REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 28, 2003	Final Report, 01 Jun 99 - 28 Feb 03

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Vision for control: optimality and usability	DAAD19-99-1-0139
6. AUTHOR(S)	
Stefano Soatto	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Washington University in St.Louis, University of California, Los Angeles	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER
U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	38929.7 - CI

11. SUPPLEMENTARY NOTES

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.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT	12 b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

This final report for the grant DAAD19-99-1-0139 and the associated instrumentation grant DURIP 442510-21338 describes research conducted both at Washington University in St.Louis and at the University of California, Los Angeles, from 1999 to 2003.

Research has been documented in 34 publications, among which two best-paper awards (the Marr Prize, the highest recognition in the field of Computer Vision, and the Siemens Prize with the Outstanding Paper Award by the IEEE Computer Society) and a book, expected to appear in June 2003.

Technology firsts include the first ever system (and still the only one) for estimating three-dimensional structure and motion of an arbitrary (static) scene in real time, and the first optimal algorithms for estimating shape from accommodation.

Given the scale of the project, consisting of 70K\$/year plus equipment funds, the results are considerable: all the goals set forth in the original proposals have been reached, and several new projects have been initiated that were not part of the forecast plan. Such new projects explore fundamental research that holds high promise for applications of strategic value in the aftermath of 9/11/01, as we describe below.

14. SUBJECT TERMS	15. NUMBER OF PAGES 26 (including this cover page)
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

Vision For Control:
Optimality and Usability

Stefano Soatto

University of California, Los Angeles
soatto@ucla.edu

Final Progress Report
January 15, 2003

ABSTRACT

This final report for the grant DAAD19-99-1-0139 and the associated instrumentation grant DURIP 442510-21338 describes research conducted both at Washington University in St.Louis and at the University of California, Los Angeles, from 1999 to 2003.

Research has been documented in 34 publications, among which two best-paper awards (the Marr Prize, the highest recognition in the field of Computer Vision, and the Siemens Prize with the Outstanding Paper Award by the IEEE Computer Society) and a book, expected to appear in June 2003.

Technology firsts include the first ever system (and still the only one) for estimating three-dimensional structure and motion of an arbitrary (static) scene in real time, and the first optimal algorithms for estimating shape from accommodation.

Given the scale of the project, consisting of 70K$/year plus equipment funds, the results are considerable: all the goals set forth in the original proposals have been reached, and several new projects have been initiated that were not part of the forecast plan. Such new projects explore fundamental research that holds high promise for applications of strategic value in the aftermath of 9/11/01, as we describe below.

Synopsis of accomplishments

In the following we list the original milestones as outlined in the original proposal, together with the publications where they are delivered, which is described in greater detail in previous interim reports and in the cited references [24, 40, 39, 27, 38, 37, 19, 9, 11, 3, 2, 23, 34, 20, 33, 7, 12, 16, 35, 26, 10, 36, 29, 17, 8, 6, 13, 4, 21, 30, 18, 5, 25, 22]. A complete list of equipment purchased under the DURIP program is included with the submission of this report.

Year I

Analysis *Analysis of optimal algorithms for reconstructing three-dimensional structure from motion (SFM). Convergence properties, region of attraction of the global minimum. Analysis of known local extrema (bas-relief ambiguity, rubbery motion). Noise sensitivity.* These goals were reached in year 1, and resulted in award-winning publications [28, 22].

Algorithms *Implementation of provably convergent algorithms for SFM. Implementation of simple outlier rejection and missing data.* This goal was achieved in year 1 and the algorithm was implemented off-line. Code has been made available to the community via the web.

Experiments *Use of pseudo real-time optical flow and feature tracking for testing implemented SFM algorithms on real image sequences. Indoor and outdoor sequences tested. Error performance provided using calibration. Tests for stability, performance and robustness.* This goal was achieved in year 1, and published in [31, 32].

Year II

Analysis *Geometric configurations corresponding to local extrema. Characterization of the relationship between local extrema and the global geometry of the optimization of SFM. Study of invariant representations. Analysis and reduction of the state-space into components that are invariant to local extrema.* This goal was achieved to completion during year 2, and has appear in print in the leading journal in the field [2, 23].

Algorithms *On-line statistical tests for outlier rejection. On-line path estimation for non-holonomic path following.* This goal was achieved during year 2, and has been presented in [4, 17].

Experiments *Testing on pseudo real-time segmentation. Use of local consistency cues (2D) as well as 3D motion cues.* This goal has been vastly exceeded during year 2: the algorithm for automatic feature detection, tracking, outlier rejection and rigid body estimation has been implemented in real time and presented in [18]. The system we developed, the first of its kind, has been made publicly available, and has been featured on the July 2000 issue of the EE Times [1].

Year III

Analysis *Observers for variable state dimension. Non-linear reduced-order observers; observability issues. Bounds on inference errors.* This was achieved and the results published in [3].

Algorithms *Implementation of a representative sample of image-based control design techniques.* Thanks to DURIP funding, we have implemented basic image-based control algorithms on a set of 3 mobile robots (Evolution Robotics ER-1); core processing is performed on a laptop PC, including frame processing, acquired via firewire IEEE 1394.

Experiments *Vision-based navigation on a remote-controlled vehicle. Indoor and outdoor unknown and unstructured environments.* This goal has been exceeded, as the full SFM system has been implemented on a laptop and tested both indoors and outdoors. No remote control has been necessary. Full integration into an autonomous prototype is still under way.

In addition to achieving the milestones outlined in the original proposal, we have been able to make significant breakthroughs in other problems that recently emerged and that were not part of the original proposal. These include the study of the accommodation cue in vision as well as the dense estimation of 3D shape using variational techniques implemented via numerical solution of partial differential equations (PDEs). A list of the accomplishments on these topics follows:

Visual accommodation and shape from defocus In a series of papers published in the most prestigious refereed conferences and journals [11, 12, 13, 14, 15, 16, 30], we have completely tackled the following problem: *given two or more images of a scene obtained with different focus settings, reconstruct the 3-D shape of the scene as well as its radiance. Furthermore, characterize the accuracy of the estimates, both analytically and experimentally.* We have presented the first optimal algorithms, some of which are implemented in real time. Application of these technology ranges from exploration of small cavities and lumens (e.g. in endoscopy or inspection) and true recognition, for instance of faces, based on 3-D shape rather than on pictorial information. Some of these results have been reported in previous interim reports; new results will be outlined below.

Particle filtering Nonlinear stochastic state estimation algorithms (or "filters") have been presented for systems evolving on Lie groups and homogeneous spaces [6].

Matching despite occlusion Optimal algorithms for region correspondence despite occlusions have been introduced in [7].

Estimation of dense shape and radiance Variational techniques for estimating the 3-D shape of a scene that do not rely on matching point features, but rather estimate a dense surface directly, have been introduced in [19, 20, 21, 33, 36, 37, 38, 39, 40]. This novel line of work shows remarkable results and is gathering considerabl attention from the scientific community. Some preliminary results were outlined in prior interim report. Here we report final results on challenging sequences where no existing method would work.

Modeling dynamic visual processes Stochastic dynamical models of visual process, such as foliage, steam, smoke, fire, have been proposed to support visual recognition tasks. Results for dynamic textures have been reported in [26]. These techniques place the difficult and important problem of detecting and recognizing dynamic "events" (e.g. the presence of a fire, or a person limping) on a solid analytical footing, and shows the first ever experimental results on this problem. We briefly describe preliminary results below.

Multiple motion segmentation Results drawn from algebraic geometry have been crucial in deriving a complete theory for segmenting multiple rigid motions in a sequence of images [34, 35]. For reasons of space, we do not further describe these results, which can be accessed through the relevant publications.

Tracking deforming target A completely novel approach to the difficult problem of visual tracking of deforming targets (where neither the shape nor the motion of the target can change over time, and they are both unknown) has been presented in [33, 37]. The approach will be outlined in more detail in this document.

Laboratory infrastructure and facilities

The first part of the research program was carried out while the PI was at Washington University in St.Louis. As July 2000, the PI has relocated the laboratory at UCLA, where he is the founder and director of the UCLA Vision Lab. The facility is hosted in about 1000 sqft of space with state of the art equipment including a full 6-camera motion-capture system (synchronized infrared marker-sensitive cameras), three robots (two Evolution Robotics ER-1 and a developer platform), purchased through ARO's DURIP. In addition, the laboratory features several PC workstations (mostly dual processors, 2GB of RAM, various processing speeds) donated by Intel, digital camera, frame grabbers, lights, stands, motorized pan-tilt units, embedded units (IQeye3 cameras + FPGAs + Ethernet), 802.11a, 802.11b and 802.11c wireless units etc.

The laboratory currently houses 12 full-time members: 2 postdocs (D. Cremers, A. Duci), one M.D. pursuing his Ph.D. (G. Scarlatis), 8 Ph.D. students (staggered as follows: 2 first-year, 1 second-year, 3 third-year, 1 fourth-year, 1 fifth-year), 1 visiting Ph.D. student (F. Guido). An additional visiting Ph.D. student is scheduled to join the lab in late January (N. Moretto), and an additional Postdoc (R. Vidal) is expected to join us in the spring. All members of the lab are supported on research funding.

List of instruments purchased under the DURIP contract

Enclosed with the submission of this report.

Training of graduate students, industry and government professionals

This project, with a budget of 70K$/year, supported part of the salary and tuition of two graduate students through most of their Ph.D. program. Hailin Jin and Paolo Favaro, currently Ph.D. candidates at Washington University, have successfully completed all the requirements for their doctorate, including the Qualifying

exam and the Candidacy exam, and they are preparing to defend their dissertation sometime between March and June of 2003.

In addition to having completed standard coursework and having performed research under my direction, both students have demonstrated leadership abilities by supervising younger students, such as summer interns in my lab, and also by initiating and completing research projects on their own, independent of my supervision. Some of this work has resulted in an independent publication that lists them as sole authors [16].

The research material developed during the course of this project is at the basis of a textbook under advanced stages of development, expected in press in June of 2003 [24]. This material has been used to develop a new graduate course at UCLA (CS268, Machine Perception), as well as a seminar course (CS269, Visual Recognition and Biometrics). Also, the PI has used this material to design a short course through the UCLA Extension on 3-D modeling and reconstruction from Video. This course, which gathered a surprising success with over 30 participants from industry and government on its first offering, will be repeated annually in late September.

Current sources of funding

The activities of the UCLA Vision Laboratory, directed by the PI, is currently funded by ONR (MURI), AFOSR, DARPA (IXO), NSF (ECS, IIS), NIH, Intel and Microsoft.

