle, this algorithm is capable of finding the global minimum of a multidimensional objective function even when local minima exist. For more details about the simulated annealing algorithm, readers are referred elsewhere.[34–36]

Other Optimization Algorithms

In addition to the iterative and simulated annealing algorithms, many other types of optimization approaches have been employed for therapeutic plan optimization. Linear programming was applied to the dose optimization of 3DCRT plans and cyberknife plans. The utility of filtered backprojection from CT image reconstruction has also been explored by several researchers.[37–39]

The constrained least square algorithm[40] was employed to optimize 3DCRT plans[41,42] and IMRT plans. Constrained optimization of a linear system can be viewed in two ways. One involves transformation of the problem into a reduced space. Another approach is to work with the lagrangian function and to obtain the solution of the system by a direct matrix manipulation. In this way, a priori knowledge of the variance of the system variables can be included as a lagrangian multiplier. Without repeatedly invoking the dose calculation, this algorithm allows one to obtain the optimal solution of the system with significantly increased computational speed, providing a fast interactive planning environment for IMRT planning.

Mixed integer programming technique was used to generate treatment plans for linear accelerator–based radiosurgery,[43] IMRT,[44] and 3DCRT.[45] Lee and Zaider also applied integer programming for permanent prostate implant planning.[46] The mixed integer programming models incorporate strict dose restrictions on the tumor volume and constraints on the desired number of beams, isocenters, couch angles, and gantry angles. The goal is to deliver the full prescription dose uniformly to the tumor volume while minimizing excess radiation to the surrounding normal tissue. Hou and colleagues used simulated dynamics in a classic system of interacting particles for IMRT optimization.[47] In this approach, an analogy is established between intensity profile optimization in IMRT and relaxation to the equilibrium configuration in a dynamic system. Dose-volume constraints are handled by placing hard constraints on partial volumes. The genetic algorithm is another widely used approach in sciences and engineering and has found some preliminary application in RT dose optimization.[48–51]

For all of their complexity, the algorithms to optimize a multidimensional function are routine mathematical procedures. In general, simulated annealing and genetic algorithms are powerful approaches, but excessive computation time is a drawback to their clinical application. Treatment planning based on filtered backprojection and direct Fourier transformation have difficulty handling the negative fluence problem and are not generally applicable for an arbitrary dose prescription and kernel. Iterative methods are widely used to optimize a multidimensional objective function by starting with an initial approximate solution and generating a sequence of solutions that converge to the optimal solution of the system.

It is useful to note that much effort has also been devoted to formulate the problem into a more effective mathematical framework. For example, Xing and Lian and their colleagues introduced a new concept of a preference function and recast the problem into the framework of Bayesian

statistical analysis.[35,52–54] In this approach, instead of a rigid prescription dose, a range of prescription doses prioritized by the preference function is allowed. The rationale here is that since a rigid prescription is not achievable and the final solution will deviate from it anyway, we would have much better chance to obtain what we want if we could inform the system with some a priori information about our preferences on different possible scenarios (instead of leaving the decision-making totally to the computer). The techniques developed over the years in statistical decision-making can be easily extended to RT plan optimization problem. The primary advantage of the technique is that it enables one to effectively incorporate the existing clinical knowledge or other prior knowledge into inverse planning. When the prescribed dose takes a single rigid value, the above formula becomes identical to the conventional least squares approach or alike. Maximum likelihood estimation[55,56] or the maximum entropy approach[57] also represents a special case of the formalism. Finally, it is interesting to point out that various techniques in related fields such as neural networks[58] and fuzzy logic[59] are also being translated for RT dose optimization.

Practical Aspects of IMRT Planning

Inverse planning is a computer-based decision-making technique that derives the optimal treatment plan by starting with a set of desired doses or DVHs prescribed to the target and normal tissues. To use an inverse planning system to generate a treatment plan, one must delineate the tumor volume and sensitive structures, for which dose avoidances are desired. This differs from conventional planning, in which the target volume is often defined directly on the portal films (see Chapter 11 "Plan Evaluation"). If target contours need to be altered after a conventional treatment plan is obtained or during a course of treatment, it is usually achieved by modifying the positions of the corresponding MLC (or by modifying a block). In inverse treatment planning, however, the beam profiles and beam apertures are derived by the system, and any change in the target volume requires reoptimization of the plan. Moreover, all of the tasks following IMRT planning, such as patient-specific QA and data entry, need to be repeated.

IMRT planning is still inherently a trial-and-error process owing to the large number of input parameters.[60] The trial-and-error process here is quite different from that in 3DCRT, in which intuition and previous experience can be easily used to guide the planning process. In an anterior-posterior treatment, for example, if the dose in the anterior region is higher than that of the posterior region, one can simply increase the weight of the posterior field. This type of guidance is lost in inverse planning, and, frequently, the trial-and-error process has to proceed in a "blind-guessing" fashion because the influence of most of the system parameters is not known until the dose optimization is complete. A good understanding of the effect of treatment planning parameters used in optimization on the resultant dose distribution is necessary to carry out the planning and the plan "tweaking" process. Recently, tools for assisting the interactive planning have emerged. The dose shaping technique described below is one example. Hopefully, this type of research will make clinical inverse planning more straightforward in the future.

Plan review is an important aspect of IMRT. In inverse planning, an objective function is constructed based on general physical, dosimetric, and biologic considerations and is defined as a global quantity.[52] The translation of the treatment objectives to a single objective function is at best an approximation. Just like any data reduction or compression scheme, there is a loss of information with regard to the characteristic of the individual data point. Even with the best possible objective function, the optimal solution may still not represent the best clinical solution in every aspect. It is important to review the plan to ensure that the final solution is consistent with clinical judgment. IMRT plan evaluation tools vary from one commerical planning system to another. Typically, they include isodose distributions in axial, coronal, and sagittal planes; DVHs; and maximum, minimum, and average target and sensitive structure doses. A description of plan evaluation methods is presented in Chapter 12 "Delivery Systems."

The dose inhomogeneity of an IMRT plan is usually higher than that in 3DCRT as a consequence of increased conformity. Any deviation from a conventional uniform dose scheme should be carefully evaluated to ensure its clinical acceptability. If hot or cold spots are unavoidable, efforts should be made to ensure that they are not located in undesirable locations. For example, a cold spot in the center of the target or a hot spot outside the target should be avoided. Even a hot spot inside the target volume may not be desirable. For example, for prostate cancer, a hot spot close to the urethra is usually not acceptable, particularly if the total dose is escalated. The dose gradient of an IMRT plan near the boundary of the target or OAR can be very high. If the structure(s) is susceptible to the setup uncertainty and/or organ motion, the actual dose received by the target or OAR may be significantly lower or higher than that shown in the plan.[61,62] In this case, an adequate margin for the structure is important to ensure that the planned dose distribution can be achieved in a clinical setting.[63]

Beam placement in IMRT is worth discussing. Generally speaking, the beam configuration may have significant influence on the quality of an IMRT plan even when a large number of incident beams (eg, nine beams) are used.[64–69] Clinically, however, beam orientations are selected on a trial-and-error basis. To obtain an optimal beam configuration, in principle, one can simply add the degree of freedom of beam angles into the objective function and optimize them together with the beamlet weights.[65,70] Although this does not pose any conceptual challenge, the computational time becomes excessive because of the greatly enlarged search space and the coupling between the beam profiles and the

beam configurations. The beam intensity profiles have to be optimized for every trial beam configuration because the influence of a set of gantry angles on the dose distribution is not known until beam intensity optimization is performed. A computationally efficient optimization algorithm is necessary to have a clinically practical beam orientation optimization tool. Some progress has been made toward this goal.[71,72] But before commercial companies implement clinically practical tools for automated or semiautomated beam placement, alternative techniques or even some general guidelines would be useful to facilitate IMRT planning.

One of the appealing approaches is the class-solution method.[73] The basic idea is to construct a representative beam configuration based on previous experience for a given disease site and then use this "class-solution" for subsequent treatment planning. Schreibmann and Xing systematically investigated the issue and proposed a set of class-solutions for IMRT prostate treatment.[74] To derive a population-based beam orientation class-solution, a beam orientation optimization algorithm was used to derive the optimal solutions for each individual in a group of 15 patients with prostate cancer. Figure 2-11 shows the distributions of optimal beam angles for five, six, seven, and eight beams for the 15 patients studied. The colored short lines represent directions found in individual cases, and the red bold long lines represent the directions identified as the class-solutions. These results indicate that the beam orientations for a certain incident direction are confined in a certain range and that beam orientation class-solutions may be a reasonable compromise between what is practical and what is optimal for prostate IMRT. For other disease sites, beam orientation class-solutions may not exist because the geometric variations among the patient population are too large.

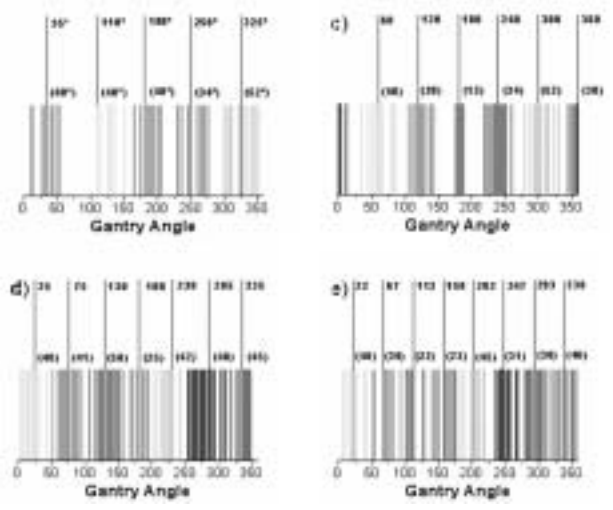


FIGURE 2-11. Distributions of beam angles for five, six, seven, and eight beams, respectively. The short colored lines represent directions found in individual cases, and the long red (bold) lines represent the directions identified as class-solutions. Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

On approval of the plan by the physician, an RT plan file or a DICOM-RT file is generated containing all of the relevant machine parameters for IMRT treatment. The IMRT plan file can be complex. For example, it may contain hundreds of MLC segments. Consequently, manual delivery is not an option. Instead, delivery is usually accomplished by the computer-controlled systems, including the record and verify system, linear accelerator control, and MLC control software. The detailed treatment settings contained in the RT plan or DICOM file are transferred from the planning system to the record and verify system. Normally, redundant checksums are also in place for each record, ensuring the safe transfer of data over the computer network. Although it is perhaps not necessary to list all of the information about the treatment in the chart, the chart should contain concise information about the treatment that can be easily verified, for example, the treatment machine, energy, number of beams, gantry and couch angles, monitor units (MUs) for each beam, number of fractions, and fraction doses. The plan output, such as isodose lines for a selected plan in different views on CT images, DVHs, and the QA report, should also be documented. The intensity maps for each beam should be included if possible.

Advanced Topics in IMRT Treatment Planning

Inverse planning is at the foundation of IMRT, and its performance critically determines the success of an IMRT treatment. Unfortunately, the currently available inverse planning formalism is not satisfactory, and the solutions out of so-called "optimization" systems are often suboptimal. Considerable effort may be required to compute a clinically acceptable plan, and the final results may strongly depend on the planner's experience and understanding of the planning system. These shortcomings of the existing systems are familiar to anyone engaged in clinical IMRT treatment planning. In addition to the prescription doses, the current planning system requires the user to preselect the angular variables (gantry, couch, and collimator angles) and the weighting factors of the involved structures. These variables and parameters constitute an additional multidimensional space, which is coupled to the beam profiles.

A survey carried out by us indicates that there are five major problems with current inverse planning systems: (1) no effective mechanism for incorporating prior experience into plan optimization; (2) lack of direct control over the regional dose or, more generally, lack of interactive tools to guide the planning process; (3) no effective tools for aiding beam placement in IMRT planning; (4) inability to incorporate organ motion directly; and (5) inefficient inter-

face between planning and delivery systems. Toward establishment of a clinically efficient and robust inverse planning system, many investigators have attacked the problems mentioned above, some of which are the subject of the following sections.

Statistical Analysis–Based Formalism for Therapeutic Plan Optimization

An important element that is missing in the current inverse planning formalism is a mechanism for incorporating prior knowledge into the dose optimization process. In image analysis and many other fields, it has proven valuable to include partial knowledge of the system variables into the optimization process[35,54,75] because it provides guidance in the search for the truly optimal solution. Statistical analysis formalism, which appears in virtually all branches of the sciences and engineering, affords a natural basis for this type of application and provides a powerful vehicle to achieve the goal of treatment plan optimization. Using this approach, Lian and colleagues demonstrated the feasibility of incorporating a range of prioritized dose prescriptions into the planning process.[52,76] The approach is based on a newly introduced concept of a preference function, whose role is to relax our requirement of a rigid dose prescription, to allow a range of doses to be considered, and to quantify the willingness to accept a dose in that range. In addition, to make the system less ill-defined, this new scheme can be used to formalize our clinical knowledge (such as outcome data[19,77]) and incorporate them into dose optimization. In Figure 2-12, we show the preference function derived using published data from Eisbruch and colleagues[78] for parotid glands (four different irradiation volumes). Coupled with the statistical inference techniques,[53,54] this should make the inverse planning process more computationally intelligent.

Another application of the formalism is to include model parameter uncertainties into dose optimization. For example, the radiobiologic formalism involves the use of model parameters that are of considerable uncertainty. Biologic "margins" have been used to account for the variability in radiation sensitivity. This method assumes the patient to be more sensitive than the mean value for normal tissues and more resistant than the tumor. EUD-based optimization with the incorporation of model parameters has been demonstrated through the use of a statistical inference technique.[79] Because currently available models for computing the biologic effects of radiation are simplistic and the clinical data used to derive the models are sparse and of questionable quality, the technique is valuable to minimize the influence of statistical uncertainties.

Multiobjective Optimization

Radiation dose optimization is intrinsically a multiobjective problem because of the existence of multiple conflicting objectives in the system. In the conventional approach, the multiple objectives are combined to form an overall objective function through the use of so-called importance factors.[25,80] Contrary to this, the dose delivered to each structure constitutes one of the objectives in multiobjective optimization, which is an alternative way to deal with the trade-offs of multiple conflicting objectives. The method attempts to obtain all efficient solutions and provide the planner with a more thorough picture of the possible options or the trade-offs between the different objectives. Here an efficient solution (often called the Pareto solution in multiobjective optimization theory) is defined as a plan with a good compromise of all of the objectives involved in the problem or, more precisely, a plan that cannot be further improved without significantly deteriorating the dose distribution in one or more organs. Mathematically, the multiobjective optimization (or vector optimization)[81–84] is to determine a set of decision variables that optimizes a vector function whose elements represent *M* objective functions without violating the system's constraints. The collection of all efficient solutions is named the Pareto front. Although the approach is conceptually appealing, practical issues, such as the enormous computing time required to obtain the Pareto solution and how to effectively select a plan from the Pareto front, must be resolved before it finds widespread application in RT plan optimization. Perhaps a hybrid of single- and multiobjective techniques is a viable option.

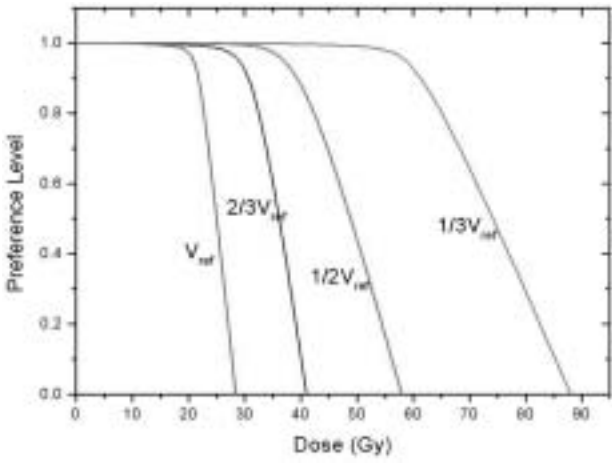


FIGURE 2-12. Preference function of parotid salivary glands for four fixed partial volumes (V_{eff}). In constructing the assumed dose-volume preference function, the preference level is assumed to be 1 for normal tissue complication probability (NTCP) < 5% and 0 for NTCP > 50%. The reference volume (V_{ref}) refers to the volume of the gland. Adapted from Eisbruch A et al.[78] (To view a color version of this image, please refer to the CD-ROM.)

Integration of IMRT Planning and Delivery

IMRT planning is generally performed in two steps: calculation of the intensity maps of the incident beams and decom-

position of each intensity map into a series of MLC-shaped segments using a leaf sequencing algorithm. In practice, the decoupling of dose optimization and leaf sequencing has a number of consequences. In addition to the need for the development of complicated MLC leaf sequencing software, the number of segments resulting from the approach is often unnecessarily large. The leaf sequencing algorithm sometimes has to go through additional steps to accommodate some special hardware constraints of the MLC delivery system that can be easily dealt with at the stage of dose optimization. To improve the efficiency of the interface between the inverse planning and the dynamic MLC delivery systems, attempts have been made to incorporate machine constraints and other physical aspects of the delivery system into dose optimization. The most effective method is perhaps the aperture- or segment-based optimization, which optimizes directly the objective function with respect to the shapes and weights of the segmented fields.[44,51,85,86] In this approach, the number of segments for each incident beam is prespecified instead of left "floating." Generally speaking, it is more computationally involved to optimize an objective function with respect to the segment shapes and weights because of the nonlinear dependence of the dose on the leaf coordinates. However, the benefits gained by eliminating the extra leaf sequencing step and the associated drawbacks outweigh the slight computational cost.

