AD________

GRANT NUMBER DAMD17-94-J-4196

TITLE: Digital Mammographic Image Compression

PRINCIPAL INVESTIGATOR: Jun Wang
Dr. H.K. Huang

CONTRACTING ORGANIZATION: University of California,
San Francisco
San Francisco, California 94143-0962

REPORT DATE: August 1996

TYPE OF REPORT: Final

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE August 1996

3. REPORT TYPE AND DATES COVERED Final (15 Jul 94 - 15 Jul 96)

4. TITLE AND SUBTITLE
Digital Mammographic Image Compression

5. FUNDING NUMBERS
DAMD17-94-J-4196

6. AUTHOR(S)
Jun Wang
Dr. H.K. Huang

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, San Francisco
San Francisco, California 94143-0962

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 2170-25012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

13. ABSTRACT *(Maximum 200*

This research includes two parts. In the first part of the study, we proposed a method to combine the proposed lossless compression algorithm with a digitization process to increase the overall throughput of the digitization and compression operations. Conventionally, digitization and compression of mammograms are done separately. Compression is applied to a full digital image after it has been acquired through digitization. In this research, the proposed lossless compression process is implemented in the acquisition computer connected to the scanner, running concurrently with the digitization process. The implementation of the compression algorithm is such that there is no extra time required for compression.

In the second part of the study, we first evaluated a group of wavelet filter banks and selected the optimal filter bank for mammogram compression. We then compared different quantization strategies for different wavelet subbands. The optimal quantization parameters were proposed.

14. SUBJECT TERMS
Breast Cancer, Digital Mammograms, Image Compression
Wavelet Transform, On-line Compression

15. NUMBER OF PAGES 17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Wagner

PI - Signature Date

Table of Contents

1. Introduction

Digital mammography has certain advantages over conventional screen-film mammography[1,2]. These include the ability to manipulate images by changing the look-up-table, zooming and scrolling, imaging processing, and rapidly sending the images to remotely located experts for consultation.

There are two ways to generate digital mammograms: a) secondary digitization, in which conventional mammography films are digitized; and b) direct acquisition of primary digital images. Currently, high resolution (50 μm) film digitizers are commercially available for secondary digitization and the image quality satisfies most diagnostic requirements in mammography. However, primary full-field digital mammography units are not yet commercially available.

In secondary digitization, it has been shown that a 50-100 μm sampling distance is required in order to achieve an adequate resolution for digital mammography[3,4]. At these high resolutions, the digitization time is long and the images created are large, normally between 10-40 MBytes/image. Both the storage and transmission cost of such large images will increasingly become a problem as more digital mammograms are created.

Image compression reduces data storage requirements while preserving useful information[5,6,7]. It provides a solution to handling such large digital images efficiently. Most existing compression methods are developed for engineering purposes. They do not consider the image characteristic and stringent requirements of image quality in medicine. The purpose of this research is to develop loss and lossless compression techniques for digital mammograms.

In the previous work, we investigated both lossless and lossy compression methods[8]. In the lossless compression, we developed a structure lossless compression method for mammograms. The algorithm utilizes the unique shape characteristics and image characteristics of mammograms. The combination of the segmentation and the prediction coding enables us to achieve high compression ratios without losing any useful information. We also developed a lossless compression method using a wavelet transform. The optimal wavelet filter and the quantization strategy were not optimized in the preliminary study.

The research of the past year includes two parts. In the first part of the study, we proposed a method to combine the proposed lossless compression algorithm with a digitization process to increase the overall throughput of the digitization and compression operations. Conventionally, digitization and compression of mammograms are done separately. Compression is applied to a full digital image after it has been acquired through digitization. In this research, the proposed lossless compression process is implemented in the acquisition computer connected to the scanner, running concurrently with the digitization process. The implementation of the compression algorithm is such that there is no extra time required for compression. In the second part of the study, we investigated the optimal wavelet filter bank and the optimal quantization strategy for the wavelet compression.

2. Body

2.1 Method

In this section, we will first describe the on-line structure lossless compression method and then describe the optimal parameter for the wavelet compression.