Additional funding to further the results of the current projects are sought from ARO, under a pending proposal co-authored with Prof. A. Yezzi of the Georgia Institute of Technology.

Description of research achievements during the last period (Year III)

The last period has coincided with the perfectioning of the algorithms for recovering three-dimensional structure from motion (SFM), both from the theoretical and the experimental points of view.

The theory, which includes methods to handle singular perturbations due to point features appearing and disappearing, has been summarized in the IEEE Transactions PAMI [3].

That theory is at the basis of the implementation of the first ever algorithm for SFM operating in real time. This system is under constant development, and several laboratories in the US and abroad have been able to port our code on their system and test it independently. These include Boston University, Georgia Tech, UC Berkeley, Oxford, Lund, Padova etc.

As promised in the last Interim Report, we have purchased two reconfigurable robots from Evolution Robotics, INC., a Pasadena Company, and, in addition, we have received a third robot as part of their developer network. We have implmented the structure from motion algorithms discussed in the previous Interim reports on a portable laptop platform, and embedded them in the mobile platform to be used for autonomous guidance algorithms in a vision-bsed control scenario. As part of a beginning AFOSR project, we plan to implement these algorithms on mobile platform for landing and low-altitude flight in swarm configurations.

A version of the algorithm followed by dense uncalibrated shape estimation is at the basis of novel algorithms for 3-D terrain mapping and super-resolution image registration. Figure 1 shows the results of applying our algorithms to enhance a small area visible from four images. As a side benefit, the three-dimensional terrain relief is also computed. In addition, we have carried our experiments on tracking non-rigid objects, especially human motion, further. Thanks to MURI funds, we have purchased a full 6-camera motion capture system which we have used to acquire gait data for human subjects (walk, run, limp, skip etc.) Results are presented in [26], and funding is currently being sought for furthening this project.

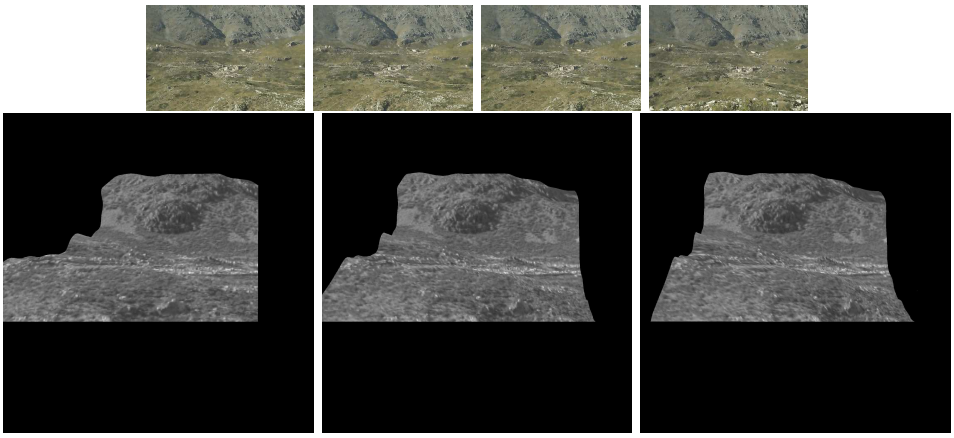

Figure 1: **Dense shape estimation and registration example.** *A collection of aerial images (top) can be used to estimate a dense 3-D model, which supports a radiance function ("texture map") that can be used to generate novel images at an arbitrary resolution from an arbitrary viewpoint (bottom). Data courtesy of M. Pollefeys and L. van Gool.*

Selected preliminary results of new projects and future research plans

In this section we summarize preliminary results of ongoing projects that sprung up from the current project, which constitutes our on-going work. A proposal by the PI and his collaborator Dr. A. Yezzi from Georgia Tech is pending with ARO and include plans to further some of these issues.

Tracking deforming targets

In this section, mostly taken from [37], we report the results obtained with our novel approach to tracking deforming targets. The motion of a target is defined by a group action, and its deformation by a diffeomorphism. Both are unknown, and both are inferred from data using robust variational region-based techniques.

Fig. 2 illustrates the difference between the motion and shape average computed under the Euclidean group, and the affine one. The three examples show the two given shapes γ_i, the mean shape registered to the original shapes, $g_i(\mu)$ and the mean shape μ. Notice that affine registration allows us to simultaneously capture the square and the rectangle, whereas the Euclidean average cannot be registered to either one, and is therefore only an approximation.

Fig. 3 shows the results of tracking a storm. The affine moving average is computed, and the resulting affine motion is displayed. The same is done for the jellyfish in Fig. 4.

Fig. 5 and 6 are meant to challenge the assumptions underlying our method. The pairs of shapes chosen, in fact, are not simply local deformations of one another. Therefore, the notion of shape average is not meaningful *per se* in this context, but serves to compute the change of (affine) pose between the two shapes (Fig. 5). Nevertheless, it is interesting to observe how the shape average allows registering even apparently disparate shapes. Fig. 6 shows a representative example from an extensive set of experiments. In some cases, the shape average contains disconnected components, in some other it includes small parts that are shared by the original dataset, whereas in others it removes parts that are not consistent among the initial shapes (e.g. the tails). Notice that our framework is not meant to capture such a wide range of variations. In particular, it does not possess a notion of "parts" and it is neither hierarchical nor compositional. In the context of non-equivalent shapes (shapes for which there is no group action mapping one exactly onto the other), the *average shape serves purely as a support to define and compute motion in a collection of images of a given deforming shape.*

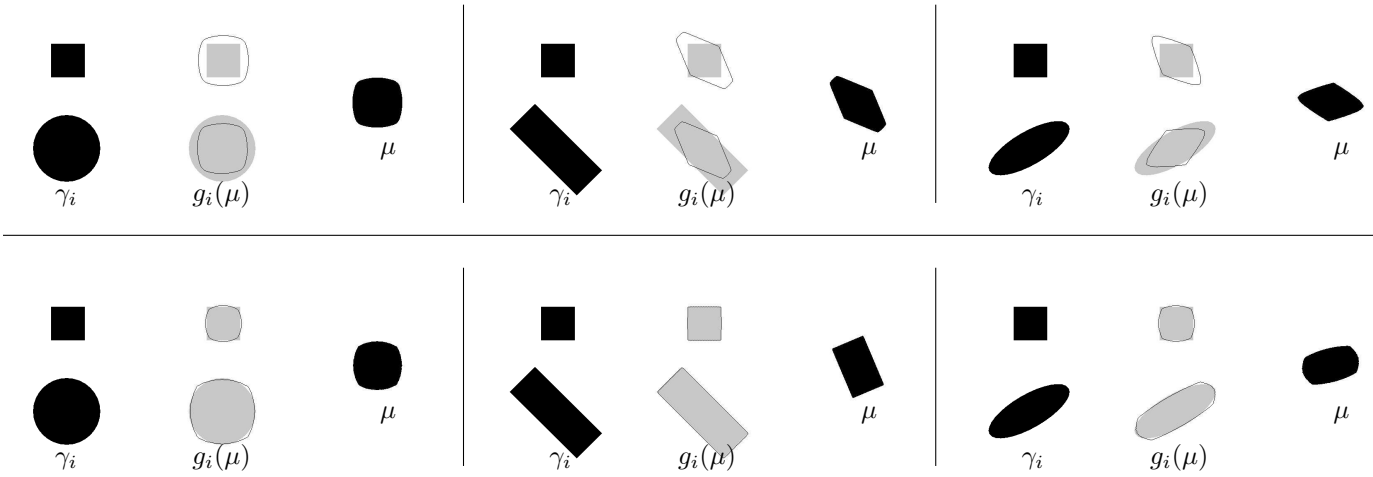

Figure 2: **Euclidean (top) vs. affine (bottom) registration and average.** *For each pair of objects γ_1, γ_2, the registration $g_1(\mu), g_2(\mu)$ relative to the Euclidean motion and affine motion is shown, together with the Euclidean average and affine average μ. Note that the affine average can simultaneously "explain" a square and a rectangle, whereas the Euclidean average cannot.*

Fig. 7 shows the results of simutaneously segmenting and computing the average motion and registration for 4 images from a database of magnetic resonance images of the corpus callosum.

Finally, Fig. 8 shows an application of the same technique to simultaneously register and average two 3D surfaces. In particular, two 3D models in different poses are shown. Our algorithm can be used to register the surfaces and average them, thus providing a natural framework to integrate surface and volume data.

Stereoscopic segmentation

In this project we are developing techniques for inferring shape and radiance of scenes under assumptions that prevent current stereo algorithms to work. These include scenes with no visible "features" (photometrically distinct points), or scenes with dense "texture", that causes local image matching methods to be trapped in local minima. The discussion follows the theory presented in [38].

In figure 9 we show 4 of 22 calibrated views of a scene meant to illustrate the domain of applicability of our algorithm. The scene contains three objects: two shakers and the background. The shakers exhibit very little texture (making local correspondence ill-posed), while the background exhibits very dense texture (making local correspondence prone to local minima). In addition, the shakers have a dark but shiny surface, that reflects highlights that move relative to the camera since the scene is rotated while the light is kept stationary. In figure 10 we show the surface evolving from a large ellipse that neither contains nor is contained in the shape of the scene, to a final solid model. Notice that the parts of the initial surface evolve outwards, while parts evolve inwards in order to converge to the final shape. This bi-directionality is a feature of our algorithm, which is not shared - for instance - by shape carving methodologies. There, once a pixel has been deleted, it cannot be retrieved. In figure 11 we show the final result from various vantage points. In figure 12 we show the final segmentation in some of the original views (top). We also show the segmented foreground superimposed to the original images. Two of the 22 views were poorly calibrated, as it can be seen from the large reprojection error. However, this does not significantly impact the final reconstruction, for there is an averaging effect by integrating data from all views. In figure 13 we show an image from a sequence of views of a watering can, together with the initial surface. The estimated shape is shown in figure 14 The results shown were obtained using a C++ implementation running on a $700MHz$ laptop. For 22 640×480 images and a cubic grid of $128 \times 128 \times 128$ the algorithm takes about 20 minutes to converge (tested by threshold on the iteration residual).