Interactive Planning Tools for IMRT

The interactive process of IMRT planning is less intuitive than that of forward planning because of the involvement of a large number of parameters whose roles in the final solution are not explicitly known until the completion of a dose optimization calculation. There is a need for the planner to adaptively modify or fine-tune a solution toward the desired direction. For example, frequently after optimization, the dose in only a few small regions is not satisfactory. Currently, plan modification is achieved by adjusting structure-dependent system parameters (eg, prescription, importance factors), which influence not only the dose in the region of interest but also in other areas. To modify the dose in a specific region, in principle, one can use ray-tracing to find the beamlets intercepting the area and adjust their intensities accordingly. The problem is that there are numerous ways to modify this intensity and the optimal arrangement of the beamlet intensities is not obvious. Cortrutz and Xing pointed out that local dosimetric behavior can be more effectively controlled by introducing a region-dependent penalty scheme and demonstrated the utility of this approach using a model system and clinical examples.[87,88] After the conventional planning is done, they identify the subvolumes on isodose layouts or the dose interval on the DVH curve in which the fractional volume needs to be changed. The local penalty (eg, local importance factor or local prescription) is then adjusted, and the dose is reoptimized. The fine-tuning of doses is manually iterative in nature, and the process can be easily accomplished using a graphic user interface. Using this technique, it has been shown that one can eliminate hot and cold spots. Generally, in dose optimization, there is no net gain. That is, the improvement in the dose to a region is often accompanied by a dosimetrically adverse effect(s) at another point(s) in the same or different structures. Practically, however, some dose distributions are more acceptable than others. The important issue here is to find the solution that improves the dose(s) at the region of interest with a clinically insignificant or acceptable sacrifice.

It is useful to mention that some "hot spot editor" tools have recently been implemented in commercial systems. These editors rely primarily on a rudimentary ray-tracing and is done as follows: (1) visually locating the hot/cold spot; (2) finding the corresponding beamlets that contribute to the dose at the point of interest (POI); (3) decreasing/increasing the intensities of one or more of the beamlets; and (4) updating the dose distribution. In reality, there are multiple beamlets contributing to the dose at the POI and the problem is determining the optimal way to modify them so that the doses at other points are compromised minimally. In the current commercial systems the hot/cold spot is improved by decreasing/increasing the beamlets with a predesigned updating method, which is rarely optimal and often causes new cold/hot spots somewhere else within the patient. The approach described by Cortrutz and Xing allows optimal adjustment of the beamlet intensities, thus avoiding the aforementioned problem. Since the re-optimization is done on top of the existing solution, it requires only very limited additional computing time. Furthermore, it is done in the background (just like the recalculation of dose in step (4) is done in the background) and the user does not need to take any additional action.

Automated and Semiautomated Beam Placement

Clinically, gantry angles are selected empirically, and there is no guarantee that the beam configuration is optimal for a given patient. Many investigations are exploring the role of beam configuration selection in IMRT[39,65,69,89–91] and developing tools for beam placement. A promising technique uses beam's eye view dosimetrics (BEVD).[71,92] The central idea of this single-beam scoring technique is that the merit of a beam direction should be measured by what that beam could achieve dosimetrically without exceeding the dosimetric or dose-volume constraints of the system. For computational purposes, a beam portal is divided into a grid of beamlets. Each beamlet crossing the target is assigned the maximum intensity that could be used without exceeding the dose tolerances of the sensitive structures and normal tissue. A forward dose calculation using the "maximum" beam intensity profile is then performed, and the score of the given beam direction (indexed by i) is calculated according to[92]

$$S_i = \frac{1}{N_T} \sum_{\text{voxTarget}} \left(\frac{d_{ni}}{D_T^p} \right)^2 \quad (6)$$

where d_{ni} is the maximum dose delivered to voxel by the beam from the direction indexed by i, N_T is the number of voxels in the target, and D_T^p is the target prescription. The BEVD score function captures the main feature of a planner's judgment about the quality of a radiation beam and allows one to select beam orientations without excessive computational time. For a given patient, the score function for every possible beam direction is evaluated and the directions with the highest BEVD scores are identified. Although the technique does not yield the final beam configuration in a multifield IMRT treatment, it provides useful information with regard to which beam directions are potentially good or bad. During planning, the beams with the highest scores are considered favorable for the treatment. It is also illustrative to point out that the BEVD information can also be integrated into beam orientation optimization program to improve the convergence behavior and computational speed.[72]

Hybrid Treatment of IMRT with Other Modalities

IMRT affords one the ability to produce not only spatially uniform but also purposely nonuniform doses. A natural application of the feature is to combine IMRT with other RT modalities to generate a dose distribution that would otherwise be impossible. Along this line, IMRT has been considered a method of salvaging suboptimal prostate implants.[93] The combination of IMRT with conventional electron beam(s) for improving the photon-electron field matching and for treatments of certain specially shaped targets has also been investigated.[94] Figure 2-13 illustrates a combined head and neck boost treatment using electron and intensity-modulated photon beams. In head and neck cancer, treatment initially involves the irradiation of the primary tumor and the cervical lymph nodes. After the tolerance dose of the spinal cord is reached using opposed lateral photon beams, the lateral fields are reduced off the spinal cord. The treatment of the anterior neck along with the primary tumor is continued using the reduced photon beams, whereas the posterior neck overlying the spinal cord is treated with lateral electron fields. This requires the matching of an electron field with two opposed photon fields. For comparison, the combined treatment using the conventional technique (electron + conventional photon beams) is shown in Figure 2-13. When the electron and unmodulated photon beams are matched directly, hot spots greater than 15% above the prescribed dose are seen in the abutting region. This is reduced to 5% when the proposed technique was used, with markedly better dose homogeneity in the abutting region. Furthermore, because of the broadened photon penumbra, the dose homogeneity in the junction region becomes less sensitive to patient setup errors.[94]

Gated or Synchronized IMRT

IMRT can produce highly conformal doses to targets and a sharp dose gradient between targets and surrounding critical structures.[60,95–98] Together with improved patient immobilization, the target margins can be reduced to facilitate dose escalation. However, margin reduction remains challenging for treatments in the thoracic regions and other sites in which intrafraction respiratory motion is significant.[99,100] The delivery of IMRT in either dynamic or static mode can cause unexpected high- and low-dose regions owing to the interplay between the movements of the tumor and the MLC leaves.[101] This issue can, in principle, be improved through

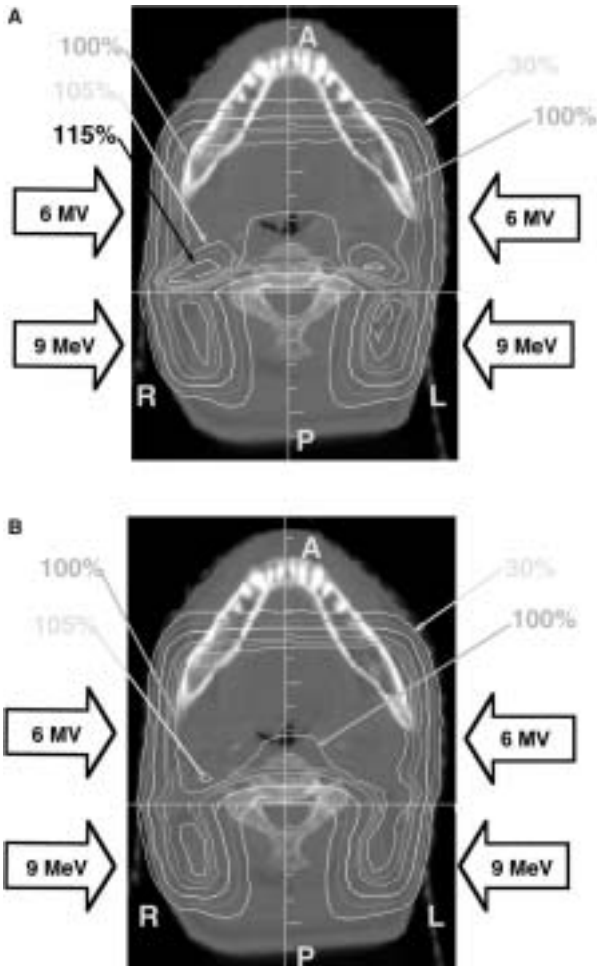


FIGURE 2-13. Comparison of the isodose distributions of the treatment plans in a transverse section of a patient with head and neck cancer when the electron and photon beams were matched directly (*A*) and when dynamic intensity modulation was used for the photon beams (*B*). Isodose levels are shown at 30%, 50%, 70%, 90%, 95%, 100%, 105%, 110%, and 115%. Note that the 110% and 115% isodose lines are not present in (*B*). Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

the use of gating or respiration synchronization.[102–106] (See Chapter 9 "Respiratory Motion Management" and Chapter 19.4 "Intrafractional Organ Motion and Planning: Emerging Technology".)

In gating, tracking, and breath-hold, the treatment machine is switched "on" or "off" in response to a signal that is representative of a patient's breathing motion. Both passive and active devices can be used to monitor the respiratory motion. Ideally, the beam is on only during portions of the breathing cycle when motion is small. The disadvantage of this technique is the prolonged treatment time compared with that of nongating approaches.

Motion-synchronized RT is based on two assumptions: (1) tumor motion is considered to be predictable, and the model of motion can be established prior to the treatment and is assumed to be the same (or at least adaptively predictable) throughout the treatment course, and (2) the treatment delivery system, either the MLC leaves or the treatment couch, can be instructed to precisely move to certain locations to adapt to the motion of the tumor. The main advantage of the motion-synchronized RT is that the radiation beam is on all of the time; therefore, there is no treatment time prolongation. However, several major technical difficulties must be overcome. Inverse treatment planning in this case must take into account the functionality of the delivery system.

Another issue is that respiratory motion exists in all stages of the RT process, including preplanning imaging and treatment planning and delivery. If respiratory motion is not accounted for during image acquisition, artifacts may arise during the image acquisition, leading to the distortion of the target volume. For gated treatment, the same window should be used for imaging and planning and delivery so that tumor positions and patient anatomy can be reproduced accurately. For motion-synchronized RT, several sets of CT images representing different phases of the breathing cycle need to be acquired through either a high-speed multislice CT scanner or by postprocessing software to sort the images. All of these images will be used for IMRT treatment planning, and the resulting MLC leaf sequences need to be multiplexed for delivery.

Biologically Conformal IMRT

Although the biology of tumors plays a crucial role in the success of RT, commonly used CT and magnetic resonance images provide few metabolic data and have significant shortcomings in characterizing benign and malignant tumors. Recent advancements in functional imaging make it possible to noninvasively obtain a patient's metabolic distribution. Coupled with the technical capability of IMRT in generating customized three-dimensional dose distributions with subcentimeter resolution, this may afford a significant opportunity to improve conventional RT by producing doses in accordance with biologic requirements.[107–110] Research effort is focused on integrating functional data into IMRT treatment planning to improve clinical cancer management. In general, functional imaging suggests nonuniform dose distributions to meet the heterogeneous biologic requirements. Xing and colleagues identified some relevant issues and developed a preliminary four-dimensional inverse planning scheme for functional imaging-guided IMRT.[109] The metabolic and functional data are incorporated effectively by modulating the prescription doses in the target voxels. This algorithm enables one to produce a high dose where there is resistance and/or where tumor burden is large and to differentially spare the sensitive structures with more emphasis on functionally important regions.

Delivery Techniques for Fixed-Gantry IMRT

IMRT delivery with MLC is based on the simple principle that moving jaws or leaves can be employed to control the dose delivered to a point. As mentioned previously, IMRT planning is currently performed in two steps: optimization of intensity maps and MLC leaf sequencing. The latter is to convert an intensity map into an MLC leaf sequence file, which specifies the leaf positions as a function of the fraction of MUs delivered. For the delivery, the two-dimensional beam fluence is divided into strips corresponding to the projection of each leaf pair of the MLC. Each MLC leaf pair is then required to modulate the fluence along its projection (see Chapter 12 "Delivery Systems"). For Varian linear accelerators (Varian Medical Systems, Palo Alto, CA), the leaf pairs are independent, reducing the conversion of two-dimensional fluence profiles into a collection of one-dimensional problems. As a result, the problem becomes finding a series of leaf positions (coordinates of leading and trailing leaves) to cover the area under a one-dimensional fluence function. There is no unique solution to this problem, leading to a number of ways to accomplish beam modulation.

MLC-based delivery is generally divided into static step-and-shoot[111–116] and dynamic modes.[117–119] A step-and-shoot leaf sequence file consists of alternatives of dose-only and motion-only instances. The step size of MLC movement in this mode is determined by the dimension of the beamlet in the leaf movement direction. Dynamic delivery differs from the step-and-shoot mode in that leaf movement and dose delivery are realized simultaneously. These algorithms are described below.

Step-and-Shoot Delivery

In step-and-shoot delivery, the total dose at a spatial point is the superposition of contributions from a series of segment fields (typically, the number of segments is between 20 and 100). The x-ray beam is off when the MLCs travel from one segment to another. This is perhaps the most intuitive technique to deliver intensity-modulated fields using MLC. The QA procedure for this delivery mode is relatively simpler (than

dynamic delivery) because there is no correlation between the leaf speed and the dose. Instead of describing the algorithm generally, an example is used to illustrate how the step-and-shoot leaf sequencing methods work.[111]

Figure 2-14 depicts a simple example of an intensity pattern. The intensity in a 6×4 cm field is expressed using five discrete intensity levels. Four 1 cm–wide leaves (numbered 12, 13, 14, and 15) are to be used to generate the intensity pattern. The profiles that each leaf pair must generate are shown in the individual graphs (the profile required by leaf pair 12 in this example).

For a modulated field to be delivered at a gantry angle, each component profile along the center of the j^{th} leaf pair must be rendered into a leaf trajectory. Intensity modulation along the profile, as shown in Figure 2-15, is obtained by sweeping the leading leaf, 12B, and the following leaf, 12A, from left to right along the x-axis. The first step in this procedure is to divide the total relative beam intensity into a number of equal intervals of width $\Delta\Phi$, as indicated in the illustration. The number of intervals selected to span the range of the intensity is NI. The second step in the procedure is to find the intersection of the centers of these profile increment bins with the profile. These points are

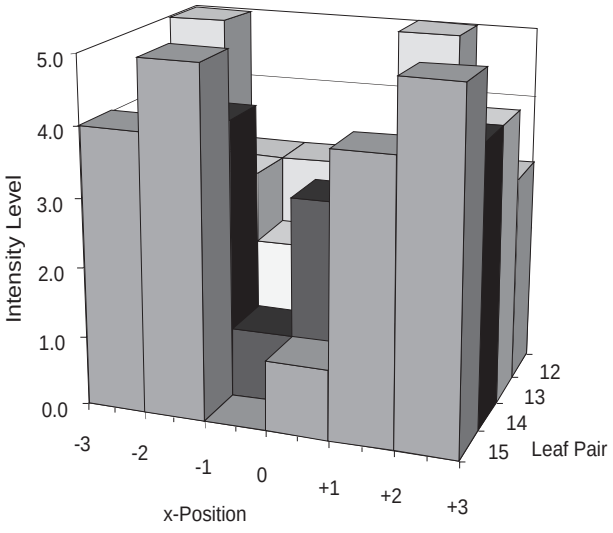


FIGURE 2-14. Intensity map used for illustrating the step-and-shoot leaf sequencing algorithm. (To view a color version of this image, please refer to the CD-ROM.)