2.1.1 On-line Compression and Digitization Processes

The on-line lossless compression method was derived from the previous work. It consists of two steps. The first step segments the breast image from its background. The pixels beyond the boundary of the breast are discarded. The second step compresses the remaining portion of the image using a predictive lossless compression technique. Both segmentation and lossless compression can be applied to one line of the image data at a time.

The segmentation process scans one line of the image at a time to detect the boundary of the breast image. Each line of data is scanned to find the boundary points. The information of the previous lines of the image are used to determine the current boundary point. The next step of the structure-lossless compression uses predictive coding[9] to compress the segmented breast image. The prediction starts from the boundary of the breast and moves to the chest wall direction. The current pixel is predicted from the previous pixel in the same line. Since the neighboring pixels are correlated, the difference between adjacent pixels is generally small.

Huffman coding[10] with a pre-determined table is used to code the resulting differences of the pixels. The Huffman table is determined by a set of randomly selected mammograms. If an individual pixel is not in the table, it is recorded separately.

A film digitizer is controlled via an acquisition computer. The digitizer scans (or digitizes) a film one line at a time from the left to the right, advancing line by line from the top to the bottom of the film. The data generated is first stored in an internal buffer in the digitizer. When the buffer is full, the data is transferred to the acquisition computer which has a larger size memory buffer that will hold the entire image. This scanning and transferring process continues several lines at a time, until the whole film is scanned.

During the conventional digitization process the sole responsibility of the acquisition computer is to read the data from the digitizer's memory buffer into its own memory buffer. So the acquisition computer CPU is mostly idle when no data is being transferred. The time to digitize a film is determined by the digitizer scanning speed and the film size. The scanning speed is normally 50-100 lines/second, depending on the sampling distances.

We utilize the computer's idle CPU time to compress the digitized image during digitization. The compression is implemented as a second process running concurrently with the digitization process in the acquisition computer (Figure 1). Both processes share the same image buffer. A semaphore is used to control the access of the shared image memory by the two processes.

First, the scanning process reads several lines of data from scanner and writes into the acquisition memory buffer whenever there is data ready in the scanner's temporary buffer. After writing data each time, the scanning process increases the semaphore by one.

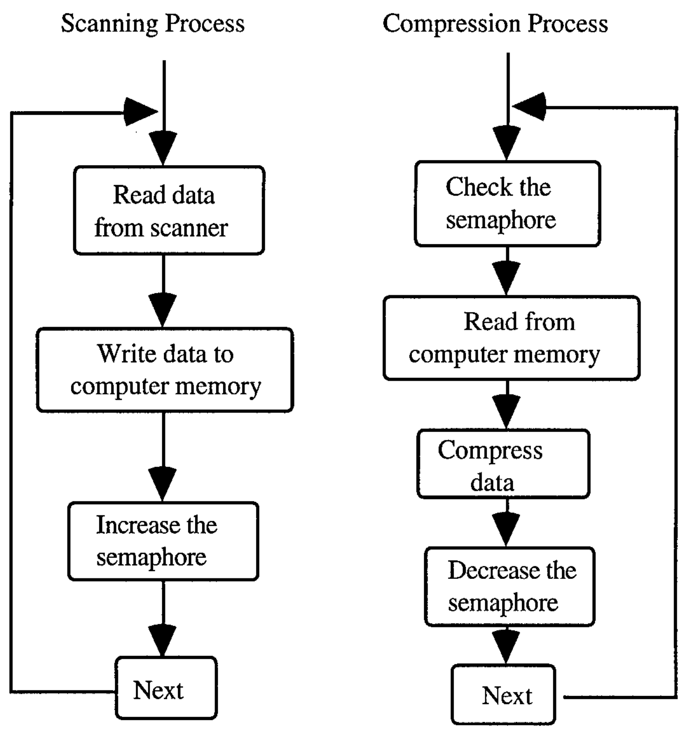

Figure 1. Data flow of the concurrent scanning and compression processes

Meanwhile, the compression process checks the semaphore. If the semaphore indicates new data in the acquisition computer memory, the compression process reads these several lines of data from the image buffer, finds the boundary of the breast and compresses this breast image portion with predictive coding method as described in the previous section. The compression process then decreases the semaphore by one. If there is no new data in the memory, the semaphore blocks the memory and stops the compression process until new data arrives in the memory. These processes continue until an entire film has been scanned and the image is compressed.