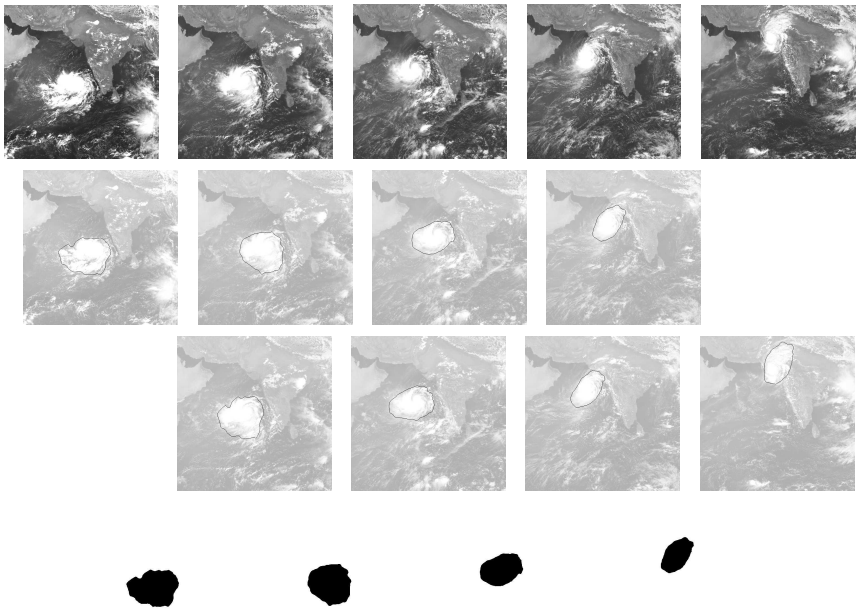

Figure 3: **Storm** *(first row) a collection of images from EUMETSAT ©2001, affine motion of the storm based on two adjacent time instances, (bottom) moving average of order 1.*

Dense 3D shape estimation with with challenging photometry

In this section we report some preliminary experiments on a novel approach to jointly estimate dense shape and non-Lambertian photometry. This allows us to reconstruct objects tha have shiny or translucent surfaces, a case for which all previous passive vision algorithms are at a loss. The algorithm is currently being patented, and its description has been submitted for publication. In this section we test the algorithm on the two objects shown in Figure 16, both courtesy of (whitheld during review). Van Gogh is made of polished metal, and is highly specular. Pseudo-ground truth has been generated by laser scanning followed by manual mesh polishing (Figure 17). Buddha is actually a synthetic scene, meant to simulate translucent material. Ground truth is available (Figure 20). In Figure 17 we show the estimates of shape produced by the algorithm described in [7], together with the estimates obtained by assuming a diffuse + specular reflection model, both compared with pseudo ground truth, obtained with a laser scanner. Our estimate is obviously not as crisp as the ground truth, but it does capture important details on the face. The evolution of the estimate of shape can be seen in Figure 21, as well as in the uploaded movies.

In Figure 22 we show synthetic images generated using the radiance map. Note that the specularities move with the viewpoint. This is also visible in the uploaded movies. In Figure 18 we show a few synthetic images compared with the real images from the same vantage point. In Figure 20 we show the estimated shape for the Buddha in Figure 16. In this case, ground truth is available since the images are synthetic. We also show the results obtained by assuming Lambertian reflection. In Figure 19 we show images synthesized from the model, compared with corresponding true images. In Figure 21 we show the evolution of shape, and in Figure 22 we show several novel views.

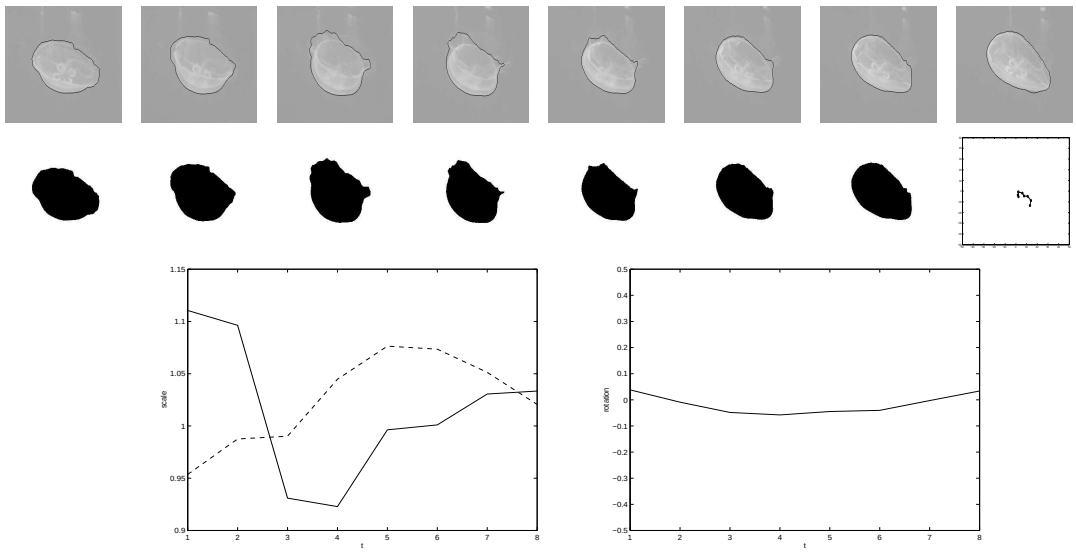

Figure 4: **Jellyfish.** *Affine registration (top), moving average and affine motion (bottom) for the jellyfish. Last row: affine scales along x and y, and rotation about z during the sequence.*

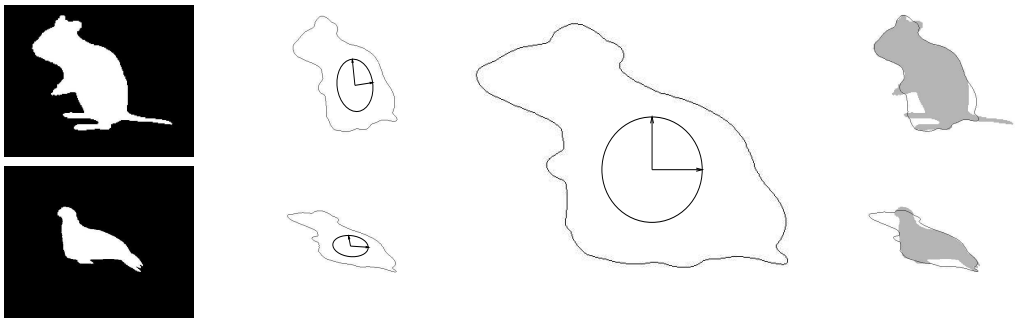

Figure 5: **Registering non-equivalent shapes.** *Left to right: two binary images representing two different shapes; affine registration; corresponding affine shape; approximation of the original shapes using the registration of the shape average based on the set-symmetric difference. Results for the signed distance score are shown in Fig. 6.*

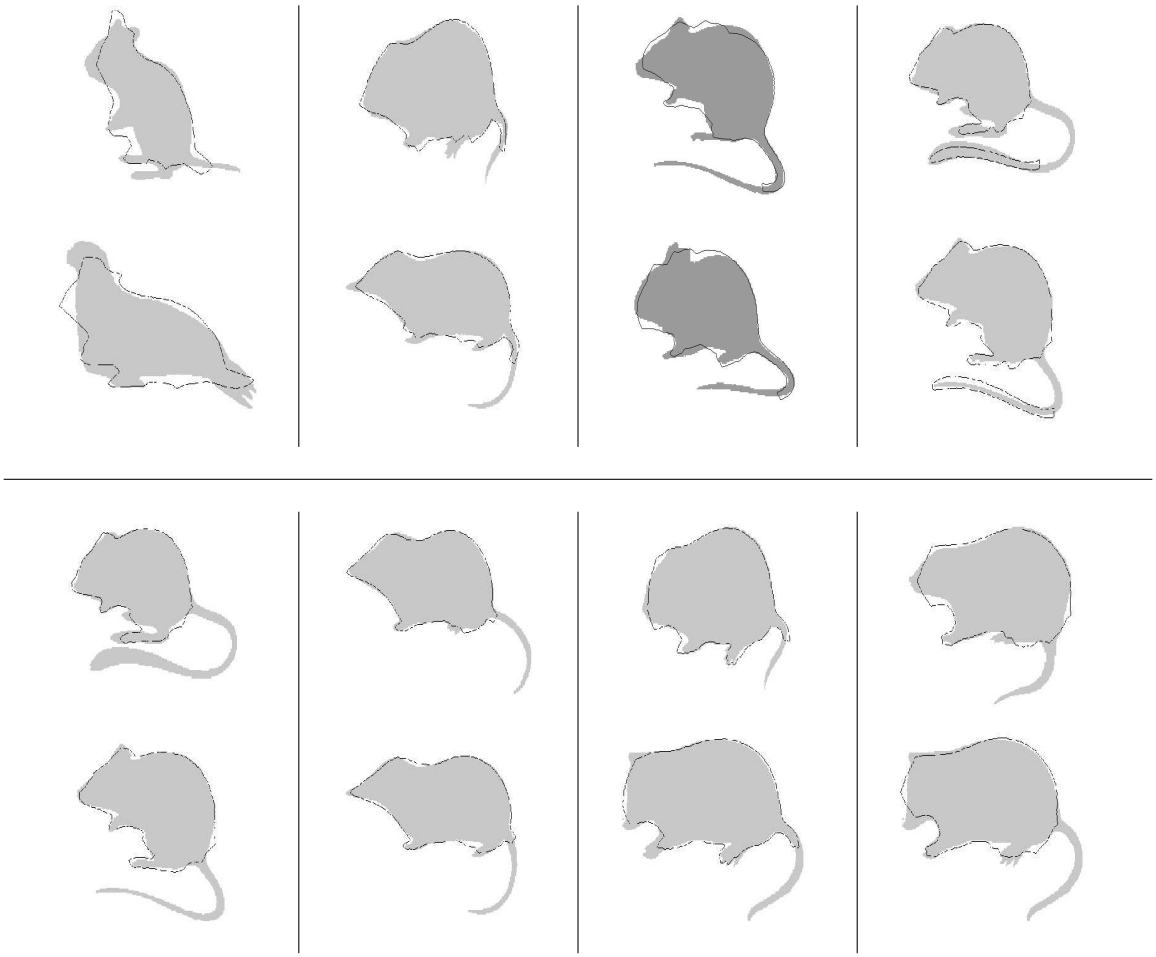

Figure 6: **Biological shapes** *For the signed distance score, we show the original shape with the affine shape average registered and superimposed. It is interesting to notice that in some cases the average shape is disconnected.*