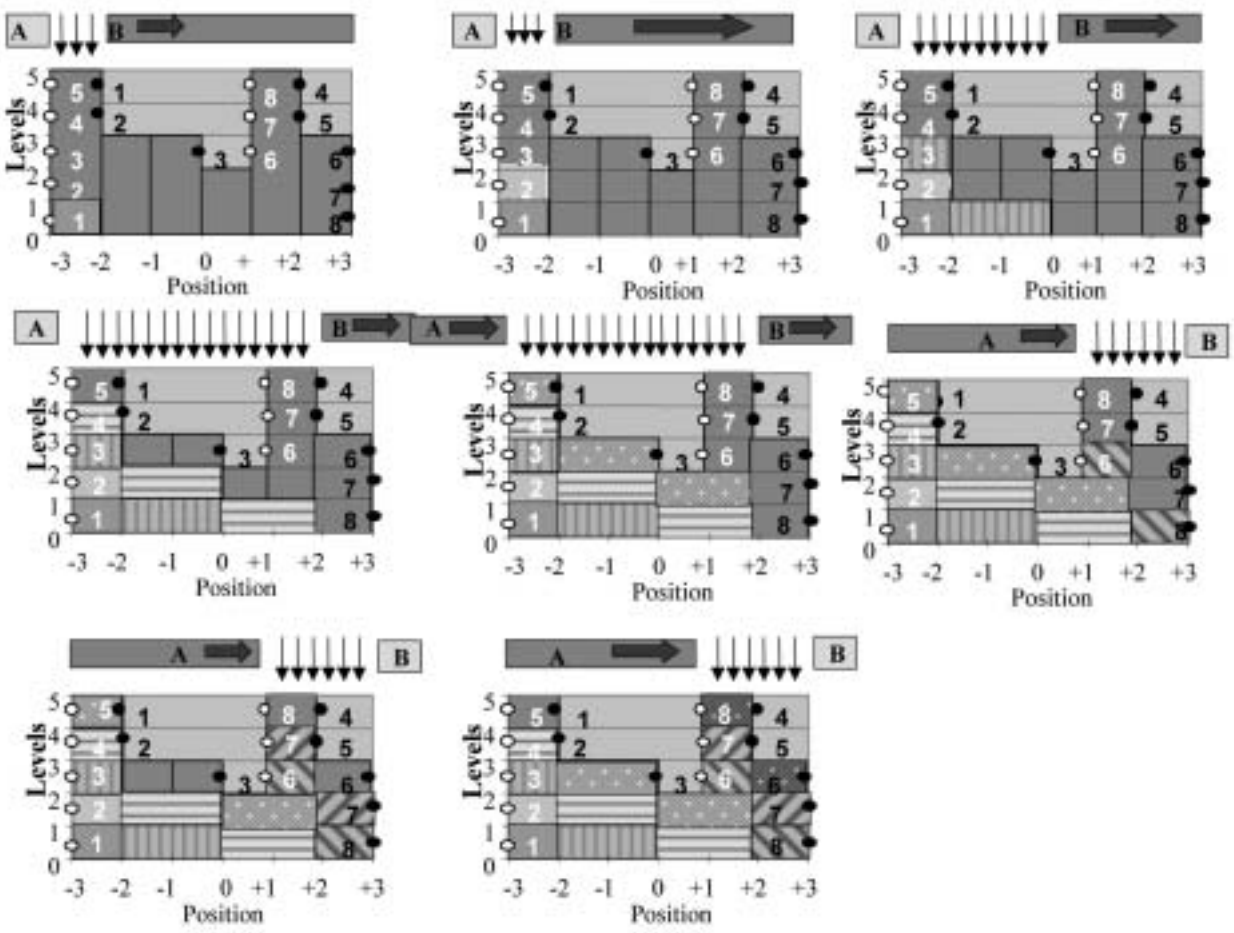


FIGURE 2-15. Intensity profile to be produced by leaf pair 12. Reproduced with permission.

indicated by circles in Figure 2-15. The algorithm requires that an even number of such points be found. The third step is to divide the coordinate points into two groups. One group consists of those points lying on an ascending slope of the profile in which there is a positive gradient (open circles in Figure 2-15), and the other group consists of those points lying on a descending slope of the profile in which there is a negative gradient (filled circles). The fourth step is to rank the points in each group. The numbers indicated are the i-index for the sequence for the twelfth pair of leaves. Pairing together the coordinates of equal rank order and assigning the coordinates to each pair of leaves produces the desired leaf sequence for the k^{th} gantry angle position, $\{xA_{i,j,k}, xB_{i,j,k}\}$, where the index i ranges from 1 to NI_j. The number of steps required to create the trajectories will not be the same for all profiles that make up a field. Steps must be added to the shorter sequences with the leaves abutting beneath a jaw at one end of the profile so that all sequences for a field will have the same number of steps.

Another type of step-and-shoot delivery is based on the sequential reduction of intensities according to a prespecified scheme.[112,113,120] The pattern of integers in Figure 2-16 represents an intensity pattern to be delivered using this leaf setting sequence. The 5×5 cm field is to be delivered with a maximum beamlet intensity of 10 and a minimum beamlet intensity within the field of 1. The underlying principle of the algorithm for determining the sequence is that the most efficient way to subdivide a sequence is by halves. The sequence is to be delivered by increments that are powers of 2. In this case, the increments are 8, 4, 2, and 1. The first step is to set the leaves in a pattern that can deliver an exposure of 8. There are four beamlets with intensities of 8 or more. They are not contiguous, but leaves can be set to form two windows around the two regions that each deliver an intensity of 8. This is step 1 in Figure 2-17. After this exposure, all but one of the beamlet positions still require an exposure of 1 or more. A leaf pattern can then be found that exposes beamlets that require a residual exposure of 4 or more. However, two such regions exist that require two separate sets of leaf settings. These are steps 2 and 3 in Figure 2-17. The residual intensity then contains values up to 3, which can be reduced by exposures of 2. Again, to expose all of the beamlets, two leaf patterns are required, each delivering exposures of 2. These are steps 4 and 5 in Figure 2-17. Then all of the beamlet positions have either received their full exposure are have a residual value of 1. Two more leaf patterns are required to reach all of the 1 positions and reduce the residual intensity to zero. In all, seven steps are required to deliver the intensity pattern. The single-profile step-and-shoot leaf setting algorithm requires 13 steps to deliver this pattern.

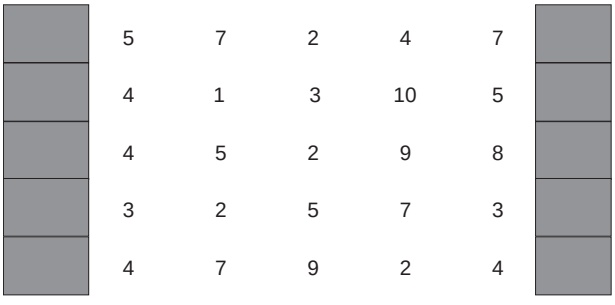


FIGURE 2-16. Example intensity map used for illustrating the "areal" leaf sequencing algorithm.

Dynamic Delivery

Let $\Phi(x)$ be the fluence along the trajectory of the leaf pair. An example profile is shown in Figure 2-18. To deliver the fluence, one must determine the arrival times at x, $t_A(x)$ for leaf A and $t_B(x)$ for leaf B. The units of the arrival times can be seconds, or they can be expressed as MUs. The irradiation time interval at x between the opening of the ray by leaf B and the shielding of the ray by leaf A is indicated by

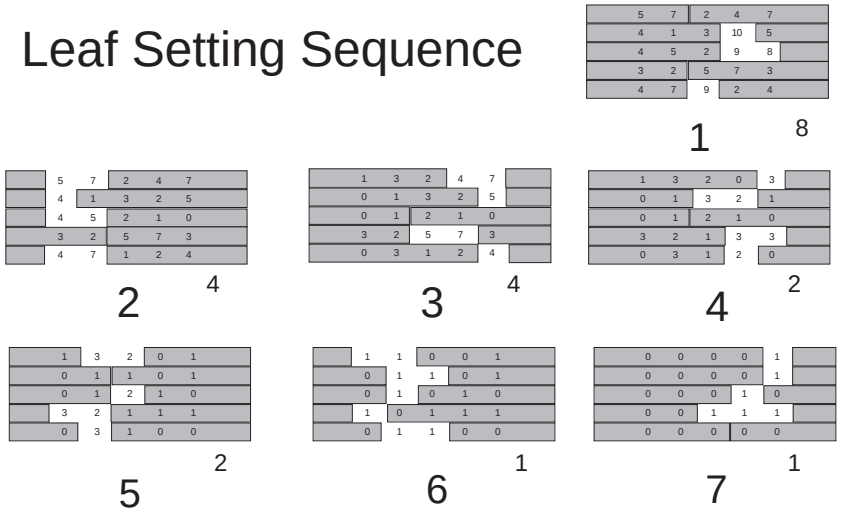


FIGURE 2-17. Leaf sequencing steps involved in decomposing the intensity map shown from Figure 2-16. The "areal" algorithm described by Xia and Verhey[112] is used here. Reproduced with permission.

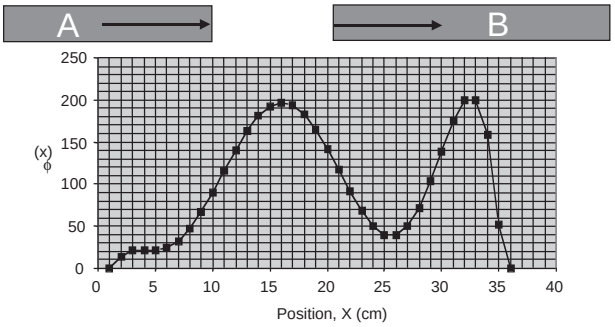


FIGURE 2-18. Example intensity map used for illustrating the dynamic leaf sequencing algorithm. Reproduced with permission.

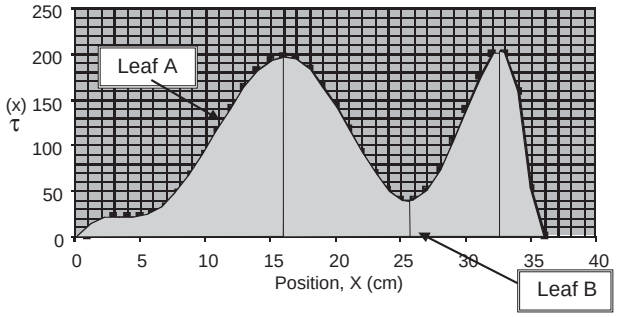


FIGURE 2-19. Time-position graph for the two leaves during the dynamic delivery process. Reproduced with permission.

$$\tau(x) = t_A(x) - t_B(x) = \Phi(x)/\Phi_0. \quad (7)$$

Figure 2-19 can then be considered to be a time-position graph for the two leaves. The upper border of the shaded area is the leaf A trajectory, and the lower border, or x-axis, is the leaf B trajectory. The problem with this interpretation is that it requires leaf B to travel with infinite velocity and leaf A to travel backward in time! The dilemma can be resolved by applying a sequence of operations that transform the two trajectories such that they become deliverable. Note that there are four regions marked along the fluence profile in which the gradient is either positive or negative. To remove the time reversal from the continuous fluence profile, a reflection operator is introduced and is defined by

$$\tau' = \Delta\,\tau_{R1} \pm [\tau(x) - \Delta\,\tau_{R1}] \quad (8)$$

$$\tau' = \Delta\,\tau_{R2} \pm [\tau(x) - \Delta\,\tau_{R2}] \quad (9)$$

where $\Delta\,\tau_{R1}$ is the average value of the portion of the profile with a negative gradient within the R1 region around the first maxima and $\Delta\,\tau_{R2}$ is the average value of the portion of the profile with a negative gradient within the R2 gradient region around the second maxima. The positive sign is applied when there is a positive gradient and the negative sign when there is a negative gradient. The reflection operator is applied to the curves in the negative gradient regions to yield curves that do not require the leaves to travel backward in time. The results are shown in Figure 2-22.

The operations have introduced a discontinuity in the leaf sequence curves that can be removed by applying a translation operator defined in region R1 by

$$\tau''(x) = \tau'(x) + \Delta\,\tau \quad (10)$$

where the increments are selected to remove the discontinuity between region R1 and R2, as illustrated in Figure 2-20. For the sake of generality, in region R1, the translation constant is zero. However, now there are still horizontal portions of the curves that represent infinite velocity of the leaves. There is always a horizontal segment occurring in either leaf A or leaf B trajectories across the entire sequence. To remove the infinite velocity, some additional slope is introduced to each leaf trajectory. This can be achieved by applying a shear operator to the entire lengths of both leaf trajectories. The shear operator is defined by

$$\tau'''(x) = \tau''(x) + x/\upsilon_{max} \quad (11)$$

This operator tilts the upper and lower horizontal bounds of each segment of the sequence by an amount determined by the maximum leaf velocity, resulting in a sequence that can be practically delivered. The slope of the shear is the inverse of the maximum velocity that the leaves can move, υ_{max}. The resulting leaf setting sequence is shown in Figure 2-20. The leaves begin the sequence closed at the left side of the field and end the sequence closed together at the right side of the field. In a region in which the original fluence gradient is positive, the leading leaf, leaf B, moves with a constant speed determined by the maximum velocity, and the trailing leaf A moves along the trajectory given in equation 11. In those regions in which the fluence gradient is negative, the trailing leaf, leaf A, moves with the maximum velocity, whereas the leading leaf moves along the trajectory given in equation 11.

The algorithm used to calculate the velocity modulation of the slower leaf can be derived by differentiating equation 12 with respect to distance:

$$d\tau'''/dx \equiv 1/\upsilon\,(x) = d\tau''/dx + 1/\upsilon_{max} \quad (12)$$

The derivative of τ'' with respect to x can be obtained from equation 10 and is simply the derivative of τ' in all subdivisions of the trajectory. The derivative of τ' can be obtained from equations 8 and 9 and depends on the sign of the fluence gradient:

$$\nabla\Phi < 0 \Rightarrow d\tau'/dx = -d\tau/dx \quad (13a)$$

$$\nabla\Phi > 0 \Rightarrow d\tau'/dx = +d\tau/dx \quad (13b)$$

The derivative of τ can be seen in equation 7 to be

$$d\tau/dx = (d\Phi/dx)/\Phi_0 \quad (14)$$

assuming that the variation in the incident fluence is negligible with respect to x. Using these results in equation 12, the velocity modulation equation becomes

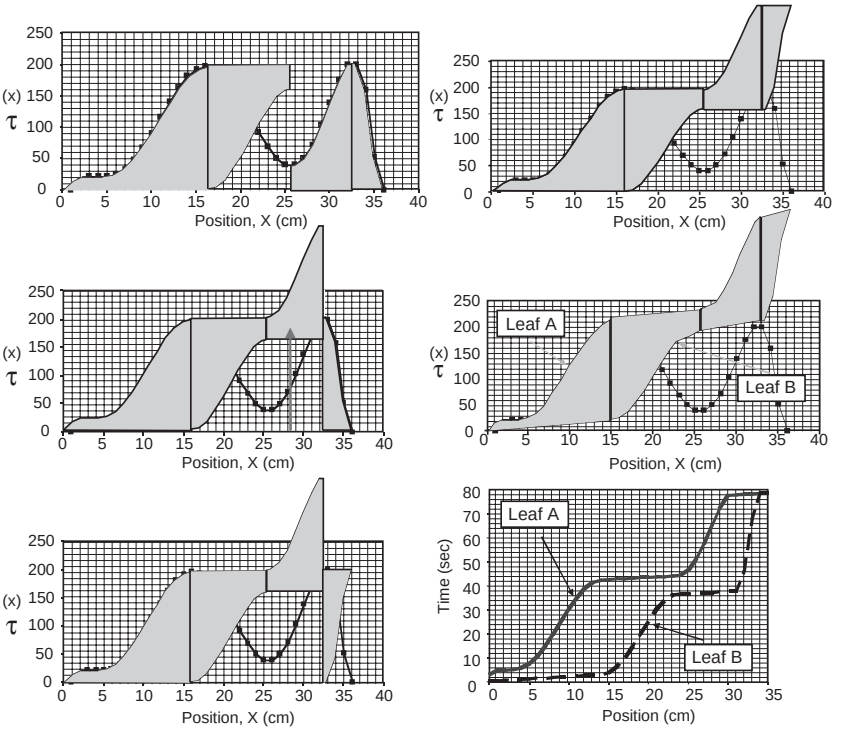


FIGURE 2-20. Trajectories of leaf A and leaf B during the dynamic delivery process. Reproduced with permission.

$$1/\upsilon(x) = \pm\,(d\Phi/dx)/\Phi_0 + 1/\upsilon_{max} \qquad (15)$$

where the positive sign applies to positive fluence gradient regions and the negative sign applies to negative fluence gradient regions. By rearranging, one arrives at

$$\upsilon(x) = \upsilon_{max}\,/[1 \pm \upsilon_{max} \cdot (d\Phi/dx)/\Phi_0] \qquad (16)$$

This equation can be used to generate the velocity modulation required to deliver the fluence profile, starting with the leaves closed together at one side of the profile and ending with the leaves closed together again at the other. The leaf setting sequence computed by the velocity equation for the original fluence in Figure 2-18 is shown in Figure 2-20. The results are exact.

Our experience with both step-and-shoot and dynamic delivery indicates that there is no clear-cut advantage for any one of the methods except in some special situations. The main disadvantage of the step-and-shoot method is a sacrifice of accuracy in the delivery of beam profiles that have steep gradients. The dynamic method delivers the required distribution by sweeping the leaf pair across the beam and becomes inefficient in producing "large and flat" fluence segments.[121] It is possible to implement an algorithm combining the step-and-shoot and the dynamic deliveries to use the advantages of each.[122] This scheme would determine the slope of each segment of the intensity profile and then choose the suitable delivery method.

Related Issues in MLC-Based Delivery

Unlike conventional RT with static MLC fields, significant dosimetric issues must be addressed when IMRT delivery is used. Most of the algorithms in the literature, however, assume an ideal MLC and ignore the influence of many physical effects and the mechanical constraints of a realistic MLC, such as transmission and head scatter, tongue-and-groove effects, and collision constraints for adjoining leaf pairs.

Yang and Xing proposed an algorithm to account for the leaf transmission and head scatter effects in step-and-shoot leaf sequencing.[123] In their approach, an error function, defined as the least square difference between the desired and the delivered fluence maps, is introduced. Mathematically, this function is expressed as

$$F = \sum_{i,j} [\varphi(i,j) - \varphi_d(i,j)]^2 \qquad (17)$$

where $\varphi(i,j)$ and $\varphi_d(i,j)$ are the calculated and the desired fluences of beamlet (i,j), respectively. In equation 18, only those beamlets with nonzero fluences in the desired intensity map are considered because one cannot physically produce a beamlet with zero fluence.