If the compression is faster than the scanning, compression can be done as soon as the scanning finishes. The scanning rate is normally 40 to 100 lines/sec. We have implemented this on-line compression algorithm in a film digitizer. In this prototype system, the compression rate achieved is faster than the scanning rate.

2.1.2 The Optimal Wavelet and Quantization Parameters

A diagram of a wavelet compression[11,12] is shown in Figure 2. A two-dimensional (2D) wavelet transform is first applied to a image data resulting in a multiresolution representation of the image. Then the wavelet coefficients are quantized using scalar quantization. Finally, run-length and Huffman coding are used to code the quantized data.

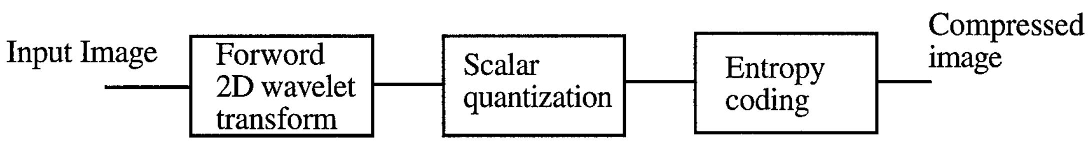

Figure 2. Wavelet compression scheme

Choice of wavelet filter bank is very important for high performance image compression. We evaluated the compression performance of a group wavelet filter banks, and determined the best filter bank for mammograms. A total of 26 filter banks were selected and the performance of compression ratio versus peak signal to noise ratio (PSNR) was compared. The optimal wavelet filter bank was selected according to the compression performance.

A wavelet transform separates an image into different frequency subbands. The previous study showed that four levels of wavelet decomposition yields the optimal compression results. The quantization step among each subband was selected in order to determine the optimal quantization strategy for a uniform quantization. An adaptive quantization according to the standard deviation of the each subband was compared with the optimal uniform quantization.

2.2. Results

2.2.1 On-line Compression

We first compared the times required to digitize and to compress an image in two situations: (1) where digitization and compression processes were done separately, (2) and where digitization and compression were combined. A mammogram was digitized at different sampling distances ranging from 50 μm to 200 μm with a 50 μm increment. The times required for each process were measured.

Table 1 shows the results. The second and third rows are the times required to digitize films and to compress images, respectively, when they are done separately. The fourth row shows the time required when digitization and compression are combined. Comparing the second and the fourth row, there is no extra time required to compress an image when the compression is done on-line.

Table 1. The scanning and compression time at different sampling distances

Sampling Distance (μm)	200	150	100	50
Scanning (sec)	15	15	28	103
Compression (sec)	7	10	22	102
Scanning with On-line Compression (sec)	15	15	28	103

We also compared the compression ratios for 46 randomly selected mammograms. The compression ratio is defined as full image size (digitized as a full size film) divided by the compressed image size. The compression ratios for 46 mammograms are ranged between 3.2 : 1 to 8.9 : 1 with an average ratio of 5.65 : 1 and a standard deviation of 1.46.

The distribution of compression ratios for 46 mammograms is shown in Figure 3. About 70% of the images have compression ratios between 4 : 1 and 7 : 1. About 20 % of the images have compression ratios larger than 7 : 1. Only 10% of the images have compression ratios between 3 : 1 and 4 : 1.

We compared the compression ratios achieved at the different sampling distances. Five mammograms were digitized with different sampling distances. The compression ratios were averaged at each sampling distance. Table 2 lists the results.