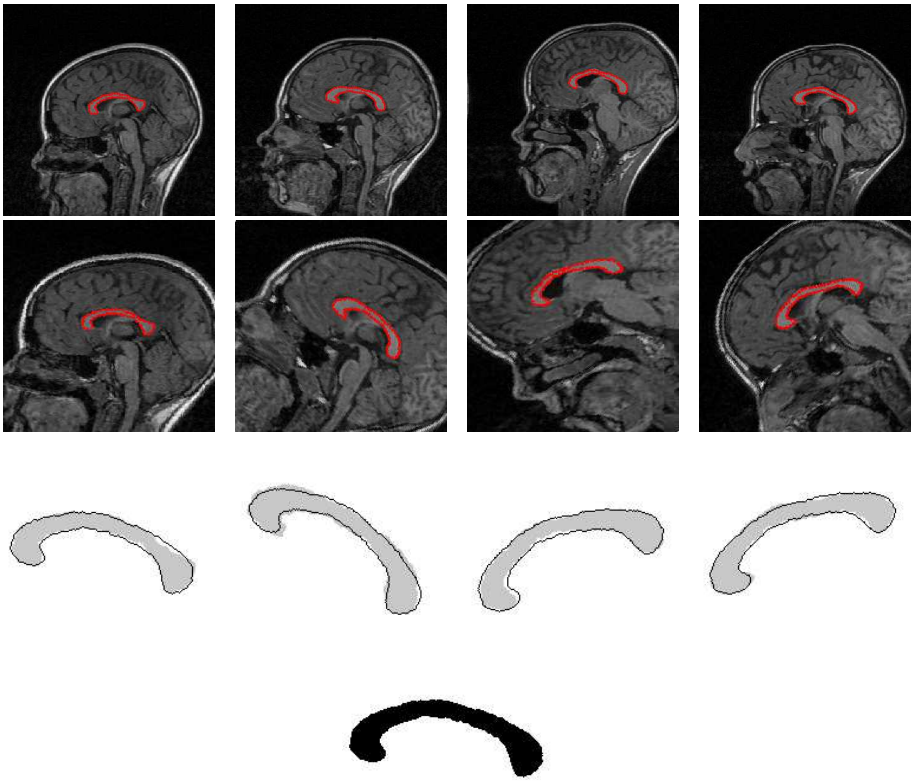

Figure 7: **Corpus Callosum** *(top row) a collection of (MR) images from different patients (courtesy of N. Dutta and A. Jain), further translated, rotated and distorted to emphasize their misalignment, alignment and (bottom) average template corresponding to the affine group.*

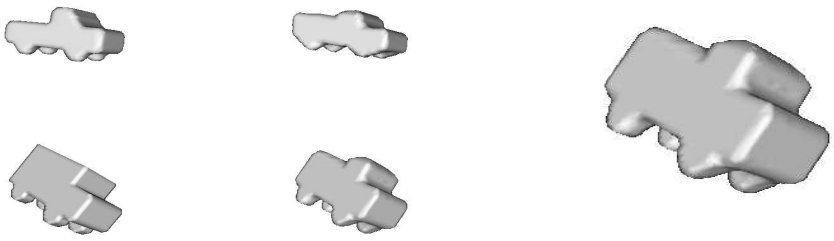

Figure 8: **3D Averaging and registration** *(left) two images of 3D models in different poses (center) registered average (right) affine average. Note that the original 3D surfaces are not equivalent. The technique presented allows "stitching" and registering different 3D models in a natural way.*

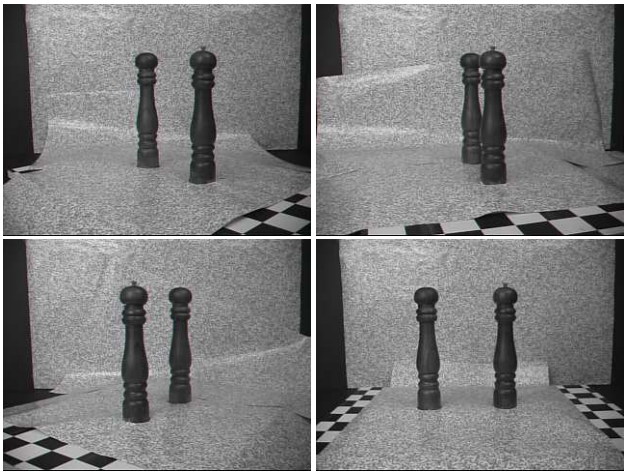

Figure 9: *Original "salt and pepper" sequence (4 of 22 views).*

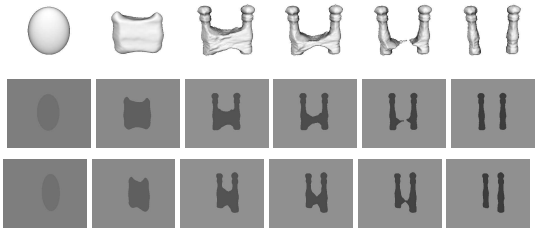

Figure 10: *(top) Rendered surface during evolution (6 of 800 steps). Notice that the initial surface is neither contained nor contains the actual scene. (Bottom) segmented image during the evolution from two different viewpoints.*

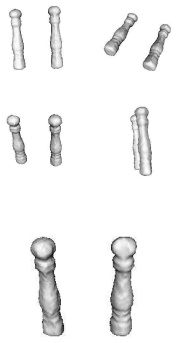

Figure 11: *Final estimated surface, seen from several viewpoints. Notice that the bottoms of the salt and pepper shakers are flat, even though no data was available. This is due to the geometric prior, which in the absence of data results in a minimal surface being computed.*

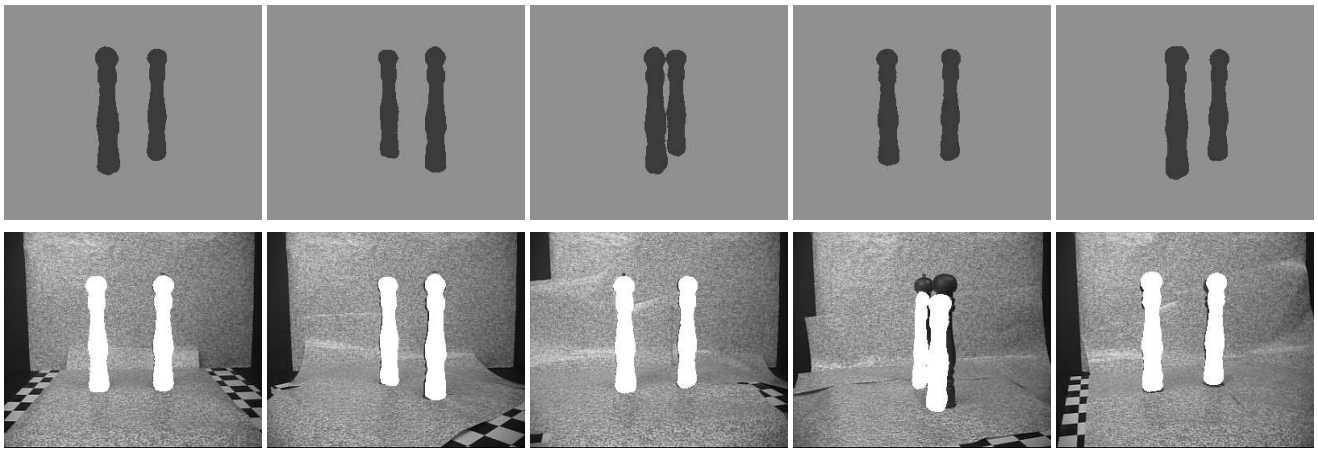

Figure 12: *(Top) image segmentation for the salt and pepper sequence. (bottom) Segmented foreground superimposed to the original sequence. The calibration in two of the 22 images was dramatically wrong. However, the effect is mitigated by the global integration, and the overall shape is only marginally affected by the calibration errors.*

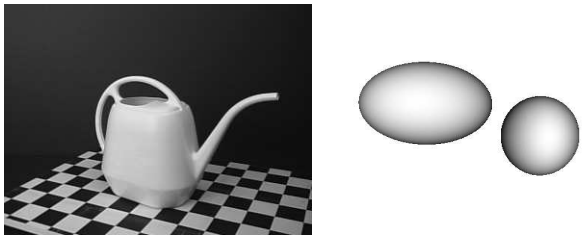

Figure 13: *The "watering can" sequence and the initial surface. Notice that the initial surface is not simply connected and does not include and is not included by the shape. In order to capture a hole it is necessary that it intersects the initial surface. One way to guarantee this is to start with a number of small surfaces.*

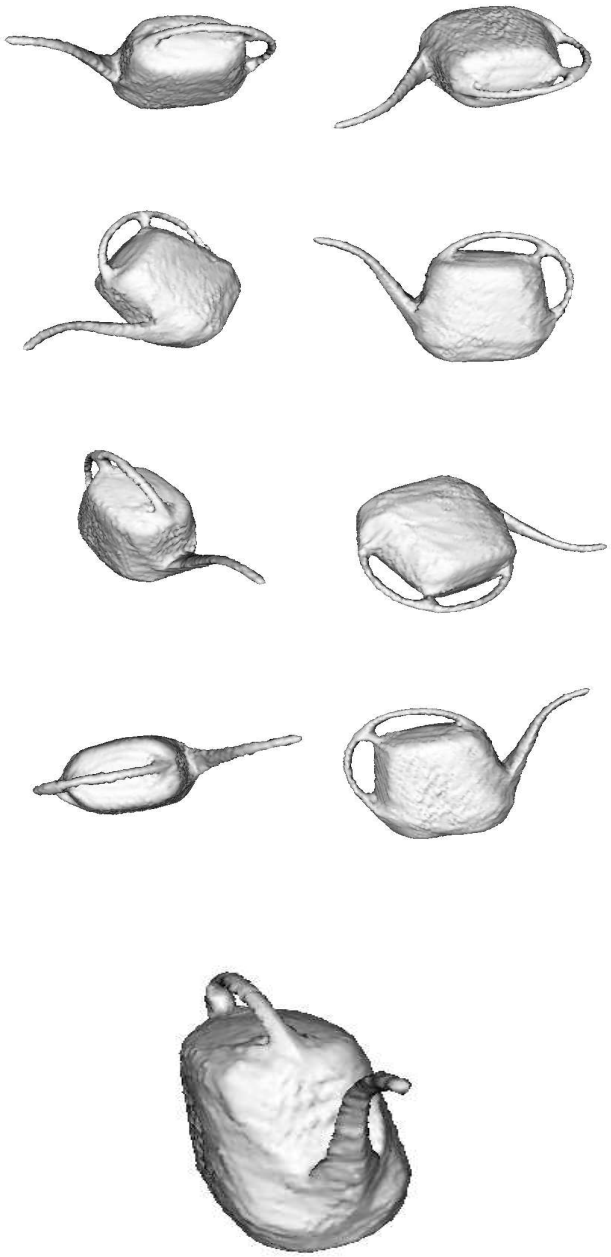

Figure 14: *Final estimated shape for the watering can. The two initial surfaces have merged. and the topology and geometry of the watering can has been correctly captured.*

Figure 15: *(top) Rendered surface during evolution for the watering can.*

Figure 16: *Scenes with strong specularities (left) or made of translucent materials with no distinct point features are a challenge to most stereo algorithms.*

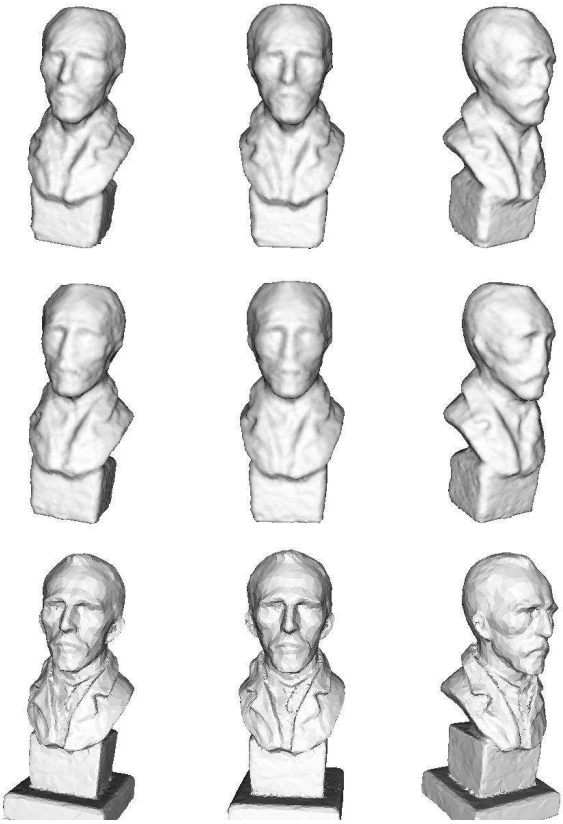

Figure 17: Estimated shape (top), compared with pseudo-ground truth (bottom), obtained with a 3D laser scanner and manual mesh cleaning. Our results improve those obtained with a purely diffuse + specular model (middle).

Figure 18: *Synthetic images using the estimated radiance tensor (top) compared with the true images taken from the same vantage point.*

Figure 19: *Synthetic images obtained with the estimated radiance tensor field (top) compared with the true images taken from the same vantage point.*

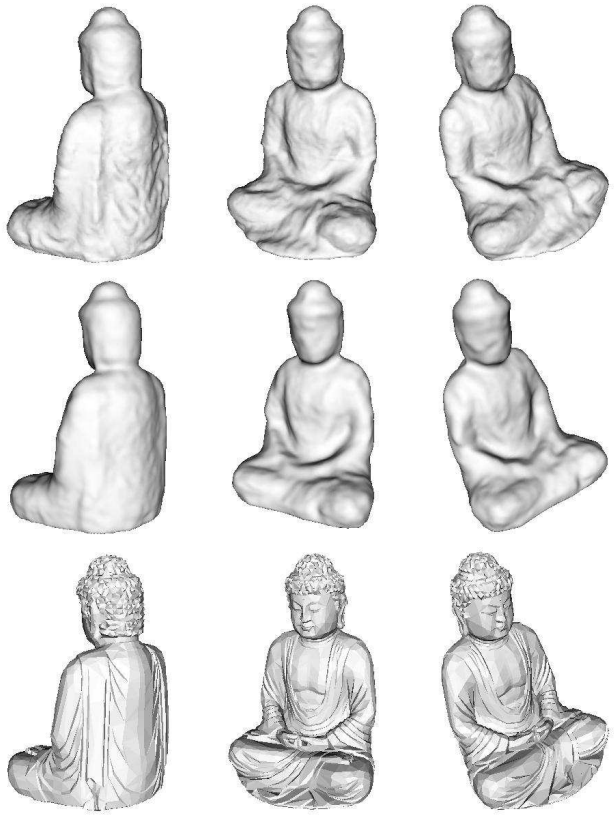

Figure 20: Estimated shape (top), compared with ground truth (bottom). Also compare with the results obtained by assuming Lambertian reflection (middle).

Matching with missing parts and occlusions

In this section we report some preliminary experiments, documented in [7], to match image structures and shapes despite missing parts.

Modeling and recognition of human gaits

In this section we give some very preliminary results on modeling human gaits using stochastic dynamical systems. A stochastic model is identified from data, acquired using DURIP funding. The model is then simulated in order to ascertain whether it captures the crucial statistical features, for instance whether it allows to visually discriminate between a normal walk and limping. We are currently in the process of studying techniques to exploit the higher-order statistics inferred from data sequences in order to recognize walking gaits. The ultimate goal is to identify classes of motions (e.g. walking vs. running vs. limping) as well as individuals from their walking gaits.

Acknowledgments

We wish to acknowledge the generous support and encouragement by Program Managers Dr. Linda Bushnell and Dr. Hua Wang.

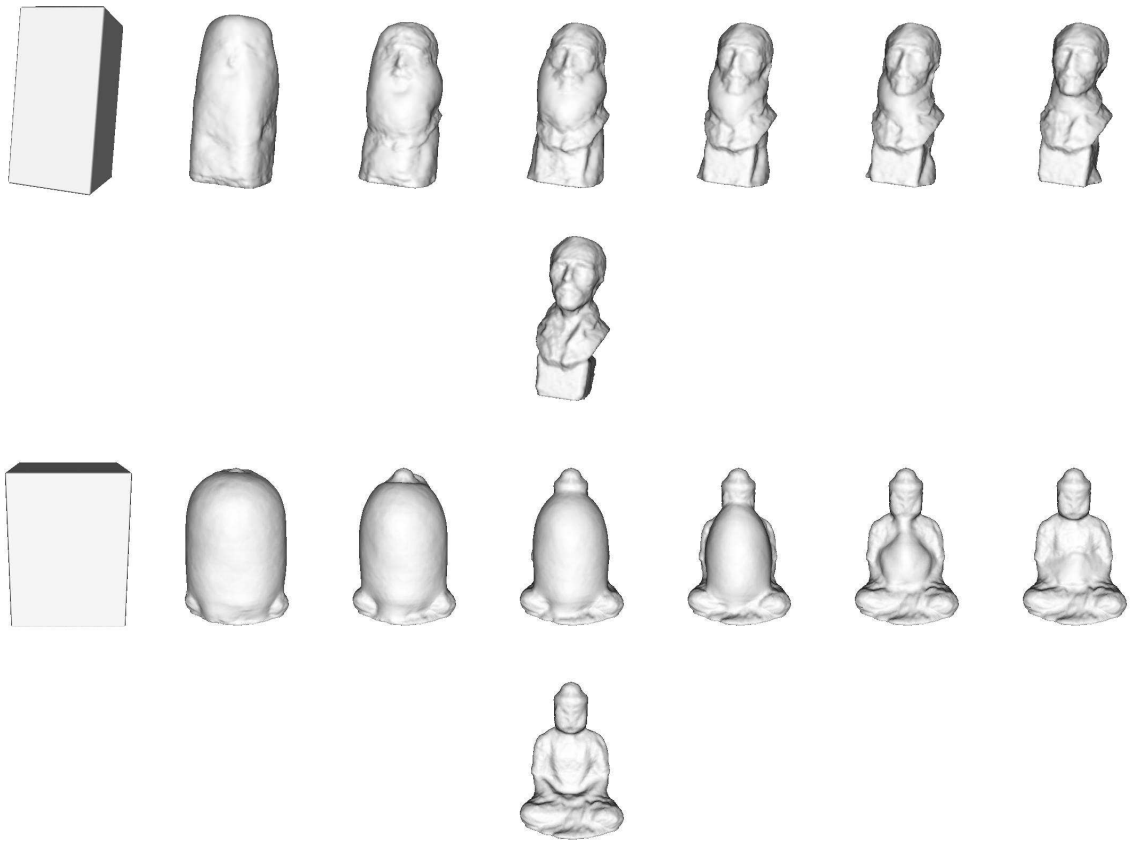

Figure 21: *Shape evolution for Van Gogh (top) and Buddha (bottom).*

References

[1] *Conference tracks advances in computer vision.* EE Times, http://www.eet.com/story/technology/OEG20000626S0073, July 11, 2000.

[2] A. Chiuso, R. Brockett, and S. Soatto. Optimal structure from motion: local ambiguities and global estimates. *Intl. J. of Computer Vision*, 39(3):195–228, September 2000.

[3] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. Motion and structure causally integrated over time. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24 (4):523–535, 2002.

[4] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. 3-d motion and structure causally integrated over time: Implementation. In *European Conference on Computer Vision*, pages 735–750, June 2000.

[5] A. Chiuso and S. Soatto. 3-d motion and structure causally integrated over time: Theory. In *Tutorial lecture notes of IEEE Intl. Conf. on Robotics and Automation*, April 2000.

[6] A. Chiuso and S. Soatto. Monte carlo filtering on lie groups. In *Intl. Conf. on Decision and Control*, pages 304–309, December 2000.

[7] A. Duci, A. Yezzi, S. Mitter, and S. Soatto. Region matching with missing parts. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 3, pages 28–64, 2002.

[8] P. Favaro, H. Jin, and S. Soatto. Real-time virtual object insertion. In *Proc. of the Intl. Conf. on Computer Vision*, page 749, 2001.