The calculation starts with the MLC leaf sequence file derived from the desired fluence map without considering

transmission and head scatter. The effects of transmission and head scatter are minimized by iteratively adjusting the fractional MUs in the initial MLC leaf sequences using a downhill simplex optimization method. A three-source model[124] is used to evaluate the relative head scatter contribution for each segment. The three effective sources are the source for the primary photons from the target and two extrafocal photon sources for the scattered photons from the primary collimator and the flattening filter, respectively. The algorithm has been assessed by comparing the dose distributions delivered by the corrected leaf sequence files and the theoretical predication, calculated by Monte Carlo simulation using the desired fluence maps and several clinical IMRT cases. The deviations between the desired fluence maps and the ones calculated using the corrected leaf sequence files are less than 0.3% of the maximum MU for the test field and less than 1.0% for the clinical IMRT cases. The experimental data demonstrate that both absolute and relative dose distributions delivered by the corrected leaf sequences agree with the desired ones within 2.5% of the maximum dose or 2 mm in high–dose gradient regions. It is found that the influence of the two effects is more pronounced in the absolute dose than in the relative dose. Figure 2-21 illustrates a measured absolute dose profile for a test field. In performing the measurements, MLC leaf sequences with different correction schemes described in the figure caption were used.

The influences of rounded leaf ends and interleaf transmission,[125] tongue-and-groove effect,[126] and the effect of back-scattered photons from the moving jaws and MLC leaves on the monitor chamber signal[127] have been studied using accurate models with realistic MLC geometries. It has been pointed out that the tongue-and-groove effect may be significant when underdosage occurs between two adjacent leaf pairs owing to the fact that the region between is always covered by the tongue, the groove, or both.[128,129] Algorithms have been developed to either minimize or remove this effect when MLC leaf sequences are generated.[114,130–133] Many researchers have shown that the tongue-and-groove effect can result in an underdose of as much as 10 to 15% in some special situations.[128,129,134,135] However, a Monte Carlo study by Deng and colleagues suggested that the difference between the dose distributions with and without the tongue-and-groove effect was hardly visible for an IMRT treatment with multiple gantry angles in a clinical setting.[126] More thorough investigations on the tongue-and-groove effect and other physical factors are needed to understand the system and to determine better solutions.

Finally, the inclusion of various physical factors is made simple if segment-based inverse planning is used. This represents one of the major advantages of the new type of inverse planning approach with integration of machine constraints.

Quality Assurance

IMRT adds a new degree of freedom to conventional RT and allows one to tune the dose distribution on an individual beamlet level. At the same time, it significantly increases the level of sophistication and complexity of the planning and delivery systems. With more and more institutions starting IMRT programs, it becomes increasingly important to have robust and efficient QA tools for clinical use. Otherwise, the gain from IMRT may be lost in a nonoptimal QA procedure and/or be offset by the increased cost of treatment. In general, IMRT QA has three aspects: commissioning and testing of the inverse treatment planning and IMRT delivery system, routine QA of the MLC delivery system, and patient-specific validation of each treatment plan. The first task is mainly concerned with the integrity of the IMRT system. The second involves the normal operation of the dynamic delivery system, and the third

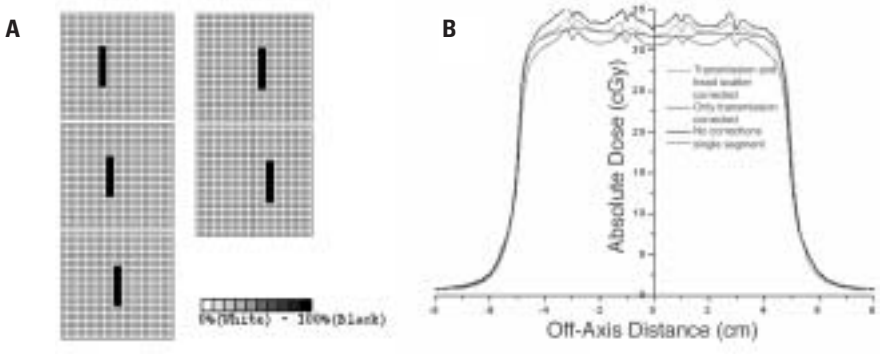


FIGURE 2-21. (*A*), A schematic diagram of the test field. The field includes five consecutive 2.0×10 cm^2 segments and attempts to produce a 10×10 cm^2 open beam. (*B*) The measured absolute dose profiles along the midline of leaf pair 21A–21B in the isocenter plane at a depth of 5 cm in solid water for the test field are shown on the left. The measured results of the single-segment 10×10 cm^2 open field with 30 monitor units are shown in the central black curve as a benchmark. The red curve is obtained with correction of head scatter and transmission. The top black and bottom blue curves represent the calculated dose profiles with only head scatter or transmission considered.[123] Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

task ensures accurate and safe treatment of the patient. Recently, there have been many excellent reports on IMRT QA–related issues.[60,136–139] Some practical aspects of IMRT QA are also discussed in Chapter 13 "Commissioning and Dosimetric Quality Assurance" and Chapter 14 "Quality Assurance Processes and Future Directions". In this section, the QA procedure and some recent advancements are summarized.

Commissioning and Testing

To ensure that the system can be used safely and accurately, the inverse treatment planning system must be commissioned prior to clinical use. Commissioning and testing consist of four separate but related steps. The first is concerned with the system's ability to accurately compute series of broad beam data. This type of testing is rudimental but useful to identify potential problems quickly. The second study tests the dose model and the delivery system with several specially designed intensity patterns. The accuracy of dose calculation for intensity-modulated beams can thus be assessed. The third type of study examines the system's functionality and dosimetric correctness for a number of hypothetical phantom cases. In addition to dose calculation, the functionality of dose optimization is evaluated at this level of tests. Figure 2-22 shows two examples of this type of measurement using a cylindrical water phantom and ion chamber.[140] The last type of study is to test the system using clinical cases to ensure the dosimetric accuracy and integrity of the system. This study evaluates the combination results of image acquisition and segmentation, geometric and dosimetric calibration of the planning system, planning and dose calculation, and data transfer. The dose distributions for single or multiple fields are usually done using an ion chamber and films in a phantom. Other dosimeters, such as thermoluminescent dosimeters and semiconductor detectors, can also be employed. The American Association of Physicists in Medicine (AAPM) Task Group 40 and Task Group 53 reports provide guidelines on this topic and remain the benchmark documents on the subject.[141,142] This subject has also been discussed extensively in recent publications.[136,137,143]

Routine Machine QA

Intensity modulation is achieved with computer-controlled MLC using either static[112,144–146] or dynamic delivery techniques. To ensure that the planned dose distributions are safely and accurately delivered, an important requisition is the normal operation of the delivery system, which is warranted by routine machine QA. The principles and practice of QA for RT can be found in the classic documents of Van Dyk and Purdy,[147] as well as the report of AAPM Task Group 40.[141] For IMRT, several things specific to the IMRT MLC control system need to be checked periodically. Currently, the routine accuracy check of MLC leaf positioning in most clinics is performed using radiographic films with specially designed MLC leaf sequences.[148,149] Besides being time consuming, the results of film measurements are difficult to quantify and interpret. A few research groups have attempted to use an electronic portal imaging device (EPID) for quantitative verification of MLC leaf positions with edge detection algorithms.[150–153] The detection precision is limited to ~ 1 mm owing to the finite pixel size and the signal-to-noise ratio of the EPID, which is clearly insufficient for routine QA of the MLC delivery system. Here we describe a quantitative technique[154] for MLC leaf positioning QA developed recently at Stanford University. Given its simplicity, efficiency, and accuracy, we believe that the technique is ideally suitable for routine MLC QA and should have widespread clinical application in the future.

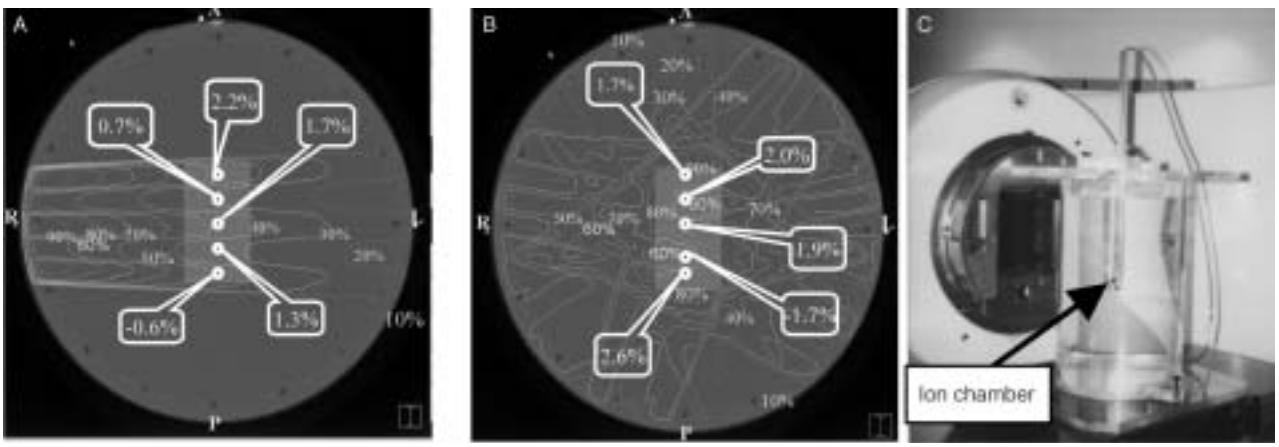


FIGURE 2-22. Hypothetical intensity-modulated radiation therapy plans generated for a cylindric water phantom and the measured dose distributions. The phantom is positioned with its axis perpendicular to the couch top and is supported by a bearing, allowing for rotation about its axis. This allows for measuring the dose of a multifield plan without gantry rotation. The measurements were made using the Varian dynamic multileaf collimator modulating 4 MV x-ray beams. The plans were generated using the *CORVUS* system. Reproduced with permissionfrom Xing et al. [140] (To view a color version of this image, please refer to the CD-ROM.)

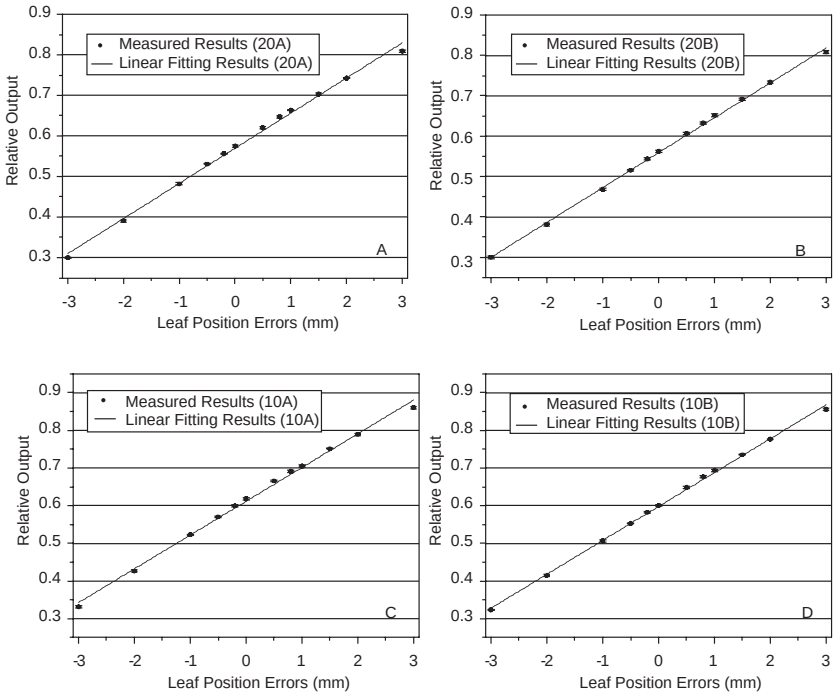


FIGURE 2-23. Relative output versus the displacement of leaves 20A, 20B, 10A, and 10B from their desired positions. The symbols are the measured data, and the solid lines are the least square fitting of the corresponding data sets. Reproduced with permission from Yang Y and Xing L.[154]

The Stanford MLC QA technique uses the fact that when a finite-sized detector is placed under a leaf, the relative output of the detector will depend on the relative fractional volume irradiated. A small error in leaf positioning would change the fractional volume irradiated and lead to a deviation of the relative output from the normal reading. For a given MLC and detector system, the relationship between the relative output and the leaf displacement can be easily established through experimental measurements and used subsequently as a quantitative means for detecting possible leaf positional errors. Figure 2-23 illustrates a set of calibration curves for different leaves obtained using an ion chamber and a Varian CL 2300C/D accelerator with an 80-leaf MLC.[154] Our results indicate that the method could accurately detect a leaf positional change of ~ 0.1 mm. The principle of the method is independent of the type of MLC and detector. The method overcomes the previously stated shortcomings of both film measurement and edge detection techniques and provides a reliable means for quantitative examination of MLC positional accuracy.

The principle has also been applied to MLC leaf positioning QA using an EPID,[155] which has the advantage of simultaneously detecting positional errors of any leaf at any point. In this technique, the active imaging region of an EPID is divided into a number of small rectangular regions of interest, each of which is centered at a point at which the leaf positioning accuracy is to be examined (Figure 2-24). Every region of interest here acts as a finite-sized detector, and the integral signal from it can be processed based on the pre-established relation between the integral signal and the leaf displacement at the point. The EPID-based system also allows us to take the dosimetric influence of the adjacent leaves into account. For this purpose, the integral signal at a region of interest is expressed as a weighted sum of the contributions from the displacements of the leaf above the point and the adjacent leaves. The linear coefficients of the system equations are determined by fitting the integral signal data for a group of predesigned MLC leaf sequences to the known leaf displacements that are intentionally introduced during the creation of the leaf sequences. Once the calibration is done, the system can be used for routine MLC

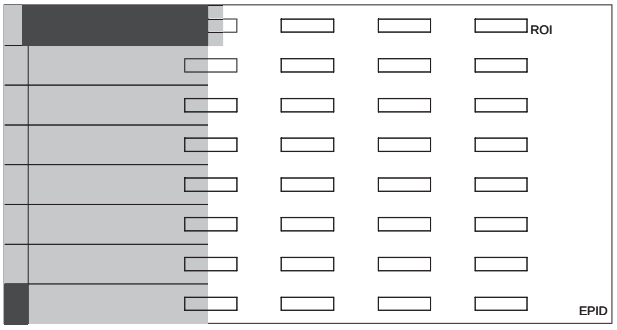


FIGURE 2-24. Diagram of the geometric setup using an electronic portal imaging device (EPID) to examine leaf positioning accuracy. (To view a color version of this image, please refer to the CD-ROM.)

leaf positioning QA to detect possible leaf errors. Table 2-1 shows a set of test data obtained using the technique. Overall, our results show that the proposed technique is superior to the conventional edge-detecting approach in two aspects. First, it deals with the problem in a systematic approach and allows one to take into account the influence of the adjacent MLC leaves effectively. Second, it has a much higher signal-to-noise ratio and is thus capable of quantitatively measuring extremely small leaf positional displacements. The technique can effectively detect a relative lead positional error as small as 0.1 mm at an arbitrary point within the field in the absence of an EPID setup error and 0.3 mm when this uncertainty is considered.

IMRT Treatment Plan Validation

The tasks of patient-specific QA can be divided into geometric and dosimetric verification. The former is concerned with the geometric accuracy of the IMRT beams, including isocenter and portal verification. The dosimetric verification includes a quantitative check of fluence maps, radiation doses at multiple points, and, in some cases, the dose distribution. Currently, the dosimetric verification is primarily done experimentally.

Geometric Verification

A pair of orthogonal simulation films (or digital reconstruction radiographs [DRRs]) is used to verify the patient position by comparison with portal films. In 3DCRT, a portal image is taken using the double-exposure technique, one with the customized radiation port and the other with a larger rectangular open field, so that both the field boundary and selected patient anatomy can be visualized. A simulation image for an IMRT field can be created as well using the MLC boundary as the port of the radiation field. An example of such a portal image for an IMRT head and neck treatment is shown in Figure 2-25. For portal image exposure, an MLC field that defines the field boundary needs to be extracted from the IMRT leaf sequence file. The MLC-defined field aperture can be appended to the DRR to be displayed together with the patient's anatomy. The DRR in the beam's eye view, as shown on the left in Figure 2-25, is used as a reference for comparing with the portal image for target localization during the treatment.[156]

TABLE 2-1. Detected Leaf Positional Errors with Different Intentionally Introduced Errors for Different B-Bank Leaves at Two Locations

Location	*Leaf Number*	*IIE, mm*	*DLPE, mm*
X = –10 cm	16B	0.3	0.31
	20B	1.6	1.55
	23B	–0.5	–0.51
	31B	0.1	0.11
	33B	0.4	0.42
	34B	0.5	0.49
	37B	–0.4	–0.42
	44B	–0.2	–0.18
	45B	2.3	2.37
	50B	–0.5	–0.50
X = 0 cm	13A	0.8	0.76
	14A	–0.2	–0.19
	19A	–1.2	–1.20
	22A	–0.1	–0.09
	27A	2.7	2.66
	31A	0.1	0.11
	39A	–0.8	–0.79
	44A	0.2	0.22
	45A	–0.3	–0.28
	46A	–1.8	–1.81

DLPE = detected leaf positional error; IIE = intentionally introduced error.