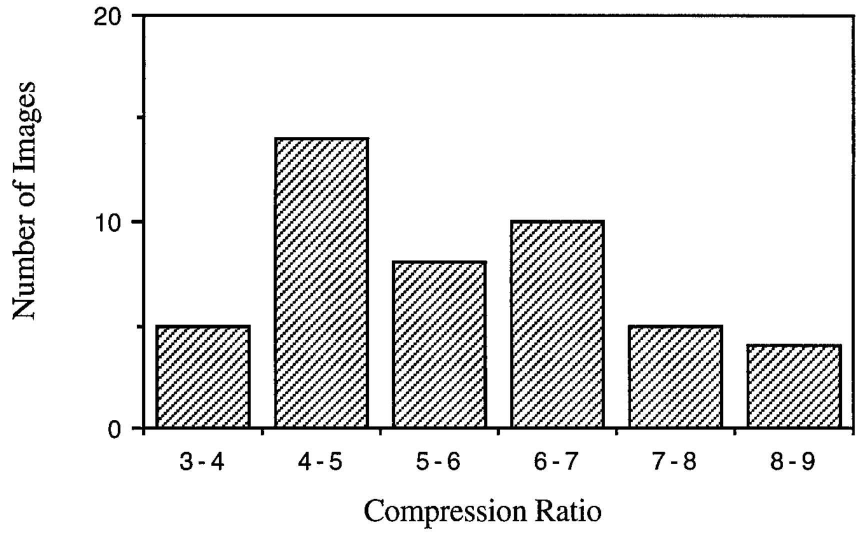

Figure 3. The distribution of the compression ratios

The compression ratio increases slightly as the resolution increases (sampling distance decreases). This is because as the sampling distance decreases, the correlation between pixels becomes higher, the predicted error becomes smaller, and the compression ratio is higher.

Table 2. The compression ratio at different sampling distances

Sampling Distance (μm)	200	150	100	50
Image Size (Mb)	2.2	3.84	8.6	34.6
Compression Ratio	7.12	7.24	7.39	7.83

2.2.2 Wavelet Compression

Among 26 filter banks evaluated, five of them gave good compression performance. We selected the shortest filter bank, 9/7 tap filter, among the five best filter banks and compared it with the popular Daubechies4 (D) and Haar filters. Figure 4 shows the results. At the same PSNR of 44 dB, the compression ratio using the 9/7 filter banks is about 40 % higher than that of the Haar filter bank.

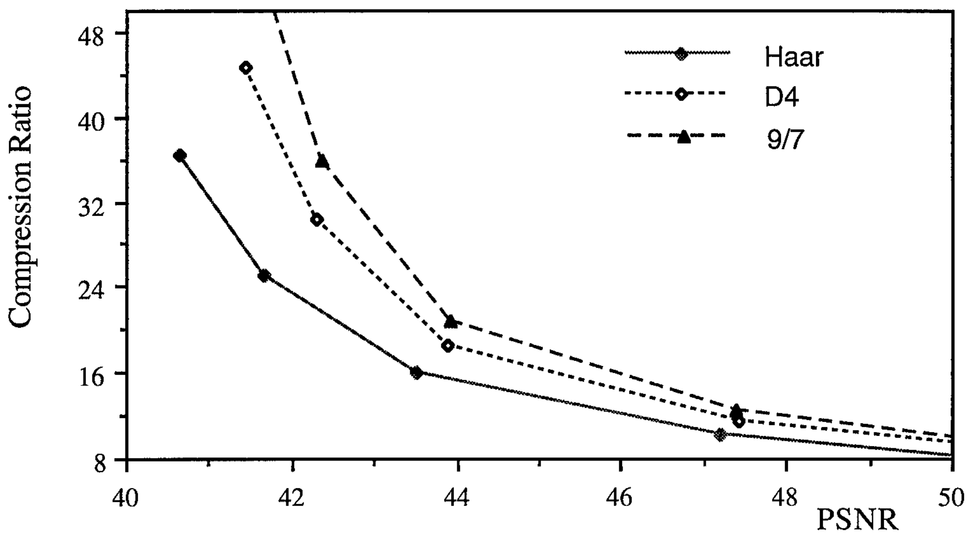

Figure 4. The performance of the different wavelet filter banks.