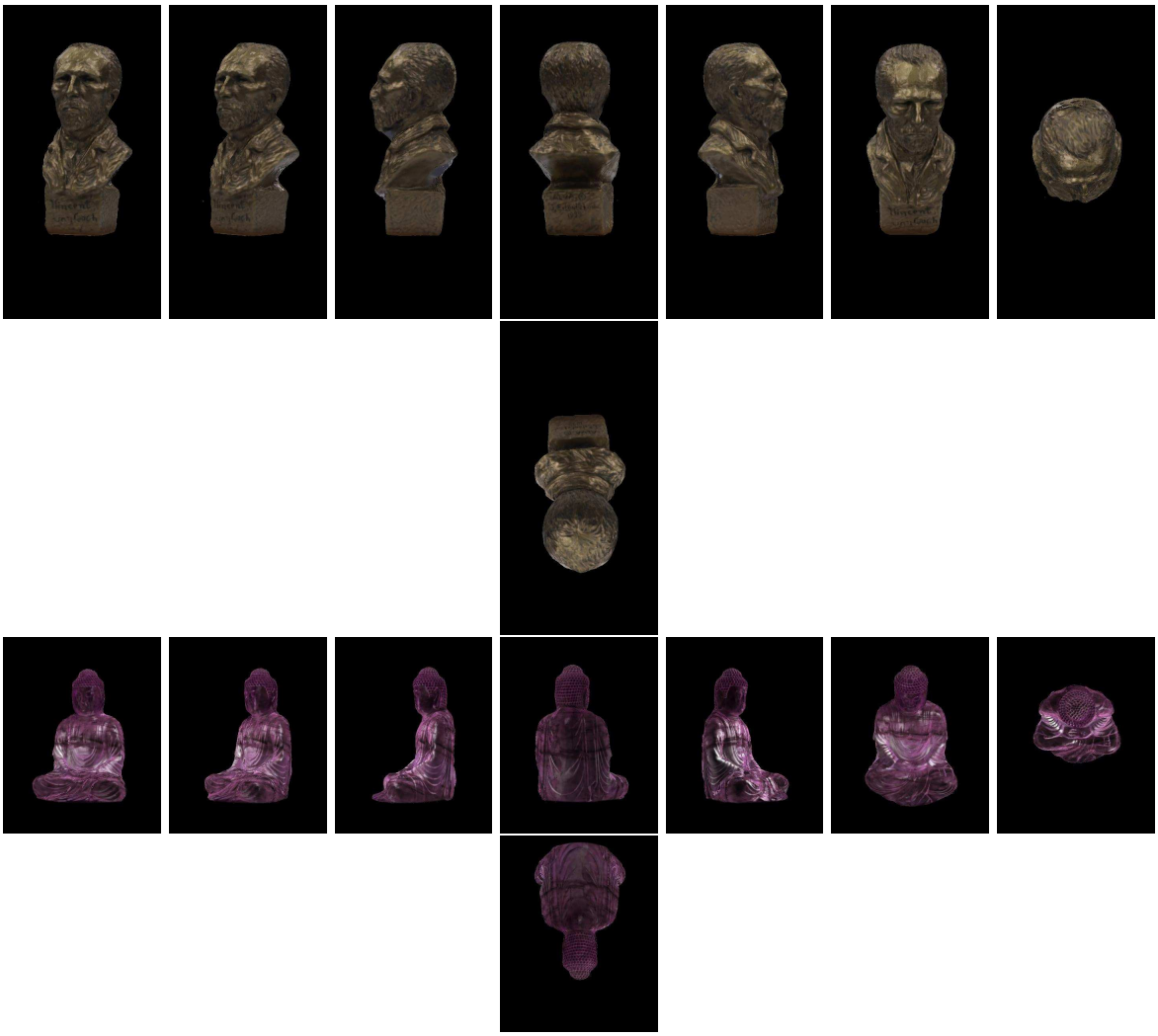

Figure 22: *Synthetic images obtained from the estimated radiance. As it can be seen, the appearance changes with the vantage point. See also the uploaded movies.*

[9] P. Favaro, H. Jin, and S. Soatto. A semi-direct approach to structure from motion. page The Visual Computer, (in press) 2002.

[10] P. Favaro, H. Jin, and S. Soatto. A semidirect approach to structure from motion. In *IEEE Intl. Conf. on Image Analysis and Processing*, pages 250–255, Palermo– Italy, September 2001.

[11] P. Favaro, A. Mennucci, and S. Soatto. Observing shape from defocused images. 51(2), in press, 2003.

[12] P. Favaro and S. Soatto. Learning shape from defocus. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 2, pages 735–745, 2002.

[13] P. Favaro and S. Soatto. Shape and reflectance estimation from the information divergence of blurred images. In *European Conference on Computer Vision*, pages 755–768, June 2000.

[14] P. Favaro and S. Soatto. Seeing beyond occlusions (and other marvels of a finite lens aperture). In *UCLA CSD-020041*, November 1, 2002.

[15] P. Favaro and S. Soatto. Shape from anisotropic diffusion. In *UCLA CSD-020042*, November 1, 2002.

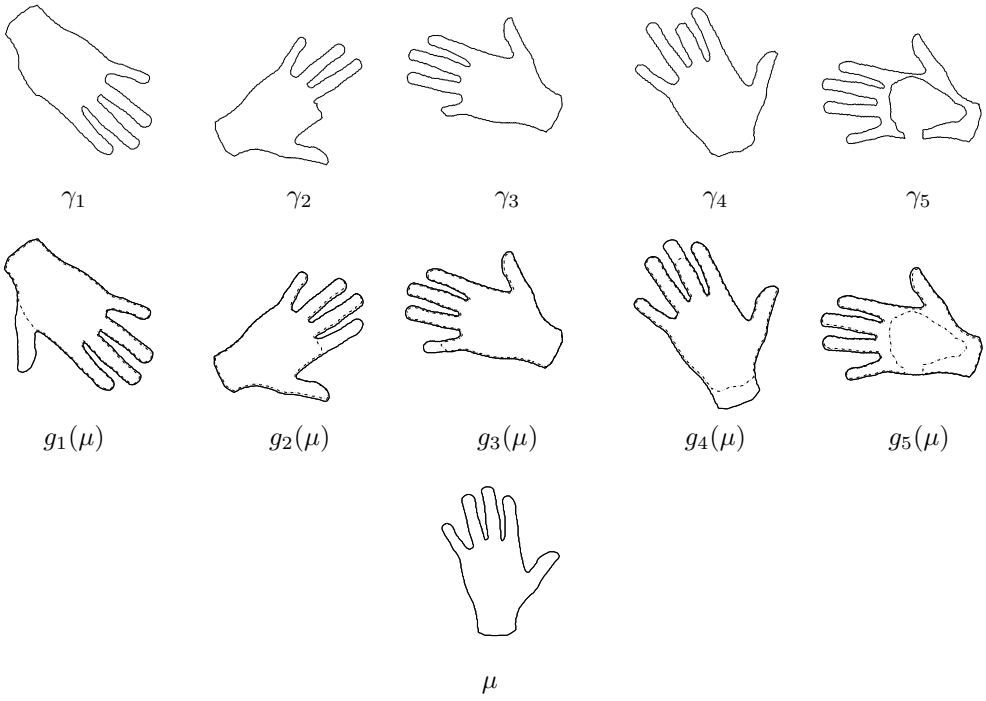

Figure 23: **Hands.** *(Top) a collection of images of the same hand in different poses with different missing parts. The support of the missing parts is unknown. (Middle) similarity group, visualized as a "registered" image. (Bottom) estimated template corresponding to the similarity group ("complete shape").*

[16] H. Jin and P. Favaro. A variational approach to shape from defocus. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 2, pages 18–30, 2002.

[17] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *Proc. of the Intl. Conf. on Computer Vision*, pages 684–689, 2001.

[18] H. Jin, P. Favaro, and S. Soatto. Real-time 3-d motion and structure from point features: a front-end system for vision-based control and interaction. In *Computer Vision and Pattern Recognition; code available from* `http://vision.ucla.edu`, pages 778–779, June 2000.

[19] H. Jin, R. Tsai, L. Chen, A. Yezzi, and S. Soatto. Estimation of 3d surface shape and smooth radiance from 2d images: A level set approach. *J. of Comp. Physics*, accepted, 2002.

[20] H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *Proc. of the Intl. Conf. on 3DPVT*, pages 626–630, June 2002.

[21] H. Jin, A. Yezzi, and S. Soatto. Stereoscopic shading: integrating shape cues in a variational framework. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pages 169–176, June 2000.

[22] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. Euclidean reconstruction and reprojection up to subgroups. In *Proc. of the IEEE Intl. Conf. on Computer Vision*, pages 773–780, 1999.

[23] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. Euclidean reconstruction and reprojection up to subgroups. *Intl. J. of Computer Vision*, 38(3):219–229, July 2000.

[24] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3D vision, from images to models*. Springer Verlag, (in preparation).

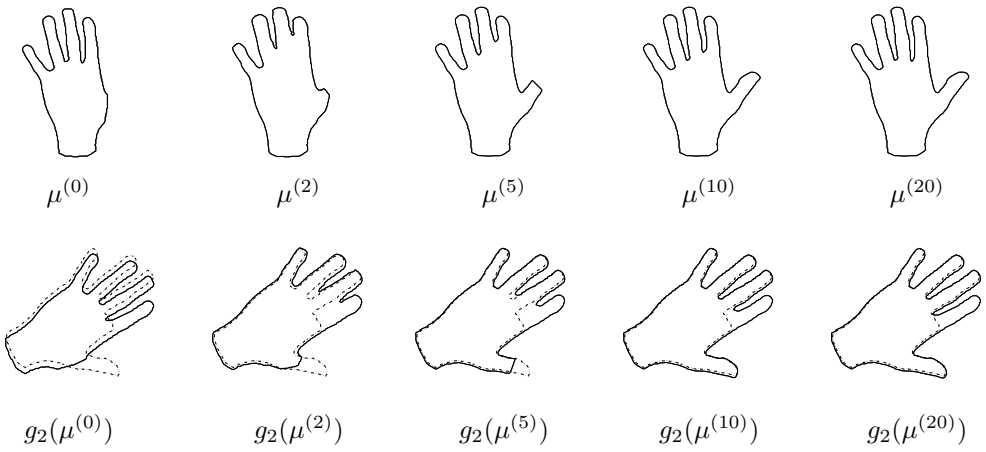

Figure 24: **Hands Evolution.** *(Top) evolution of the complete shape for* $t = 0, \ldots, 20$. *(Bottom) evolution of* $g_2(\mu)$ *for* $t = 0, \ldots, 20$.

[25] A. Mennucci and S. Soatto. On observing shape from defocused images. In *Proc. of the Intl. Conf. on Image Analysis and Processing*, pages 550–555, 1999.

[26] P. Saisan, G. Doretto, Y. Wu, and S. Soatto. Dynamic texture recognition. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages II 58–63, December 2001.

[27] S. Soatto. *Progress in systems and control theory: Dynamical systems, control, coding, computer vision*, chapter The accommodation cue in vision, pages 432–448. Birkhouser, 1999.

[28] S. Soatto and R. Brockett. Optimal structure from motion: Local ambiguities and global estimates. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 282–288, Santa Barbara, CA, June 1998.

[29] S. Soatto, G. Doretto, and Y. Wu. Dynamic textures. In *Proc. of the Intl. Conf. on Computer Vision*, pages 439–446, 2001.