Dosimetric Verification and Independent Dose and Fluence Map Calculations

No consensus has emerged regarding what dosimetric quantities need to be examined to validate an IMRT treatment plan. Patient-specific dosimetric QA typically consists of dose measurements at multiple points and fluence map measurements. Some institutions also perform film dosimetry for each patient treatment. Because of the inherent complexity of the problem, it may be some time before definitive recommendations come from national organizations. In general, the goal of the dosimetric verifications is to ensure that the delivered dose distribution agrees with the one from the treatment planning system. The descriptions on equipment and procedure for these measurements have been the subject of a few recent review articles. The fundamental philosophy of IMRT QA and our experience with computer-based patient-specific QA are presented here.

First, one should note that the 3DCRT approach based on point dose verification is insufficient to validate an IMRT plan because of the independence of the involved beamlets. In 3DCRT, verification is mainly concerned with the MU calculation for each incident field. An independent

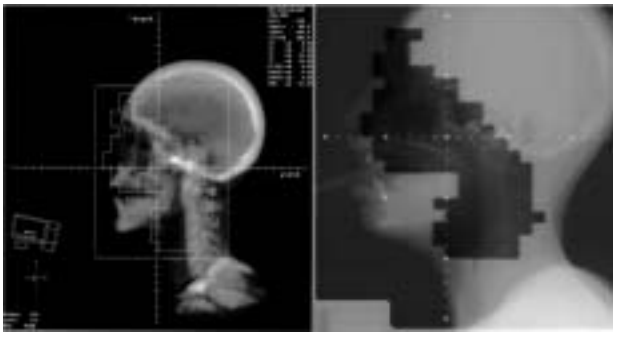


FIGURE 2-25. Top: A left digitally reconstructed radiograph *left* with the field boundary of the IMRT field. A double-exposure portal image for the same field is shown on the right. Bottom: An anterior posterior digitally reconstructed radiograph *left* with the field boundary of the IMRT field. A double-exposure portal image for the same field is shown on the right. These images are used by physicians to verify the maximum extent of the IMRT treatment fields.

calculation of the dose or MU at a point based on primitive machine data is recommended by AAPM Task Group 40.[141] Because the fluence of a uniform or wedged field is spatially correlated, information of the dose at a point can, in principle, be used to estimate the dose in other points provided that the off-axis information is known. However, this is not the case for intensity-modulated fields because the weights of the beamlets across a field are independent. The correctness of the dose at a spatial point warrants, at most, only the correctness of the beamlets passing through or nearby that point. To validate an IMRT treatment plan, the spatial distributions of the beamlets must be verified in addition to the point dosimetric check.

In practice, the above two tasks can be achieved by the verification of point dose(s) and fluence maps. The fluence map of an incident beam is usually normalized to the maximum beamlet weight in the beam. For a given intensity-modulated field, the verification of the fluence map or beamlet correlation ensures the correctness of the doses at other points once the dose(s) at one or more points inside the field is examined. Together with the point doses, they provide information on the integrity of the IMRT fields.

We now discuss how to efficiently carry out the two types of tests. Obviously, the most robust method is to measure the point doses and fluence maps to validate an IMRT plan. As depicted as the dashed lines in Figure 2-26, the approach checks both planning and delivery. Its drawback is that an intensive effort is needed to carry out the measurement for each field or patient. Alternatively, one can separate the QA of the delivery and planning systems, as illustrated by the solid lines at the bottom. Although QA of the delivery system is imperative, its goal should be practically achievable by periodical checks rather than actual measurement before each patient's treatment. The division of IMRT QA into machine QA and patient-specific QA allows us to check the integrity of an IMRT treatment plan by using computer calculation, simplifying the pretreatment QA. In fact, the same philosophy has been used in 3DCRT over the years, in which a manual calculation is often used instead of actual point dose measurement to validate the patient-specific MU settings.

Algorithms to perform the independent point dose and fluence map calculations for IMRT have been reported recently.[157–162] Here a general formalism for the IMRT point dose check used at Stanford University Hospital (the software, *IMSure*, has been commercialized by Prodigm Inc., Chico, CA) is described. In this approach, the dose at an arbitrary spatial point is expressed as a summation of the contributions from all of the beamlets with the amplitude of each beamlet modulated by a dynamic modulation factor. The dynamic modulation factor represents the fractional time that the beamlet is "open" during the dynamic delivery process and can be computed once the MLC leaf sequences are known.[159] The dose at a point (x, y, z) is written as

$$D(x,y,z) = MU\sum_{m}^{M} C_m D_m^0 \tag{18}$$

where the D^0_m is the dose contribution to the calculation point per MU from the m-th beamlet when it is open, MU is the total monitor unit, and C_m is the dynamic modulation factor. When the MLC leaf transmission and head scatter effects are taken into account, C_m can be calculated by[160]

$$C_m = \sum_{k}^{K}[Sc_{m,k} + \alpha Sc'(1-\delta_{m,k})]f_k \tag{19}$$

with

$$\delta_{m,k} = \begin{cases} 1 & \text{if} \quad m \in A_k \\ 0 & \text{if} \quad m \notin A_k \end{cases} \tag{20}$$

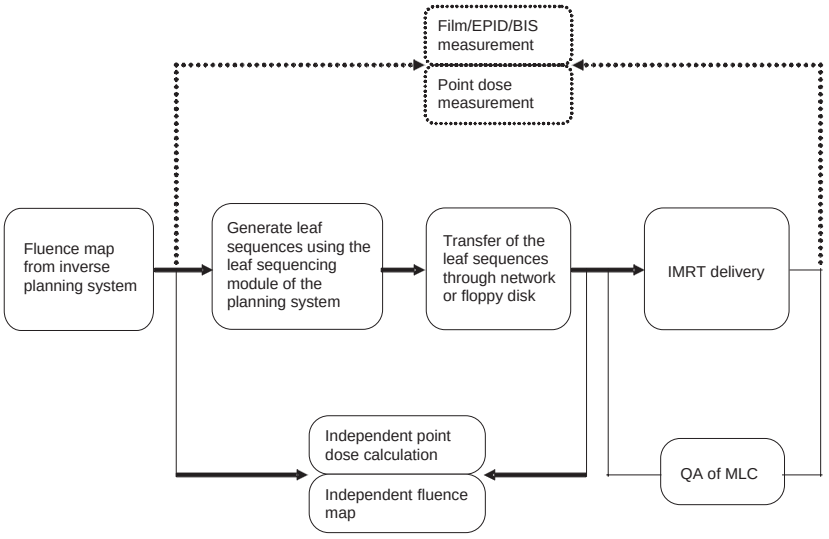


FIGURE 2-26. Intensity-modulated radiation therapy (IMRT) plan validation process. The fluence map/point dose verification is depicted by the dashed line on the top. The computer-based approach is outlined as the solid lines at the bottom of the figure. BIS = beam imaging system ; EPID = electronic portal imaging device; MLC = multileaf collimator; QA = quality assurance.

where f_k is the fractional MU of the k-th segment and A_k is the radiation field shape of the k-th segment. $Sc_{m,k}$ is the head scatter factor of the beamlet m in the k-th segment, Sc' is the head scatter factor for the rectangular field defined by the jaws, and α is the average transmission factor. The head scatter factor $Sc_{m,k}$ for each beamlet in a segment is calculated using the three-source model described earlier.[124]

Computer verification of the fluence maps or the MLC leaf sequences can be done similarly. The software reads in the leaf sequences and simulates the motion of the MLC leaves.[163] The computed fluence map is then compared quantitatively with the intended map from the treatment planning system. A set of predefined QA indices are introduced to measure the "closeness" between the computed and the reference maps. The implication of the simulation is twofold. By comparing the recalculated fluence map with that from the planning system, it examines the functionality of the leaf sequencer of the planning system and ensures that the leaf sequence is executable and correct. It can also detect possible errors that occur during the transfer process of the leaf sequence file from the planning computer to the MLC workstation. The goal of the simulation is to warrant that, assuming that a rigorous independent QA of the MLC system has been performed so that the dynamic MLC can accurately execute the instruction of a leaf sequence file, the execution of the leaf sequence will generate the desired fluence map should it pass the simulation test.

Because of the simplicity and reliability of computer-based IMRT plan validation, it becomes clinically practical to enforce QA of the point doses and fluence maps on an individual patient or field basis. Furthermore, the method is valid for both step-and-shoot and dynamic deliveries. The utility of the computer verification has been demonstrated by the many clinical IMRT cases at many institutions, and its widespread use should simplify the QA procedure. However, it is important to keep in mind that experimental measurement is the only reliable source for IMRT plan validation. Any computer-based validation tool must be validated by experimental means before its clinical use.

Special IMRT Techniques and Machine Limitations

Concurrent Boost

One of the advantages of IMRT is its ability to deliver different dose levels to different regions simultaneously so that target volumes with different prescription doses can be planned and delivered together (see Chapter 18.7, "Simultaneous Integrated Boost: Emerging Technology").[96,164] This approach has several potential advantages. Besides the efficiency of planning and delivery with a single plan, the resulting dose distribution can be more optimal. The conventional sequential boost strategy employs two or more independent plans in which the initial fields cover the elective regions and smaller boost fields focus on the primary target. The boost dose is often limited by the tolerances of nearby OAR, which have been given a significant amount of radiation. If planned simultaneously using IMRT, it is possible to distribute the dose evenly among fractions, and the system also has a greater degree of freedom to optimize the intensity among many beams. There are biologic advantages as well; for example, the shortened treatment course and increased dose per fraction to primary tumors can often be translated into a higher biologically equivalent dose, thus increasing the probability of local control.

Treatment of Large Tumors

The treatment of large tumors necessitates the use of large treatment fields. Depending on the implementation of the MLC by the linear accelerator vendors, the maximum field size formed by dynamic MLC may be different from those imposed by the collimators (jaws). Typically, the maximum field size is much smaller. For example, in the Varian MLC, the jaws and the MLC carriages do not move with the leaves. The leaf length in the current model of the MLC is 14.5 cm (projected at the isocenter). Given that each leaf pair must travel from the left boundary to the right boundary of the beam aperture and the back end of any leaf cannot travel past the edge of the jaw, the maximum width of the field aperture that can be accommodated in one sweep of leaves is also limited to 14.5 cm. The maximum IMRT field size that can be delivered in one sweep is 40×14.5 cm for an 80-leaf MLC (or a 120-leaf MLC) or 26×14.5 cm for a 52-leaf MLC.

To treat a target volume wider than 14.5 cm, an incident field must be split into two or more subfields unless some special techniques are used.[165] A simple step "break" in the middle, as is usually done for static treatments with MLC, may be implemented. Although this is certainly feasible, it could lead to field matching problems because uncertainties in patient setup and leaf positioning may cause undesirable hot or cold spots in the junctioning region. Given that the intensity varies across the field in IMRT anyway, it is natural to consider splitting the beam into components with overlap between them having variable intensity in the overlap region. A simple dynamic "feathering" technique for splitting large fields has been proposed by Wu and colleagues.[166] In this method, the intensity-modulated field is divided into two (or more) components. The components overlap each other, and the intensity gradually decreases in the overlap region for one component and increases for the other. The sum of intensities remains the same as for the original field. Each component is delivered using the sweeping window technique with the dynamic MLC. This method provides a smooth transition from one field component to the next, thereby eliminating the field junction problems. The dynamic feathering technique may also be applied to split large static fields to minimize the junction problem.[94,167] The feathering technique has been

extended to treat the whole abdomen area, in which splitting into more than two beams in the leaf motion direction may be necessary. Also, field sizes larger than 40 cm may be required in the cephalad-caudad direction, leading to the use of multiple isocenters, and feathering (not splitting) is helpful.[168]

Dose Matching of an IMRT Plan with a 3DCRT or an IMRT Plan

One of the important problems in RT of breast cancer, Hodgkin's disease, head and neck cancer, and cervical cancer is the matching of an IMRT dose distribution for the treatment of part of the target volume(s) with a conventional 3DCRT or IMRT plan for the treatment of a different portion of target volume(s). Ideally, dose optimization of the second part should take into account the existing dose from the previous plan to optimally match the two dose distributions.[94,169]

The two treatment plans that need to be matched are generally produced sequentially. The first plan used for treating part of the tumor volume(s) is obtained with the consideration of the second plan. To reduce the sensitivity of potential setup errors, an attempt needs to be made to "blur" the penumbra or dose gradient in the direction perpendicular to the matchline. Specifically, instead of a sharp dose gradient, the dose is allowed to extend by an additional 1.5 to 5 cm in the direction perpendicular to the matchline. In this transition region, the dose is forced to fall off linearly. The overlap is generally determined by the desired sensitivity against setup error. After the first plan is done with the extended transitional dose gradient region, the second IMRT plan is optimized with consideration of the existing doses of the first plan. The goal of the second dose optimization is to obtain an IMRT plan that yields a uniform composite dose distribution in the target volume(s) and (including the transitional regions) while sparing the sensitive structures. The approach takes advantage of the state-of-the-art intensity modulation and dose optimization techniques and provides an effective solution to the timely clinical problem of IMRT dose matching. In addition to better dose uniformity in the target volumes in the matchline region, it reduces the sensitivity of the doses to setup uncertainties in the matchline region. The technique is not yet available in commercial planning systems but should be implemented in the near future.

Radiation Protection Issues

Generally speaking, IMRT tends to use more beams (than traditional approaches) to conform the isodose curves to the shape of the tumor volume. As a consequence, a larger volume of normal tissue is exposed to lower doses as opposed to a smaller volume of normal tissue irradiated by higher doses in 3DCRT. In addition, the number of MUs is often increased by a factor of 2 to 3 owing to dynamic intensity modulation, increasing the total-body exposure, which may increase the risk of secondary malignancies.[170,171] Hall and Wuu theoretically compared IMRT and 3DCRT and suggested that both factors tend to increase the risk of secondary cancers.[172] Altogether, IMRT is likely to almost double the incidence of secondary malignancies compared with conventional RT (from about 1 to 1.75% for patients surviving 10 years or more). The risk may be larger for patients with longer survival rates (and for younger patients), but the ratio should remain the same (see Chapter 3 "Radiobiology of IMRT" and Chapter 30 "Pros and Cons of IMRT").

Reduction of the number of segments using more advanced dose optimization techniques and/or appropriate shielding of the treatment room are crucial to reduce the potential risks to hospital personnel. The National Council on Radiation Protection and Measurements has developed an empiric method for designing shielding against ionizing radiation that will protect workers and the general public from harmful radiation exposures.[173,174] These methods have been used for several decades, and additional information that can be used in conjunction with these methods has since been published.[175–177] A thorough study of IMRT shielding design has been presented by Mutic and colleagues[178] and Low.[179]

Summary

Institutions worldwide are attempting or planning to integrate IMRT technology into their clinics. Before IMRT implementation, it is important to understand the physical principles behind the overall process of inverse planning and dynamic deliveries. This will help in making better decisions regarding which system best suits each clinical environment and facilitates the implementation process. The efficiency and quality of IMRT treatment depend on many factors. At this point, it seems that timely developments of inverse planning and QA techniques are highly desirable to make IMRT a truly superior and robust treatment modality. With these advancements, it is anticipated that IMRT will provide improved dose distributions with less effort in treatment planning, delivery, and verification.

Acknowledgments

We would like to thank J. G. Li, A. Pugachev, S. Crooks, C. Cotrutz, J. Lian, S. Hunjan, Z. Shou, E. Schreibmann, Y. Chen, G. Luxton, T. Pawlicki, Q. Le, S. Hancock, C. King, I. Gibbs, and B. Loo for many useful discussions. We also wish to acknowledge the support from the American Cancer Society (RSG-01-022-01-CCE), National Cancer Institute (5R01CA098523-02), Department of Defense (DAMD17-03-1-0023 and DAMD17-03-1-0019), and Vadasz Family Foundation. Last but not least, we wish to thank International Journal of Radiation Oncology, Biology, Physics, Medical Physics, and Physics in Medicine and Biology for permis-

sion to use their copyrighted materials.

References

1. Webb S. Intensity-modulated radiation therapy. Bristol (UK): Institute of Physics Publishing; 2001.
2. Webb S. Historical perspective on IMRT. In: Palta JR, Mackie TR, editors. AAPM intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 1–24.
3. Verhey LJ. Comparison of three-dimensional conformal radiation therapy and intensity-modulated radiation therapy systems. Semin Radiat Oncol 1999;9:78–98.
4. Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. Phys Med Biol 1995;40:1435–49.
5. Yu CX, Li XA, Ma L, et al. Clinical implementation of intensity-modulated arc therapy. Int J Radiat Oncol Biol Phys 2002;53:453–63.
6. Wong E, Chen JZ, Greenland J. Intensity-modulated arc therapy simplified. Int J Radiat Oncol Biol Phys 2002;53:222–35.
7. Duthoy W, De Gersem W, Vergote K, et al. Whole abdominopelvic radiotherapy (WAPRT) using intensity-modulated arc therapy (IMAT): first clinical experience. Int J Radiat Oncol Biol Phys 2003;57:1019–32.
8. Crooks SM, Wu X, Takita C, et al. Aperture modulated arc therapy. Phys Med Biol 2003;48:1333–44.
9. Crooks S, King C, Pawlicki T, et al. Towards an optimal conformal technique for prostate radiotherapy: a dosimetric comparison between 3D CRT, IMRT, and IMAT. 2002. [In press].
10. Yang Y, Xing L. Clinical knowledge-based IMRT plan optimization. [Submitted]
11. Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. Med Phys 1997;24:103–10.
12. Wu Q, Mohan R, Niemierko A, et al.Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. Int J Radiat Oncol Biol Phys 2002;52:224–35.
13. Thieke C, Bortfeld T, Niemierko A, et al. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. Med Phys 2003;30:2332–42.
14 Brahme A. Optimized radiation therapy based on radiobiological objectives. Semin Radiat Oncol 1999;9:35–47.
15. Niemierko A, Goitein M. Modeling of normal tissue response to radiation: the critical volume model. Int J Radiat Oncol Biol Phys 1993;25:135–45.
16. Niemierko A. Radiobiological models of tissue response to radiation in treatment planning systems. Tumori 1998;84:140–3.
17. Schultheiss TE, Orton CG, Peck RA. Models in radiotherapy: volume effects. Med Phys 1983;10:410–5.
18. Schultheiss TE, Orton CG. Models in radiotherapy: definition of decision criteria. Med Phys 1985;12:183–7.
19. Martel MK, Ten Haken RK, Hazuka MB, et al. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. Lung Cancer 1999;24:31–7.
20. Martel MK. NTCP modeling for normal lung and tumors: analysis of clinical data. 3D conformal and intensity modulated radiation therapy: physics & clinical applications. Middleton (WI): Advanced Medical Physics Publishing; 2001. p. 489.
21. Kwa SL, Lebesque JV, Theuws JC, et al. Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients. Int J Radiat Oncol Biol Phys 1998;42:1–9.
22. Mohan R, Wang X, Jackson A, et al. The potential and limitations of the inverse radiotherapy technique. Radiother Oncol 1994;32:232–48.
23. Wang XH, Mohan R, Jackson A, et al. Optimization of intensity-modulated 3D conformal treatment plans based on biological indices. Radiother Oncol 1995;37:140–52.
24. Kallman P, Lind BK, Brahme A. An algorithm for maximizing the probability of complication-free tumour control in radiation therapy. Phys Med Biol 1992;37:871–90.
25. Xing L, Li JG, Donaldson S, et al. Optimization of importance factors in inverse planning. Phys Med Biol 1999;44:2525–36.
26. Xing L, Chen GTY. Iterative algorithms for inverse treatment planning. Phys Med Biol 1996;41:2107–23.
27. Xing L, Hamilton RJ, Spelbring D, et al. Fast iterative algorithms for three-dimensional inverse treatment planning. Med Phys 1998;25:1845–9.
28. Brooks RA, DeChiro G. Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. Phys Med Biol 1976;21:689–732.
29. Jones L, Hoban P. A method for physically based radiotherapy optimization with intelligent tissue weight determination. Med Phys 2002;29:26–37.
30. Luenberger D. Introduction to linear and nonlinear programming. Reading (MA): Addison-Wesley; 1973.
31. Kirkpatrick S, Gelatt C, Vecchi M. Optimization by simulated annealing. Science 1983;220:671–80.
32. Metropolis N, Rosenbluth A, Rosenbluth M, et al. Equation of state calculation by fast computing machines. J Chem Phys 1953;21:1087–91.
33. Rosen, II, Lam KS, Lane RG, et al. Comparison of simulated annealing algorithms for conformal therapy treatment planning. Int J Radiat Oncol Biol Phys 1995;33:1091–9.
34. Webb S. Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator. Phys Med Biol 1991;36:1201–26.
35. Winkler G. Image analysis, random field and dynamic Monte Carlo methods. Berlin: Springer-Verlag; 1995.
36. Webb S. Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator: II. Inclusion of two-dimensional modulation of the x-ray intensity. Phys Med Biol 1992;37:1689–704.
37. Bortfeld T, Burkelbach J, Boesecke R, et al. Methods of image reconstruction from projections applied to conformation radiotherapy. Phys Med Biol 1990;35:1423–34.
38. Holmes T, Mackie TR. A filtered backprojection dose calculation method for inverse treatment planning. Med Phys 1994;21:303–13.

39. Pugachev AB, Boyer AL, Xing L. Beam orientation optimization in intensity-modulated radiation treatment planning. Med Phys 2000;27:1238–45.
40. Andrew HC, Hunt BR. Image restoration. Upper Saddle River (NJ): Prentice Hall; 1977.
41. Starkschall G. A constrained least-squares optimization method for external beam radiation therapy treatment planning. Med Phys 1984;11:659–65.
42. Starkschall G, Eifel PJ. An interactive beam-weight optimization tool for three-dimensional radiotherapy treatment planning. Med Phys 1992;19:155–63.
43. Lee EK, Fox T, Crocker I. Optimization of radiosurgery treatment planning via mixed integer programming. Med Phys 2000;27:995–1004.
44. Bednarz G, Michalski D, Houser C, et al. The use of mixed-integer programming for inverse treatment planning with pre-defined field segments. Phys Med Biol 2002;47:2235–45.
45. Langer M, Morrill SS, Brown R, et al. A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy. Med Phys 1996;23:957–964.
46. Lee EK, Zaider M. Intraoperative dynamic dose optimization in permanent prostate implants. Int J Radiat Oncol Biol Phys 2003;56:854–61.
47. Hou Q, Wang J, Chen Y, et al. An optimization algorithm for intensity modulated radiotherapy—the simulated dynamics with dose-volume constraints. Med Phys 2003;30:61–8.
48. Langer M, Brown R, Morrill S, et al. A generic genetic algorithm for generating beam weights. Med Phys 1996;23:965–71.
49. Ezzell GA, Gaspar L. Application of a genetic algorithm to optimizing radiation therapy treatment plans for pancreatic carcinoma. Med Dosim 2000;25:93–7.
50. Wu X, Zhu Y. A mixed-encoding genetic algorithm with beam constraint for conformal radiotherapy treatment planning. Med Phys 2000;27:2508–16.
51. Cotrutz C, Xing L. Segment-based dose optimization using a genetic algorithm. Phys Med Biol 2003;48:2987–98.
52. Xing L, Li JG, Pugachev A, et al. Estimation theory and model parameter selection for therapeutic treatment plan optimization. Med Phys 1999;26:2348–58.
53. Lian J, Cotrutz C, Xing L. Therapeutic treatment plan optimization with probalistic dose prescription. Med Phys 2003;30:655–66.
54. Winkler RL. An introduction to bayesian inference and decision. New York: Holt, Rinehart & Winston; 1972.
55. Llacer J. Inverse radiation treatment planning using the dynamically penalized likelihood method. Med Phys 1997;24:1751–64.
56. Llacer J, Solberg TD, Promberger C. Comparative behaviour of the dynamically penalized likelihood algorithm in inverse radiation therapy planning. Phys Med Biol 2001;46:2637–63.
57. Wu X, Zhu Y. A maximum-entropy method for the planning of conformal radiotherapy. Med Phys 2001;28:2241–6.
58. Wu X, Zhu Y. A neural network regression model for relative dose computation. Phys Med Biol 2000;45:13–22.
59. Li RP, Yin FF. Optimization of inverse treatment planning using a fuzzy weight function. Med Phys 2000;27:691–700.
60. Ezzel G. Clinical implementation of IMRT treatment planning. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Colorado Springs (CO): Medical Physics Publishing; 2003. p. 475–94.
61. Xing L, Lin Z, Donaldson SS, et al. Dosimetric effects of patient displacement and collimator and gantry angle misalignment on intensity modulated radiation therapy. Radiother Oncol 2000;56:97–108.
62. Manning MA, Wu Q, Cardinale RM, et al. The effect of setup uncertainty on normal tissue sparing with IMRT for head-and-neck cancer. Int J Radiat Oncol Biol Phys 2001;51:1400–9.
63. International Commission on Radiation Units and Measurements. Recording and reporting photon beam therapy. Report 50. Washington (DC): International Commission on Radiation Units and Measurements; 1993.
64. Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: some theoretical considerations. Phys Med Biol 1993;38:291–304.
65. Pugachev A, Li JG, Boyer AL, et al. Role of beam orientation optimization in intensity-modulated radiation therapy. Int J Radiat Oncol Biol Phys 2001;50:551–60.
66. Pugachev A, Xing L, Boyer AL. Beam orientation optimization in IMRT: to optimize or not to optimize? Presented at the XII International Conference on the Use of Computers in Radiation Therapy; 2000; Heidelberg, Germany.
67. Soderstrom S, Brahme A. Selection of suitable beam orientations in radiation therapy using entropy and Fourier transform measures. Phys Med Biol 1992;37:911–24.
68. Rowbottom CG, Nutting CM, Webb S. Beam-orientation optimization of intensity-modulated radiotherapy: clinical application to parotid gland tumours. Radiother Oncol 2001;59:169–77.
69. Djajaputra D, Wu Q, Wu Y, et al. Algorithm and performance of a clinical IMRT beam-angle optimization system. Phys Med Biol 2003;48:3191–212.
70. Stein J, Mohan R, Wang XH, et al. Number and orientations of beams in intensity-modulated radiation treatments. Med Phys 1997;24:149–60.
71. Pugachev A, Xing L. Computer assisted beam orientation selection in IMRT. Phys Med Biol 2001;46:2467–76.
72. Pugachev A, Xing L. Incorporating prior knowledge into beam orientation optimization. Int J Radiat Oncol Biol Phys 2002;54:1565–74.
73. Xing L, Pugachev A, Li JG, et al. A medical knowledge based system for the selection of beam orientations in IMRT. Int J Radiat Oncol Biol Phys 1999;45:246.
74. Schreibmann E, Xing L. Feasibility study of beam orientation class-solutions for prostate IMRT. Medical Physics; 2004. [In p
75. Kay [illegible]undamentals of statistical signal processing: estimation theory. Upper Saddle River (NJ): Prentice Hall; 1993.
76. Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system. Med Phys 2000;27:701–11.
77. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 1991;21:109–22.
78. Eisbruch A, Ten Haken RK, Kim HM. Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. Int J Radiat Oncol Biol Phys 1999;45:577–87.

79. Lian J, Xing L. Incorporating model parameter uncertainty into inverse treatment planning. Med Phys 2004. [In Press]
80. Chen Y, Michalski D, Houser C. A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. Phys Med Biol 2002;47:1647–58.
81. Yu Y, Zhang JB, Cheng G, et al. Multi-objective optimization in radiotherapy: applications to stereotactic radiosurgery and prostate brachytherapy. Art Intel Med 2000;19:39–51.
82. Cotrutz C, Lahanas M, Kappas C, et al. A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. Phys Med Biol 2001;46:2161–75.
83. Lahanas M, Schreibmann E. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. Phys Med Biol 2003;48:2843–71.
84. Bortfeld T. Physical optimization. In: Palta JR, Mackie TR, editors. AAPM intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 51–76.
85. De Gersem W, Claus F, De Wagter C, et al. Leaf position optimization for step-and-shoot IMRT. Int J Radiat Oncol Biol Phys 2001;51:1371–88.
86. Shepard DM, Earl MA, Li XA, et al. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. Med Phys 2002;29:1007–18.
87. Cotrutz C, Xing L. Using voxel-dependent importance factors for interactive DVH-based dose optimization. Phys Med Biol 2002;47:1659–69.
88. Cotrutz C, Xing L. IMRT dose shaping using regionally variable penalty scheme. Med Phys 2003;30:544–51.
89. Rowbottom CG, Webb S, Oldham M. Improvements in prostate radiotherapy from the customization of beam directions. Med Phys 1998;25:1171–9.
90. Das S, Cullip T, Tracton G, et al. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. Int J Radiat Oncol Biol Phys 2003;55:215–24.
91. Bedford JL, Webb S. Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy. Med Phys 2003; 30:1788–881.
92. Pugachev A, Xing L. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. Int J Radiat Oncol Biol Phys 2001;51:1361–70.
93. Holt R, Xing L. Salvage of suboptimal prostate seed implants using IMRT. Med Phys 2001;28:1308.
94. Li JG, Xing L, Boyer AL, et al. Matching photon and electron fields with dynamic intensity modulation. Med Phys 1999;26:2379–84.
95. Chao KS, Majhail N, Huang CJ, et al. Intensity-modulated radiation therapy reduces late salivary toxicity without compromising tumor control in patients with oropharyngeal carcinoma: a comparison with conventional techniques. Semin Radiat Oncol 2002;12(1 Suppl 1):20–5.
96. Wu Q, Mohan R, Morris M, et al. Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas. I: dosimetric results. Int J Radiat Oncol Biol Phys 2003;56:573–85.
97. Mundt AJ, Lujan AE, Rotmensch J, et al. Intensity-modulated whole pelvic radiotherapy in women with gynecologic malignancies. Int J Radiat Oncol Biol Phys 2002;52:1330–7.
98. Forster KM, Smythe WR, Starkschall G, et al. Intensity-modulated radiotherapy following extrapleural pneumonectomy for the treatment of malignant mesothelioma: clinical implementation. Int J Radiat Oncol Biol Phys 2003;55:606–16.
99. Hanley J, Debois MM, Mah D, et al. Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. Int J Radiat Oncol Biol Phys 1999;45:603–11.
100. Murphy MJ, Martin D, Whyte R, et al.The effectiveness of breath-holding to stabilize lung and pancreas tumors during radiosurgery. Int J Radiat Oncol Biol Phys 2002;53:475–82.
101. Bortfeld T, Jokivarsi K, Goitein M, et al. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. Phys Med Biol 2002;47:2203–20.
102. Ramsey CR, Scaperoth D, Arwood D, et al. Clinical efficacy of respiratory gated conformal radiation therapy. Med Dosim 1999;24:115–9.
103. Wong JW, Sharpe MB, Jaffray DA, et al. The use of active breathing control (ABC) to reduce margin for breathing motion. Int J Radiat Oncol Biol Phys 1999;44:911–9.
104. Vedam SS, Keall PJ, Kini VR, et al. Determining parameters for respiration-gated radiotherapy. Med Phys 2001;28:2139–46.
105. Keall PJ, Kini VR, Vedam SS, et al. Motion adaptive x-ray therapy: a feasibility study. Phys Med Biol 2001;46:1–10.
106. Kini VR, Vedam SS, Keall PJ, et al. Patient training in respiratory-gated radiotherapy. Med Dosim 2003;28:7–11.
107. Ling CC, Humm J, Larson S, et al. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. Int J Radiat Oncol Biol Phys 2000;47:551–60.
108. Rosenman J. Incorporating functional imaging information into radiation treatment. Semin Radiat Oncol 2001;11:83–92.
109. Xing L, Cotrutz C, Hunjan S, et al. Inverse planning for functional image-guided IMRT. Phys Med Biol 2002;47:3567–78.
110. Alber M, Nusslin F. An objective function for radiation treatment optimization based on local biological measures. Phys Med Biol 1999;44:479–93.
111. Bortfeld T, Boyer AL, Schlegel W, et al. Realization and verification of three-dimensional conformal radiotherapy with modulated fields. Int J Radiat Oncol Biol Phys 1994;30:899–908.
112. Xia P, Verhey LJ. Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments. Med Phys 1998;25:1424–34.
113. Crooks SM, McAven LF, Robinson DF, et al. Minimizing delivery time and monitor units in static IMRT by leaf-sequencing. Phys Med Biol 2002;47:3105–16.
114. Siochi RA. Minimizing static intensity modulation delivery time using an intensity solid paradigm. Int J Radiat Oncol Biol Phys 1999;43:671–80.
115. Langer M, Thai V, Papiez L, et al. Improved leaf sequencing reduces segments or monitor units needed to deliver IMRT using multileaf collimators: the influence of irradiation and packaging on the quality of prepacked vegetables. Med Phys 2001;28:2450–8.

116. Beavis AW, Ganney PS, Whitton VJ, et al. Optimization of the step-and-shoot leaf sequence for delivery of intensity modulated radiation therapy using a variable division scheme. Phys Med Biol 2001;46:2457–65.
117. Convery DJ, Rosenbloom ME. Generation of intensity modulated fields by dynamic collimation. Phys Med Biol 1992;37:1359–74.
118. Svensson R, Kallman P, Brahme A. An analytical solution for the dynamic control of multileaf collimators. Phys Med Biol 1994;39:37–61.
119. Spirou SV, Chui CS. Generation of arbitrary intensity profiles by dynamic jaws or multileaf collimators. Med Phys 1994;21:1031–41.
120. Dai JR, Hu YM. Intensity-modulation radiotherapy using independent collimators: an algorithm study. Med Phys 1999;26:2562–70.
121. Beavis AW, Ganney PS, Whiton VJ, et al. Slide and shoot: a new method for MLC delivery of IMRT. In: Proceedings of the XIII International Conference on the Use of Computers in Radiation Therapy, Heidelberg, Germany, May 2000.
122. Beavis A, Ganney P, Whitton V, Xing L. Slide and shoot: a new method for MLC delivery of IMRT. The Use of Computers in Radiation Therapy, Heidelberg, Germany, 2000.
123. Yang Y, Xing L. Incorporating leaf transmission and header scatter corrections into MLC leaf sequences for IMRT. Int J Radiat Oncol Biol Phys 2003;55:1121–34.
124. Yang Y, Xing L, Boyer A, et al. A three-source model for the calculation of head scatter factors. Med Phys 2002;29:2024–33.
125. Chen Y, Boyer AL, Ma CM. Calculation of x-ray transmission through a multileaf collimator. Med Phys 2000;27:1717–26.
126. Deng J, Pawlicki T, Chen Y, et al. The MLC tongue-and-groove effect on IMRT dose distributions. Phys Med Biol 2001;46:1039–60.
127. Hounsell AR. Monitor chamber backscatter for intensity modulated radiation therapy using multileaf collimators. Phys Med Biol 1998;43:445–54.
128. Chui CS, LoSasso T, Spirou S. Dose calculation for photon beams with intensity modulation generated by dynamic jaw or multileaf collimations. Med Phys 1994;21:1237–44.
129. Wang X, Spirou S, LoSasso T, et al. Dosimetric verification of intensity-modulated fields. Med Phys 1996;23:317–27.
130. van Santvoort JP, Heijmen BJ. Dynamic multileaf collimation without 'tongue-and-groove' underdosage effects. Phys Med Biol 1996;41:2091–105.
131. Webb S, Bortfeld T, Stein J, et al. The effect of stair-step leaf transmission on the 'tongue-and-groove problem' in dynamic radiotherapy with a multileaf collimator. Phys Med Biol 1997;42:595–602.
132. Dirkx ML, Heijmen BJ, van Santvoort JP. Leaf trajectory calculation for dynamic multileaf collimation to realize optimized fluence profiles. Phys Med Biol 1998;43:1171–84.
133. Ma L, Boyer AL, Ma CM, et al. Synchronizing dynamic multileaf collimators for producing two-dimensional intensity-modulated fields with minimum beam delivery time. Int J Radiat Oncol Biol Phys 1999;44:1147–54.
134. Galvin JM, Smith AR, Lally B. Characterization of a multi-leaf collimator system. Comment in: Int J Radiat Oncol Biol Phys 1993 Jan 15;25(2):373–5. Int J Radiat Oncol Biol Phys 1993;25:181–92.
135. Sykes JR, Williams PC. An experimental investigation of the tongue and groove effect for the Philips multileaf collimator. Phys Med Biol 1998;43:3157–65.
136. IMRT Collaborative Working Group. Intensity-modulated radiotherapy: current status and issues of interest. Int J Radiat Oncol Biol Phys 2001;51:880–914.
137. Sharpe MB. Commissioning and quality assurance for IMRT treatment planning. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 449–74.
138. Xia P, Chuang C. Patient-specific quality assurance in IMRT. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 495–514.
139. Moran JM. Dosimetry metrology. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 415–38.
140. Xing L, Curran B, Hill R, et al. Dosimetric verification of a commercial inverse treatment planning system. Phys Med Biol 1999;44:463–78.
141. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: report of AAPM Radiation Therapy Committee Task Group 40. Med Phys 1994;21:581–618.
142. Fraass B, Doppke K, Hunt M, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. Med Phys 1998;25:1773–829.
143. Ezzell G, Galvin J, Low D, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. Med Phys 2003;30:2089–115.
144. Bortfeld TR, Kahler DL, Waldron TJ, et al. X-ray field compensation with multileaf collimators. Int J Radiat Oncol Biol Phys 1994;28:723–30.
145. Ma L, Boyer AL, Xing L, et al. An optimized leaf-setting algorithm for beam intensity modulation using dynamic multileaf collimators. Phys Med Biol 1998;43:1629–43.
146. Crooks S, Pugachev A, King C, et al. Examination of the effect of increasing the number of radiation beams on a radiation treatment plan. Phys Med Biol 2002;47:3485–501.
147. Van Dyk J, Purdy JA. Clinical implementation of technology and the quality assurance process. In: Van Dyk J, editor. The modern technology of radiation oncology. Madison (WI): Medical Physics Publishing; 1999. p. 19–51.
148. Chui CS, Spirou S, LoSasso T. Testing of dynamic multileaf collimation. Med Phys 1996;23:635–41.
149. LoSasso TJ. IMRT delivery system QA. In: Palta JR, Mackie TR, editors. AAPM intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 561–91.
150. Herman MG, Balter JM, Jaffray DA, et al. Clinical use of electronic portal imaging: report of AAPM Radiation Therapy Committee Task Group 58. Med Phys 2001;28:712–37.
151. Chang J, Mueller K, Sidhu K, et al. Verification of multileaf collimator leaf positions using an electronic portal imaging device. Med Phys 2002;29:2913–24.

152. Samant SS, Zheng W, Parra NA, et al. Verification of multileaf collimator leaf positions using an electronic portal imaging device. Med Phys 2002;29:2900–12.
153. Vieira SC, Dirkx ML, Pasma KL, Heijmen BJ. Fast and accurate leaf verification for dynamic multileaf collimation using an electronic portal imaging device. Med Phys 2002;29:2034–40.
154. Yang Y, Xing L. Using the volumetric effect of a finite-sized detector for routine quality assurance of MLC leaf positioning. Med Phys 2003;30:433–41.
155. Yang Y, Xing L. Quantitative measurement of MLC leaf displacements using an electronic portal image device. Phys Med Biol 2004;49:1251-1533.
156. Chen Y, Xing L, Luxton G, et al. A multi-purpose quality assurance tool for MLC-based IMRT. ICCR, Heidelberg, Germany, May, 2000.
157. Boyer A, Xing L, Ma CM, et al. Theoretical considerations of monitor unit calculations for intensity modulated beam treatment planning. Med Phys 1999;26:187–95.
158. Kung J, Chen G. A monitor unit verification calculation in intensity modulated radiotherapy as a dosimetric quality assurance. Med Phys 2000;27:2226–30.
159. Xing L, Chen Y, Luxton G, et al. Monitor unit calculation for an intensity modulated photon field by a simple scatter-summation algorithm. Phys Med Biol 2000;45:N1–7.
160. Yang Y, Xing L, Li JL, et al. Independent dosimetric calculation with inclusion of head scatter and MLC transmission for IMRT. Med Phys 2003;30:2937–47.
161. Xing L, Yang Y, Li J, et al. Monitor unit calculation and plan validation for IMRT. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 3567–78.
162. Watanabe Y. Point dose calculations using an analytical pencil beam kernel for IMRT plan checking. Phys Med Biol 2001;46:1031–8.
163. Xing L, Li JG. Computer verification of fluence maps in intensity modulated radiation therapy. Med Phys 2000;27:2084–92.
164. Schefter TE, Kavanagh BD, Wu Q, et al. Technical considerations in the application of intensity-modulated radiotherapy as a concomitant integrated boost for locally-advanced cervix cancer. Med Dosim 2002;27:177–84.
165. Xing L, Yi BY, Li J, et al. Adaptive inverse planning with consideration of MLC field size constraint. Presented at the American Association of Physicists in Medicine Annual Meeting; 1999; Nashville, TN.
166. Wu Q, Arnfield M, Tong S, et al. Dynamic splitting of large intensity-modulated fields. Phys Med Biol 2000;45:1731–40.
167. Xing L, Boyer A, Kapp D, et al. Improving the matching of abutting photon fields by modulating photon beams. Med Biol Eng Comp 1997;35:921.
168. Hong L, Alektiar K, Chui C, et al. IMRT of large fields: whole-abdomen irradiation. Int J Radiat Oncol Biol Phys 2002;54:278–89.
169. Xing L, Yang Y, Li J, et al. Dose matching of an IMRT plan with an electron or 3D conformal treatment plan. Presented at the World Congress on Medical Physics and Biomedical Engineering; 2003 Aug 24–29; Sydney, Australia.
170. Followill D, Geis P, Boyer A. Estimates of whole-body dose equivalent produced by beam intensity modulated conformal therapy. Int J Radiat Oncol Biol Phys 1997;38:667–72.
171. Verellen D, Vanhavere F. Risk assessment of radiation-induced malignancies based on whole-body equivalent dose estimates for IMRT treatment in the head and neck region. Radiother Oncol 1999;53:199–203.
172. Hall EJ, Wuu CS. Radiation-induced second cancers: the impact of 3D-CRT and IMRT. Int J Radiat Oncol Biol Phys 2003;56:83–8.
173. Radiation protection design guidelines for 0.1-100 MeV particle accelerator facilities. Vol. 51. National Council on Radiation Protection and Measurements; Washington (DC): 1977.
174. Structural shielding design and evaluation for medical use of x-rays and gamma rays of energies up to 10 MeV. Vol 49. National Council on Radiation Protection and Measurements. Washington (DC): 1976.
175. McGinley PH. Shielding techniques for radiation oncology facilities. 2nd ed. Madison (WI): Medical Physics Publishing; 2002.
176. McGinley HM, Miner MS. A history of radiation shielding of x-ray therapy rooms. Health Phys 1995;69:759–65.
177. Rodgers JE. Radiation therapy vault shielding calculation methods when IMRT and TBI procedures contribute. J Appl Clin Med Phys 2001;2:157–64.
178. Mutic S, Low DA, Klein EE, et al. Room shielding for intensity-modulated radiation therapy treatment facilities. Int J Radiat Oncol Biol Phys 2001;50:39–46.
179. Low D. Radiation shielding for IMRT. In: Palta JR, Mackie TR, editors. AAPM intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 401–14.

Thomas Bortfeld
Rupert Schmidt-Ullrich
Wilfried De Neve
David E. Wazer
Editors

Image-Guided IMRT



Springer

Molecular/Functional Image-guided Intensity Modulated Radiation Therapy

Lei Xing, Yong Yang, Daniel M. Spielman

4

Contents

4.1 Introduction 187
4.1.1 Molecular and Functional Imaging 187
4.1.2 IMRT as a Means of Producing Biologically Conformal Dose Distributions 187
4.2 Functional and Molecular Imaging and Biologically Conformal Radiation Therapy 188
4.2.1 Integration of Functional and Molecular Imaging into IMRT Planning 189
4.2.2 Image Registration 189
4.2.3 Quality Assurance of Molecular and Functional Imaging Modalities 190
4.2.4 Inverse Treatment Planning 190
Relation Between Metabolic Abnormality Level and Radiation Dose 191
Implementation 191
Role of Intra-structural Tradeoff 193
Spectral Uncertainty 194
Biological Model for Molecular/Functional Image-guided IMRT 194
Plan Review Tools 195
4.3 Conclusion 196
References 196

4.1 Introduction

4.1.1 Molecular and Functional Imaging

For much of the last century, medical imaging has been focused on faster and more detailed anatomic pictures of the human body. The accomplishment of the visible human project of the National Library of Medicine (http://www.nlm.nih.gov/research/visible) represents perhaps one of the most important milestones in these developments. With the goal of producing a system of knowledge structures that transparently links visual knowledge forms to symbolic knowledge formats such as the names of body parts, a complete, anatomically detailed, 3D representations of the normal male and female human bodies were rendered based on transverse CT, MR and cryosection images of male and female cadavers. Medical imaging has been an integral part of radiation therapy since the discovery of X-rays and the imaging techniques, such as X-ray, CT, MRI and ultrasound (US) imaging, are the foundation for the modern radiation therapy modalities that are routinely used in clinics, such as 3D conformal radiation therapy, intensity modulated radiation therapy (IMRT), stereotactic radiosurgery, and brachytherapy. Indeed, the development of radiation therapy has strongly relied on the imaging technology and, historically, almost every major advancement in imaging science would bring radiation therapy to a new level.

In general, medical imaging is involved in all key steps of radiation treatment (Fig. 1). One of the most important uses of imaging techniques is the delineation of a tumor target. Despite the tremendous successes, the anatomic imaging techniques such as CT/MRI/US are inherently deficient in that they can only reveal spatial changes in physical properties and fail to provide basic biological information that is much needed for the optimal management of the patients. Clinically, tumor biology plays an important role in the diagnosis, treatment decision-making, and assessment of therapeutic response of various diseases. It is thus highly desirable to develop imaging techniques capable of revealing the spatial biology distribution of the patients. Toward this goal, a new branch of science, referred to as molecular imaging, is emerging as a result of research efforts in cellular biology and imaging techniques over the years. The development of cellular and molecular imaging provides significant opportunities for the radiation discipline to take the patient's biological information into the radiation therapy treatment decision-making process and to truly individualize cancer radiotherapy.

4.1.2 IMRT as a Means of Producing Biologically Conformal Dose Distributions

IMRT is an advanced form of external beam irradiation and represents a radical change in radiation oncology



Fig. 1. A schematic of radiation treatment process. Molecular/functional imaging plays an important role in each of the key steps (represented by *blocks*) of the radiation treatment process

practice [1–3]. This new process of treatment planning and delivery shows significant potential for improving the therapeutic ratio and offers a valuable tool for dose escalation and/or radiation toxicity reduction. Preliminary published results and unpublished results from several institutions indicate that with IMRT, radiation doses to sensitive structures can be reduced significantly while maintaining adequate target dose coverage [4–15]. Because in many clinical situations the dose to the tumor volume is limited by the tolerance doses of the sensitive structures, it is considered likely that IMRT will improve local control and lead to an increase in survival rate for certain cases through dose escalation. In addition, IMRT has the potential to improve the efficacy of treatment planning and delivery in routine clinical practice with the use of computerized planning and treatment process. For details about IMRT inverse treatment planning, delivery and quality assurance, we refer the readers to the related chapters of this book.

In IMRT, each incident beam is divided into a number of beamlets (typically, the size of a beamlet is in the order of 1×1 cm), allowing us to modify the dose distribution on an individual beamlet level. Using IMRT, it is possible to produce not only spatially uniform but also non-uniform dose distributions. Recently, Ling et al. and several other researchers [16–21] have emphasized the technical capability of "dose painting" and "dose sculpting" offered by IMRT, which allows customized dose delivery to the target volume(s) with centimeter or even sub-centimeter spatial resolution. Using functional and molecular imaging techniques to identify spatial metabolic distribution and hence guide the delivery of radiation represents a paradigm shift in radiation oncology and this type of "biologically" conformal radiation therapy may provide a significant opportunity to improve conventional IMRT treatment. A timely question is how to integrate the state-of-the-art functional imaging technologies into radiation therapy techniques such as IMRT to positively impact clinical cancer management. The purpose of this chapter is to review recent progress in this endeavor and identify the important issues in the development of biologically conformal radiation therapy.

4.2 Functional and Molecular Imaging and Biologically Conformal Radiation Therapy

Current IMRT treatment plan optimization is based on the assumption of uniform biology distribution within the target volume and is aimed at achieving geometrically conformal dose distributions under the guidance of CT/MRI images. In reality, it has long been recognized that the spatial distribution of biological properties in most tumors and normal tissues are heterogeneous. With the advent of various molecular and functional imaging techniques, it is now possible to map out the biology distribution on a patient specific basis. To use the spatially heterogeneous biology information derived from the new imaging modalities to guide IMRT dose painting and sculpting process, several key problems need to be resolved. In general, the molecular/functional imaging-guided IMRT generally favors non-uniform dose distributions and requires a plan optimization formalism in voxel domain to deal with the biological heterogeneity. In addition, new methods of specifying the desired doses and a mechanism for inter- and intra-structural tradeoff, which will be explained below, must be introduced to efficiently produce metabolically/functionally conformal doses. In Fig. 2 we list the general steps of biologically conformal IMRT treatment. Each of the steps in Fig. 2 is discussed below.

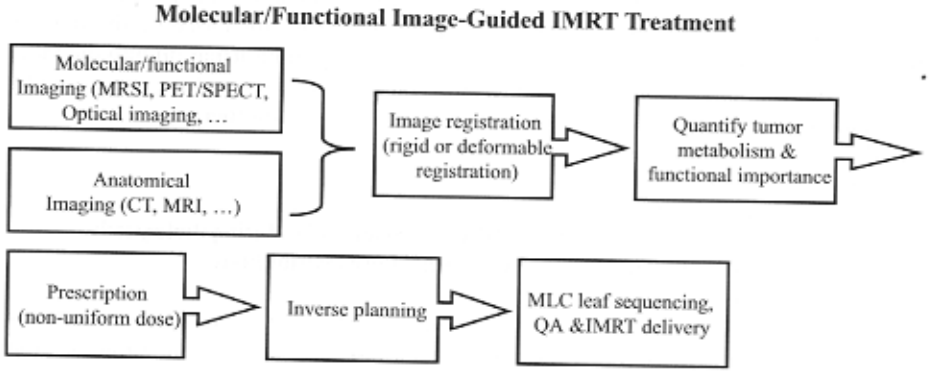


Fig. 2. Procedure of biologically conformal IMRT treatment

4.2.1 Integration of Functional and Molecular Imaging into IMRT Planning

The area of molecular/functional imaging is rapidly evolving [22–24]. Many of the molecular imaging modalities (such as fluorescent and bioluminescent imaging, optical imaging, SPECT/PET with novel isotopes/contrast agents targeting some specific molecular markers, MR spectroscopic imaging (MRSI)) are being developed for tumor specific imaging and deployed into clinical practice. Presently, MRSI, PET/SPECT and micro-bubble based ultrasound are perhaps the most mature modalities and available for guiding radiation therapy treatment. Details on various molecular and functional imaging modalities have been given elsewhere in this volume (see Verhey and van de Wiele chapters) and will not be repeated here. The remainder of this chapter will be focused on the issues related to the integration of the new imaging modalities into radiation treatment planning.

4.2.2 Image Registration

Radiation therapy treatment planning is mainly CT image-based because it provides complete geometric data and electron density information for accurate dose calculation. To utilize the biological information derived from the new image modalities, we must map the imaging data onto treatment planning CT images. The level of complexity of image coregistration depends on the imaging techniques involved and specific software tools often need to be developed in order to use some of the new imaging modalities, such as fluorescent images, endoscopic images and endorectal images. Sometimes, deformable model-based image registration is required if the shape(s) of the involved organs are deformed from its normal shape.