The optimal quantization parameters were determined experimentally. We varied the quantization steps between two adjacent levels at a constant ratio C. The best ratio was selected. The adaptive quantization where the quantization step size proportional to the standard deviation of each subband, were also compared with the constant ratio quantization. The results are shown in Figure 5. The optimal quantization is achieved when the quantization step size is the same among all subband, i.e. C=1.

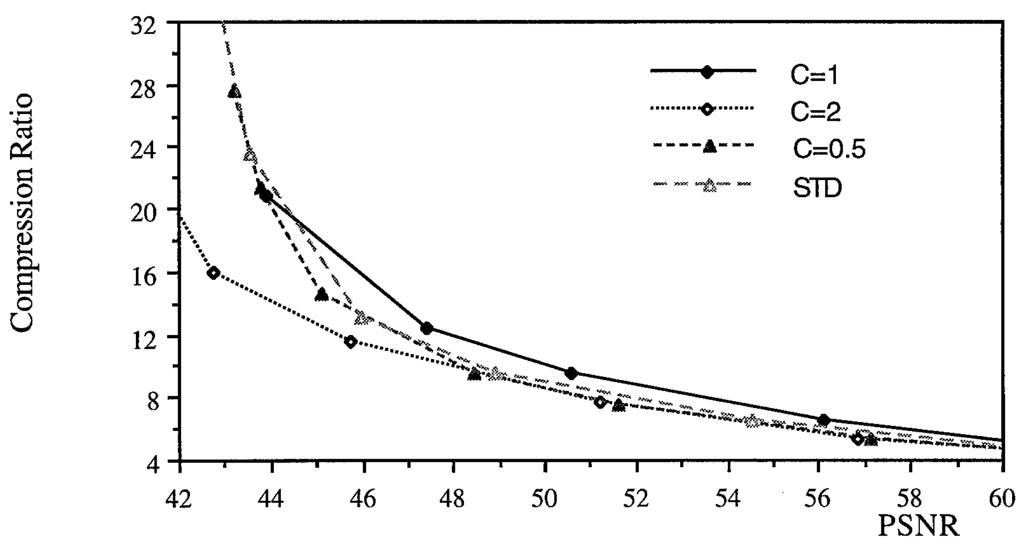

Figure 5 Comparsion of different quantization parameters

2.3 Discussions

The work of the on-line compression was not in the original proposal. However, we felt the work is important for the development of teleradiology and other applications. The digitization time is relatively long for a mammogram. The on-line compression combines the digitization and compression processes in parallel. The image is compressed where it is digitized and that cuts the digitization and the compression time up to 50%.

We completed the development of the mammogram compression using a wavelet transform stated in the original proposal. However, due to the early termination of the fellowship, the last part of evaluating compressed mammogram quality will not be completed.

3. Conclusions

We investigated both lossless and lossy compression methods for mammograms in this study. A structure-lossless compression algorithm segments a breast image from it background and only compresses the image portion. The combination of the segmentation and the prediction coding enables us to achieve high compression ratios without losing any useful information. This structure-lossless compression was also modified and implemented concurrently with a digitization process in an acquisition machine . The result is that a mammogram can be compressed at the same time when it is digitized. There is no extra time required for compression. This on-line compression could potentially be used for telemammography and other digital applications to improve overall digitization and compression performance.

A wavelet compression method was developed for mammograms. A 2D wavelet transform was first applied to a digital mammogram. A uniform quantization was applied to subband image. Run-length coding followed by Huffman coding was applied to quantized data. The wavelet filter bank and the quantization parameters were optimized for mammograms.

4. References

1. Anderson LF. Large-Scale Effort Tests Digital Mammography's Potential. Journal Of The National Cancer Institute. Vol. 86(8), pp.580-2, 1994.