[30] S. Soatto and P. Favaro. A geometric approach to blind deconvolution with application to shape from defocus. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pages 10–17, June 2000.

[31] S. Soatto and P. Perona. Reducing "structure from motion": a general framework for dynamic vision. part 1: modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(9):993–942, September 1998.

[32] S. Soatto and P. Perona. Reducing "structure from motion": a general framework for dynamic vision. part 2: Implementation and experimental assessment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(9):943–960, September 1998.

[33] S. Soatto and A. Yezzi. Deformotion: deforming motion, shape average and the joint segmentation and registration of images. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 3, pages 32–47, 2002.

[34] R. Vidal, S. Soatto, and S. Sastry. A factorization method for 3d multi-body motion estimation and segmentation. In *Proc. of the Allerton Conference on Signals and Systems*, October 2002.

[35] R. Vidal, S. Soatto, and S. Sastry. Segmentation of dynamic scenes from the multi-body fundamental matrix. In *Proc. of the Workshop on Analysis of Dynamic Scenes*, June 2002.

[36] A. Yezzi and S. Soatto. Stereoscopic segmentation. In *Proc. of the Intl. Conf. on Computer Vision*, pages 59–66, 2001.

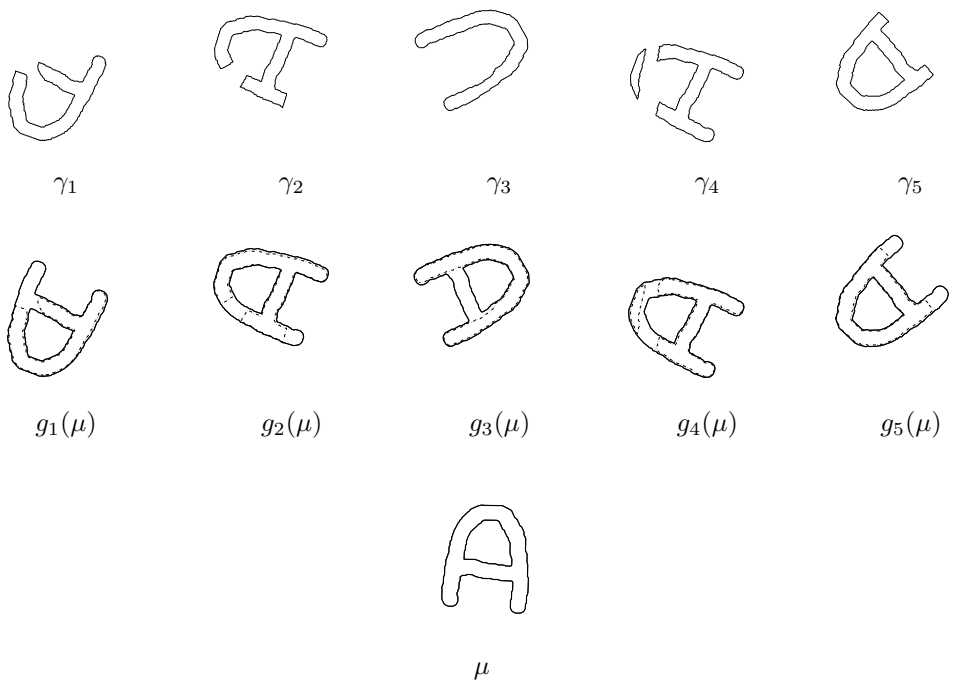

Figure 25: **Letter "A."** *(Top) a collection of images of the letter "A" in different poses with different missing parts. The support of the missing parts is unknown. (Middle) similarity group ("registration"). (Bottom) estimated template corresponding to the similarity group ("complete shape").*

[37] A. Yezzi and S. Soatto. Deformotion: deforming motion, shape average and the joint segmentation and registration of images. *Intl. J. of Comp. Vis.*, accepted, 2003.

[38] A. Yezzi and S. Soatto. Stereoscopic segmentation. *Intl. J. of Computer Vision*, (accepted) 2003.

[39] A. Yezzi, S. Soatto, A. Tsai, and A. Willsky. *Mathematics and Multimedia*, chapter Curve and Surface Evolution for Image Segmentation and Stereo Reconstruction Using the Mumford-Shah Functional. IMA, 2002.

[40] A. J. Yezzi, S. Soatto, H. Jin, A. Tsai, and A. Willsky. *Geometric Level Set Methods in Imaging, Vision and Graphics*, chapter Stereoscopic segmentation. Springer Verlag, 2003.

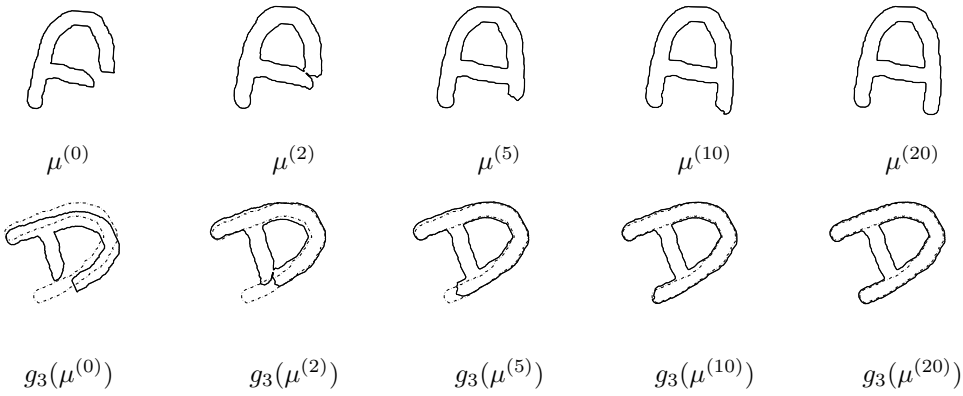

Figure 26: **Letter "A" Evolution.** *(Top) evolution of the complete shape for* $t = 0, \ldots, 20$. *(Bottom) evolution of* $g_3(\mu)$ *for* $t = 0, \ldots, 20$.

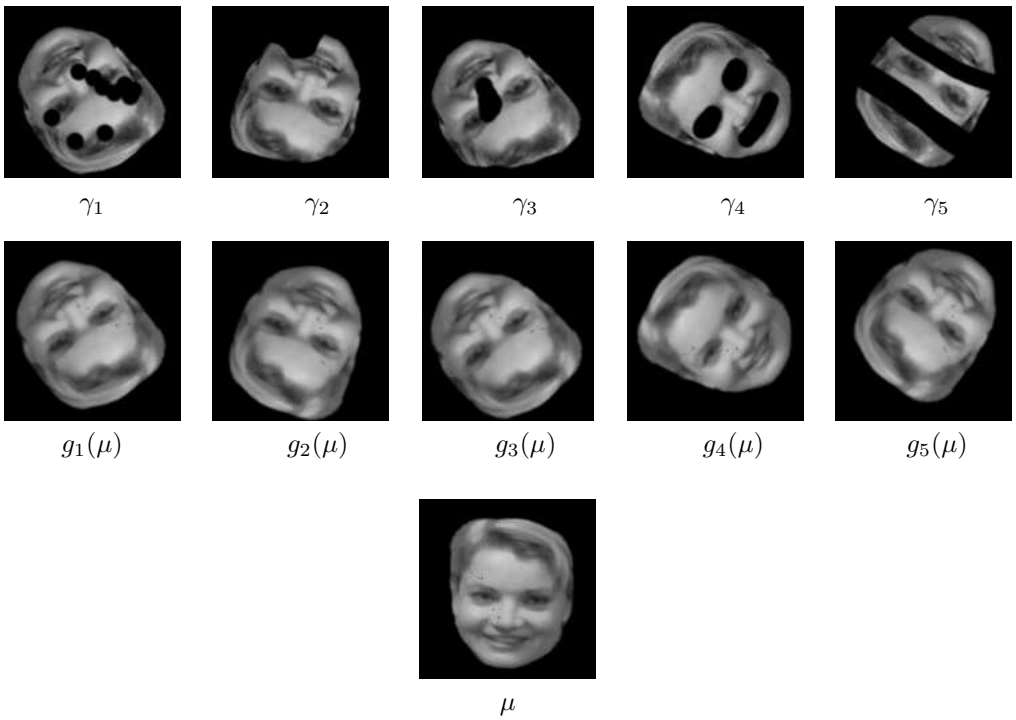

Figure 27: **Faces** *(Top) a collection of images of the same face in different poses with different missing parts. The support of the missing parts is unknown. (Middle) similarity group, visualized as a "registered" image. (Bottom) estimated template corresponding to the similarity group ("complete image").*

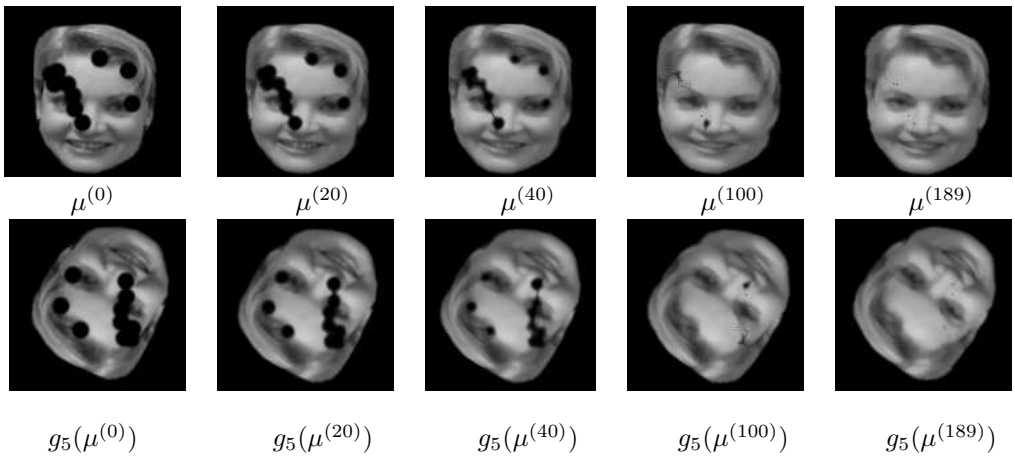

Figure 28: **Face evolution.** *(Top) evolution of the complete image for* $t = 0, \ldots, 189$. *(Bottom) evolution of* $g_5(\mu)$ *for* $t = 0, \ldots, 189$.

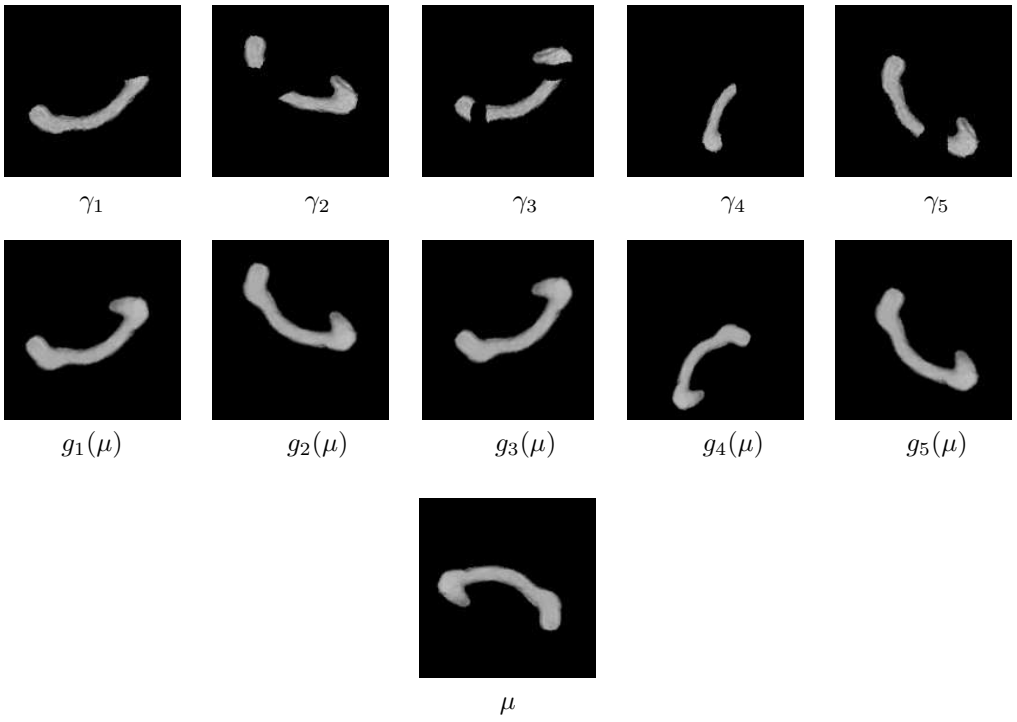

Figure 29: **Corpus Callosum.** *(Top) a collection of images of the same corpus callosum in different poses with different missing parts. The support of the missing parts is unknown. (Middle) similarity group, visualized as a "registered" image. (Bottom) estimated template corresponding to the similarity group ("complete image").*

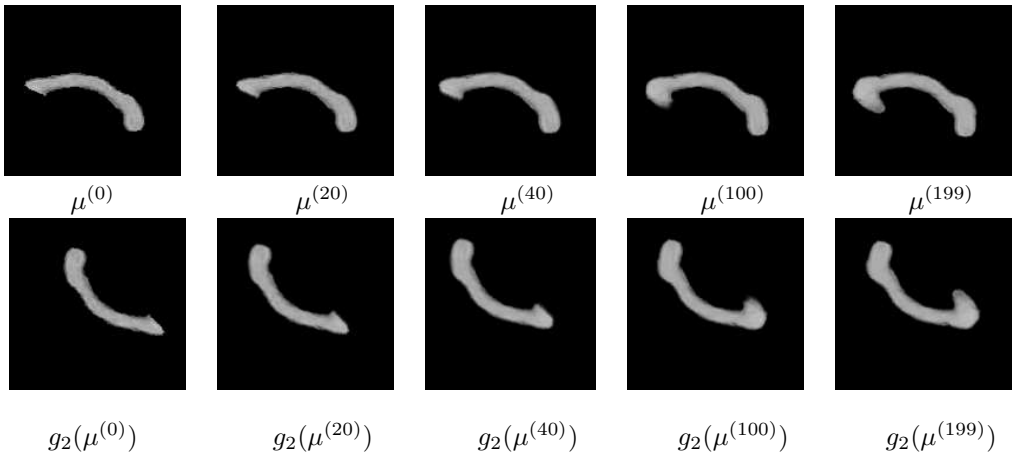

Figure 30: **Corpus Callosum evolution.** *(Top) evolution of the complete image for* $t = 0, \ldots, 199$. *(Bottom) evolution of* $g_2(\mu)$ *for* $t = 0, \ldots, 199$.

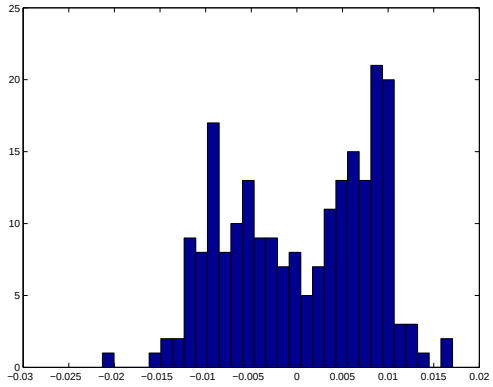

Figure 31: Histogram of the values assumed by the first component of the residual of the learning applied to the walk sequence

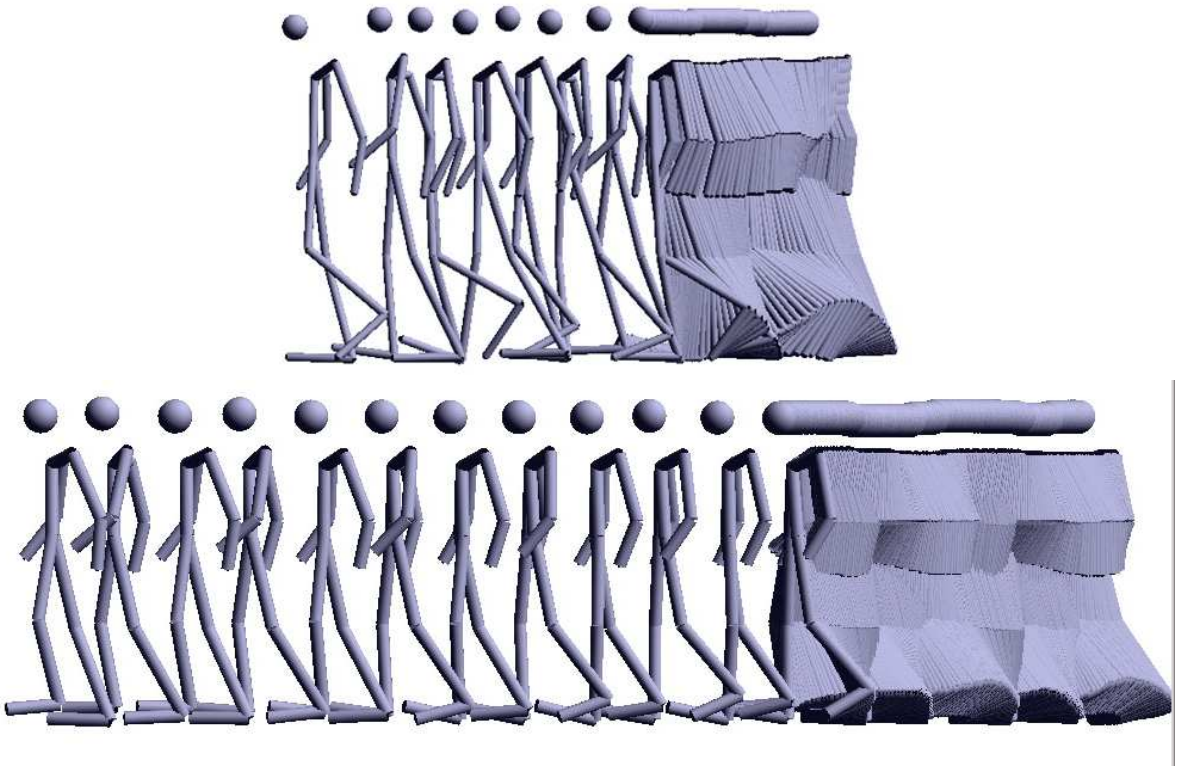

Figure 32: Motion sequences for walking. First row is the original walk data, second row is the synthesized sequence.

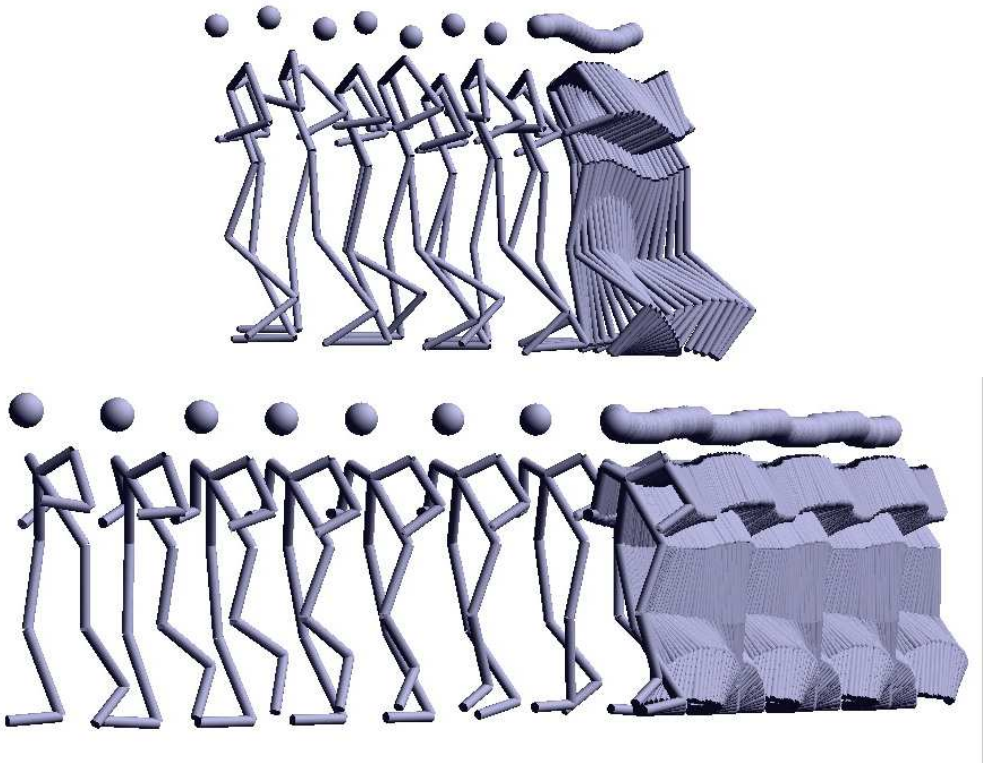

Figure 33: Motion sequences for running. First row original data, second row synthesized motion.