Let us take endorectal MRSI as an example. The introduction of endorectal surface coils significantly improves spatial resolution and signal-to-noise ratio (SNR) of prostate MR imaging and allows evaluation of tumor location, tumor volume, capsular penetration, invasion of neurovascular bundle, and seminal vesicle involvement, which is crucial for accurate treatment planning. Endorectal-coil based MRSI has also been shown effective in distinguishing between areas of cancer and normal prostatic epithelium through differences in [choline + creatine]/citrate ratio [25–28]. However, the use of endorectal probe inevitably distorts the prostate and other soft tissue organs, making it impossible to fuse directly the acquired image data onto treatment planning CT. In Fig. 3 we show the difference between endorectal coil-based MRI defined and CT-defined prostate volume [29]. In order to fuse MRI/MRSI with treatment planning CT, it is necessary to develop an effective deformable image registration procedure. Otherwise, the gain from the use of the state-or-the-art imaging techniques may be lost due to the inferior performance of image registration.

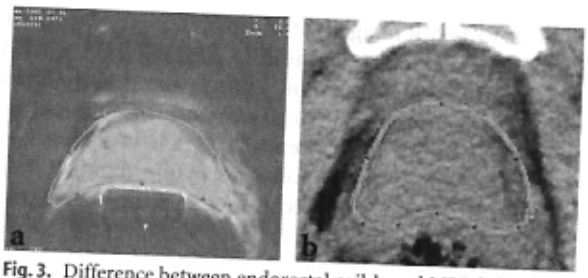


Fig. 3. Difference between endorectal coil-based MRI defined and treatment planning CT-defined prostate volumes

Zaider et al. [30] have reported a translation and scaling based registration method to map MRS positive volumes onto the CT and ultrasound images. In their approach, the coordinates of the boundary and the center of mass were used to linearly interpolate the positions of the mapped voxels. A larger discrepancy was found for regions with more severe distortion (4 mm). Lian et al. [29, 31] have developed an effective deformable image registration algorithm to map the MRI/MRSI information obtained using a rigid or inflatable endorectal probe onto CT images and to verify the accuracy of the registration by phantom and patient studies. For this purpose, a thin plate spline (TPS) transformation first introduced by Bookstein [32] was implemented to establish voxel-to-voxel correspondence between a *reference* image and a *floating* image with deformation. The idea is to find a continuous transformation to minimize the landmark difference in two images. The detailed description of the TPS transformation can be found in Bookstein's original paper [32]. To access the quality of the registration, an elastic phantom with a number of implanted fiducial markers was designed. Radiographic images of the phantom were obtained before and after a series of intentionally introduced distortions. After mapping the distorted phantom to the original one, the displacements of the implanted markers were measured with respect to their ideal positions and the mean error was calculated. Phantom studies showed that using the deformable registration method the mean landmark displacement error was 0.62 ± 0.39 mm when the distortion was of the order of 23.07 mm. A deformable model seems to be necessary to map faithfully the metabolic information onto the treatment CT images. When a non-deformable method based on a rigid-body transformation and scaling was used for the same distortion, the mean displacement of the fiducials with respect to their actual positions was found to be as large as 12.95 ± 0.57 mm. In patient studies, CT images of two prostate patients were acquired, followed by

3-Tesla (3 T) MR images with a rigid endorectal coil. For both patient studies, significantly improved registration accuracy was achieved. The prostate centroid position displacement was 0.58 ± 0.10 mm and the coincidence index was $92.6 \pm 5.1\%$ when a TPS transformation was used. Different from the non-deformable approach, the TPS-based registration accommodates the organ distortion and enables us to achieve significantly higher MRI/MRSI and CT image registration accuracy. More advanced finite element method is also developed to attack the problem [33].

4.2.3 Quality Assurance of Molecular and Functional Imaging Modalities

Any new imaging modality requires validation and quality assurance to ensure that the obtained images faithfully reflect the reality. In anatomical imaging, surrogate phantoms have been widely used for assessing the geometric and physical (e.g., electron density) properties of the images. For radiation therapy application, Mutic et al. have reported a simple design of a PET phantom to validate the image registration of PET and CT images [34]. Generally speaking, for a biological imaging modality, validation of geometric accuracy represents only one facet of the problem. The accuracy of the pixel values of the imaging modality also needs our attention. While the specific meaning of the pixel values depends on the modality, let us take an MRSI phantom (Fig. 4) constructed by Hunjan et al. as an example to illustrate the basic idea. The multi-modality, multi-purpose phantom is suitable for quality assurance testing of fusion data from MRI, MRSI and CT images [35]. The phantom contains fiducial markers that are simultaneously MR, MRS, and CT-visible. To examine the accuracy of MRSI for brain tumor, the phantom was filled with a brain-mimicking solution with an insert holding eight vials containing calibrated solutions of precisely varying metabolite concentrations that emulated increasing grade/density of brain tumor. Metabolite ratios calculated from fully relaxed 1D, 2D and 3D MRS data for each vial were compared to calibration ratios acquired in vitro using a 9.4-Tesla MR spectrometer. Figure 5 shows an axial scout scan of the MRS metabolite ratio quantitation standard showing the calibration vials 1–8. The resulting single voxel MR spectra are shown inset next to corresponding vials and a linear fit between the Choline/NAA ratio (NAA: *N*-acetyl-aspartate, see Verhey chapter, this volume) of the calibration solutions obtained at 9.4 T vs the calibration-solution-filled vials inside the phantom obtained at 1.5 T. For detailed information on the design of the phantom and measurements, please refer to [35,36].

4.2.4 Inverse Treatment Planning

In general, molecular/functional imaging could impact the current radiation therapy treatment in two fundamental aspects [16, 20, 37]. First, it offers an effective means for us to delineate more accurately the tumor and define better the treatment volume. Second, it provides valuable spatial metabolic information in the tumor and sensitive structures. While it is straightforward to modify the radiation portals to accommodate any changes in treatment volume, new methods of dose optimization and medical decision-making must be developed

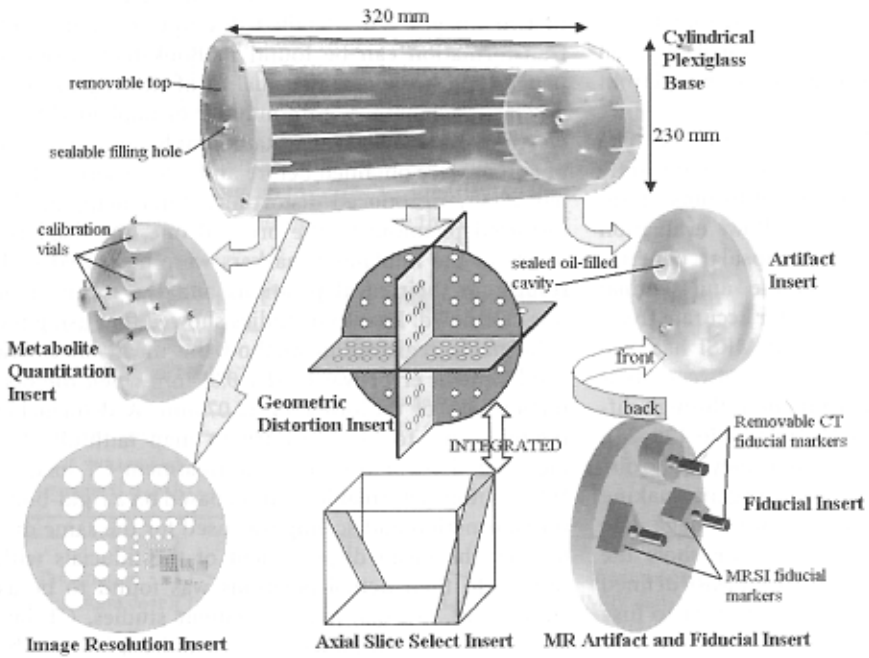


Fig. 4. A photo of quality assurance phantom built for testing/validating the geometric and metabolic accuracy of the endorectal coil-based MRI/MRSI (from Hunjan S et al. (2003) IJROBP 57:1159–1173, with permission)

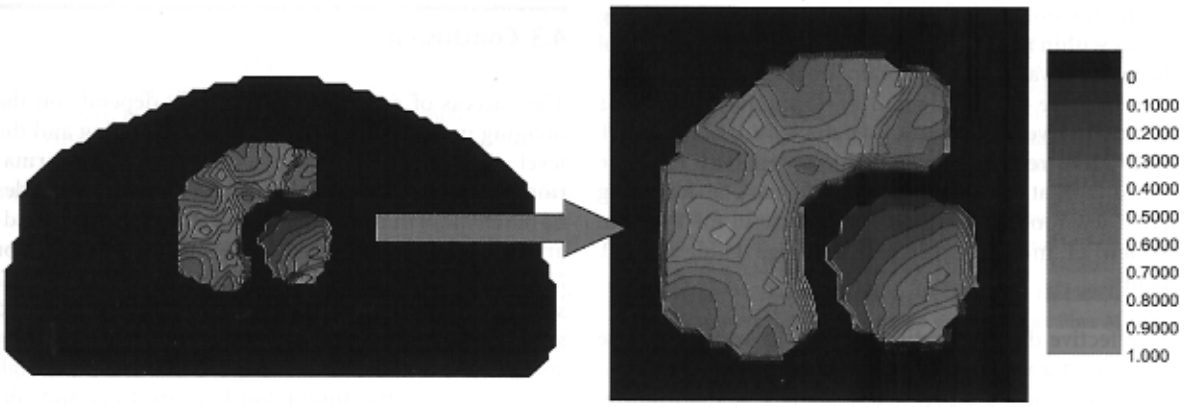


Fig. 9. Dosimetric capability map of the target and sensitive structure for a hypothetical case with five equally spaced beams. The data for each structure is normalized to unity. The *left panel* shows the complete geometry of the hypothetical structures. The capability maps of the target and sensitive structure are enlarged and shown in the *right panel*

modeling of the system are [45]: (i) how to determine the non-uniform dose prescription provided that the biology distribution is known; and (ii) how to find the optimal solution. While the latter problem is similar to that in conventional IMRT inverse planning, the solution to the first problem entails some theoretical considerations. Earlier we used the metabolic abnormality index to characterize phenomenologically the tumor burden. Using a radiobiological model, it is possible to relate the prescription dose to the more fundamental radiobiology parameters to optimize the cell killing.

Let us start with the linear quadratic (LQ) model. We include the effect of tumor cell proliferation but ignore the quadratic term. The model parameters include clonogen density (ρ), radiosensitivety (α), and proliferation rate (γ). The time dependence of the parameters are ignored. The tumor control probability, TCP_i, for a tumor voxel i, can be expressed as

$$TCP_i = \exp\left[-\rho_{0i} V_i \exp\left(-\alpha_i D_i + \gamma_i \Delta T\right)\right] \tag{4}$$

where V_i is the volume of voxel i, ρ_{0i}, α_i and γ_i represent the initial clonogen density, radiosensitivety and proliferation rate in voxel i, respectively, D_i is the dose received by voxel i, and ΔT is the overall treatment time. In Eq. 4, $\gamma_i = \ln 2/T_{pi}$ where T_{pi} is the potential doubling time in voxel i. TCP for the tumor is given by

$$TCP = \prod_i TCP_i \tag{5}$$

A constraint from the normal cells within the tumor volume given by

$$\sum_i m_i D_i = E_t \tag{6}$$

should be applied to determine the tumor dose prescription, where m_i is the mass of voxel i, E_t is the integral dose in tumor. The problem now becomes to maximize the TCP under the constraint of Eq. 6, which can be solved using the method of Lagrange multipliers [46]. When the mass and volume are equal for all tumor voxels, the desired prescription dose of a voxel is given by

$$D_i = \frac{\alpha_r}{\alpha_i} D_r - \frac{1}{\alpha_i}\left(\gamma_r - \gamma_i\right)\Delta T - \frac{1}{\alpha_i}\ln\left(\frac{\alpha_r \rho_{0r}}{\alpha_i \rho_{0i}}\right) \tag{7}$$

where D_r is the reference dose for the voxel with reference radiobiological parameters (ρ_{0r}, α_r, γ_r). In general, D_r should be set to a value that yields a clinical sensible TCP at the reference voxel. For a given disease site, the radiation dose used in current clinical practice with "intent to cure" can be used as a good starting point in selecting the value of D_r. Once the desired dose prescription distribution is determined, IMRT inverse planning optimization can proceed by numerically maximizing the TCP while maintaining the NTCP below a certain limit. One can also take a "hybrid" approach by using the conventional objective function with the above dose prescription.

Plan Review Tools

The sheer volume of information inherent in 3D treatment designs and the corresponding dose distributions make display and objective assessment problematic. Details of a dose-distribution's spatial characteristics can be obtained by examining 2D isodose curves in a slice-by-slice fashion; however, this is a quasi-quantitative, time-consuming process and is not an efficient way to compare competing plans even for conventional IMRT. In the presence of an additional degree of freedom (metabolic abnormality), the problem is exacerbated by the breakdown of uniform dose assumption within the target volume. One of the commonly used approaches is the reliance on data reduction techniques in the quantitative assessment of alternative plans. DVH is one of the most widely used data reduction techniques. This technique enables the ready reduction of the complex 3D data set of a treatment design into the 2D display

of the fractional volume of a given structure receiving doses within a particular range. Unfortunately, this tool becomes invalid for metabolic/functional plan evaluation because of possibly non-uniform biological status of the involved structures. Metabolic/functional IMRT techniques require new plan review tools to facilitate the quantitative comparison of plans. The following are a few tools that are potentially useful for the plan review of molecular/functional image guided IMRT plans:

1. Effective dose ratio distribution: The effective dose ratio at a voxel is defined as the ratio of the physical dose and the prescribed dose. This distribution considers both the spatial dose distribution and the metabolic map and provides intuitive information on the geometric location of underdosing or overdosing regions. In this way, we can use conventional wisdom to evaluate a metabolic/functional based dose distribution. The DVH corresponding to the effective dose ratio distribution is also useful.
2. DVH clusters: In practice, not all underdosing/overdosing are equally significant and underdosing/overdosing at a certain metabolic level maybe more acceptable than at other metabolic levels. A cluster of DVHs, each corresponding to an incremental range of metabolic activity of interest, may provide useful tool to address the issue. The cluster of DVHs can be used to check the overall dosimetric behavior at an individual metabolic level. Figure 7 represents an example of a three-level DVH cluster. For a sensitive structure with functional data available, similar techniques apply.
3. Functional dose-volume histogram (FDVH): Distribution of functional importance appears to be heterogeneous in some normal organs and functional imaging modalities such as MRSI or PET/SPECT may provide valuable information about the spatial distribution of the functional importance. The FDVH, originally introduced by Lu et al. [47], Marks et al. [48], and Alber and Nusslin [49] may prove to be a useful plan review tools. A similar histogram function can be introduced for the tumor, but its usefulness needs to be justified.
4. Modified TCP and NTCP calculation tools: The conventional TCP and NTCP formula [38, 39, 50, 51] need to be modified to take into account the heterogeneous biology distribution [52–54]. This modification should be straightforward if the spatial distributions of radiobiological parameters are known. Although it is difficult to obtain quantitative results from the model calculation because of the uncertainties in the parameters, qualitative conclusions regarding the deliberately non-uniform irradiation scheme can be drawn and may shed useful insight into the problem [54].

4.3 Conclusion

The success of radiotherapy critically depends on the imaging modality used for treatment planning and the level of integration of the available imaging information. The use of functional/metabolic imaging provides us much more than a tool to delineate better the boundary of a tumor target. Together with anatomical CT or MRI images, functional imaging affords valuable 3D structural plus 1D metabolic data for both tumor and sensitive structures, valuable for guiding us to design spatially non-uniform dose distributions to deliver high doses to where the tumor burdens are high and differentially spare the sensitive structures according to the functional importance distributions. The integration and utilization of the functional data in radiation therapy treatment planning become increasingly important to improve clinical cancer management. While it is straightforward to modify the radiation portals to accommodate any changes in treatment volume, new methods of dose optimization and medical decision-making must be developed to take full advantage of the metabolic information and IMRT. How to achieve biologically conformal doses, instead of the geometrically conformal dose distribution, presents a new challenge to radiation oncology discipline. Hopefully, with the efforts from multiple institutions, the new approach of imaging, planning and decision-making will be resolved. Ultimately, whether using deliberately inhomogeneous dose distributions obtained under the guidance of functional imaging such as MRSI can improve patient survival and reduce the side effects associated with radiation treatment should be established through extensive clinical trials.

Acknowledgements. We would like to thank Drs. A. Boyer, S. Hunjan, J. Lian, C. Cotrutz, Z. Shou, P. Maxim, E. Schreibmann, Q. Le, S. Hancock, I. Gibbs, K. King, B. Daniel, D. Kim and E. Adalsteinsson for many useful discussions. We also wish to acknowledge grant support from the National Cancer Institute (1 R01 CA98523-01), National Institute of Health (P41 RR09784), the Department of Defense (DAMD17-03-1-0023), and the Vadasz Family Foundation. The authors are also grateful to the publishers of Physics in Medicine and Biology, Medical Physics, and International Journal of Radiation Oncology, Biology and Physics for their permission for using the copyrighted figures.

References

1. Webb S (2001) Intensity-modulated radiation therapy. Institute of Physics Publishing, Bristol, pp xv, 633