2. Wu Y, Doi K, Giger ML, Nishikawa M. Computerized Detection of Clusters Microcalcification in digital Mammograms: Applications of artificial Neural networks. *Med. Phys,* Vol. 19(3), pp.555-560, 1992.
3. Karssemeijer N, Frieling JTM, Hendriks, JHCL, Spatial Resolution in Digital Mammography. *Investigative Radiology*, Vol. 28(5), pp. 413-419.
4. Nawano S, Evaluation of Digital Mammography in Diagnosis of Breast Cancer. *Journal of Digital Imaging,* Vol. 8 (1) Suppl 1, pp. 67-69, 1995.
5. Margos PA, Schafer RD and Mersereau RM, Two-Dimensional Linear Prediction and Its Application to Adaptive Predictive Coding of Images. *IEEE Trans. on Acoustics, Speech, and Signal Processing*. Vol. ASSP-32(6), pp.1213-1229, 1984.
6. Wintz PA, Transform Picture Coding. Proceedings of IEEE. Vol. 60(7). pp. 809-820, 1972.
7. Woods JW and O'Neil SD, Subband Coding of Images. *IEEE Trans. on Acoustics, Speech, and Signal Processing*. Vol.ASSP-34(5), pp.1278-1288, 1986.
8. Jun Wang and H.K. Huang, Digital Mammographic Image Compression, Annual Report: U.S. Army Breast Cancer Research.
9. Margos PA, Schafer RD and Mersereau RM, Two-Dimensional Linear Prediction and Its Application to Adaptive Predictive Coding of Images. *IEEE Trans. on Acoustics, Speech, and Signal Processing*. Vol. ASSP-32(6), pp.1213-1229, 1984.
10. Huffman DA, A Method for the Construction of Minimum-redundancy Codes. *Proceedings of the IRE,* Vol. 40(10) pp. 108-1101, 1952.
11. Wang J, and Huang HK, Three-Dimensional Medical Image Compression Using A Wavelet Transform With Parallel Computing, *SPIE Medical Imaging 95*, Vol. 2431, p162-172.
12. Wintz PA, Transform Picture Coding. Proceedings of IEEE. Vol. 60(7). pp. 809-820, 1972.

Appendix A: Bibliography

1. Jun Wang and H.K. Huang, Three-Dimensional Medical Image Compression Using A Wavelet Transform With Parallel Computing. *SPIE Medical Imaging 95*, Vol. 2431, p162-172.
2. Jun Wang and H.K. Huang, Real-Time Structure-Lossless Mammographic Image Compression. RSNA'95 (Abstract).
3. Jun Wang and H.K. Huang, On-line Structure-Lossless Mammographic Image Compression. Proceedings of SPIE, vol. 2707, pp.456-462, 1996.
4. Jun Wang and H.K. Huang, 3D Medical Image Compression by Using 3D Wavelet Transformation. IEEE Trans. on Medical Imaging (to be appear).

Appendix B: Personnel Receiving Pay

Ms Jun Wang received total of $19,000/year from July 15, 1994 to July 14, 1996, including $5,000 tuition and fees each year.

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

93014508 of 8

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO   SANTA BARBARA • SANTA CRUZ

LABORATORY FOR RADIOLOGICAL INFORMATICS
DEPARTMENT OF RADIOLOGY
SAN FRANCISCO, CA 94143-0628

TELEPHONE (415) 476-5941
FAX (415) 502-3217

June 24, 1996

Mr. Brain Martin
Department of Army
U.S. Army Medical Research Acquisition Activity
Fort Detrick, MD 21702-5014

Dear Mr. Brian Martin,

This is a request to terminate U.S. Army predoctoral fellowship training Grant No. DAMD17-94-J-4196 effective **July 15, 1996.**

The grant was awarded to the University of California, San Francisco for a three year period from July 15, 1994 to July 14, 1997. Since the principal investigator, Jun Wang, has completed her doctoral study on June 14, 1996, it is therefore requested to terminate the grant after July 15, 1996. The final month from June 15 to July 14 will be used to produce the final report.

If you have any questions, please do not hesitate to contact the undersigned.

Principle Investigator: Jun Wang

HK Huang

Sponsor: Dr. H.K. Huang

JOAN KAISER, Contracts and Grants Officer: