U C L A
MEDICAL IMAGING DIVISION
PACS/TELERADIOLOGY

Research and Development
Progress Report

February, 1995

University of California, Los Angeles
Department of Radiological Sciences
Los Angeles, CA 90095-1721

TABLE OF CONTENTS

Foreword

In this booklet, you will find reports of our most recent research work in PACS related areas, including image archive, network manager, image compression, image navigation, medical database modeling, virtual view reality, and fault tolerant technique. In many ways, we have already left behind the original intent of PACS implementation, which is simply the replacement of film. By emphasizing cost-effectiveness, fault tolerance, automation, information indexing, innovative user interface and quantitation, our research work aims at making radiologic practice more economical, reliable, efficient, and accurate.

There are two notable new interests in the areas of teleradiology and network supercomputing, both direct extensions of a departmental-wide PACS. Our first teleradiology imaging center in Melbourne Florida has just began clinical operation this month. We are optimistic that this effort will expand to a global scale in the near future. This project aims at providing subspecialty radiologic expertise to remote areas while containing the rising cost of health care for the served population.

Biomedical supercomputing promises to deliver image processing to routine clinical practice. Our PACS infrastructure provides an excellent testbed for this concept. We are looking forward to many exciting years of activity in this new direction.

Medical Imaging Division
Picture Archiving and Communication System/Teleradiology

Department of Radiological Sciences, Chairman
Professor Richard J. Steckel, M.D.

FACULTY

Matti Al-Aish, Ph.D.
Professor

Jeffry R. Alger, Ph.D.
Associate Professor

Moses A. Greenfield, Ph.D.
Professor Emeritus
BioMed Physics Grad Program

Bruce K.T. Ho, Ph.D.
Assistant Professor and
Director, SIL/Global Teleradiology

Carolyn Kimme-Smith, Ph.D.
Associate Professor

Nicholas J. Mankovich, Ph.D.
Associate Professor

Michael F. McNitt-Gray, Ph.D.
Assistant Professor

James W. Sayre, Ph.D.
Associate Professor in Biostatistics

Shantanu Sinha, Ph.D.
Assistant Professor

Ricky K. Taira, Ph.D.
Assistant Professor

M. Albert Thomas, Ph.D.
Assistant Professor

Daniel J. Valentino, Ph.D.
Adjunct Assistant Professor and
Manager, PACS/Teleradiology

STAFF

Hing M. Chan
Programmer Analyst

Woodrew Chao
Software Engineer

Lu J. Huang
Technical Mgr, Operations

Sandra L. Johnson
Clinical Manager, Operations

Kinchi Kong
PACS Network Administrator

Kelvin T. Leung
Technical Engineer

Y.C. Lin, Technical Mgr,
Software Engineering

W. Keith Wingate
PACS Systems Administrator

Gwen Worthy-Smith
Administrative Coordinator

Richard Yaker
Programmer Analyst

George Zachariah
Software Engineer

POST-DOCTORAL SCHOLARS

Zheng Chen, Ph.D.

George J. Hademenos, Ph.D.

GRADUATE STUDENTS

Kevin Chang
Michael Harreld
Brent J. Liu
Pongskorn (Ko) Saipetch

Doris T. Chen
Zhimei Jiang
Marco Ma
Reza Sadri
Jun Wei

Amir Zarkesh
Calvin Leung
Ramesh Panwar
Serge Van Kriekinge

CLINICAL COLLABORATORS

Hooshang Kangarloo, M.D.
Professor

Zoran Barbaric, M.D.
Professor

Denise Aberle, M.D.
Associate Professor and Chief,
Thoracic Imaging

John Bentson, M.D.
Professor and Chief, Neuroradiology

M. Ines Boechat, M.D.
Associate Professor

Katherine Brown, M.D.
Associate Professor

Poonam Batra, M.D.
Associate Professor

Theodore Hall, M.D.
Assistant Professor

Robert Lufkin, M.D.
Associate Professor

Lossless image compression using wavelet-like filter banks and arithmetic coding

Pongskorn Saipetch, Ramesh K. Panwar*, Bruce K. T. Ho, Marco Ma

University of California at Los Angeles, Department of Radiological Sciences, Los Angeles, CA 90024
*Dapartment of Computer Sciences

ABSTRACT

We have developed a fast, reversible compression algorithm based on wavelet-like filter banks and Arithmetic Coding (AC) for primary diagnostic teleradiology using inexpensive, off-the-shelf hardwares. This new method offers routine 2:1 to 4:1 compression ratios, while requires only a few seconds processing time on a microcomputer. This compression performance will make affordable primary diagnostic teleradiology over standard phone lines or ISDN possible.

Keywords: image compression, lossless compression, filter bank, arithmetic coding, teleradiology, microcomputer.

1. INTRODUCTION

Applications in primary-diagnosis teleradiology require that the original radiographic data are transmitted over a baud-rate limited communication line without any degradation in quality. Affordability and practicality of viewing stations also requires that a high-speed reconstruction of the compressed images can be implemented using inexpensive, widely available hardwares. We have developed a fast reversible compression method that give typically between 2:1 to 4:1 compression ratios using standard, off-the-shelf microcomputer hardwares.

We use the fact that arithmetic coding (AC) is a very efficient compressor and can approach the entropy bound defined by any arbitrary model in compressing a sequence of symbols [1][2]. When a particular model predicts that most of the message will be made of a few of the symbols (narrow histogram), the entropy bound will decrease. AC can approach this entropy bound better than any other compression methods such as Huffman or Lempel-Ziv encoding. More sophisticated model can give a narrower histogram at the expense of computational time and memory. We found that our computing resource of the viewing workstations is limited to an order-1 model, predicting only the probabilities of symbols without using any conditional probability. Therefore, we need to construct a reversible transform that can transform our image into another version whose histogram is extremely narrow. We construct the transform by factoring the transfer function of a maximally decimated filter bank while imposing the constraints that the filter coefficients must be rational (to get perfect reconstruction using a few bit of precision) and that the analysis filters must have 2nd or 3rd zero moment (to approximate the low-order polynomials well enough to make an extremely narrow histogram).

2. WAVELET-LIKE FILTER BANKS

To compress an image, it is necessary to have a good model to predict the occurrence of pixel graylevels. Sophisticated models usually have more predictive power, resulting in a better compression performance than simpler models. However, sophisticated models require more processing power to calculate. Therefore, there is a definite trade-off between compression performance and computational resource requirement. For the present, affordability of radiological viewing stations limits our available computing resources and forces us to use an order-1 model, predicting only probabilities of graylevels without any conditional probability. To maximize compression performance from an order-1 model, we need to construct a reversible transform that change the image's representation into an alternative form that has large probabilities for a few transformed data points (narrow histogram.)

From our previous work on radiological image compression [3], we know that wavelet transform can change the presentation of data so that the resulting histogram is very narrow. However, the algorithm must be modified to allow lossless compression. The resulting transform is similar to our previous wavelet transform-based compression, but we require the filter coefficients to be rational for exact arithmetic. Furthermore, the quantization step is eliminated since we do not want to lose any information. The main idea of the transform is to subdivide arbitrary signal into constant-Q frequency bands using a

recursive filter banks generated from a small number of prototype filters. The filtering process is equivalent to decomposing the signal using a set of basis functions which are localized in both space and frequency and are scaled and shifted versions of a prototypical function. The description of wavelet transform ideas can be found in [4][5][6][7][8]. We choose to implement our transform using a filter bank in a wavelet-transform-like fashion because we can easily adapt the storage format developed in our wavelet transform research to perform progressive transmission over slow communication lines.

Fig. 1 shows two-level analysis and synthesis of one-dimensional data using a maximally decimated filter bank. H_0 and H_1 are low-pass and high-pass filters for analysis. F_0 and F_1 are low-pass and high-pass filters for synthesis. The four filters are related in z-domain as followed:

$$F_0(z) = H_1(-z)$$
$$F_1(z) = -H_0(-z)$$
$$H_0(z)H_1(-z)-H_1(z)H_0(-z) = 2$$

The first two conditions eliminate aliasing; the last one eliminates phase and amplitude distortion in the reconstructed signal [9]. We design the filters such that all the coefficients of all the filters are rational number with small denominators using continued fraction and Diophantine techniques [8], so that all the operations are done with 32 and 64 bits integers. Furthermore, the analysis filters, H_1 and H_0, are designed to have 2 or 3 consecutive moments zero to well approximate low degree polynomials which make up most images [4]. The transform is a linear phase to eliminate the need for phase compensation in cascade pyramidal filter structure.

We transform the two-dimensional images using one-dimensional row and column transforms. The process is similar to row and column FFT of images. The edges of the image are handled by performing the convolutions on symmetric extensions of the images to avoid boundary artifacts introduced by simple periodization. The types of extensions and the associated parameters needed to perform the symmetric wavelet transform are described in details by Brislawn in [10]. In our experiments, the images are transformed to 3 or 4 levels depending on their sizes. For 2048x2048 images, we choose to perform 4 levels of transform so that the low resolution parts of the images are 128x128, a small enough size for thumb-nail picture indexing. For smaller images, a 3-level transform is sufficient.

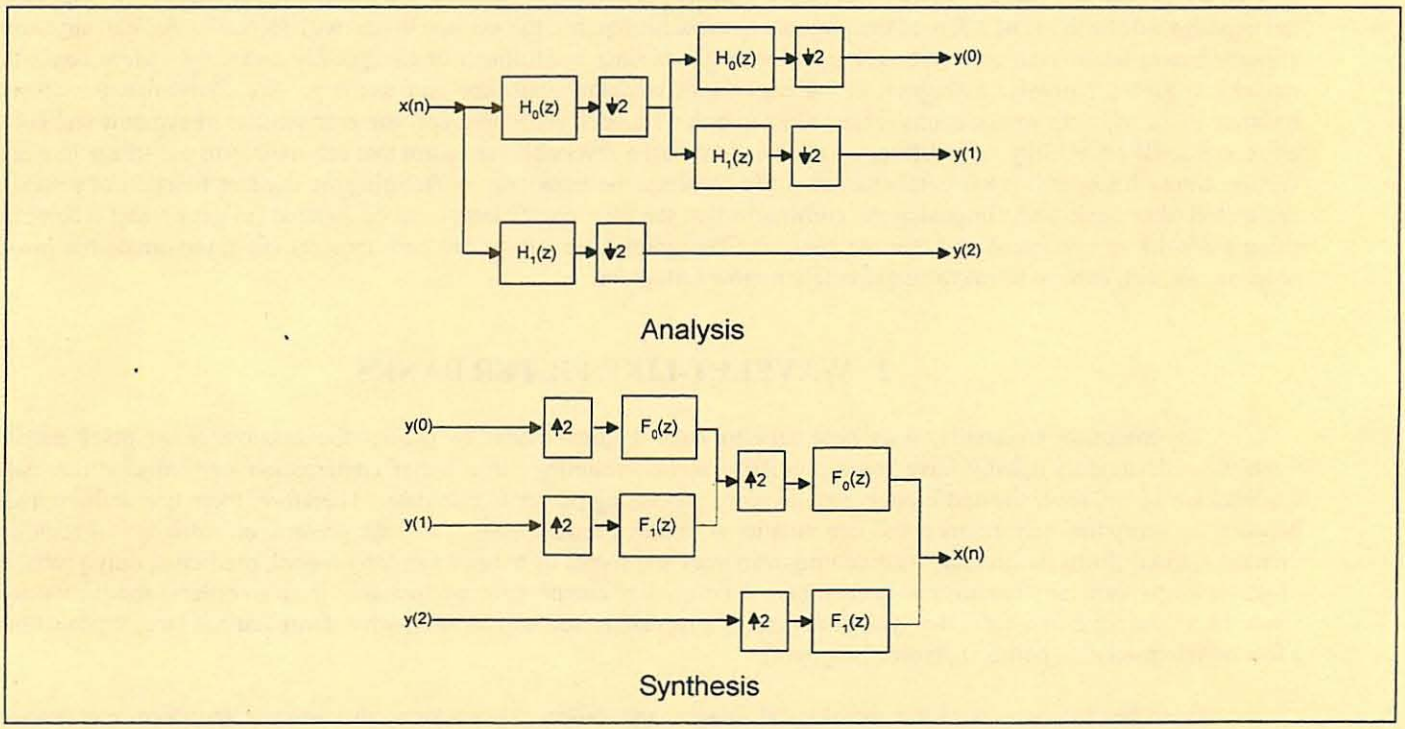


Fig. 1 Two stages of analysis and synthesis of a wavelet transform of one-dimensional data.

3. ARITHMETIC CODING

After transforming the image using our filter bank. The resulting coefficients is compressed by arithmetic coding. To the arithmetic coder, the coefficients look just like a stream of symbols with various probabilities. The function of the coder is to represent the symbol stream using as few bits as possible.

Arithmetic coding encodes the whole stream of symbols with a real interval between 0 and 1. Before the arrival of the stream, the initial interval is [0,1). For every symbol the coder receives, the interval narrows down by a factor equal to the probability of the symbol. As the stream become longer, the interval needed to encode it becomes smaller. The more likely symbols reduce the range by less than the less likely symbols. The number of bits per symbol after the complete encoding can approach the non-integral entropy bound. Arithmetic coding is chosen over Huffman's coding because its compression performance is at least as good and sometimes, when one of the symbols' probability is close to 1, far exceeds it because of Huffman's requirement that the number of bits per symbols must be integral. As an extreme example, a message consisting of 999,999 zeroes follow by a 1 is represented by the interval:

$$[(0.999999)^{999999}(0.999999),(0.999999)^{999999}) \approx [e^{-1},\frac{e^{-1}}{0.999999})$$

while a message with one 1 followed by 999,999 zeroes is represented by the interval:

$$[0.999999,0.999999+\frac{e^{-1}}{1,000,000})$$

Both messages require only 22 bits to represent. On the other hand, typical unblocked Huffman encoding needs 1,000,000 bits to represent the same messages. Analysis of the entropy of the messages shows that we need only 21.37 bits to represent them.

Arithmetic coding's independence from the modeling of data allow us to use the same coder for more sophisticated models when the computing resource become available.

The transform coefficients of the image are grouped according to their resolutions and orientations to skew the histogram of the coefficients. The more concentrate the histogram, the better the compression. Each group is compressed individually using arithmetic coding as described in [1] and [2]. Comparison of actual compression and entropy bound for high-frequency coefficients is shown in (Fig. 2.) Note that our arithmetic coding implementation approaches the entropy bound to within 1%

4. EXPERIMENTS

A set of radiological images consisting of five CR, two chest CTs, three abdominal CTs, one head CT, three abdominal MRI, three head MRI, and ten mammograms is used in our experiment. The radiographs are 512x512x8-bit, 1024x1024x10-bit, and 2048x2048x10-bit. The CT images are 512x512x12-bit. The MRI images are 256x256x12-bit and 512x512x12bit. The mammograms are 1024x1024x10bit and 2048x2048x10bit These images are transformed into either 3 or 4 levels. The resulting coefficients are encoded with an arithmetic coder. Typical images are shown in Fig. 3.

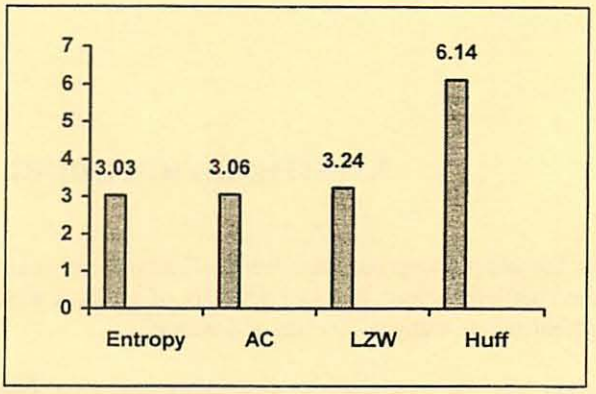


Fig 2. Comparison of compression performance between AC, LZW, and Huffman. Entropy is the theoretical lower bound of an order-1 model. AC is our implementation of arithmetic coding. LZW is done using Unix *compress* program. Huffman is an adaptive Huffman program from [2]. Data from a 512x512x12bit MRI image.

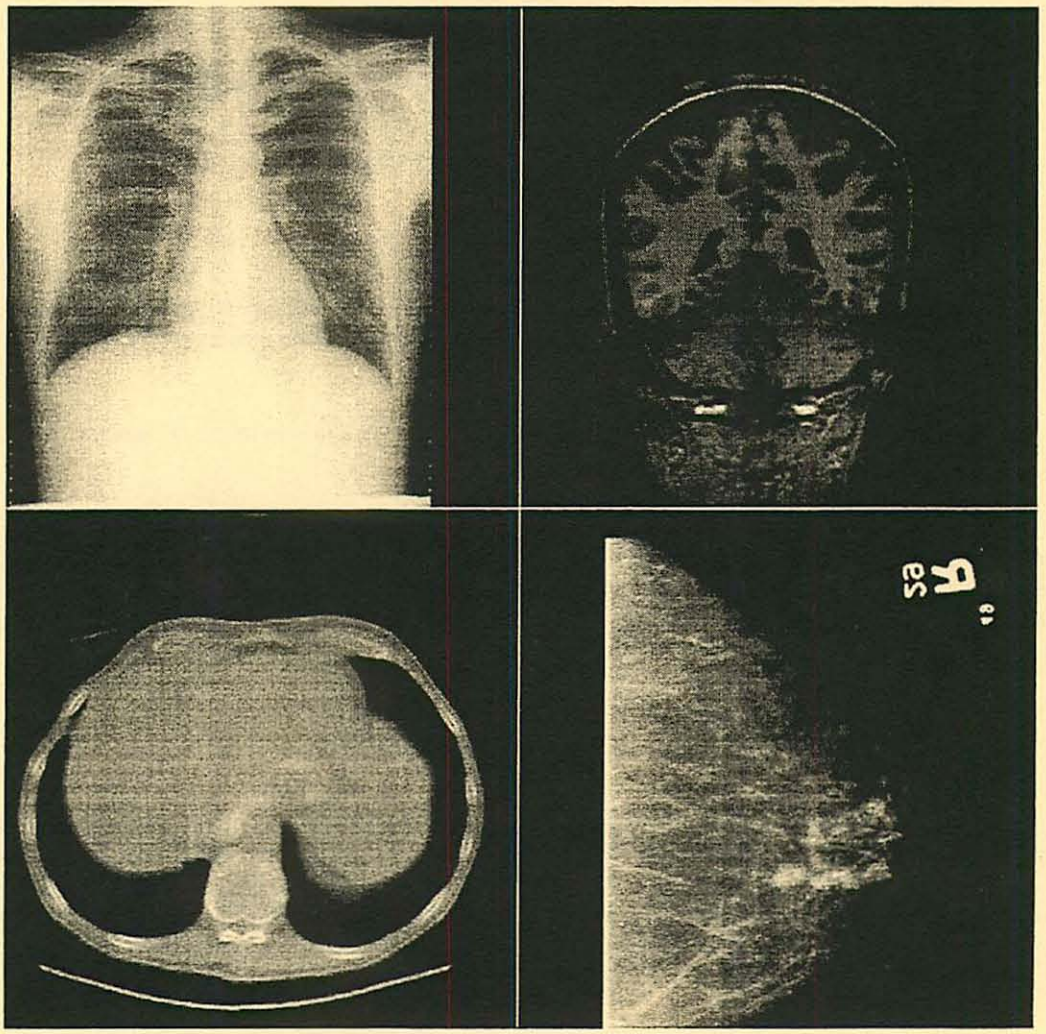


Fig. 3. Typical images in our experiment. Clockwise from upper left corner: Chest CR, Head MRI, Mammogram, and Abdominal CT.

5. RESULTS

The result of compression experiment is shown in Table 1, Fig. 4 and Fig. 5. The lowest compression ratio of 2.2 is obtained from a 512x512x8bit chest CR. This is to be expected since the low resolution image of the chest contains more high frequency information than the high resolution image; the distance between adjacent pixels in the low resolution image is larger than that of the high resolution images implying larger changes in gray level between pixels. The highest compression ratio of 4.8 is obtained from a breast biopsy at 1024x1024x10bit. This is because the image zooms in a small physical area with relatively large constant gray level regions.

The data shows that compression ratio increases when the original dimension of the image increase. This is explained by the low-resolution/high frequency-content argument above. The data also shows that large constant graylevel areas around the anatomy in CT, MRI, and mammograms are redundant and contribute to high compression ratio (Fig. 4 and Fig. 5.)

Image	Type	Description	Bits/Pixel	Size(Pixels)	Compressed Bits/Pixel	Compression Ratio
CR1	CR	Chest	8	512	3.6	2.2
CR2	CR	Chest	10	2048	2.9	3.5
CR3	CR	Chest	10	2048	3.0	3.3
CR4	CR	Extremity	10	1024	3.3	3.0
CR5	CR	Extremity	10	512	4.3	2.3
CT1	CT	Chest	12	512	3.8	3.2
CT2	CT	Chest	12	512	4.0	3.0
CT3	CT	Abdomen	12	512	3.9	3.1
CT4	CT	Abdomen	12	512	3.5	3.4
CT5	CT	Abdomen	12	512	3.5	3.4
CT6	CT	Head	12	512	3.6	3.3
MR1	MRI	Abdomen	12	512	3.1	3.9
MR2	MRI	Abdomen	12	512	3.5	3.4
MR3	MRI	Abdomen	12	256	4.3	2.8
MR4	MRI	Head	12	256	3.6	3.3
MR5	MRI	Head	12	256	3.9	3.1
MR6	MRI	Head	12	256	4.3	2.8
MAM1	Mammogram	Screen	10	2048	2.8	3.6
MAM2	Mammogram	Screen	10	2048	2.7	3.7
MAM3	Mammogram	Screen	10	2048	2.6	3.9
MAM4	Mammogram	Screen	10	2048	2.9	3.4
MAM5	Mammogram	Screen	10	2048	2.6	3.9
MAM6	Mammogram	Screen	10	1024	3.4	2.9
MAM7	Mammogram	Screen	10	1024	3.3	3.0
MAM8	Mammogram	Biopsy	10	1024	2.1	4.8
MAM9	Mammogram	Biopsy	10	1024	2.4	4.2
MAM10	Mammogram	Biopsy	10	1024	2.3	4.4

Table 1. Results of compression of test images.

The images used in our experiment are all square because we acquire them that way, not because of the limitation of our transform. Using symmetric extension of the data, our transform can also be performed on images of arbitrary sizes as shown by Brislawn in [10]. This property is especially important in radiographs of extremities, where cropping is frequently done. The compression time is 8 seconds for 1024x1024x2byte image on Pentium-90. The time is linear with the number of pixels.

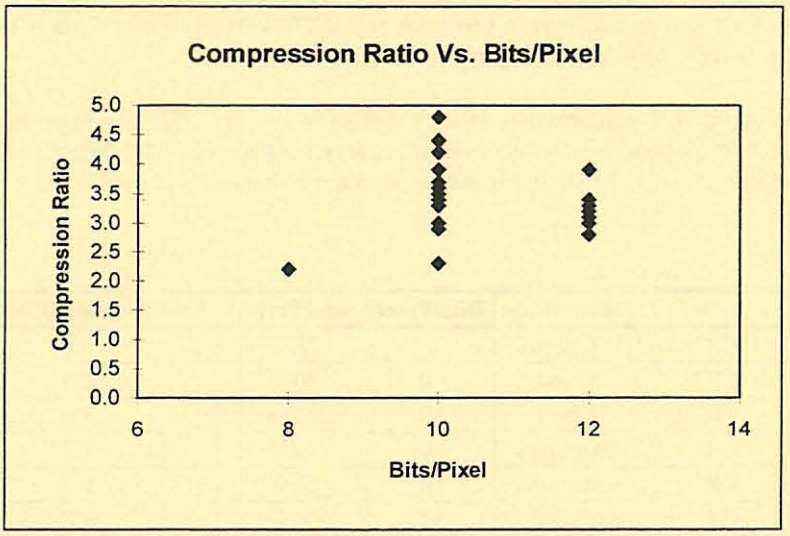


Fig. 4 Compression Ratio Vs. Bits/Pixel in the images.
The top three data points at 10 bit/pixel are biopsy mammograms which contain large, slowly varying regions

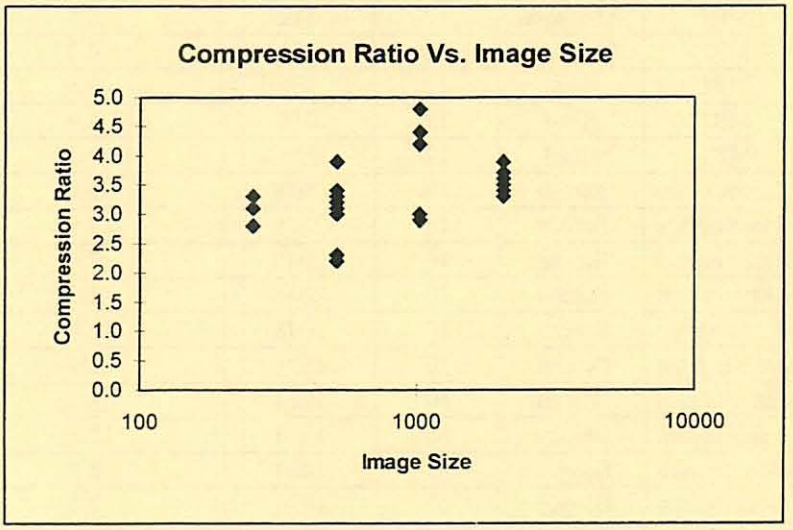


Fig. 5 Compression Ratio Vs. Image Size.
The top three data points at 1024 are biopsy mammograms which contain large, slowly varying regions

6. REFERENCES

[1] I. H. Witten, R. M. Neal, and J. G. Cleary: Arithmetic coding for data compression. *Comm. of the ACM* 30(6): 520-540, June 1987.

[2] M. Nelson: *The Data Compression Book,* M&T Books, 1992.

[3] P. Saipetch, B. K.T. Ho, M. Ma, K.S. Chuang, J. Wei: Radiological image compression using wavelet transform with arithmetic coding. *Proc. SPIE / Med. Imaging, Image Capture, Formatting, and Display.* Vol. 2164, Newport Beach CA, Feb 1994.

[4] M. Antonini *et al*: Image coding using wavelet transform. *IEEE Trans. Image Proc.* 1(2): 205-220, April 1992.

[5] I. Daubechies: Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math* 41: 909-996, 1988.

[6] A. Cohen, I. Daubechies, and J. C. Feauveau: Biorthogonal bases of compactly supported wavelets: AT&T Bell Lab., Tech. Rep., TM 11217-900529-07, 1990.

[7] S. Mallat: A theory for multiresolution signal decomposition: The wavelet representation: *IEEE Trans. Patt. Anal. Mach. Intel.* 11:674-693, July 1989.

[8] M. Vetterli: Wavelets and filter banks: Theory and design: *IEEE Trans. Signal Proc.* 40(9): 2207-2232, Sept. 1992.

[9] P. P. Vaidyanathan: *Multirate Systems and Filter Bank*, P T R Prentice-Hall, 1993.

[10] C. M. Brislawn: Classification of symmetric wavelet transform. Group C-3, Computer Research, Los Alamos National Laboratory, March 1993.

Medical Image Supercomputing in a PACS Infrastructure

Bruce K.T. Ho, Zheng Chen, Ramesh Panwar*, Reza Sadri*,
Pongskorn Saipetch, Marco Ma*

AR-264 Center for Health Sciences
Department of Radiological Sciences,
*Department Computer Science
University of California, Los Angeles

ABSTRACT

We have done a feasibility study for real-time image processing capabilities in the PACS environment with the intent to facilitate development of digital based clinical protocols. Four aspects of a central compute server are investigated to address the computational need of clinical viewing: 1) fast computation, 2) efficient memory caching, 3) direct data access to the central file server, and 4) fast output to the local workstation. Potential clinical applications include computational and data intensive procedures such as image segmentation, image registration, vascular flow study, functional image study, etc.

INTRODUCTION

Up till now, PACS is intended for acquisition, archiving, communication and display functions of radiological images [1]. These goals are consistent with the traditional roles of the analog film. As we enter an advanced era of PACS research, it is time to leap beyond the traditional clinical functions and explore the untapped potentials of digital tools, that being the power of image processing. The opportunities for computer processing abound in both diagnostic and therapeutic imaging: accurate estimate of blood flow rate and pressure during an interventional procedure [2], three-dimensional registration of CT and MR images for radiation therapy planning [3], PET and MR registration for functional imaging studies [4], semi-automatic segmentation [5] and volume measurement during diagnostic readings, three-dimensional visualization of anatomy for surgical planning [6], etc.

The PACS infrastructure provides a large archive of clinical data, continuous updates of new examinations, distribution of workstations around the hospital, and basic level of network communications. This infrastructure can be integrated with a central compute server to provide real-time image processing for clinical use. We believe that immediate turn-around of processed data is one of the most important criterion for acceptance of advanced image processing

algorithms as a part of a routine clinical protocol.

CENTRAL COMPUTE SERVER USING PARALLEL PROCESSING

Providing lots of computing power at individual PACS workstations would be prohibitively expensive and extremely wasteful. In addition to the high investment needed for the high performance processing hardware, the cost of the network and input/output (I/O) components needed to handle the huge data volume also escalates. An idle workstation can not be effective utilized by members elsewhere in the hospital. A powerful centralized compute server which can be shared by many users spread out in the network is much more suitable in this case. Dynamic resource allocation allows maximum throughput for a given number of users at any one time.

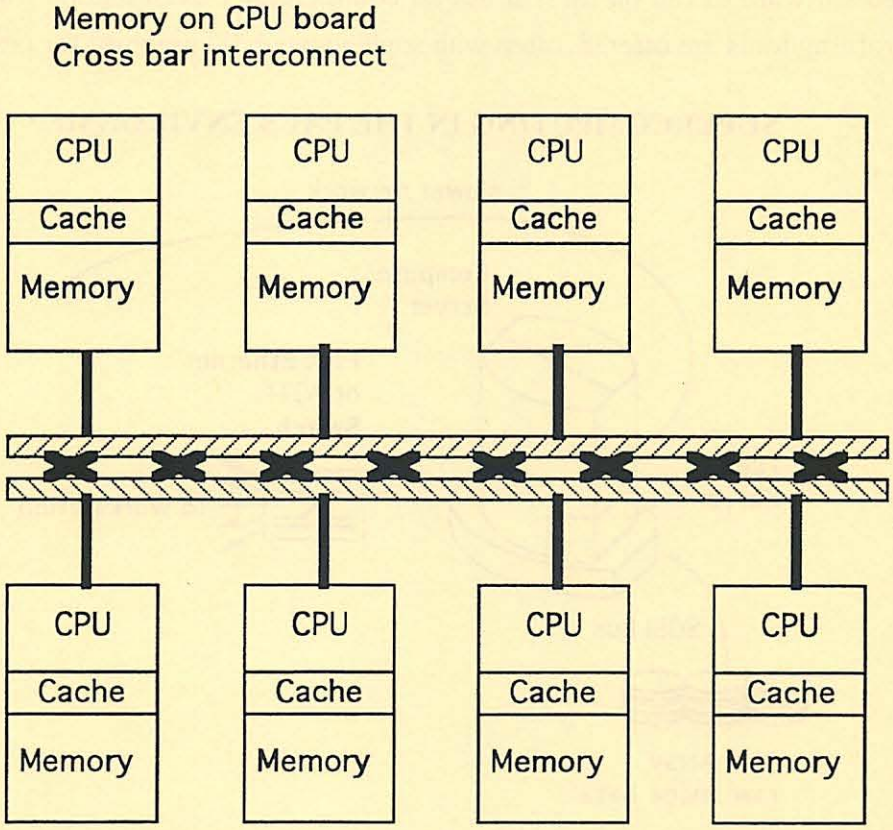


Figure 1. Parallel processing architecture with cross bars for interprocessor communication.

Real-time processing speed, once a domain of supercomputers in national laboratories, can now be achieved using commonly available parallel processor workstations. Reduced Instruction Set Computing (RISC) processors are achieving ever-higher processing speed and memory bus bandwidth, using optimized instruction pipelines, redundant integer and floating point units, scaleable internal register sets and data/instruction caching. In addition to improvements in processor design, memory access technology is another key factor which made

low cost supercomputing possible. Innovative memory bus architecture such as cross bar, gigaswitch, and wide width cacheless system memory significantly reduced the interprocessor communication overhead and memory contention.

The degree to which a software can be parallelized depends the nature of the algorithm. Fortunate for the radiology community, most image processing algorithms can be parallelized by simply partitioning the input image volume, and distribute each data block among available processors. Although a parallelizing compiler that automatically parallelizes single-processor programs is in general inefficient or unavailable, many software tools address needs for the portability, management and performance evaluation of parallel programming. For example, Parallel Virtual Machine (PVM) is a defacto programming interface standard which allows a parallelized software to run on most if not all commercially available hardware. Many vender specific profiling tools are offered, often with sophisticated 3D graphics for presentation.

SUPERCOMPUTING IN THE PACS ENVIRONMENT

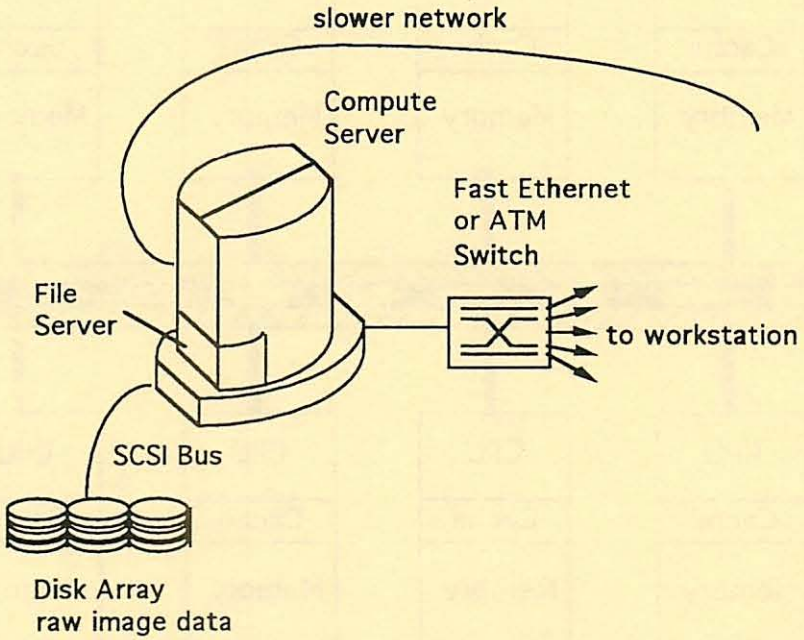


Figure 2. One node in the compute server can be dedicated as a fileserver. routine PACS image file requests are serviced by this node and a slower network. Fast network link connects the compute server to individual workstations for real-time applications only.

The main challenge in integrating supercomputing capability into a PACS environment is getting large volume of image data into the compute server and the processed result out to the workstations where the viewer is waiting. In PACS architectures with a centralized file server, the compute server can dedicate part of its I/O capabilities towards file service. The same

magnetic disk storage provides raw data for both routine clinical reading and real computation needs. The commonly availble Fast and Wide SCSI (Small Computer System Interface) interface can support 20 MB/sec. of throughput. Multiple SCSI interfaces can be coupled together for higher speed along with disk striping. Faster speed can also be achieved by high end solutions like Fiber Channel. When an user wishes to activate certain image processing on the set of image data he is viewing, there is no need to send the entire set of data to the compute server over the network, since the data should (if recently acquired) already resides on the fileserver's disks. Only the command packet of a few bytes (command name, file name(s), parameters, user identifiers, etc.) need to be send to the compute server. In PACS with distributed fileserver design, high speed network must be installed to connect the compute server to the other fileserver magnetic disks. This may incur some cost but is undoubtedly more economical to implement than installing a comparable network to every single workstation.

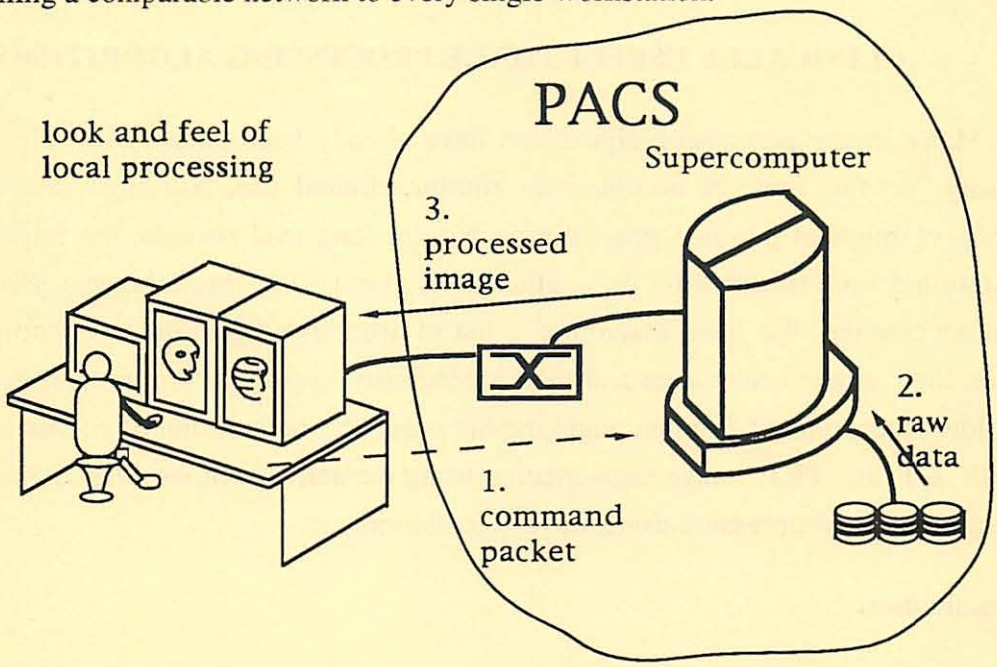


Figure 3. To request an image processing function, only a command packet is send from the workstation to the compute server. Raw data is retrieved from disks local to the computer server. Processed image is transferred back to the workstation through high speed network.

The output of the image processing is generally a much smaller set of data than the input (e.g. a cutting plane or surface rendered picture from a 3D image set, flow quantitation from an angiographic sequence, coordinates from image registration of multi-modality image volumes, etc.). There should be a reasonably fast network to carry the output data to the originating workstation. If 1K x 1K images are produced as the output, a 100 Mbit/sec. network can transfer back images at a rate of more than 12 frames/sec. Fast Ethernet, which is expected to become widely and economically implemented soon, is a strong candidate which can be very

economically deployed throughout a hospital. ATM (Asynchronous Transfer Mode) is another network option, but may be less attractive for cost-sensitive PACS applications. The workstation display program and hardware must be optimized for pixel transfer from the network to the monitor, since that is the more likely place for a bottleneck than the network itself.

By directly connecting the compute server to the PACS fileserver, and providing adequate speed network for sending back the output data, coupled with the compute server's fast number crunching, the user will indeed be able to request an image processing function on an image set and be looking at the result instantly on the same screen. The speed and appropriate user interface design will make the processing to appear local. For users already well versed in softcopy viewing, adding image processing to his/her repertoire only involves learning a few more buttons.

CLINICALLY USEFUL IMAGE PROCESSING ALGORITHMS

Many image processing algorithms have already been successively developed in the laboratory. So far, few are accepted for routine clinical use. Although innovations in the conservative medical practice generally do require long trial periods, the impractically long computational time incurred by these algorithms may be the single biggest handicap in their clinical acceptance. We have assembled a list of attractive algorithms as candidates for trial based on their clinical relevance and their amenability to parallel programming. They include blood flow measurement from an angiographic sequence, registration of volumetric image sets from MR, CT, and PET, image segmentation using the active contour method, surface shading, volumetric image compression using wavelet transform, etc.

3D visualization

A wealth of algorithms and software are available in this category. Three types of visualization software are of immediate interest: slicing plane, surface shading, and MIPS.

Slicing plane using Affine transform

The ability to view a CT or MR volumetric image set at an arbitrary angle allows radiologists and surgeons a more close up view of the anatomy of interest. This is especially true for helical CT data, where multi-planar reconstruction in non-axial direction produces high quality images without interpolation artifacts. As an exemplary application, during interventional procedures, investigators are developing a techniques for catheter localization by viewing the plane containing the needle. Other uses including tracing of vessel continuation, lesions

extension, etc.

Arbitrary angle slicing plane can be generated using the Affine transform [7], which combines the rotation matrix of rotations around x, y, and z axes of the initial volume. Parallelization is possible by separating the matrix operator. Curved surfaces can also be generated with higher order polynomials [8]. It is important to have an intuitive user interface when surface geometry gets more complicated.

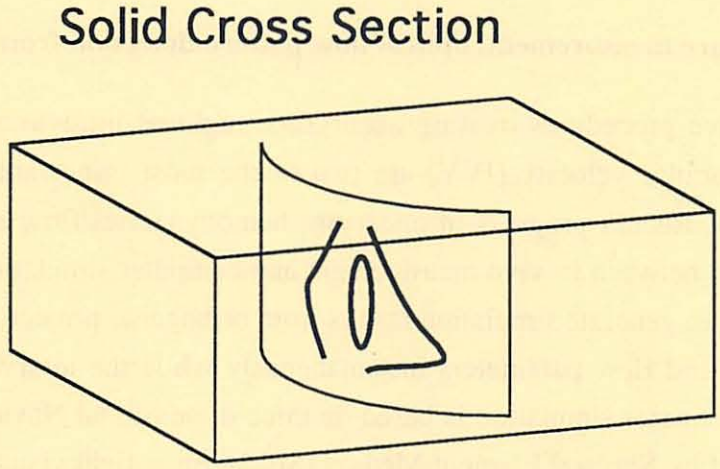


Figure 4. Slicing plane of arbitrary curvature can be generated on the fly based on user input.

Surface shading

Surface shading in combination with the active contour segmentation described later allows viewers a 3D perspective around any anatomical region (tumor, thermal oblation region, hemisphere, etc.) which has a intensity differentiable object boundary. Surface shading involves projecting a light source upon the predefined surfaces each of which takes one of six possible orientations (up, down, and 4 sides) [6]. The incident light is modeled to bounce off these surfaces based on the surface, viewer and incident light directions. In Phong shading, the angle of reflectance is calculated using interpolation of adjoining vertex normals [9]. Parallelization of the shading algorithm is simply done by partitioning the vertices.

MIPS

One of the innovative application of the helical CT is to generate 3D angiogram using before, during, and after contrast fine-slice studies. By aligning the volumes acquired in different stages of contrast perfusion, subtract away the background, only the vessel regions differentiated by contrast movement remain. The fine resolution of the helical scan provides angiograms comparable in quality to that of conventional angiography but offers the benefits of 3D viewing,

scatter and background removal, and true 3D data for analysis not possible on 2D angiograms. The computational steps involved are selection of landmarks, registration based on these landmarks, voxel subtraction, and a visualization technique such as regional slicing thin slab-maximum intensity projection (STS-MIP) [10] or surface rendering. The whole process is best done interactively, so that the user can fine tune the viewing options. A new projection view in response to user input must be generated instantaneously for a clinician to use this feature as a dynamic clinical tool. Supercomputing speed is therefore a necessity.

Flow and pressure measurement, optical flow pattern detection from angiograms

In invasive procedures treating aneurysms, regional intravascular pressure (IVP) and regional intravascular velocity (IVV) are two of the most important parameters needed for decision support. Recent progress in aneurysm hemodynamics/flow mechanics have achieve strong agreement between in-vivo measurement and computer simulation modeling [11]. In the near future, we can generate simulation results from orthogonal projection (OP) angiograms and extract pressure and flow parameters instantaneously while the interventional procedure is in progress. The computer simulation is based on three-dimensional Navier-Stokes and continuity equations, solved by Spectral Element Method (SEM). Flow field visualization is accomplished by constructing streamline, particle tracing, and IVP/IVV isocontour maps.

There is also a technique for instantaneous blood flow measurement using digital videodensitometry (also known as optical flow) without bolus tracking [2]. The governing equations uses fluid equations of continuity and incompressibility. This technique can be used with low frame rates where matching of distance-density curves fail. The technique involves applying the following equation to consecutive images:

$$\int_{\eta_1}^{\eta_2} d\eta\big(P(\eta,t_2)-P(\eta,t_1)\big)+\Gamma\int_{t_1}^{t_2} dt\left(\frac{P(\eta_2,t)}{A(\eta_2)}-\frac{P(\eta_1,t)}{A(\eta_1)}\right)=0$$

where Γ is the flow rate, P is the radially integrated contrast density, and η parametrizes the vessel length. The first term consists of integration along the vessel at two different times, whereas the second term consists of integration over time for measurements at two segment endpoints.

This application requires fiber connection to transfer the captured angiograms to the central compute server at the rate angiogram frames are generated. Because this application aims at decision support with a patient on the procedure table, it is the premier example of real-time clinical supercomputing.

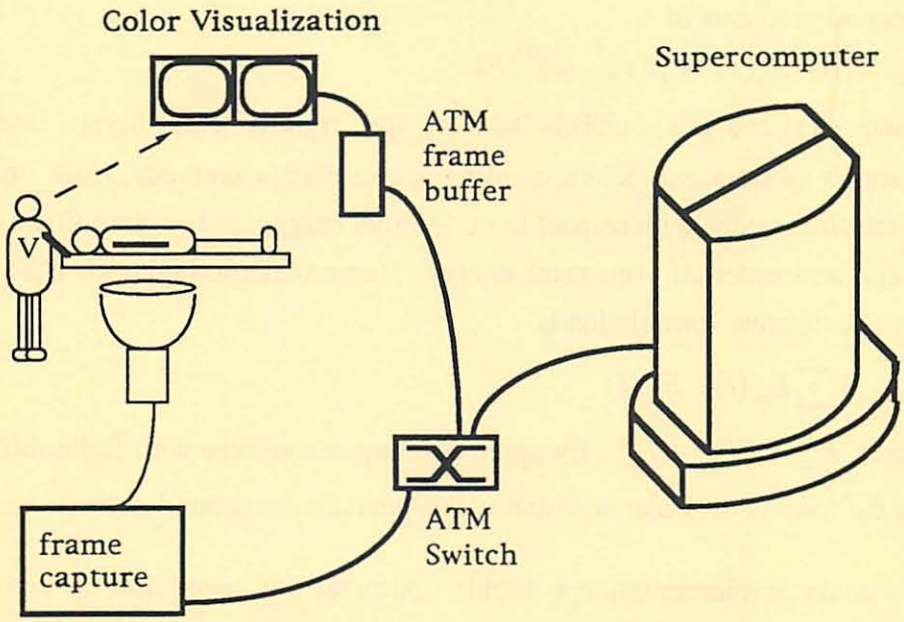


Figure 5. Real time blood flow measurement for decision support during an interventional procedure in an angio suite. Angiographic sequences are captured live off the camera and fed to the compute server through high speed network.

Anatomical object segmentation

Segmentation of anatomical objects are useful to radiologists and clinicians for the purpose of area and volume measurements and 3D visualization. An easy to use tool which can accurately delineate an organ based on user entered region of interest (ROI) has a strong clinical demanded. We have developed an active contour segmentation technique which is easy to use and accurate.

Snake or active contour model

The snake is a controlled continuity spline guided by external constraint forces (user input) and influenced by image forces (gray level gradients) that align it to edge features by energy minimization [12]. The viewer starts the segmentation process by placing the snake near the desired contour using a mouse. Our method allows the viewer to either surround the objects with a starting contour or simply click on the object [13].

Representing the snake parametrically by $v(s) = (x(s), y(s))$, the governing energy equation is: $E^*_{snake} = \int_0^1 E_{\text{int}}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))ds$, where the terms inside the integral are internal energy of the spline, energy due to image force (from pixel gradients), and energy due to external constraint forces. The internal spline energy can be written as a function of the

first and second gradients of v:

$$E_{\text{int}} = \left(\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2\right)/2$$

where $\alpha(s)$ and $\beta(s)$ controls "tension" and "rigidity" respectively. Energy minimization is done through an iterative technique using sparse matrix methods. Each iteration effectively takes implicit Euler steps with respect to the internal energy and explicit Euler steps with respect to the image and external constraint energy. Numerical methods are used to compute this procedure. The discrete formulation is

$$E^*_{snake} = \sum_{i=1}^{n} E_{\text{int}}(i) + E_{ext}(i)$$

where $E_{ext} = E_{image} + E_{con}$. By approximating derivatives with finite differences, the final form takes the form of an Euler equation with a pentadiagonal band matrix.

Our snake implementation is highly optimized for speed and convergence, capable of locating edges in deep trenches. Classically, the snake starts out and remains a closed contour throughout the iterations. We introduce a floating snake which can track partial edge feature with two open ends. Each end can overlap with neighboring snakes at the final stage of object identification and segmentation. This open ended snake is an critical step in making this type of algorithm parallelizable.

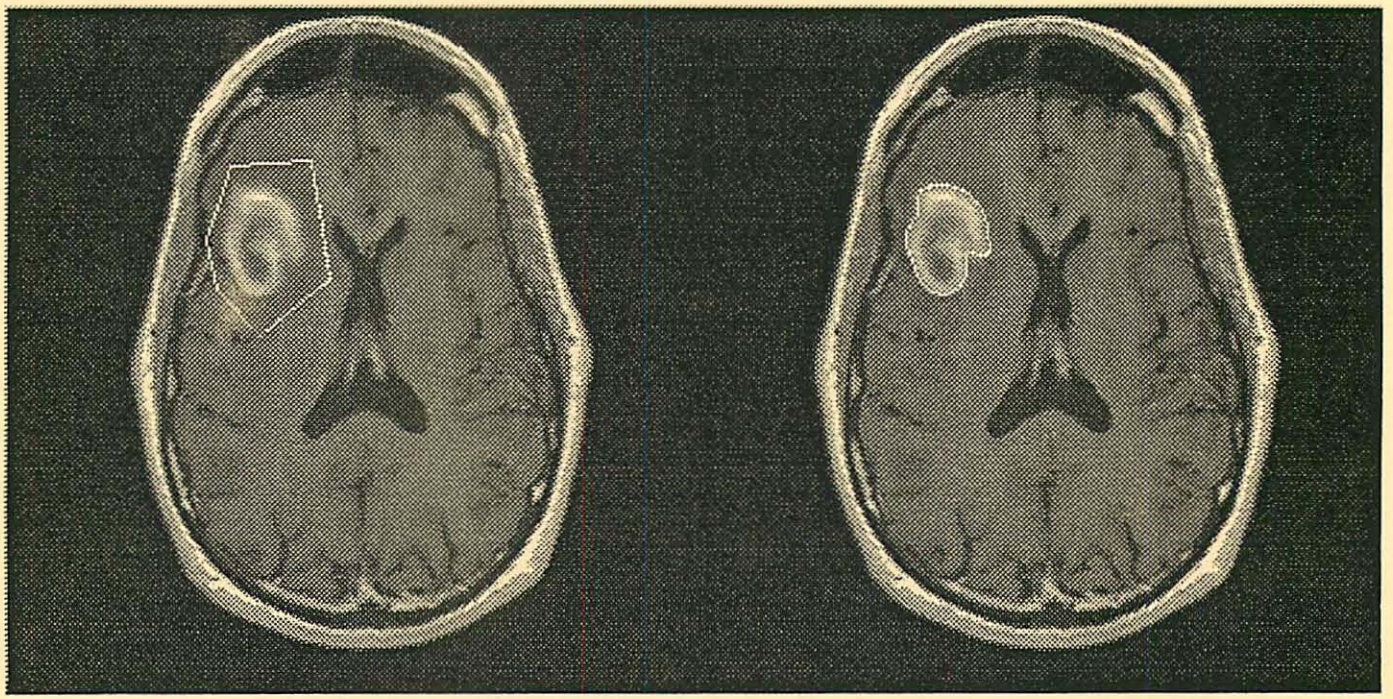


Figure 6. The user draws the initial contour around the object of interest.

Image fusion, registration of multi-modality images in 3D

The possible ways to apply image fusion include same modality over time (to measure tumor growth), different modalities (MR vs. PET) and partial anatomical overlaps (small brain section mapped to longer brain section). Image registration techniques by anatomical features have much wide application than those by stereotactic frames. Some of the clinical utility for 3D registration include 1) physiological studies by mapping PET or SPECT to MR, 2) temporal studies of lesion change by mapping current and prior image sets, 3) radiation therapy planning by mapping diagnostic MR images to CT images used for dose calculation.

Pellizari et.al. [14] developed a technique using a head vs. hat method. More recently, Wood et.al. developed another technique using voxel variance minimization. Both claim accuracy of a few mm. We parallelized Wood's technique [4] by partitioning an image volume into smaller chunks and calculate voxel variances for each separately. A designated central node takes results from each parallel node and complete the variance summation. We are conducting clinical validation of this method using a custom stereotactic frame which is both CT and MR sensitive.

BENCHMARK TESTS

We are conducting a benchmarking test to evaluate high performance computing hardware from several vendors. Various different architectures are represented by these chosen vendors, including shared memory, tightly-coupled clustering by crossbars, and loosely-coupled clustering by gigaswitches. Identical software based on PVM (parallel virtual machine) is used on all platform in order to make valid performance comparisons. Five application programs are tested: 1) active contour segmentation, 2) surface shading, 3) blood flow measurement, 4) wavelet transform based 3D image compression, and 5) 3D image registration. Each program is written so that it can run on 1, 4, 16, or 20 processors. The benchmark result will clear show how the execution time scales with increasing number of processors. When a program is perfectly partitioned (very little interprocessor overhead), a linear speed increase will be achieved.

ACKNOWLEDGMENT

We would like to acknowledge Drs. Robert Close, Roger Woods, and Keh-shi Chuang for contributing their software in the benchmark test.

REFERENCES

1. H.K.Huang, N.J.Mankovich, R.K.Taira, P.S.Cho, B.K.Stewart, B.K.T.Ho, K.K.Chan, and Y.Ishimitsu, "Picture Archiving and Communication Systems (PACS) for Radiological Images: State of the Art," CRC Critical Review in Medical Informatics, 28, Issue 4, 383-427, 1988.

2 R.A.Close, G.R.Duckwiler, F.Vinuela, "Fluid equations applied to blood flow meeasurement using digital videodensitometry," *Investigative Radiology*, 27:504-509, 1992.

3 D.L.Hill, D.J.Hawkes, M.J.Gleeson, T.C.Cox, A.J.Strong, W.L.Wong, C.F.Ruff, N.D.Kitchen, D.G.Thomas, A.Sofat, et.al. "Accurate frameless registration of MR and CT images of the head: applications in planning surgery and radiation therap," *Radiology*, 191:447-54, 1994.

4 R.P.Woods, J.C.Mazziotta, S.R.Cherry, "MRI-PET Registration with Automated Algorithm," *J.computer Assisted Tomography*, 17:536-546. 1993.

5 Z.Chen, M.Ma, P.Saipetch, and B.K.T.Ho, "A method for fast and convergent active contour in medical image segmentation", submitted to *IEEE trans. PAMI.*

6 J.K.Udupa, H.M.Hung, K.S.Chuang, "Surface and volume rendering in three-dimensional imaging: a comparison," JDI, 4:159-60, 1991.

7 P.Hanrahan, "Three-pass affine transforms for volume rendering," Computer Graphics, 24:71-77, 1994.

8 D.R.Peachey, "Solid texturing of complex surfaces," ACM, 19:279-286, 1985.

9 B.T.Phong, "Illumination for computer generated images," Comm Associate for Computer Machinery, 18:311-317, 1975.

10 S.Napel, G.D.Rubin R.B.Jeffrey, Jr., "STS-MIP: A new reconstruction technique for CT of the chest," J.Computer Assisted Tomography, 17:832-838, 1993.

11 J.C.Chaloupka, M.R.Harreld, F.Vinuela, W.J.Karplus, J.Robrets, G.Guglielmi, "Multidisciplinary approach to the study of hemodynamics/flow mechanics of lateral wall aneurysms using invivo and computer simulation modelling," Proceedings of the American Society of Neuroradiology, 31 annual meeting, p109, 1993.

12 M.Kass, A.Witkin, D.Terzopoulos,"Snake: active contour models," *Int.J.Comput.Vission*, 1:321-331, 1987.

13 Z.Chen, M.Ma, P.Saipetch, B.K.T.Ho, "A method for fast and convergent active contour in medical image segmentation," submitted to *IEEE trans. PAMI.*

14 C.A.Pelizzari, G.T.Y.Chen, D.R.Spelbring, R.R.Weichselbaum, "Accurate three-dimensional registration of CT, PET, and/or MR Images of the brain," *J.Computer Assisted Tomography*, 13:20-26, 1989.

Volumetric Image Compression by 3D Discrete Wavelet Transform (DWT)

Jun Wei, Pongskorn Saipetch, Remash Panwar*, Doris Chen, Bruce K.T. Ho

AR-277 CHS, Department of Radiological Sciences, University of California, Los Angeles, CA 90024
* Department of Computer Sciences, UCLA, CA 90024

ABSTRACT

The newly developed Discrete Wavelet Transform (DWT) compression method is far superior to previous Full Frame Discrete Cosine Transform (FFDCT) as well as industrial standard JPEG[1]. Due to its localization properties both in spatial and transform domain, the quantization error introduced in DWT will not propagate globally as in FFDCT. Also DWT transform is a global technique that avoids the JPEG type block artifacts. As in all techniques, correlation among pixels makes compression possible. In volumetric image sets, such as CT and MR, inter-slice correlation can be exploited in addition to in-slice correlation. In this 3D DWT study, inter-slice correlation has also been investigated for CT and MR image set. Different numbers of slices are grouped together to perform wavelet transform in the transaxiale direction as a mean of testing relationship between correlation and compression efficiency. The 3D DWT is developed on UNIX platform. Significant higher compression ratio is achieved by compressing CT data as a volume versus one slice at a time. DWT is an excellent technique for exploiting inter-slice correlation to gain additional compression efficiency.

1. INTRODUCTION

Large volumes of high resolution radiological images acquisition by a picture archiving and communication system (PACS) require large capacity archival storage and high network bandwidth. In large hospital such as UCLA Medical Center, the amount of image information needed to be stored exceeds two terabytes per year[3]. Network traffic load due to retrieve large volume of image data daily becomes heavy and thus brings down the response time significantly. The problems in storage and network load become even serious when dealing with large volumetric image data sets. There arises the need for more efficient compression algorithm. Newly developed DWT compression method has shown superior advantage over other conventional methods in the 2D image compression in terms of compression efficiency. Furthermore, we applied multiresolution levels of wavelet transform in the transaxial direction to exploit inter-slice correlation for volumetric image sets, such as CT and MR images. We also introduced an optimal arithmetic coding to achieve higher compression efficiency. The arithmetic coding is done for each three dimensional block with different resolution levels.

In this 3D DWT study, inter-slice correlation has been investigated for volumetric CT and MR image data sets. Different numbers of slices are grouped together to perform wavelet transform in the z direction to demonstrate the relationship between correlation and compression efficiency. Normalized mean squared error (NMSE) and root mean squared root error (RMSE) are used to evaluate the quality of reconstructed images.

2. THREE DIMENSIONAL DISCRETE WAVELET TRANSFORM

In previous studies[1], we have transformed the two-dimensional images using one-dimensional row (x direction) and column (y direction) wavelet transforms. The process is similar to two dimensional FFT of images. The transform is carried out using a multi-scale pyramidal decomposition[4,5]. The wavelet coefficients are obtained by convolution with the one-dimensional low-pass filter h and high-pass filer g followed by decimation (see Fig. 1). The reconstruction is performed by the convolution with reconstruction filters $\tilde{h}$ and $\tilde{g}$. The resulting transformed image consists of sub-images with different resolution levels and orientations[4].

Three dimensional wavelet transform is similar to the transform in two dimensional image. Multi-scale pyramidal decomposition process is extended to the transaxial direction (z direction) for a volumetric image set. After transform in x and y direction for each image slice in the volumetric set, another transform is performed along the z direction (see Fig. 1). Correspondingly, one more upsampling and convolution in z direction are needed before y and x directions (see Fig. 2). The additional wavelet transform enable possible decorrelation among transaxial image slices.

The edges of the image are handled by performing the convolutions on symmetric extensions of the images to avoid

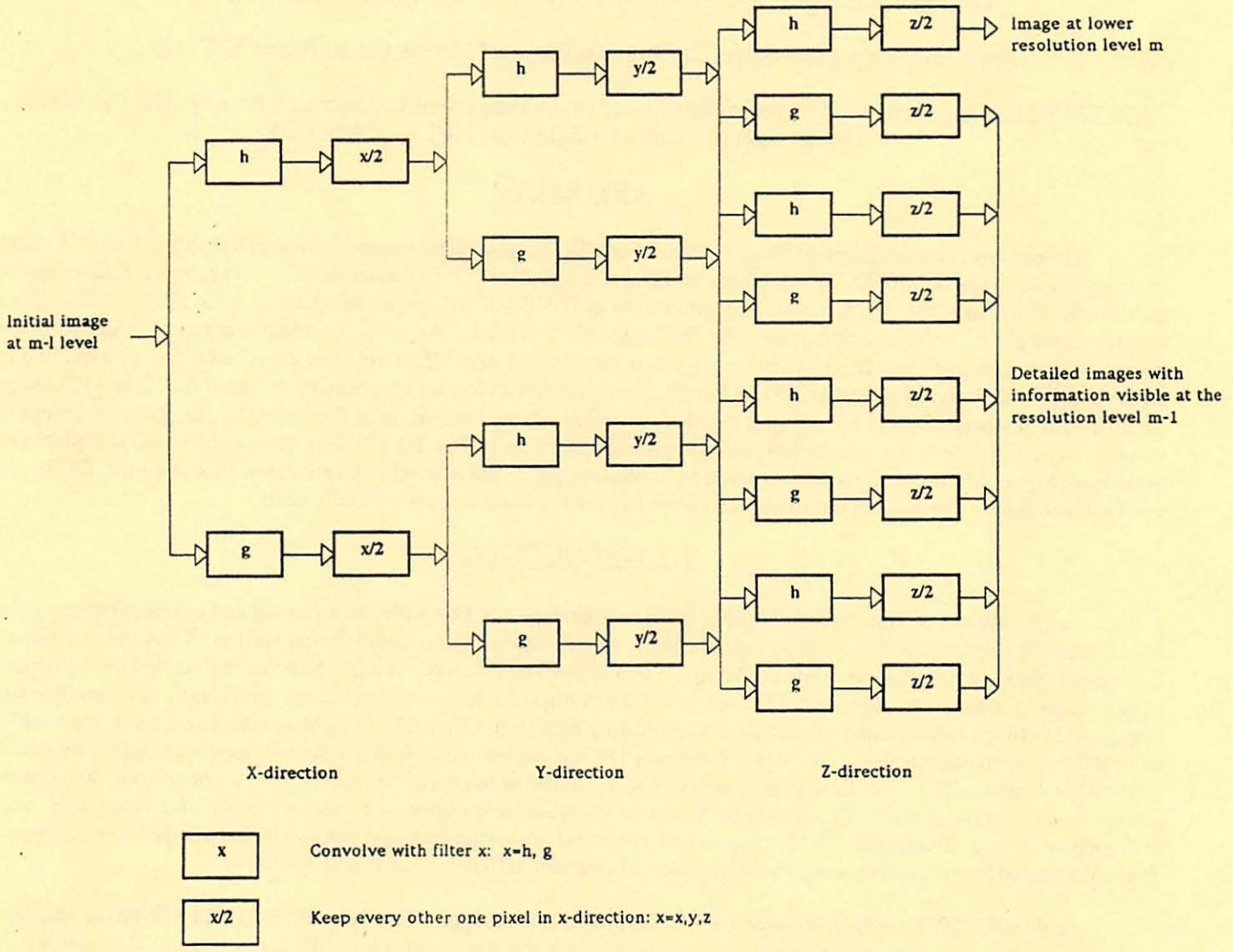


Figure 1. One stage in a multiscale pyramidal decomposition

boundary artifacts introduced by simple periodization. The types of extensions and associated parameters needed to perform the symmetric wavelet transform are described in details by Brislawn[6].

The DWT method is not bound by the power of 2 format constraint as in the case of FFDCT. However, the finite filter length used in DWT leads to some complexity in image size. The symmetric extension at the image boundary places a minimal dimensional constraint for each resolution level of DWT. Normally, there are much more than enough number of pixels in x, y direction for up to 4 resolution levels. For the z direction, the number of the slices in the volumetric set limits the resolution levels at which the DWT can be performed. Table 1. gives the minimum number of slices at each resolution level.

3. BLOCK-BASED ARITHMETIC CODING

Arithmetic coding is a model-based, lossless data compression algorithm. A string of symbols ("message") is represented by an interval of a real number between 0 and 1. To start the coding process, the number is assigned to be in [0,1). For every symbol the coder receives, the interval is scaled down by the factor according to the model. A chose of the

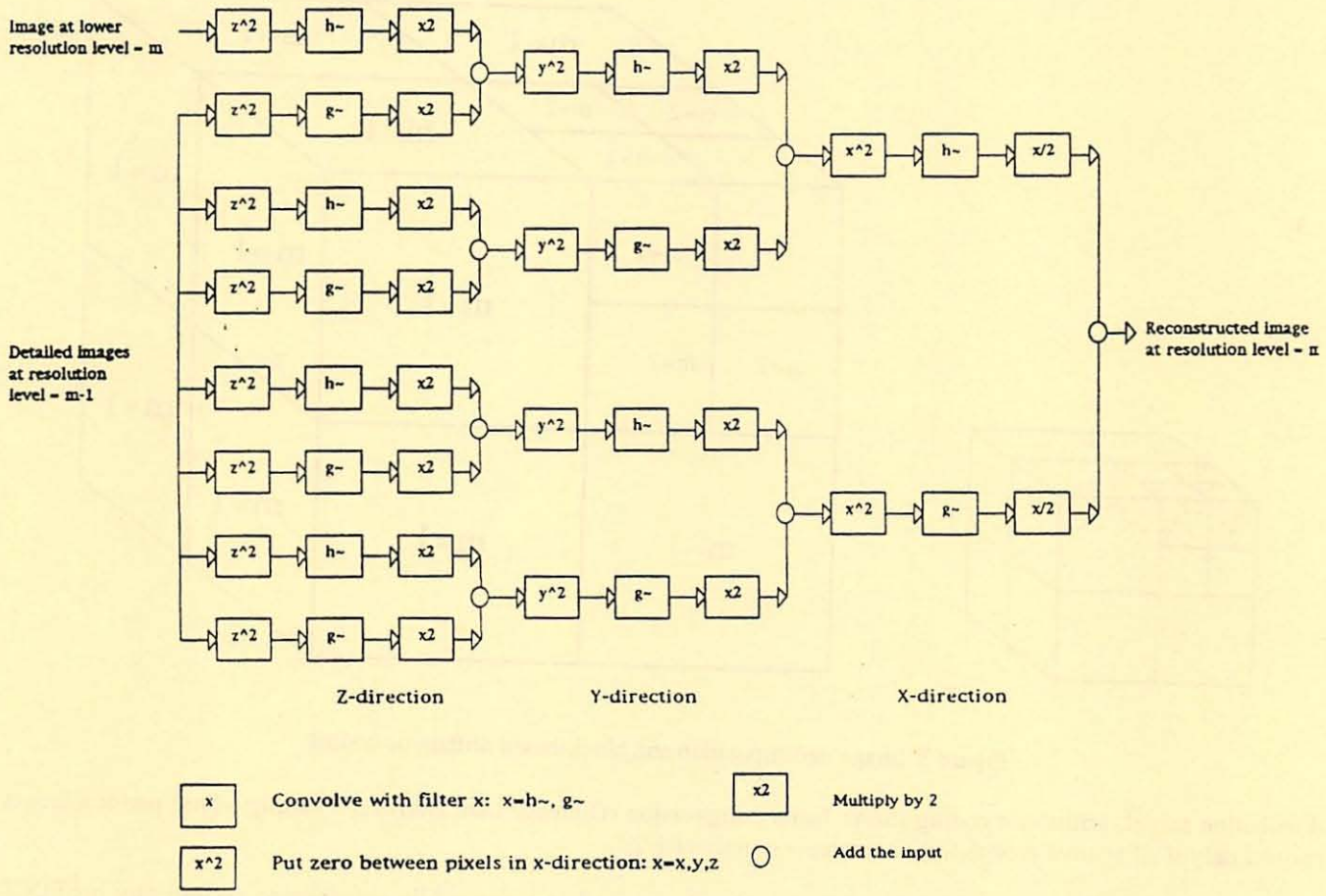


Figure 2. One stage in a multiscale pyramidal reconstruction

Resolution level (m)	Number of slices in image set
1	6
2	12
3	24
4	48

Table 1. Minimum number of slices in volumetric image set for resolution level m

model is the probability distribution of the symbols. As the message becomes longer, the interval becomes smaller. The more frequent appearance of a symbol gives a high probability which will reduce the encoded interval by less amount. Thus fewer bits are needed to represent the symbol. The number of bits per symbol for the encoded message can approach the theoretic entropy bound[7].

In this study, the symbols are the values of the quantized wavelet coefficients, the quantized wavelet coefficients of the image are grouped according to their resolutions and orientations to skew the magnitude histograms. The probability distributions of the quantized coefficients for each resolution level are concentrated to form a peak shape. For the peak shape

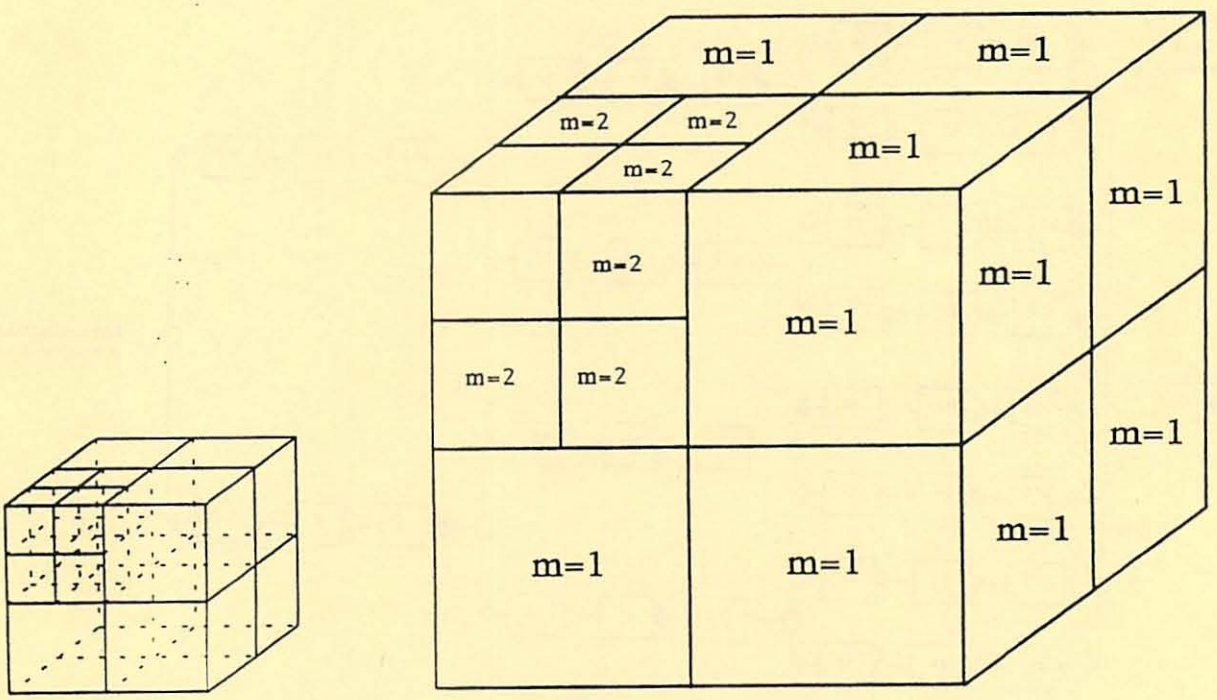


Figure 3. Image decomposition and block-based arithmetic coding

distribution model, arithmetic coding shows better compression efficiency than Huffman's coding whose performance is optimal only if all symbol probabilities are integer powers of 1/2.

Previous study has shown that the block-based arithmetic coding optimized the compression performance in FFDCT compression method[8]. We applied the similar scheme in this three-dimension DWT study. The quantized wavelet transform coefficients are partitioned into cubic blocks according to resolution levels and orientations (see Fig. 3). Each block is encode individually by arithmetic coding.

Block-based arithmetic coding will introduce headers for each block in encoded bit string. However the size of header depends on the probability distribution of the quantized wavelet coefficients in each block. A narrow peak shape probability distribution needs shorter header than a wide, spreadout distribution. In the high resolution block, the distribution is always a nice sharp peak shape which results in very short header and also much higher compression efficiency. Overall, the total header size is much smaller compared with the large gain in compression efficiency.

4. EXPERIMENTS AND METHODS

The volumetric images are transformed into optimal resolution levels for three dimensions by using the filters described by Saipetch[1]. Three directions of wavelet transform have been applied to the pixel data of images to the optimal resolution level. The wavelet transform coefficients are uniformly quantized (by scaling down to [-0.5,0.5]). The quantized coefficients are then partitioned into blocks with optimal block size according to the resolution levels and orientations. Arithmetic coding is done individually within each block. The reconstruction is basically the reverse process of the compression process (see Fig. 4). We used the normalized mean squared error (NMSE) and root mean squared error (RMSE) to evaluate the quality of the reconstructed images. The NMSE and RMSE are defined as the following:

$$NMSE = \sqrt{\frac{\sum_{m=0}^{m-1}\sum_{n=0}^{n-1}[X_r(m,n) - X_o(m,n)]^2}{N^2}}$$

$$RMSE = \sqrt{\frac{\sum_{m=0}^{m-1}\sum_{n=0}^{n-1}[X_r(m,n) - X_o(m,n)]^2}{\sum_{m=0}^{m-1}\sum_{n=0}^{n-1}[X_o(m,n)]^2}}$$

Where $X_o(m,n)$ and $X_r(m,n)$ are the pixel values from original image and reconstructed image. The image dimension is $m \times n$. The NMSE and RMSE are evaluated within slices.

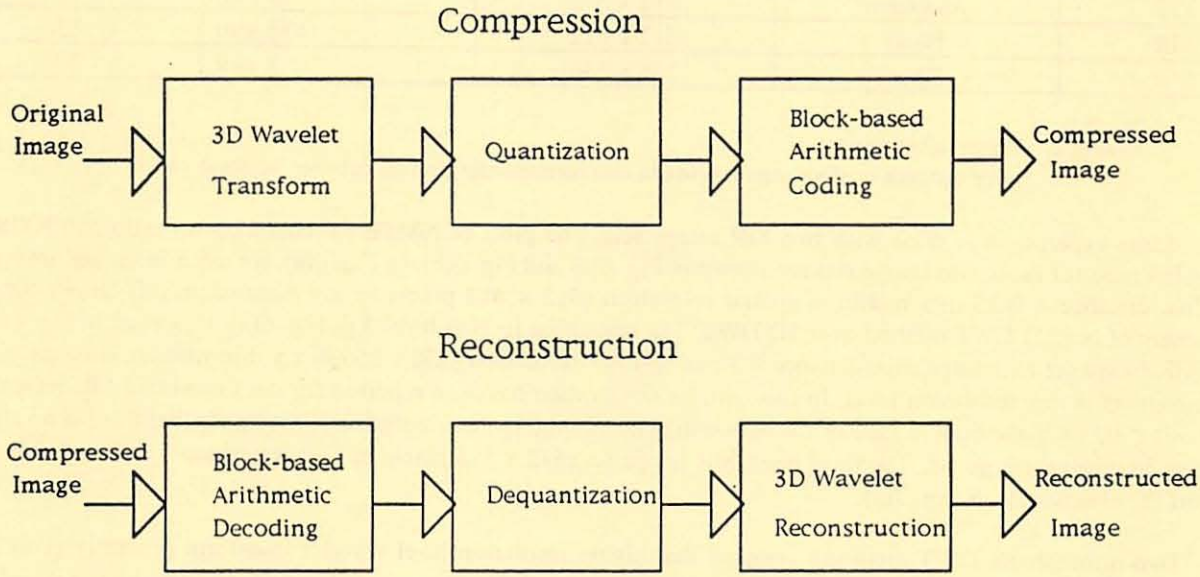


Figure 4. Compression and reconstruction procedure.

More than ten sets of CT and three sets of MR images are studied using our 3D DWT method. The NMSE and RMSE are evaluated for each image set at various resolution levels and different partition levels for block-based arithmetic coding. The NMSE vs. compression ratio and RMSE vs. compression ratio are plotted for all image sets to demonstrate reconstructed image quality.

We plotted NMSE vs. compression ratio for same image sets by applying 2D DWT to one slice at a time with the same reconstruction quality to demonstrate the performance of 3D DWT method versus 2D DWT method. We performed our 3D DWT method at resolution level m = 1,2,3 to look for the optimal levels. The partition of sub-blocks for arithmetic coding can be chosen at any resolution levels regardless which resolution level wavelet transform is performed(see Fig. 3). We have chosen different partitions of sub-blocks to look for optimal block-based arithmetic coding.

Our 3D DWT method is developed on a UNIX platform. The algorithm is amenable to parallel computing on multiprocessor workstation.

5. RESULTS

Out of ten CT and three MR image sets studied, we show the results for two CT and two MR image sets in this paper. Those two CT and two MR image sets are listed in Table 2. The results for the rest of the image sets are similar to those shown in the paper and would not be listed here.

We chose the NMSE and RMSE of the first slice in every image set to compare the performance of different compression methods. The plots of NMSE vs. compression ratio and RMSE vs. compression ratio for two CT image sets (one with slice distance = 0.49 mm and the other with slice distance = 3 mm) are shown in Fig 5(a) and 5(b). The 3D DWT

and 2D DWT results are plotted together for comparison in each plot. The cr-3D curve presents the reconstructed image quality by using our 3D DWT compression method. The cr-2D curve shows the results by performing wavelet transform for each slice individually (2D DWT for three dimensional image sets). Both methods are performed at the same resolution level (eg. level 2 in Fig. 5(a) and level 3 in Fig. 5(b)). It is clear that there is significant improvement in compression ratio upon the three dimensional wavelet transform for both CT image sets. Similar plots for the other CT image sets we have studied support the above observation.

Modality	Anatomy	x,y dimension(pixels)	slice distance*	Number of slices
CT	Brain	512 x 512	0.49 mm	63
CT	Abdomen	512 x 512	3 mm	71
MR	Brain	512 x 512	0.35 mm	83
MR	Brain	256 x 256	3 mm	30

Table 2. Image information
*slice distance = the distance between two corresponding pixels between adjacent slices

Same experiment is done with two MR image sets. The plots of NMSE vs. compression ratio and RMSE vs. compression ratio for those two image sets are shown in Fig. 6(a) and Fig. 6(b). In Fig. 6(a), we use a brain MR image with small slice distance = 0.35 mm and high spatial resolution (512 x 512 pixels in x,y dimensions). It shows the great improvement of our 3D DWT method over 2D DWT. The resolution level is level 3 in Fig. 6(a). However, in Fig. 6(b), for a brain MR image set with large slice distance = 3 mm and low resolution (256 x 256 in x,y dimensions), it seems to show no improvement at any resolution level. In fact, similar observation has been reported for the same brain MR image set in 6(b) by using 3D DCT method[9]. One of the reason may be that the spatial resolution is very poor that there is a very high noise level in original image set. The third brain MR image set (512 x 512 pixels in x, y dimensions) in our studies seems to support the observation of Fig. 7(a).

Two dimensional DWT methods suggests that higher resolution level wavelet transform generally gives higher compression efficiency with the same reconstruction quality. We applied the same scheme to our three dimensional DWT methods. To show the improvement, we plot NMSE vs. compression ratio at resolution level m = 1, 2, 3 for all the image sets. In Fig. 7(a), a plot is shown for an abdomen CT image with slice distance = 3mm and 512 x 512 pixels in x,y dimensions. A cr-2D curve (the result of 2D DWT method from Fig. 5(b)) is included. The improvement is obvious from level 1 to level 2 at high compression ratio(>40 with NMSE ~ 1.5E x -4 and RMSE ~ 9). But there is not much gain from level 2 to level 3 at all compression ratio. At low resolution level(<40), all three levels give the same compression efficiency. The studies for the other CT image sets show similar observation. The optimal resolution level for wavelet transform in 3D DWT is 1 level at low compression ratio(high quality of reconstructed image), and 2 levels at reasonably high compression ratio. For some image anatomy such as abdomen CT image set, sometimes, level 3 gives a little better performance at high compression ratio. However, it seems to be preferable for 1 or 2 levels of transform for brain CT image sets.

In Fig. 7(b), a plot for the high resolution brain MR image (512 x 512 pixels x, y dimensions) with very fine slice distance = 0.35 mm shows that even level 3 still gives great improvement over level 2. A cr-2D curve (the result of 2D DWT method from Fig. 6(b)) is also included. The performance at each resolution level varies a lot at compression ratio greater than 10. At high compression ratio(>10), 3 level transform is preferred while at low compression ratio(<10), 2 or 3 levels transform is about the same as 1 level transform. One level is sufficient as far as compression ratio is concerned. But for the lower spatial resolution brain MR image (256 x 256 pixels with slice distance = 3 mm) it does not show any improvement at any level (see Fig. 6(b)).

The optimal resolution level for wavelet transform is dependent on image modality and image anatomy. It also has an upper bound due to the symmetric extension of boundary. The upper bound is determined by the number of slices in the image set to perform wavelet transform. The Table 1 lists the minimum number of slices for each resolution level. For example, for an image set of 6 slices, it can only do up to one level transform. On the other hand, higher level transform usually implies more computing power and compression time required. In general, when there is not much improvement at higher level transform, lower level transform is preferred to same time without much loss in compression efficiency.

We have compared the compression performance with and without block-based arithmetic coding in 3D DWT. In Fig. 8, we show a plot of NMSE vs. compression ratio for an abdomen CT image set (512 x 512 pixels with slice distance = 3mm). The "cr(whole)" curve is the result from arithmetic coding for the whole image set in one block, ie. without partition of the wavelet coefficients. The "cr(block AC)" curve is the result from the block-based arithmetic coding. The partition of the sub-block of the wavelet coefficients is the same of the decimation after one level wavelet transform in Fig. 8. We also experimented different partitions, such as partition the sub-blocks after the decimation of each level transform. There is obvious improvement in compression ratio with block-based arithmetic coding from Fig. 8. Our studies for other image sets show improvement of block-based arithmetic coding. The optimal partition of the sub-blocks is the one same as the decimation after one level wavelet transform (see Fig. 3).

Finally, we studied the compression performance of 3D DWT with different number of slices in groups. For the same image set, we performed wavelet transform to the image set within different size of groups. In Fig. 9, compression ratio had been compared with transform group size of 8, 16, 32 and 48 slices for an abdomen CT image set (512 x 512 with slice distance of 3mm). The NMSE of that compression ratio is 3E x -5 and RMSE is 4. There is slight improvement of compression efficiency with large group size for all resolution levels. Notice that there is no two level or three level wavelet transform for the group of 8 slices because there are at least 12 slices for up to two level transform and 24 for three level transform. Similarly, there is no three level transform for the group of 16 slices. Fig.9 may imply that it may be benefit to compress image set in large group.

6. CONCLUSION

Experiments have shown that significant higher compression efficiency can be achieved for many volumetric CT and MR image sets by wavelet transform in three dimensions over only in two dimensions. DWT is an excellent technique to exploit inter-slice correlation to gain additional compression efficiency.

7. REFERENCE

1. P. Saipetch, B.K.T. Ho, R. Panwar, M. Ma, and J. Wei, "Application of wavelet transform with arithmetic coding in radiological image compression," IEEE Eng Medicine and Biology (in press)

2. B.K.T. Ho, J. Chao, P. Zhu, and H.K.Huang, "Design and implementation of Full Frame Bit Allocation image compression hardware module," Radiology, 179:563-567, 1991

3. B.K.T. Ho, R. Sadri, W. Chao, L. Huang, and R. Taira, "Strategy for long term PACS archive using hierarchical storage with image compression and HIS/RIS/ triggers," J. Digital Imaging (in press)

4. M. Antonini, "Image coding using wavelet transform," *IEEE Trans. Image Proc.*, Vol. 1, No. 2, pp.205-220, April 1992.

5. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Math. Intel.*, Vol. 11, July 1989.

6. C. M. Brislawn, "Classification of symmetric wavelet transform," Group C-3, Computer Research, Los Alomos Notional Laboratory, March 1993.

7. I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," Comm. of the ACM, Vol. 30(6), pp.520-540, June 1987.

8. J. Wei, M. Ma, P. Saipetch, K.S. Chuang, and B.K.T. Ho, "Block implementation of arithmetic coding for image compression," Proc. SPIE, Feb. 1994, Newport Beach, CA.

9. D.T.Chen, M. Ma, and B.K.T. Ho, "Volumetric image compression by 3D Discrete Cosine Transform (DCT)," Proc. SPIE, Feb. 1992, Newport Beach, CA.

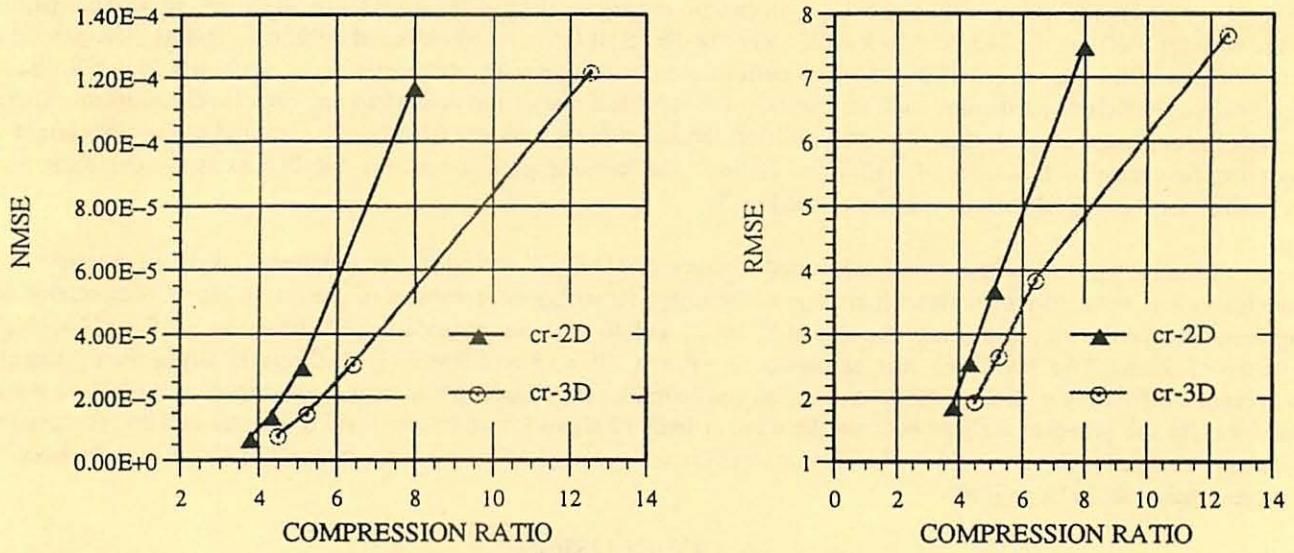


Figure 5(a). NMSE vs. COMPRESSION RATIO and RMSE vs. COMPRESSION RATIO for a brain CT image with slice distance=0.49mm. Two compression methods is compared. cr-2D shows the results from 2D DWT method. cr-3D shows the results from 3D DWT method. Wavelet transform is performed at resolution level=2. The results show that 3D DWT method improved the compression efficiency over 2D DWT method.

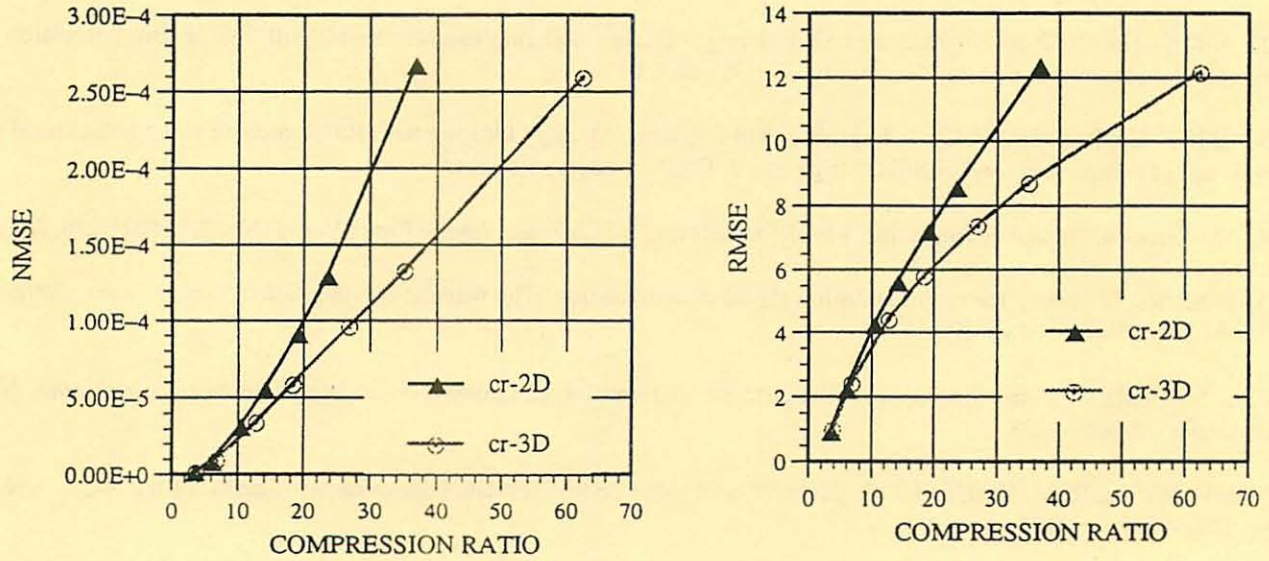


Figure 5(b). NMSE vs. COMPRESSION RATIO and RMSE vs. COMPRESSION RATIO for abdomen CT image with slice distance=3.0 mm. Two compression methods is compared. cr-2D shows the results from 2D DWT method. cr-3D shows the results from 3D DWT method. Wavelet transform is performed at resolution level=3. The results show that 3D DWT method improved the compression efficiency over 2D DWT method.

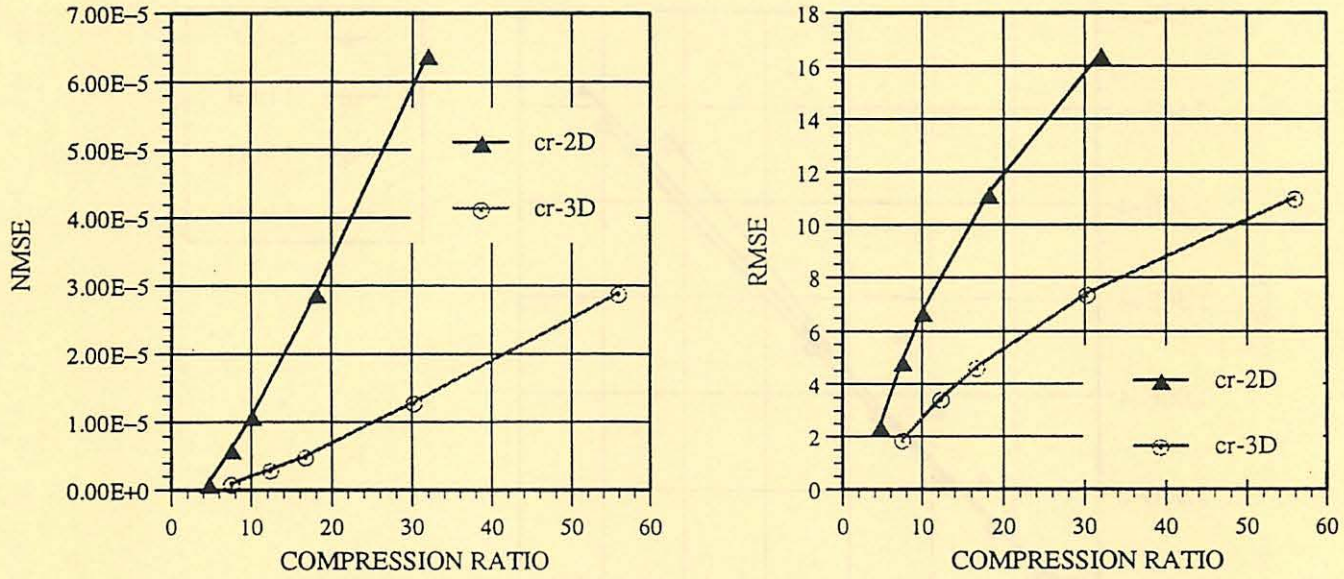


Figure 6(a). NMSE vs. COMPRESSION RATIO and RMSE vs. COMPRESSION RATIO for a brain MR image with slice distance = 0.35 mm. Two compression methods are compared. cr-2D shows the results from 2D DWT method. cr-3D shows the results from 3D DWT method. Wavelet transform is performed at resolution level=3. Image dimension is 512 x 512. The results show that 3D DWT method improved the compression efficiency over 2D DWT method.

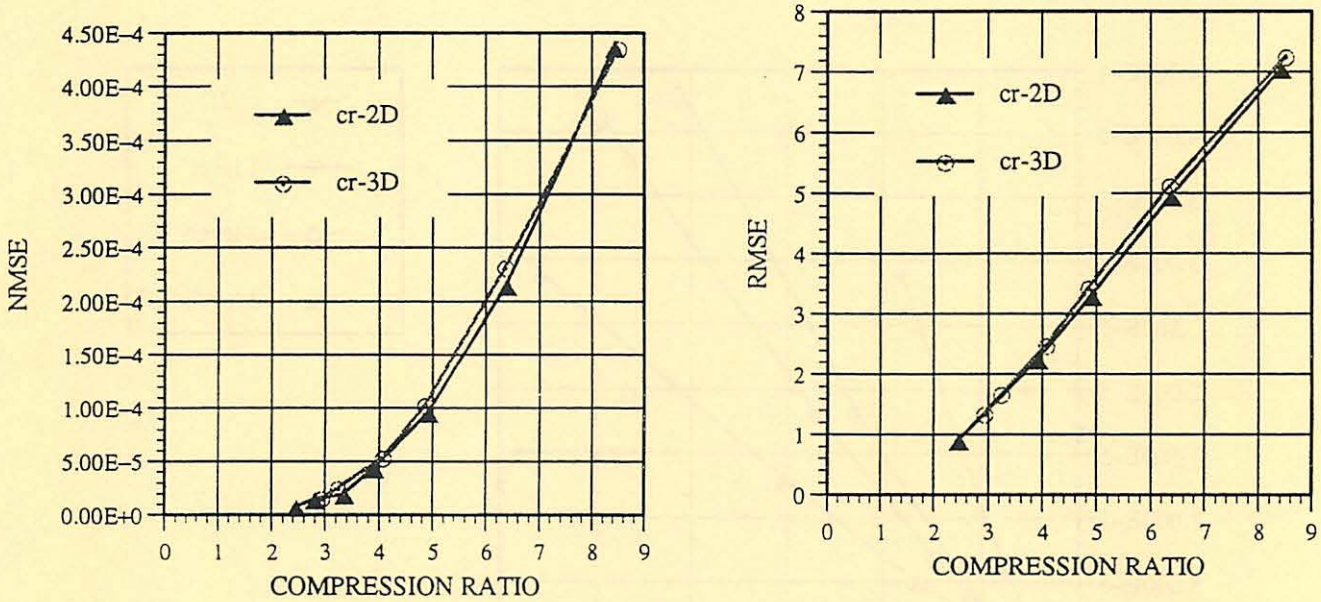


Figure 6(B). NMSE vs. COMPRESSION RATIO and RMSE vs. COMPRESSION RATIO for a brain MR image with slice distance = 3.0 mm. Two compression methods are compared. cr-2D shows the results from 2D DWT method. cr-3D shows the results from 3D DWT method. Wavelet transform is performed at resolution level=2. Image dimension is 256 x 256. It seems that 3D DWT method does not improve much in compression ratio.

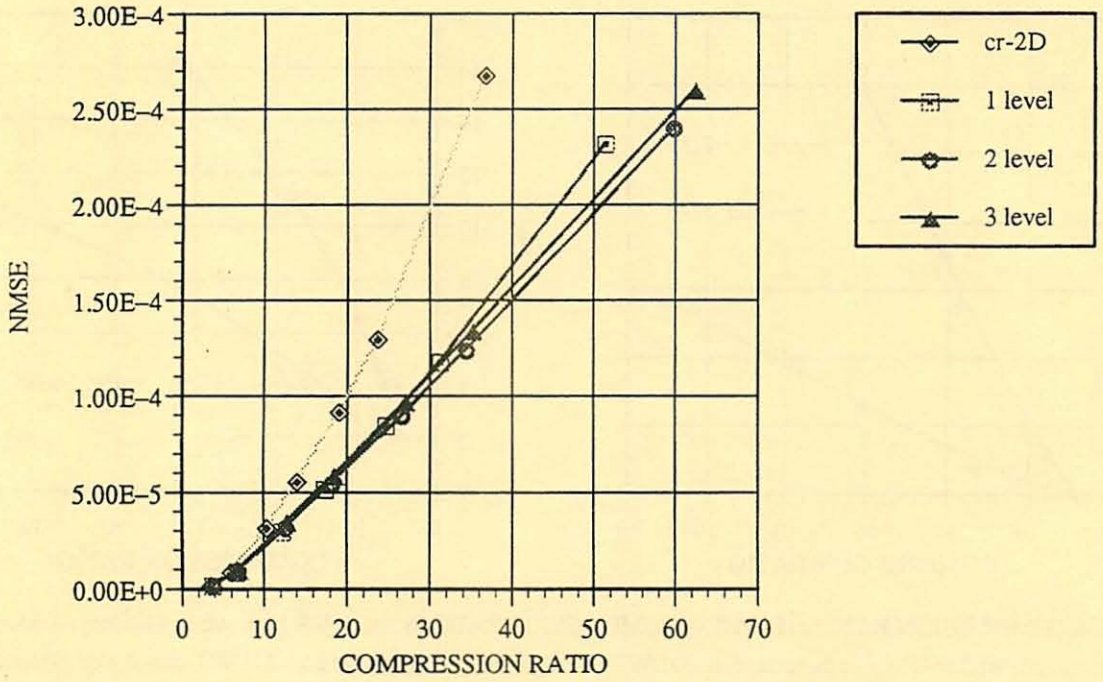


Figure 7(a). NMSE vs.COMPRESSION RATIO for abdomen CT image with slice distance = 3.0 mm.This experiment is designed to look for optimal resolution level of wavelet transform. It is shown that for high compression ratio(>40) 3 levels wavelet transform is prefered. But at low compression ratio, more than 1 level transform does not improve compression efficiency. 1 level is prefered.

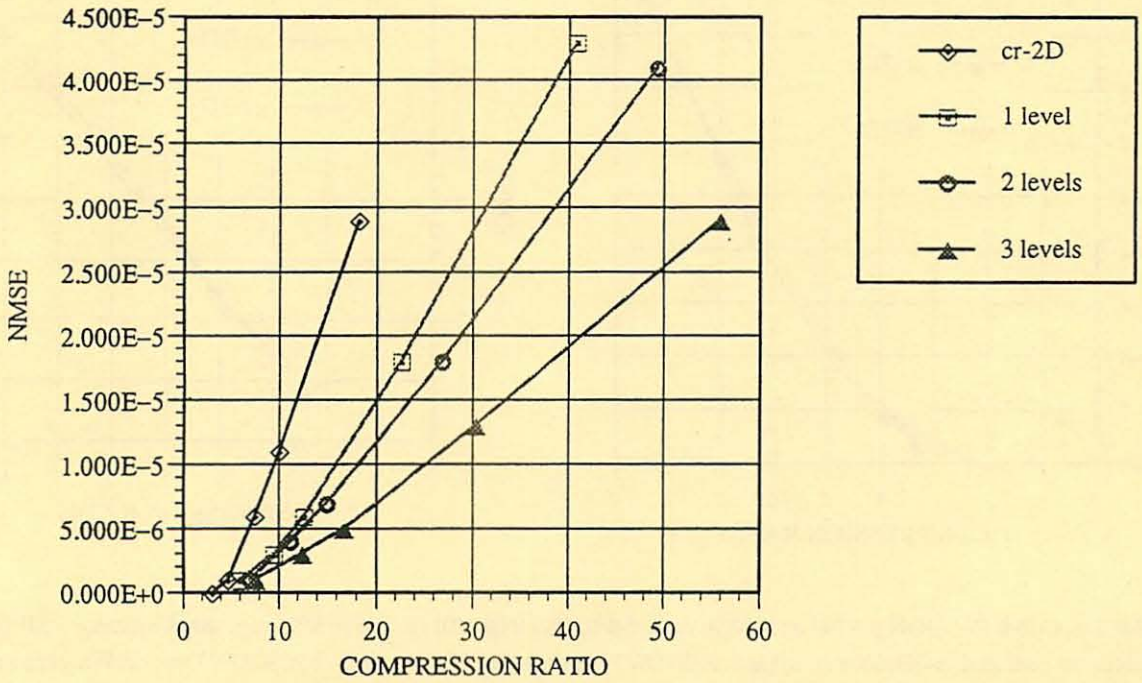


Figure 7(b). NMSE vs. COMPRESSION RATIO for a brain MR image with slice distance = 0.35 mm. This is to look for optimal resolution level of wavelet transform. It is shown that 3 levels transform provides the best compression ratio for this particular MR image set.

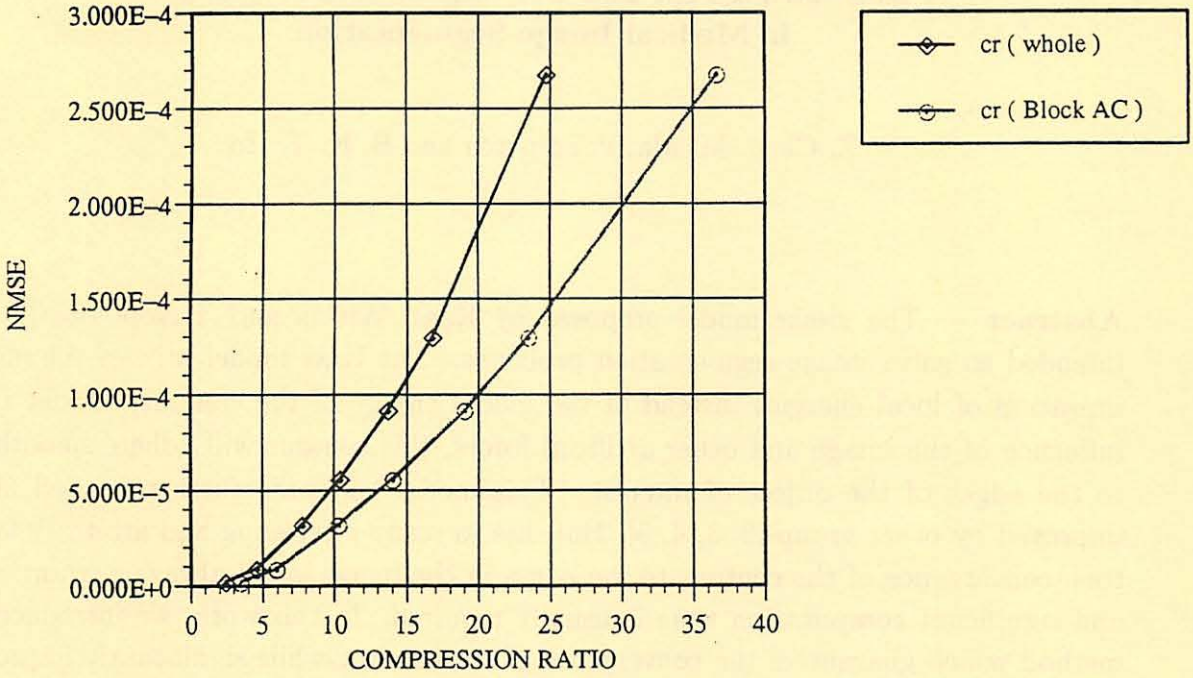


Figure 8. NMSE vs. COMPRESSION RATIO for a abdomen CT image with slice distance = 3.0 mm. This is to show that block-based arithmetic coding improves the compression ratio. The image set is transformed at 1 resolution level and arithmetical coded as a whole block versus as subblocks partitioned at 1 resolution level.

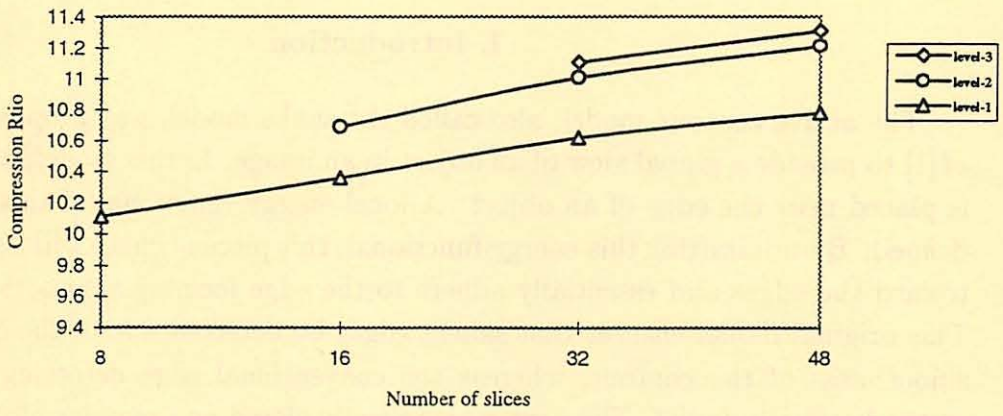


Figure 9. COMPRESSION RATIO vs. group size for 3D DWT for a abdomen CT image set. The image set is transformed in groups of 8,16,32 and 48 slices, respectively. It is shown that transform in large group results in better compression performance.

A Method for Fast and Convergent Active Contour in Medical Image Segmentation

Z. Chen, M. Ma, P. Saipetch and B. K. T. Ho

Abstract — The *snake* model proposed by Kass, Witkin and Terzopoulos [1] is intended to solve image segmentation problems. The Kass model utilizes the minimization of local energies instead of the global energy of the contour. Under the influence of the image and other artificial forces, this contour will adhere smoothly to the edges of the object of interest. This model had been further studied and improved by other groups[2, 3, 4, 5]. But due to many interfering and arbitrary factors, convergence of the contour to the edges in the image is not always guaranteed, and significant computation time is usually required. In this work, we introduce a method which guarantees the convergence of the contour while significantly improving the processing speed. Many new techniques are employed including reducing the size of the matrix, applying different forces at different stages, and rearranging the nodes of the contour during time iteration.

Index Terms: — Active contour, image segmentation, feature extraction, magnetic resonance imaging.

I. Introduction

The active contour model, also called the snake model, was proposed by Kass *et al* [1] to provide a global view of an object in an image. In this model, a piece of curve is placed near the edge of an object. A local energy functional of this curve is then defined. By minimizing this energy functional, this piece of curve will iteratively move toward the edges and essentially adhere to the edge forming a smooth edge outline. This original model ensures that salient edges be detected due to the continuity and smoothness of the contour, whereas the conventional edge detecting and tracking methods usually fail[6]. The contour is initially closed and remains closed during the entire operation. Its closedness provides an unambiguous segmentation.

In the original active contour model, the initial contour must be placed near the edge of the object so that the contour can feel the image force and move toward the edges. For faster convergence, it is desirable to introduce some artificial forces so that the contour is pushed toward the edges [2, 3]. These forces are usually image

dependent and tend to pull the contour across weak edges. Also in the active contour model, the nodes of the contour tend to either separate or congregate during the iteration. This is mainly due to the minimization of the internal energy of the contour. Attempts have been made to improve the stability of this model [4, 5]. But a very significant amount of computation time must be required, while the convergence to the desired edges is still not guaranteed.

The basic idea behind the active contour model is to locally minimize the energy functional. But the definition of this energy functional itself is, in a sense, arbitrary. There is no unique set of parameters, or even general guidelines on their ranges. Furthermore, the convergence of the contour in the edge of the image is always a concern. The contour may step over the edge and converge to unintended structures.

In this paper, we introduce a method which is convergent and significantly faster than the previous models. We also give general guidelines for the optimal range for the parameters used in the energy functional. In this method, we normalize the forces so that the contour will move toward the edges as fast as possible but never step over the edges. Instead of using a truly closed contour with the correct boundary conditions, we employ an open-ended contour and artificially force the contour to close. This reduces the computation time drastically when the number of nodes is large. We also eliminate many restrictions in the intermediate stages because only the the final position of the contour is important. The problem of node separation and congregation during the time iteration is eliminated by rearranging the nodes frequently. The result is fast computation while ensuring convergence.

II. The Active Contour Model

In a two dimensional space, a contour is described by its position on the plane, $\mathbf{v}(s) = \{x(s), y(s)\}$, where $s \in [0,1]$ is a parameter. For a closed contour, the boundary condition requires $\mathbf{v}(0) = \mathbf{v}(1)$ and $\mathbf{v}'(0) = \mathbf{v}'(1)$. The energy functional of the contour is defined as

$$E(\mathbf{v}) = \int_0^1 ds\Big\{E_{int}[\mathbf{v}(s)] + E_{img}[\mathbf{v}(s)] + E_{art}[\mathbf{v}(s)]\Big\} \tag{1}$$

where $E_{int}[\mathbf{v}(s)]$ is the internal energy of the contour, $E_{img}[\mathbf{v}(s)]$ is the image energy, and $E_{art}[\mathbf{v}(s)]$ is the artificial energy. The internal energy depends only on the curvature of the contour and is expressed as

$$E_{int}[\mathbf{v}(s)] = \frac{1}{2}\Big(\alpha(s)\,|\mathbf{v}'(s)|^2 + \beta(s)\,|\mathbf{v}''(s)|^2\Big). \tag{2}$$

The first and second derivatives are with respect to the contour parameter s. α and β are called stiffness parameters and are functions of s. But in practical applications,

they are usually taken as constants. The values of these stiffness parameters determine the smoothness of the contour. The image energy is usually taken to be proportinal to the gradient of the image intensity, $-\delta |\nabla I|^2$, although one is free to add other terms[1]. The sole purpose of the artificial energy is to attract the contour to the edges of the image[2, 3].

The final position of the contour is determined when the energy functional reaches a minimum. Applying the standard mathematical procedure to minimize the local energy, we obtain a set of Euler equations:

$$-\alpha \mathbf{v}'' + \beta \mathbf{v}'''' - \mathbf{F}_{ext} = 0 \tag{3}$$

with the boundary condition $\mathbf{v}(0) = \mathbf{v}(1)$ and $\mathbf{v}'(0) = \mathbf{v}'(1)$ for a closed contour. $\mathbf{F} = -\nabla(E_{img} + E_{art})$ is called the external force. Here α and β are taken as constants. It is most convenient to discretize Equation (3) and solve it by iteration.

$$\mathbf{v}_t = (I + \gamma \mathcal{A})^{-1} \mathbf{v}_{t-\gamma} + \gamma \mathbf{F}, \tag{4}$$

where I is an $N \times N$ identity matrix with N being the number of node of the contour, $\mathcal{A}$ is an $N \times N$ matrix, and γ is the time step of the iteration. Thus one starts from a trial contour, which can be placed away from the edges. Under the influence of the internal and external forces, the trial contour will move toward the edges. When an equilibrium, or convergence, is reached, $\mathbf{v}_t = \mathbf{v}_{t-\gamma}$. Then the solution for Equation (4) is exactly the same as that for Equation (3).

III. Forces and Motion of the Contour

The contour moves toward the edges during each iteration of Equation (4) under the influence of all forces which are derived from the energy functional, Equation (1). A convergence is reached when equilibrium is reached, *i.e.* the contour lies in such a position where the net local force is zero. In order to understand the motion of the contour, we need to examine these forces. The most important term in image energy functional is the gradient term, $E_{img} = -\delta |\nabla I|^2$, where I is the intensity of the image. As can be seen the image force is very weak except for regions near the edges in the image. If the contour is initially located far away from the edges, one would expect a very slow convergence, if any. The purpose of the artificial force is to drive the contour toward the edges in regions where the image force is weak. The internal force prevent the contour from forming sharp corners. It is most important to balance all these individual forces because the local minimum is where the net force vanishes. All forces must be of the same order of magnitude, with the image force being the strongest for obvious reason. So α, β, γ, δ, etc. must be adjusted accordingly.

The distance the contour nodes move in one iteration, measured in pixels, depends on the strength of the force at that location. For this reason, we should normalize all the forces in terms of pixel unit. In order for the contour not to overstep the edges, the maximum image force must produce a step less than the width of the edges. However, the narrower the edge, the stronger the image force. Some artificial edges, such as frame border, can be extremely sharp. The *normalization* of the image force for this type of edges must be done separately by clamping it to a preset maximum. The artificial forces should also have magnitudes less than the width of the edges.

Since only the final contour is important, it is not necessary to enforce all the restrictions in the intermediate stage. Thus one can turn on and off some forces at different stages. In our method, we divide the whole procedure into three stages: the search stage, the filling stage and the fitting stage. The second stage may not be necessary for some applications which do not have deep features in the image and can be skipped. For the sake of simplicity, we assume that the contour is approaching the edges of the object from outside, *i.e.* the contour will shrink to adhere to the edges of the object. However, the same discussion applies to an expanding contour.

1. The search stage. Starting with a closed contour placed outside the object, we want the contour to approach the edges as quickly as possible. This is done by introducing a compressive force of magnitude between 0.75 and 1.0 pixel with the center of gravity located roughly at the object. The exact location of this center is unimportant. This compressive force will be locally turned off where the image force becomes significant, say 0.5 pixel. The image force is normalized to be about 2.0 pixels, while the width of the sharpest region is about 3.0 pixels. Note that the width of an edge is defined here as the region with significant image force. Edge widths thus defined are broader than that directly defined by intensity. This ensures that the contour will stay on the edges once it reaches there because there is no force strong enough to pull it over the edge. At this stage, a smooth contour is not necessary. Thus we can exclude the internal force by setting the $\mathcal{A}$ matrix to be zero, eliminating very time consuming matrix operations for a very clean image.

2. The filling stage. At the end of the search stage, most parts of the contour are on the object edge. There may exist some deep regions which are difficult to reach, or whose edges lie perpendicularly to the direction of the compressive force. This is especially true for irregularly shaped objects in the brain, such as where the grey matter borders a deep sulcus. In this case, we need to introduce an artificial right (or left) filling force which is perpendicular to the contour itself and toward the right (left) side of the contour, Figure 1. This normal force can be expressed as

$$F_x = \frac{F_0(y_{i+1} - y_{i-1})}{\sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2}}, F_y = -\frac{F_0(x_{i+1} - x_{i-1})}{\sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2}} \quad (5)$$

The magnitude of this force, F_0 is the same as that of the compressive force. At this stage, a smooth curve is necessary, since we need to know the direction of the contour at every nodes. Turning on the internal force will ensure the smoothness of the contour.

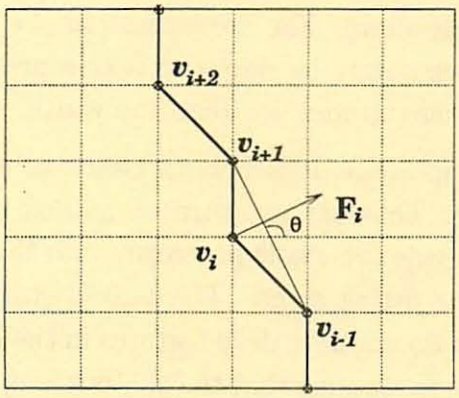


Figure 1. The artificial right-normal force on the ith node of the contour. The direction of this force is perpendicular to the line connecting the two adjacent nodes.

3. The fitting stage. When all the nodes of the contour are on the edges of the object, we need to fine tune the contour to get an pixel accurate segmentation. At this time, we turn off all the forces except for the internal force and the image force. Since the contour is already on the edges, only a few iterations are sufficient. When introducing the internal force, we need to determine the values of parameters α and β. In Figure 2, we show a portion of the contour with a sharp corner. In this case, $|\mathbf{v}_i''| = \sqrt{2}$, $|\mathbf{v}_i''''| = 2\sqrt{2}$. If such a corner is not allowed, the internal force at the corner must be bigger than the image force at that point. Since the maximum image force is, say, 2.0 pixels, we have

$$\alpha \geq \sqrt{2}, \qquad \beta \geq 1/\sqrt{2}$$

In practical applications, we would use a higher stiffness with $\alpha \approx 3$ and $\beta \approx 2$.

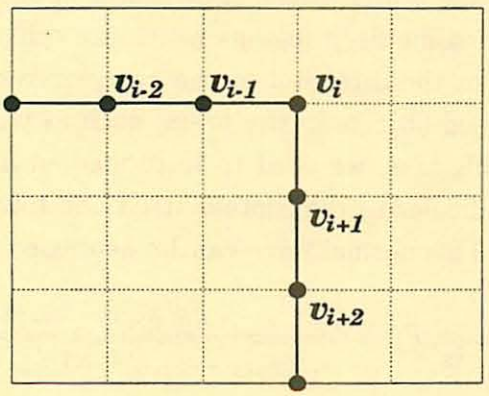


Figure 2. Part of the contour with a rectangular corner. The solid circles are the nodes of the contour. The direction of the internal force is toward the lower left corner.

As Amini *et al* [4] had pointed out, the nodes of the contour tend to either separate or congregate during the iterations. This problem can be eliminated by rearranging the nodes after a number of iterations. Nodes must be fused when they congregate, and more nodes inserted when they move apart. It is natural to allow the number of nodes to change because the final contour is expectedly very different from the initial one.

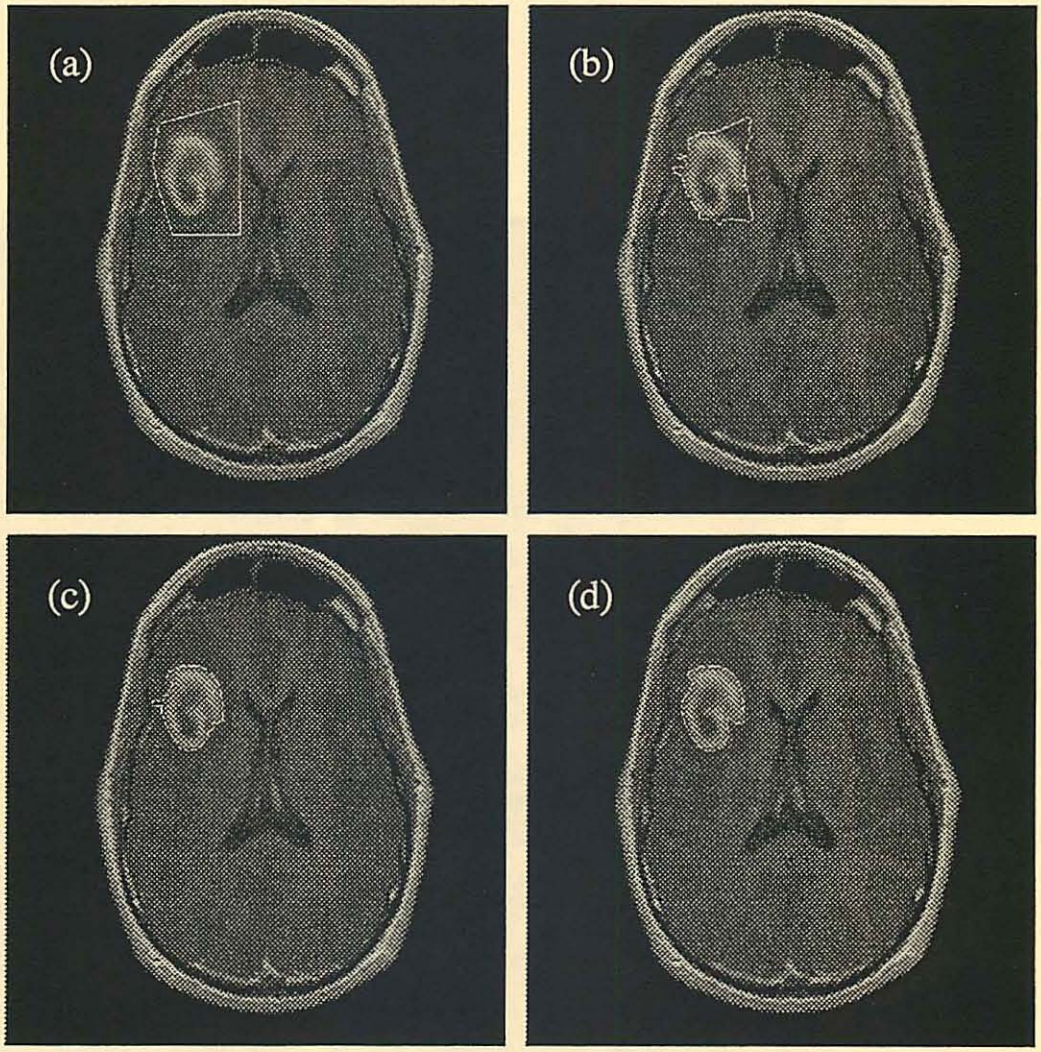


Figure 3. Segmentation of brain tumor. a) The initial contour placed around the tumor; b) intermediate contour caught by many noisy spots; c) the contour pulling over these noisy spots; d) the final contour.

IV. Results and Discussions

Brain Tumor

Tumor segmentation is important for surgical planning. A tumor is usually embedded in other tissues making the surrounding area very noisy. In addition, a tumor

consists mostly of soft tissues with relatively diffused edges. These conditions make segmentation very difficult. In Figure 3, we show an MR image of a brain containing a tumor. The initial contour is placed around the tumor (Figure 3a). As can be seen from Figures 3b and 3c, many noise spots obstruct the smoothness of the contour. It finally climbs over all of the noise spots and adheres to the edge of the tumor in Figure 3d. The number of nodes in the contour is about 300. Because of its small size of the contour, it takes only about one second to complete the whole process on a Sun SPARC 10 workstation. The difference between including and excluding internal force is insignificant.

Heart Chamber

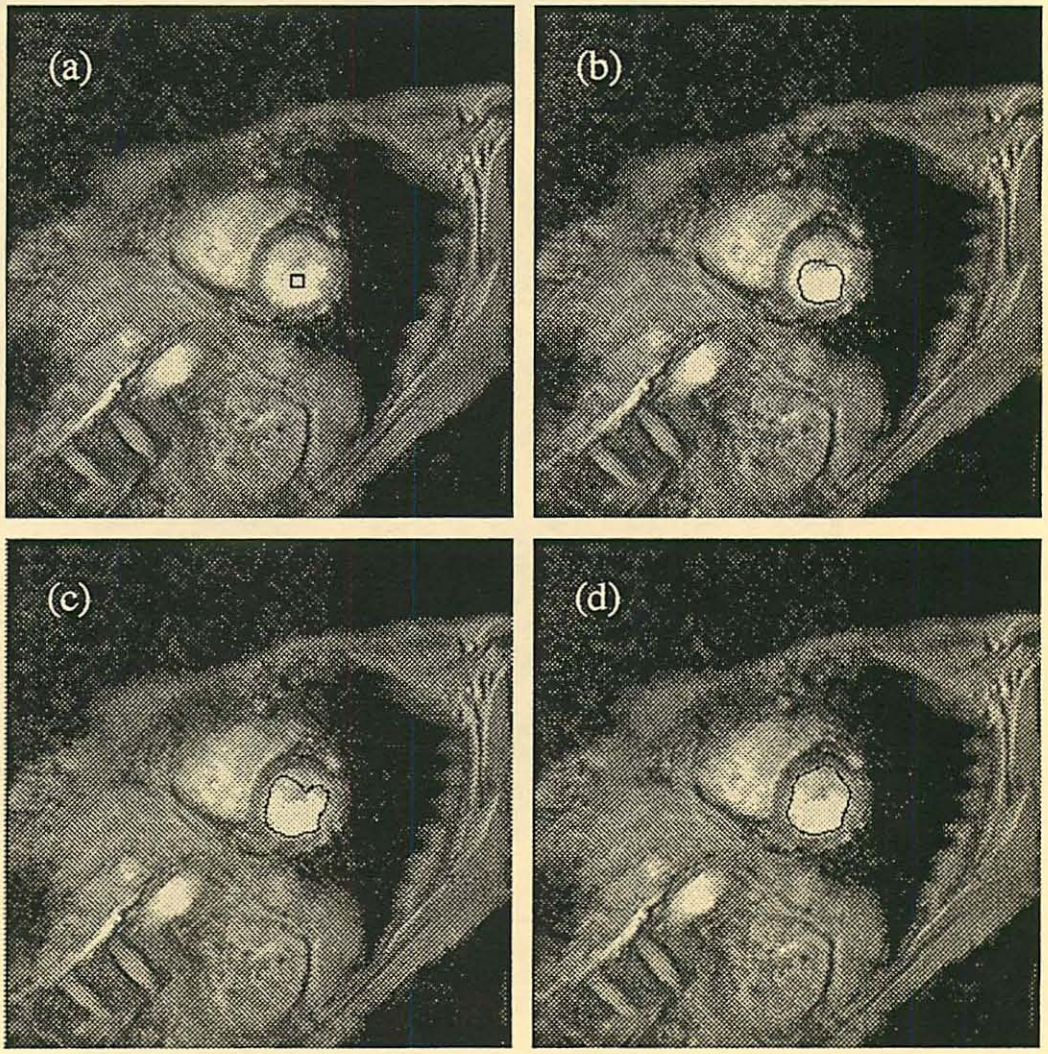


Figure 4. Expanding contour for heart chamber segmentation. a) The initial contour is placed inside the heart (the square); b) the contour expands to move toward the edges of the heart; c) the final state of the contour when it converges.

There are situations where it is desirable to place the initial contour inside the object of interest. Figure 4 shows an example of segmentation of the heart chamber in an MR image. The initial contour is first placed inside the heart chamber. An expanding force moves the contour toward the inner wall of the heart chamber. The size of the contour grows from 8 to about 300 nodes in this case. It takes less than one second to complete the segmentation.

Brain Image

Structures such as the white/grey matter edges of the brain are of high clinical interest. These structures are usually packed closely, making their segmentation very difficult. The active contour can be mistakingly attached to interfering objects near by. Furthermore, many regions in the brain are difficult for the contour to reach. For this type of problems, some preprocessing of the image is necessary. Figure 5a shows an MR image of the brain. Our goal is to segment the left cerebral hemisphere, separating it from the right cerebral hemisphere and the cerebellum.

In order to allow the contour to move close to the edges of the left cerebral hemisphere without being attracted to the other objects, we start with a binary image with portions of neighboring structures manually erased. Manual erasing can be done coarsely, making the whole contour extraction process much easier than precise tracing of the edge points by hand. The modified image is shown in Figure 5b. Binarization is applied at this point to define unambiguous structure edges. Since the sole purpose is to allow the contour to migrate close to the real edge of the left cerebral hemisphere, the binary edge does not have to exactly coincide with the edge of the object. An initial contour is placed encircling the edges of the left cerebral hemisphere (Figure 5c). Because there are many deep regions which are difficult to reach, a normal force is necessary to drive the contour into these regions. At the end of the filling stage, the contour is very close to the real edges of the left cerebral hemisphere (Figure 5d). Forces calculated from the original grey scale are used only at the last stage (fitting stage). The contour completes its evolution and converges to the the real edge of the left cerebral hemisphere Figures 5e).

In all of the above illustrations, we have chosen the node separation to be one pixel size during the entire iteration cycle for the sake of graphical esthetics. This results in a bigger matrix to be inverted, hence more computation time. In practical applications, one can subsample the contour and use larger node separation until the very last stage to reduce computation time.

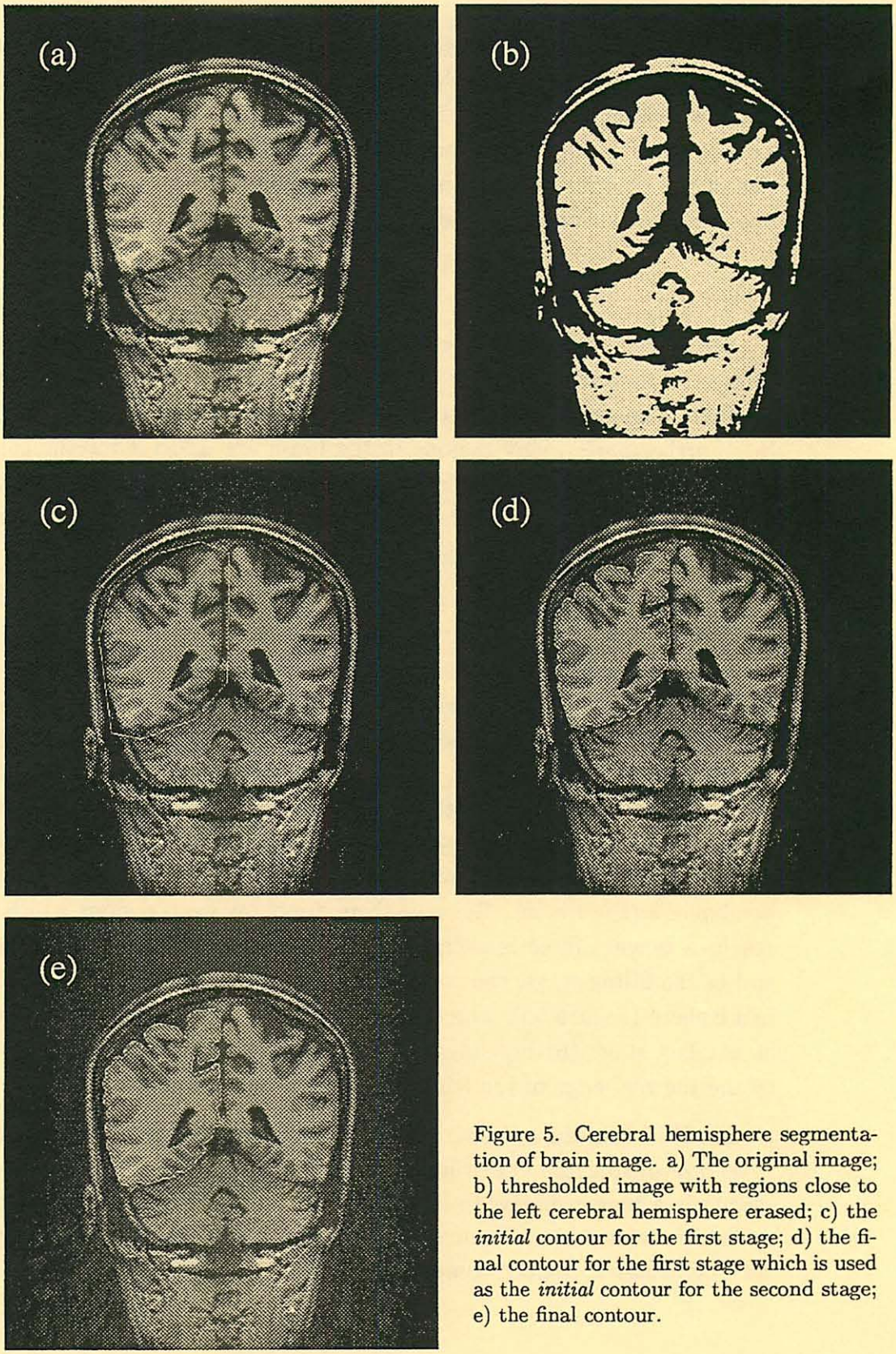


Figure 5. Cerebral hemisphere segmentation of brain image. a) The original image; b) thresholded image with regions close to the left cerebral hemisphere erased; c) the *initial* contour for the first stage; d) the final contour for the first stage which is used as the *initial* contour for the second stage; e) the final contour.

V. Conclusion

In this work, we have developed an active contour method for image segmentation based on the original model by Kass *et al* [1]. This method is significantly faster than the methods previously developed. This is mainly because of judicious selection of different forces at different stages of contour evolution, normalization of force intensity, and the elimination or size reduction of matrix inversion. With the introduction of a normal force, the contour will find edges which are normally difficult to reach. The contour is well behaved during the entire evolution process since all the forces are optimally normalized and balanced. By preprocessing some of the image using thresholding and erasing the surrounding structures, the contour is guaranteed to converge to the object of interest, even under complex situations. Finally, we also demonstrated the robustness of this method in dealing with image noise by balancing internal and external forces.

VI. Acknowledgment

The authors would like to thank Dr. Henry Shih for insights into clinical applications of this method.

References

[1] M Kass, A Witkin and D Terzopoulos, "Snake: Active Contour Models," *Int. J. Comput. Vision*, Vol. 1, pp. 321-331, 1987

[2] Laurent D Cohen, "On Active Contour Models and Balloons," *Comput. Vision, Graphics, and Image Processing: Image Understanding,* Vol. 53, pp. 211-218, 1991

[3] Laurent D Cohen and Isaac Cohen, "Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 15, pp. 1131-1147, 1993

[4] A A Amini, S Tehrani, and T E Weymouth, "Using Dynamic Programming for Minimizing the Energy of Active Contours in the Presence of Hard Constraints," *Proceedings, Second International Conference on Computer Vision,* pp. 95-99, 1988

[5] D J Williams and M Shah, "A Fast Algorithm for Active Contours and Curvature Estimation," *Comput. Vision, Graphics, and Image Processing: Image Understanding,* Vol. 55, pp. 14-26, 1992

[6] J Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 8, pp. 679-698, 1986

A Multi-client Global Teleradiology System

Bruce Ho, Woodrew Chao. Johnny Chao, Reza Sadri*,
Richard Steckel, Hooshang Kangarloo

AR-264 Center for Health Sciences
Department of Radiological Sciences
*Department of Computer Science
University of California, Los Angeles

ABSTRACT

A large scale global teleradiology project is underway linking multiple international imaging centers to the UCLA Department of Radiology. The goal is to deliver subspecialty consultation to patients in these remote areas. Technical issues in planning to establish the necessary teleradiology infrastructure include wide area network design, image compression, distributed archiving, and special viewing station features. Concepts such as teleconsultation and remote procedure monitoring are aimed at providing the same level of services at distant sites as that would be available in-house.

INTRODUCTION

A practical, clinically acceptable and scaleable teleradiology system that services a radiology subspecialty practice functions as an addition to of a department-wide PACS. Much more is involved than extending a local area network (LAN) to a wide area network (WAN). The major issues to implement a large-scale teleradiology system include distributed architecture, telecommunications, network management, strategies for distributed archiving, and efficient viewing station designs. The potential for communication bottlenecks is more severe by about an order of magnitude or more (e.g. Ethernet has a nominal signal rate of 10 Mbit/s, whereas the T1 leased line has a signal rate of 1.5 Mbit/sec.). This necessitates greater emphasis on image compression and system automation (e.g. image prefetching). Expansion of a teleradiology practice, with the incorporation of new sites over time, dictates the need for scaleable system. The geographic separation between the subspecialized radiologist and the referring physician or general radiologist in a teleradiology partnership motivates the use of innovative teleconsultation techniques. Naturally, the engineering challenges of maintaining reliable operation of a global network are daunting compared to those of a local network.

The planned teleradiology system coverage is depicted in Figure 1, showing the global

extent of the project. Latin American and Asian Pacific sites are scheduled to follow the successful demonstration of the first site in Melbourne, Florida, which is intended to demonstrate both the technical and financial viability of teleradiology as a routine clinical practice [1].

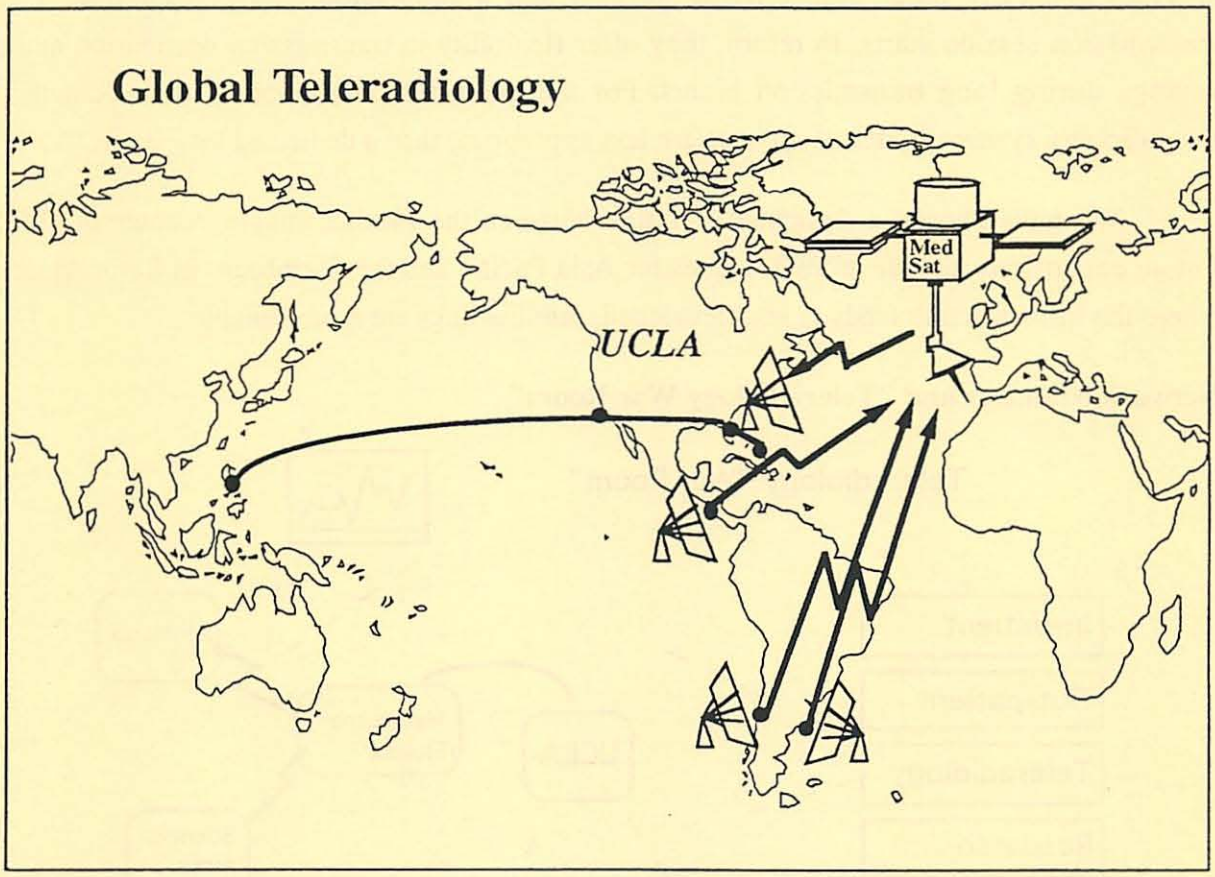


Figure 1 Global view of the UCLA multi-client teleradiology project.

TECHNICAL CONSIDERATIONS

The technologies needed in this infrastructure include many PACS components but further emphasis must be placed on overcoming the bandwidth limitations of a wide area network. The designs of viewing station interface and archive servers, for example, are centered around innovative image encoding techniques such as compression and hierarchical archiving.

Wide Area Linkage

Many leased line options exists [2, 3] for telecommunication links inside the U.S.. including 56/64 kbps line, DS-1 lines (1.5 Mbit/sec), and Fractional T-1 lines (e.g. 384 kbps, 768 kbps, Nx56). Higher bandwidth services include SMDS (Switched Multi-megabit Digital Services), ATM (Asynchronous Transfer Mode) (155 Mbps or higher), DS-3 (44.7 Mbps) and SONET. These services were not considered either due to their limited availability or excessive

cost (3).

Available switch services include switched 56, switched 384, ISDN, etc. [2]. Switched services generally incur higher hardware cost and a connection charge each time a new transmission session starts. In return, they offer flexibility in transmission destination and fee savings during long transmission hiatus. For the dedicated, continuously used link in our teleradiology system, switched services are less appropriate than a dedicated leased line.

We have chosen a dedicated T-1 link between the Florida imaging center to UCLA. Future expansions include more T-1 lines for Asia Pacific and the Caribbean. In Latin America, where the infrastructure tends to less developed, satellite links are more suitable.

Network Manager and "Teleradiology War Room"

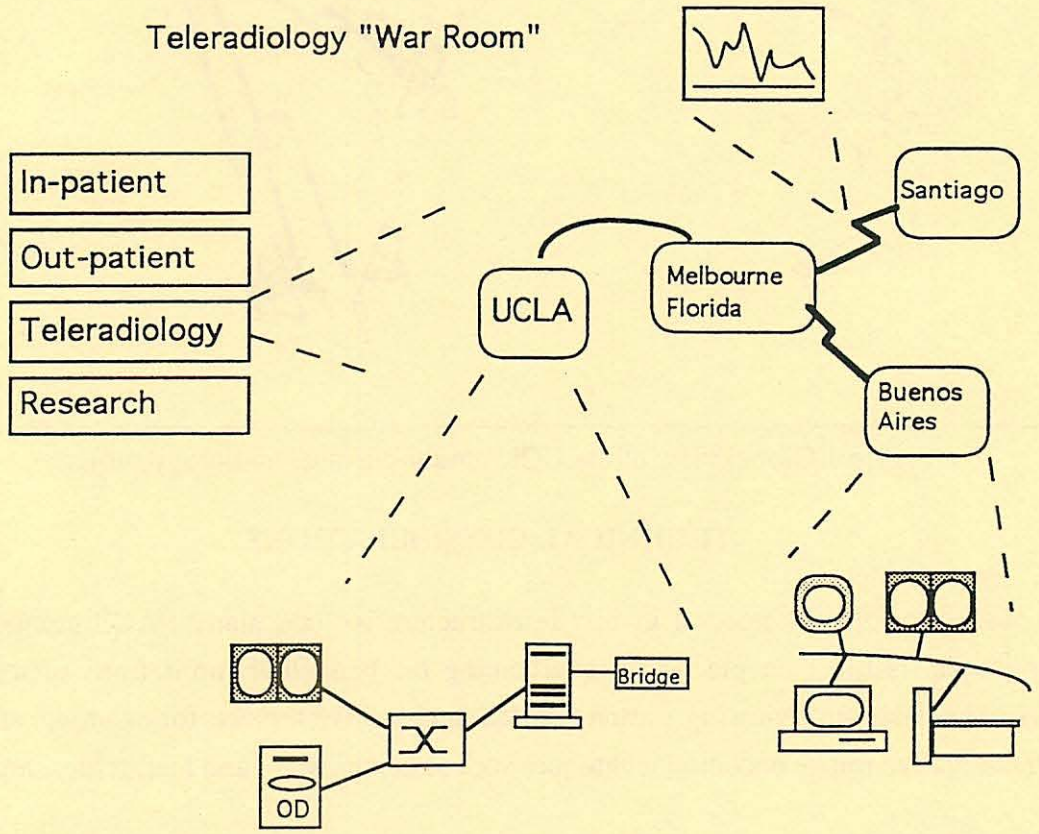


Figure 2 Teleradiology "war room" display allows hierarchical view of the entire system for centralized monitoring and control.

A large-scale international teleradiology system will require comprehensive, automated and effective network management capabilities. We are developing network management capabilities using SNMP (Simple Network Management Protocol) based network management software packages to provide instantaneous graphical views of the system components in various

levels of abstraction. We will develop application-specific management tools [4] customized for our teleradiology system, including graphical illustration of network trends and statistics, dynamic display of operating conditions, an expert system to respond automatically to "alert" messages, and recording of network problems into a trouble tracking system.

The central control room or "war room" concept [5] for our wide area teleradiology practice is illustrated in Figure 2. The graphical representation of system components and their moment-to-moment status can be displayed on a large screen for real-time monitoring. Teleradiology operation could involve minute-by-minute administrative decisions as the complexity and operating expense of the system increase. "War room" operators who can function as "resource brokers" and "traffic cops" are necessary to ensure a smooth work flow. For example, urgent image requests can be prioritized and requests for images can be redirected to other archive sites to avoid network congestion.

Diagnostic Viewing station

One of the goals of this project is to develop a teleradiology viewing station which is flexible enough for all of the radiologic subspecialties and which also incorporates special features needed for remote consultation. This viewing station has been designed to accomplish maximum viewing efficiency, utilizing high-speed hardware and an optimal user interface [6]. There are three types of user interfaces available [7]: 1) text based interface where images are accessed by clicking on a patient directory; 2) moving strip paradigm allows random access case selection by clicking on photo-icons (2); and 3) virtual panel which emulates reading from a lightbox based on a default or pre-sorted worklist.

The moving strip paradigm provides a miniaturized iconic view of the entire image content of a patient folder. The icons are arranged with one (CT or MR) study per row, and rows are sorted by modality and examination time (Figure 3). The viewer can use a dial to move the icon strips horizontally throughout the entire row, or across the rows vertically throughout the patient folder. When the image of interest is identified, clicking the mouse over the representative icon will bring up the image in full resolution on another monitor. This efficient browsing method is especially important for the teleradiology environment which might have a severe bottleneck between the viewing station and the archival site. The icons are encoded from the original images using embedded zero-tree wavelet transform (EZTWT) coding [8], with a typical compression ratio of 100:1. Based on visual cues of these iconic indices a radiologist can retrieve selected images in full resolution. In addition, these icon strips make a convenient tool for a "film" librarian to "pre-hang" the worklist in an orderly sequence preferred by the

individual radiologist. This step is analogous to hanging films on alternators in a film based practice. This in turn sets up the virtual panel mode, which emulates reading from an alternator.

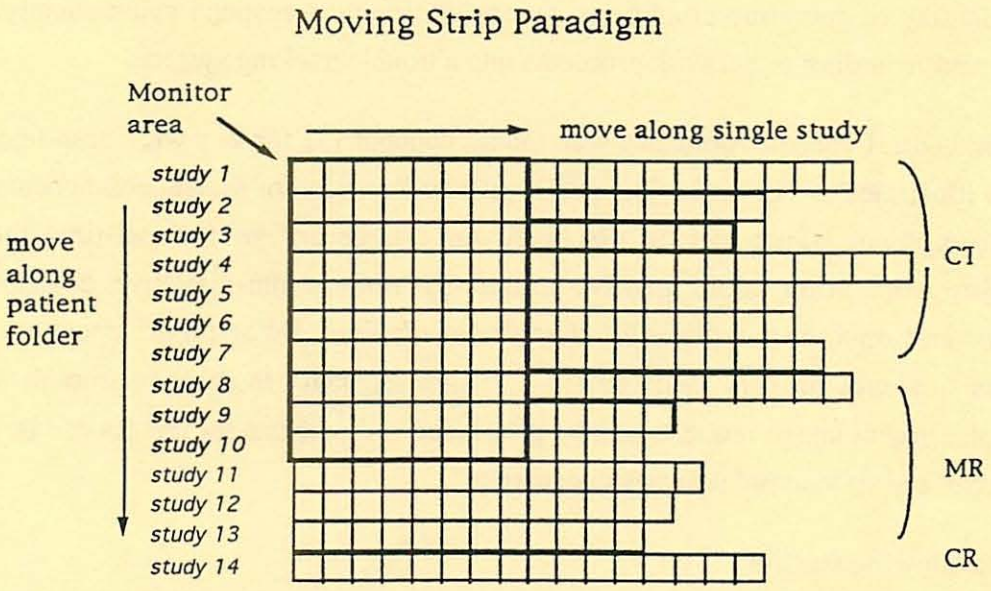


Figure 3 The moving strip paradigm allows a radiologist to browse quickly through the entire patient folder without transmitting complete image files across the WAN.

The virtual panel aims to alleviate the loss of screen space when a PACS viewing station replaces the film alternator. This potential shortcoming can be overcome by efficient methods of image navigation [7]. Using knowledge-based rules backed up by icon-directed "pre-hanging", a worklist can be generated in which new examinations and selected historical images are listed in an order dictated by each radiologist's reading habits. With the order pre-defined, the radiologist can advance through the list with a simple forward command that could be implemented with a foot pedal similar to a motorized alternator. A schematic diagram of the virtual panel concept is shown in Figure 4.

Special teleconsultation features will be added to the current display stations, including synchronized cursors, annotation, image layout, and image processing [8]. Window events which represent these activities can be captured locally and transmitted to the remote site with subsecond delay. Cursors on the screens at the sending and receiving sites can be displayed to allow both parties to initiate display functions. These commands require a very low bit-rate transmission and will use only a small fraction of the T-1 line capacity. Finally, menu-activated automatic dialing and a speaker phone will be provided with the viewing stations to facilitate discussion during the teleconsultation sessions.

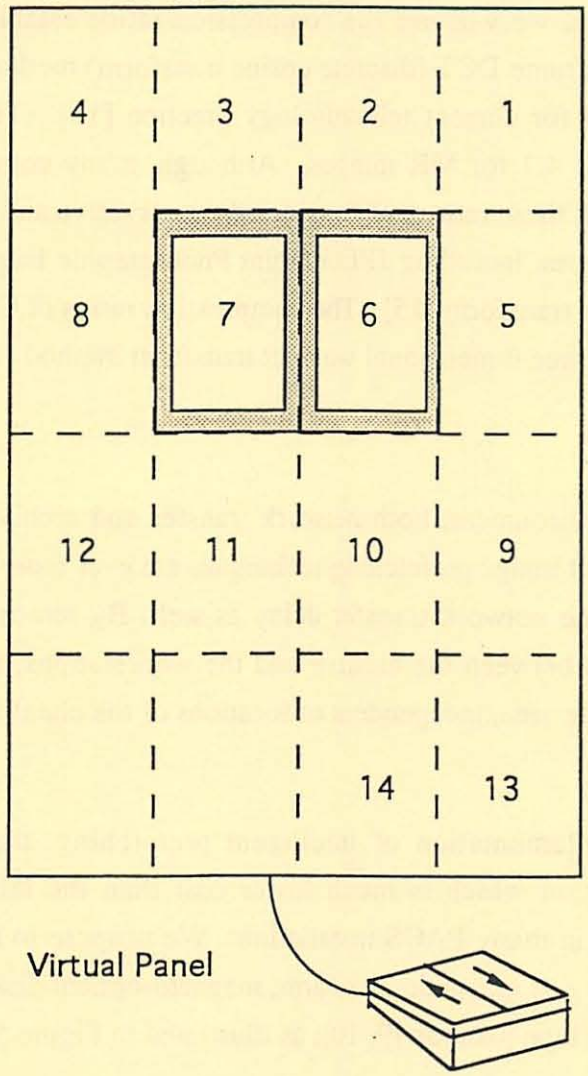


Figure 4 A virtual panel has images arranged in predefined order. When the foot pedal is depressed to move upward or downward, additional images move into viewing positions replacing currently displayed images. If all four screens of a virtual panel row contain pre-hung studies, the mouse is used to move the viewable region left or right.

Compression

The advantages of image compression can not be more convincingly demonstrated than in the band-limited environment of teleradiology. Our strategy is to use reversible compression before primary diagnosis is made and switch to irreversible compression when transferring data to the tape archive for long term on-line storage [9, 10]. Where legal or clinical concerns exist. another copy of the image which is reversibly compressed can also be cheaply stored off-line [9]. The larger compression ratios achieved by irreversible methods mean substantial reductions in transmission delay.

For long term storage, we will use the compression ratios established in our previous ROC studies [11] of the full frame DCT (discrete cosine transform) method which agree closely with the standards accepted for current teleradiology practice [12]. The ratios are 20:1 for radiographs, 8:1 for CT, and 4:1 for MR images. Although, many compression methods are being used by various groups, these ratios are considered conservative and comparable in quality among the most prevalent types, including JPEG (Joint Photographic Experts' Group) [13], full frame DCT [14], and wavelet transform [15]. The compression ratios of CT and MR images can be further improved using a three dimensional wavelet transform method [16].

Distributed Archive

In a teleradiology environment, both network transfer and archive access are potential system bottlenecks. Intelligent image prefetching techniques are ever more valuable because they can now be used to overcome network transfer delay as well. By removing the need for high speed transmission channels between the archive and the workstations, the archive site(s) can now be distributed over a wide area, independent of locations of the client imaging center and the radiologist's reading site.

With successful implementation of intelligent prefetching, the slower tape device becomes an attractive solution which is much lower cost than the large optical disk (OD) jukeboxes commonly found in many PACS installations. We propose to implement a three-tier hierarchical archive consisting of magnetic disk farm, magneto-optical disk (MOD) jukebox, and Exabyte (Boulder, Colorado) tape jukebox [9, 10], as illustrated in Figure 5.

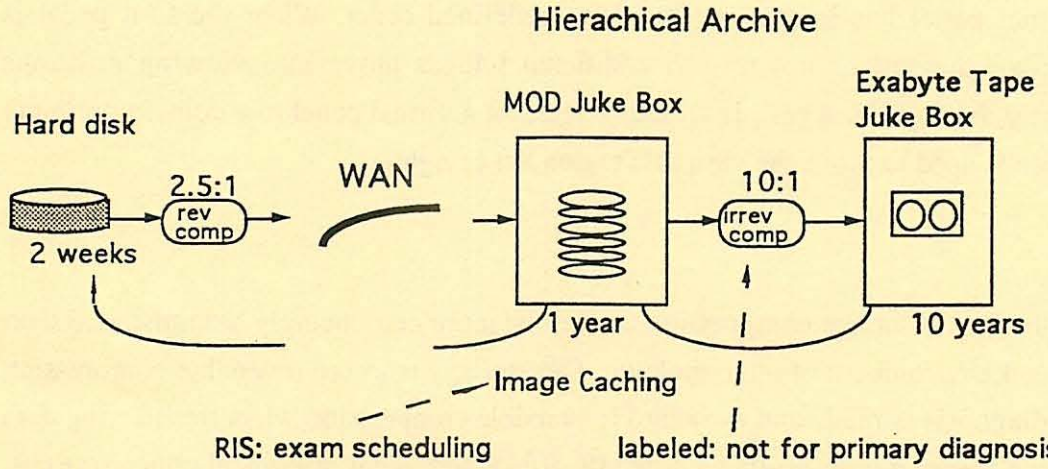


Figure 5 Three-level hierarchical archive with intelligent data migration. The magnetic disks farm holds images for approximately two weeks. MOD's store reversibly compressed images for about one year. Finally, tapes store irreversibly compressed images for ten years or more.

Using one 64 Kbits/sec channel of a DS-1 line, it takes approximately 5 seconds for one compressed (10:1) CT slice to be transmitted. Using zero-tree embedded DWT (discrete wavelet transform) encoding [19], the reconstructed image quality can be improved incrementally over the data transmission period, so that the UCLA radiologist can begin his/her monitoring activity without waiting the full 5 seconds.

CLINICAL OPERATIONS

Telemonitoring and Telediagnosis

Themonitoring, consultation, and diagnostic capabilities of the system will help UCLA radiologists to render routine teleradiology services with the same quality as in-house radiology services. Technologists performing examinations at the remote imaging centers will consult UCLA radiologists through the telemonitoring system. Our encoding scheme allows the radiologist to recognize and decide his/her interest level in a new image almost immediately. The radiologist can request immediate transmission of any image slice or display parameter by instructions over the phone.

The telediagnosis portion of the project will provide subspecialized diagnostic services for outpatients at the remote site. Preparation for a teleradiology reading session is identical to that for outpatients at UCLA. Pertinent prior studies of each patient are retrieved and transmitted along with current studies.

Once diagnostic reports are generated and transcribed, they are transmitted to the RIS and are computer-faxed to the referring physician. An important test in the initial implementation phase will be the turn-around time from a patient examination to delivery of a diagnostic report to the referring physician. This will be done by polling the queues at each stage of the imaging chain using the network manager monitoring mechanism [4, 5].

Real Time, Interactive Teleconsultation

Teleconsultation can be provided at two levels. For referring physicians who are not located at the remote imaging center, telephone consultation is provided. These physicians may obtain a hard copy of the study ahead of time. For referring physicians who can visit the remote imaging center, a complete interactive teleconsultation service will be possible using the available workstations provided at the imaging center. The on-site general radiologist can receive on-line consultation from UCLA subspecialists using the real-time interactive capabilities that include synchronized cursors, image layouts, image processing, annotation, etc. [8]

Fast and relatively expensive magnetic disks will hold urgent image data in uncompressed form for approximately two weeks. The magneto-optical disks provide intermediate storage of reversibly compressed images (about 2.5:1 ratio) for about one year. It is expected that retrieval of images less than one year old will be relatively frequent. The random access capability of MOD's is therefore advantageous relative to tape and justifies a higher storage cost. After one year, data on the MOD will be transferred to tape which costs only 0.2 cent/MB for storage medium. The MOD is reused for new data, decreasing supply costs.

An image clustering algorithm groups all studies belonging to one patient into a folder stored on a single tape. The tape jukebox can retrieve and load a new cartridge in 2.5 min., and fast-forward a tape from beginning to end in 1.5 min. Once the beginning of a file is found, the actual data transfer rate is 600 KB/sec, which is comparable to a MOD. Images on the tape are irreversibly compressed by a wavelet transform method at an averaged ratio of 10:1.

In the image caching process, requested data moves up the hierarchy from tape to MOD to the magnetic disk farm located at the reading site over the WAN. From experience with PACS, we realize that the best way to retrieve old data is through prefetching, triggered by examination scheduling information from the RIS [17]. The prefetching algorithm features intelligent patient folder composition that selects only the historical images most likely to be needed. Generally, these trigger events occur hours to one day before comparison images are needed for diagnostic reading, giving the prefetching mechanism enough lead time to overcome both WAN transfer and tape retrieval delay.

Remote Procedure Monitoring

UCLA's PACS system has a remote monitoring capability referred to as PACS Vision [18]. It captures the video signal from each scanner's console and transmits it via broadband for display on low resolution monitors in radiologists' offices. This allows the radiologist in charge to monitor and quality control each scanning procedure at various locations in the Department and pass detailed instructions to the technologist by phone while the patient is still on the scanner.

The broadband network used locally in the PACS Vision system generates a large amount of network traffic that is not suitable for wide area network implementation. For remote patient telemonitoring across the WAN, we will build a digital version of PACS Vision which reduces traffic load to a minimum through a hierarchical encoding scheme. A frame grabber installed in a PC will capture video from the scanner console. Although video capturing is continuous, only new image slices will be transmitted.

Telearchive and Teleaccess

In our teleradiology model, the client site is not required to maintain a local archive of patient image data, although there may be legal requirements that vary by jurisdiction [20]. An archival center or centers will provide many terabytes of storage capability. A dedicated remote archive may be the most economical means for on-line long term archival. Potential savings include the elimination of redundant installation, maintenance and personnel costs, and volume discounts on recording media. Compared to conventional film libraries, on-line storage is expected to decrease requirements for space and personnel as well as decrease the incidence of lost films.

In addition to serving individual client sites and their clinical needs, our distributed archive will collect valuable medical data (protected for confidentiality) from large population groups that are geographically dispersed. The demographic diversity and range of pathologies available through such a collection could facilitate clinical research on a much larger scale than heretofore possible. Implementation of this system will serve as a prototype for an international health data archive, where properly encoded and indexed global health information can be shared by research institutions for scholarly pursuits, including education and research.

SUMMARY

In summary, PACS design is evolving into a more mature field, with a number of commercially available solutions. Teleradiology is re-emerging as an area of focus in medical technology. Many of the design criteria for PACS are not directly applicable due to the severe bandwidth limitation of currently available WAN's. Throughout the architectural design phase, we have placed heavy emphasis on intelligent information use to reduce implementation cost and to minimize the need for expensive components that may become technologically obsolete even before completion. Teleradiology is much more than a research interest, as these systems offers the potential to increase access to subspecialists and decrease the cost of care [21].

REFERENCES

1. B.K.T.Ho, R.K.Taira, R.Steckel, H.Kangarloo, "Technical considerations in planning a distributed teleradiology system," Telemedicine Journal, March, 1995, (in press).

2. B.Quiat, G.A.Bolles, "WAN Services: You'd Better Shop Around", Network Computing, 4:68-80, 1993.

3. B.Quiat, "Technical Analysis: Traveling the WAN Maze...Twists and Turns in Every Trail", Network Computing, 4:83-92, 1993.

4. D.T.Chen, B.K.T.Ho, W.Chao, R.K.Taira, "Image flow management in a PACS network," PACS: Design and Evaluation, Proc. SPIE, 2165, Feb. 1994, Newport Beach, California.

5. D.T.Chen, B.K.T.Ho, W.Chao, R.K.Taira, H.Kangarloo, "PACS war room for global teleradiology," PACS Design and Evaluation, Proc. SPIE, this volume, San Diego, California, 1995.

6. B.K.T.Ho, O.Ratib, S.Horii, "PACS Workstation Design," Computerized Medical Imaging and Graphics, 15:147-155, 1991.

7. K.T.Leung, B.K.T.Ho, W.Chao, J.Chao, R.K.Panwar, V.Bhushan, Z.Barbaric, J.R.Bentsen, L.L.Seeger, "Image Naviagtion for PACS Workstations," PACS Design and Evaluation, Proc. SPIE, this volume, San Diego, California, 1995.

8. Z.M.Jiang, J.Chao, W.Chao, K.Chang, B.K.T.Ho, "teleconsultation via a wide area network, in real time and by electronic mail", PACS Design and Evaluation, Proc. SPIE, this volume, San Diego, California, 1995.

9. B.K.T.Ho, R.Sadre, W.Chao, L.Huang, R.Taira, "Strategy for long term PACS archive using hierarchical storage with image compression and HIS/RIS triggers," , J.Digital Imaging (in press).

10. W.Chao, B.K.T.Ho, J.Chao, R.Sadri, L.Huang, R.K.Taira,."implementation of system intelligence in a 3-tier PACS hierarchical storage management system", PACS Design and Evaluation, Proc. SPIE, this volume, San Diego, California, 1995.

11. D.R.Aberle, F.Gleeson, J.W.Sayre, K.Brown, P.Batra, D.A.Young, B.K.T.Ho,

B.K.Stewart, H.K.Huang, "The Effect of Irreversible Image Compression on Diagnostic Accuracy in Thoracic Imaging," Investigative Radiology, 28(5):398-403, 1993..

12. M.Cannavo, "Changes in Attitude, Changes in Lattitudes," Administrative Radiology, XIII(III): 55-61, 1994.

13. G.K.Wallace, "The JPEG still picture compression standard," Communication of the ACM, 34:30-44, 1991.

14. B.K.T.Ho, J.Chao, P.Zhu, H.K.Huang, "Design and Implementation of Full Frame Bit Allocation Image Compression Hardware Module," Radiology, 179:563-567, 1991.

15. P.Saipetch, B.K.T.Ho, R.Panwar, M.Ma, J.Wei, "Application of wavelet transform with arithmetic coding in radiological image compression," IEEE Eng Medicine and Biology (in press).

16. J.Wei, P.Saipetch, R.Panwar, D.Chen, B.K.T.Ho, "Volumetric image compression by 3D discrete wavelet transform", Image Capture, Formatting and Display, Proc. SPIE, this volume, San Diego, California, 1995.

17. H.K.Huang, R.K.Taira, "Infrastructure design of a Picture Archiving and Communication System," Am J Rad 158:743-749, 1992

18. H.K.Huang, R.H.Tecotzky, T.Bazzill, "A Fiber-optic Broadband CT/MR video communication system," J. Digital Imaging, 5(1):20-5, 1992.

19. J.M.Shapiro, "Image coding using the embedded zerotree wavelet algorithm," Proc. SPIE Conf. on Mathematical Imaging: Wavelet application in signal and image processing, Vol. 2034, July 1993.

20. American College of Radiology, "ACR Standard for Teleradiology, " 1994.

21. D.Forsberg, M.Sawchak, "Options in Teleradiology," Administrative Radiology, XIII(III): 50-54, 1994.

PACS War Room for Global Teleradiology

Doris T. Chen, Bruce K. T. Ho, Woodrew Chao, Ricky K. Taira, and Hooshang Kangarloo
Department of Radiological Sciences, Univ. of California, Los Angeles, CA 90024

ABSTRACT

UCLA is beginning teleradiology projects in Latin America, Asian Pacific, and the US. The UCLA teleradiology system communicates with remote imaging centers through a T1 based WAN and satellite technology. A network management center, i.e. PACS war room, with graphical user interface allows the system manager to monitor and control different elements of the system through various levels of abstraction from one location. The network manager software can monitor the activities of both hardware and software devices in the remote imaging centers, LAN and WAN performance, telearchiving and teleaccess pattern. Telearchiving can be monitored by graphically indicating large image movements between jukeboxes and over LAN or WAN. Once the teleradiology system is configured, the image flow pattern in a teleradiology center is predictable. Manual intervention of these teleradiology system functions is easily done through menu control in the war room display. A centralized network management with the global view of a teleradiology system has been developed. It can give commands to the elements of the system to tune the system for efficient utilization of the system resources. These managerial functions scale with the teleradiology system which is expandable to include many more potential remote imaging centers.

Key words: PACS war room, teleradiology, LAN, T1, WAN, network management, telecommunication, telearchive, remote imaging centers

1. INTRODUCTION

UCLA is starting nation-wide and international teleradiology projects in addition to continue developing its department-wide PACS. This global teleradiology is going to provide the services in Latin American, Asian Pacific and the US, in comparison UCLA's PACS is only a department-wide local area network(LAN) which spans three medical buildings about one kilometer apart.[1] System complexity arises in a large scale wide area teleradiology network, since much more geographical dispersed hardware and software elements are involved.

Challenges of operating such a large scale teleradiology are not only due to the complexity and the size of the system, but also due to the WAN's narrower bandwidth. A T1 based WAN is rated at 1.5 Mbits/sec, whereas a Ethernet is in rated at 10 Mbits/sec.[2] A dedicated T1 link WAN is almost 7 times slower than a regular Ethernet. Such a narrow bandwidth inevitably causes more system failure and bottlenecks in the remote sites. A resourceful network management is therefore urgently demanded.

A reliable teleradiology system and a smooth operation will greatly depend on how well the network can be managed. Once system problems occur in the remote sites, several questions arise. How soon would the operation personnel find out the problems? How could he/she notify the remote sites? How could the image centers solve the problem efficiently without the trained repair staffs actually physically go to every remote site? A centralized network management with the global view of a teleradiology system will be the tool to solve all the problems above.

We have developed a network management station for monitoring and controlling of the operation of a global scale teleradiology system under the concept of a PACS War Room. System conditions and network trends are displayed on a large panel, illustrated in figure 1, using color indicators and graphical menu which are easy to use for non-technical personnel. The large panel is

developed with hierarchical graphic representation of system elements including: software processes, hardware components, image file objects, patient record objects, and network links. The "War Room" concept allows the operator to visually monitor and control different elements of the entire system through various levels of abstraction from one central site.

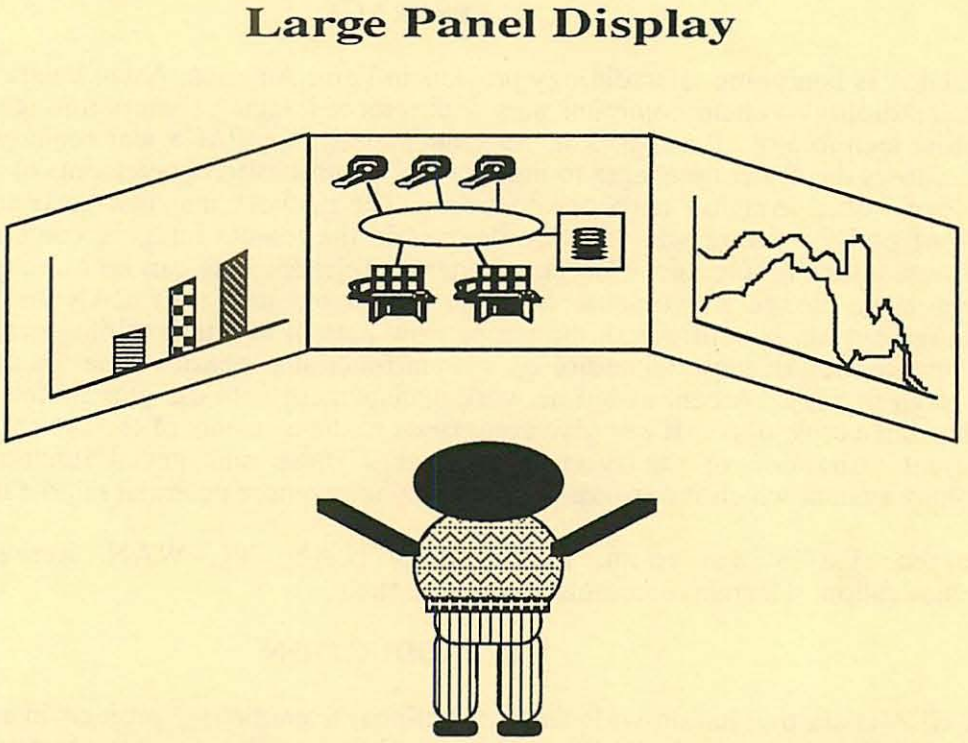


Figure 1. The operator can get hierarchical views of different aspects of the entire system at the same time.

Automation is another important feature employed in the PACS War Room. Automatic mechanisms are especially important for system repair by either calling the attention of the trained personnel or by built-in expert rules.[3] In advanced stages of development, well known system errors can be automatically corrected using rule based fault recovery mechanism. The PACS War Room applications could be applied to three categories of problems: 1) ensuring system reliability: indication of system failure in both hardware and software by color codes 2)WAN performance monitoring by suing a image flow model, and 3) fault recovery: start/stop processes by remote control and automatically paging operators. Many of these applications can become highly automated.[3]

The building tools for the war room design are an integration of three components: Harris Network Manager, IBM NetView 6000, and our own graphical tool running on Windows/NT.[4] The network management software uses Simple Network Management Protocol (SNMP).[5, 6] We have developed many specific agents for network management and front end displays based on both Harris Network Manager and IBM NetView 6000 as platforms. A network sniffer software from network general is used to collect the statistics for the entire network.[7]

The mechanisms behind the PACS War Room such as imaging chain modeling, polling mechanism, and fault recovery algorithm are discussed in detail in reference 3. In this paper, we will concentrate on presenting the War Room display for visual monitoring and central control.

2. WAR ROOM DISPLAY FOR VISUAL MONITORING

2.1 Network Topology, Hierarchical Objects and Architecture

The PACS WAR ROOM is fundamentally a display technique. It will show the map, topology, architecture, and hierarchical view of the entire network system. Color coding will be applied to the various hardware and software devices represented as icons to indicate their conditions or status. Generally, green means device is functional, yellow means marginally functional and operator should be aware, and red means failure and repair required by trained personnel or rules-based fault recovery. In the hierarchical objects construct, the error sign could propagate through many layers all the way up to the surface layer icon. The color coding scheme also propagates for intuitive display.

We developed many of the war room display techniques using the UCLA PACS system. The complete PACS network topology can be displayed on War Room panels. This has been illustrated previous articles.[1, 8, 9, 10, 11] UCLA's PACS is a large distributed network system which spans three buildings: Center for Health Sciences(CHS), MR remote site and Medical Plaza which are about one kilometer apart [1]. Three tiers of network are implemented in these three buildings: Ethernet (~10 Mbits/sec), FDDI (~100 Mbits/sec), and ultranet (~1 Gbits/sec). The global Ethernet links all PACS nodes. Capture computers for CR, CT and MR scanners transfer the images through both the FDDI and the Ethernet to the archive center. In the archive centers, two cluster controllers are connected to optical disk, juke boxes and sybase servers. For a fast image transfer, a high speed ultranet is implemented between the archive center and the 1K and 2K display stations. PACS consists of 13 acquisition computers, including 4 CT, 5 MR, 2 CR and 2 LS; 2 database servers, 2 archive centers; and 5 display stations.[1, 8, 9, 10, 11]

Similarly, a global teleradiology map can be configured with the geographical location of the remote sites in Latin American, Asian Pacific, and the US, as well as the UCLA campus. The teleradiology system communicates with remote imaging centers through a wide area network(WAN) based on T1, or satellites. A teleradiology map could be displayed as an overview of the global wide area network on one of the large panels in the PACS War Room.

Teleradiology War Room display could also illustrate the hierarchical objects of the entire system for centralized control and monitoring. The hierarchical view is depicted in figure 2. First layer is the four major divisions for the entire network system. They are in-patient, out-patient, research, and teleradiology. By click on the icon of teleradiology system, next layer under teleradiology pops out. By clicking on the icons one layer after the other, different hierarchical views can be displayed. The hierarchical structure of figure 2 is summarized in table 1. If an error occurred in one of the processes in a workstation, the error propagation indicated color coding will be triggered by that process and propagate from layer four to layer three workstation, and all the way up to the layer where the operator can monitor. Hence, whenever an error condition appears, the operator can be greatly benefit from this hierarchical graphic interface to locate the error quickly.

The network architecture overview is demonstrated in figure 3. It demonstrates the entire network architecture LAN and WAN in teleradiology. The network components involved are gateways, bridge routers, bridges, and routers. Gateways and bridge routers are implemented to connect LAN and WAN communications, and smaller bridges and routers are the connectors inside LAN of the imaging center. All the hierarchical property apply to this overview design as well.

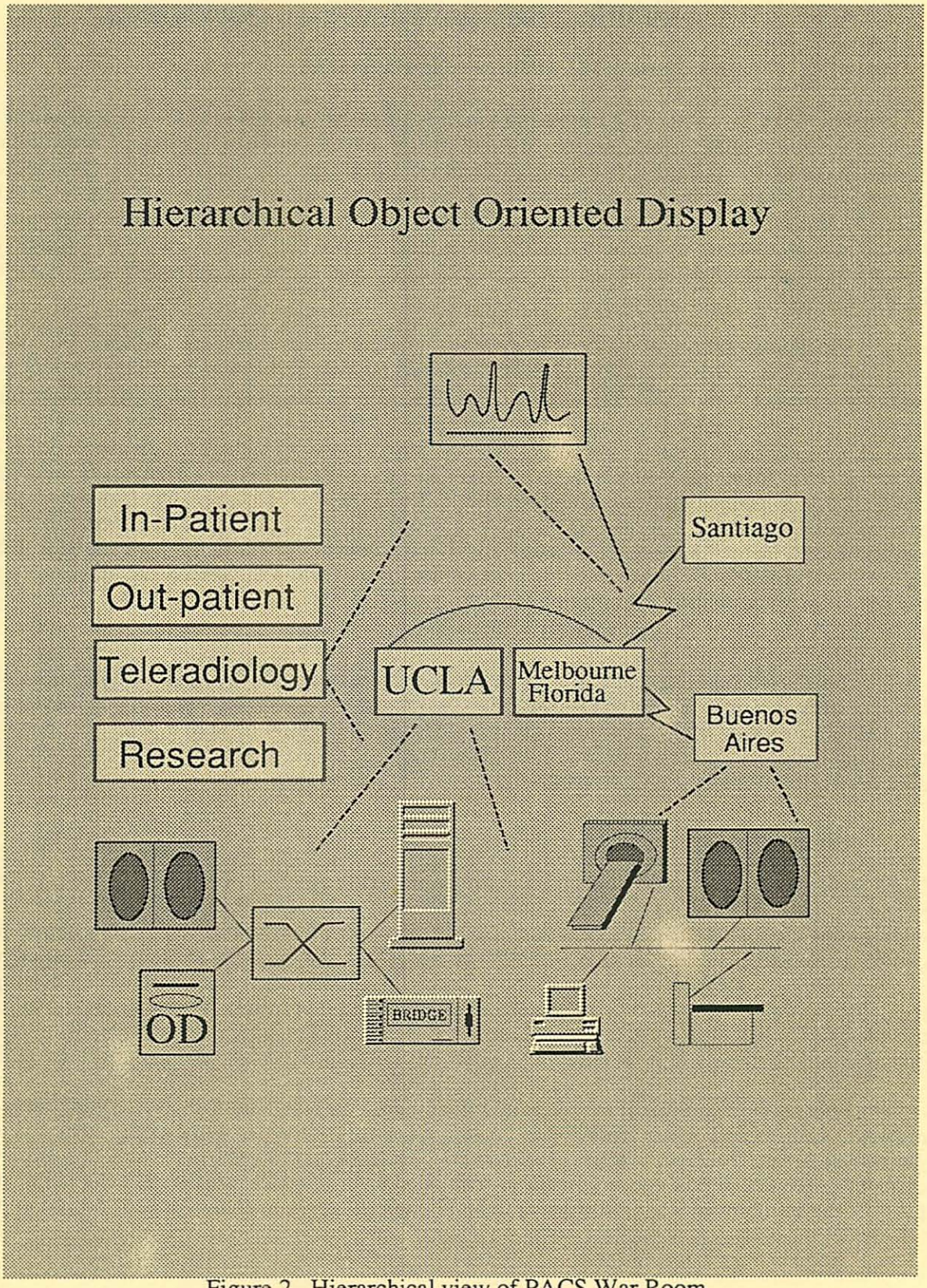


Figure 2. Hierarchical view of PACS War Room.

Layer Number	Content	One Specific Chosen View	Detail Description
one	four divisions	-	It shows the formation of the entire network system
two (teleradiology)	imaging centers, remote sites, and WAN between them T1 based	teleradiology	It shows that UCLA and Melbourne, Florida are imaging centers, and the two remote sites are Santiago and Buenos Aires. The network is T1 based WAN
three (UCLA)	imaging center: PACS architecture overview	UCLA	In PACS architecture, there are network components such as hubs, optical disk jukeboxes, display workstations, and connectors like bridges, and routers.
three (Buenos Aires)	remote site: overview of a typical remote site	Buenos Aires	An overview with components such as acquisition scanners, laser scanners, display workstations, and dedicated computers is presented.
three (WAN T1 line)	WAN performance	T1 based WAN	The traffic load of WAN performance between image centers and remote site of Santiago is presented.
four (network component)	monitoring of component, say workstation, with associated software & hardware objects (e.g. processes)	workstation	Click on any of the desired component, e.g. display workstations, and a list of the workstations in the network, and the associated software (processes) and hardware objects underneath them will be displayed [3]

Table 1. A summary of the hierarchical object oriented display

2.2 Dynamic Image Flow Display: Real Time Monitoring

We have demonstrated the capability of PACS War Room on monitoring entire the network topology , architecture, and their hierarchical property. It is also important to monitor the image flow and operation conditions with instantaneous graphical views. One War Room display of the dynamic image flow in real time is demonstrated in figure 4. Image flow throughout the chain can be visually monitored by polling the contents of various task queues.[3] We devised a graphic monitoring mechanism which shows each image file traveling from an acquisition node to an archive, and then to the destination workstations. In case of image retrieval, the same mechanism shows image flow from the archive to the requesting workstation. Color coding (in this black and white figure, color coding is represented by different patterns) identifies the queue where an image is going through processing in different stage. Dash and solid lines indicate tasks status of pending and executing. Other graphical symbols are used to indicate exceptional conditions such as retries and errors. All active images in the chain are plotted as a function of time and updated regularly. The display interval is adjustable to suit for the best viewing effect. Queue length plotted with a bar graph in any particular time slot shows the number of images awaiting service in that queue. This simple to see visual tool very effectively alerts the operator of any exceptional delays, error conditions, and

interruptions in image flows, which require immediate manual intervention. Manual intervention of these teleradiology system functions is easily done through menu control in the war room display.

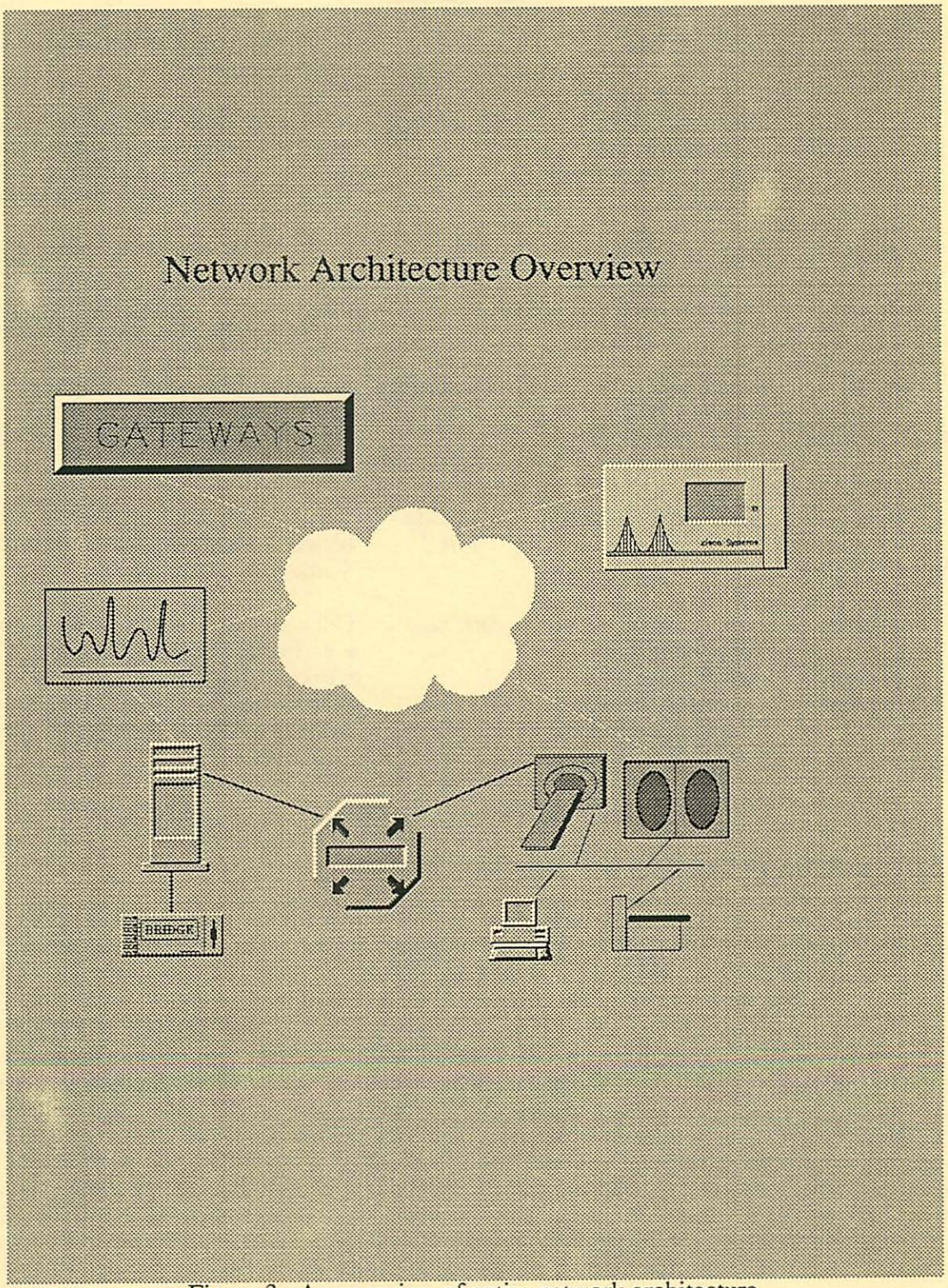


Figure 3. An overview of entire network architecture

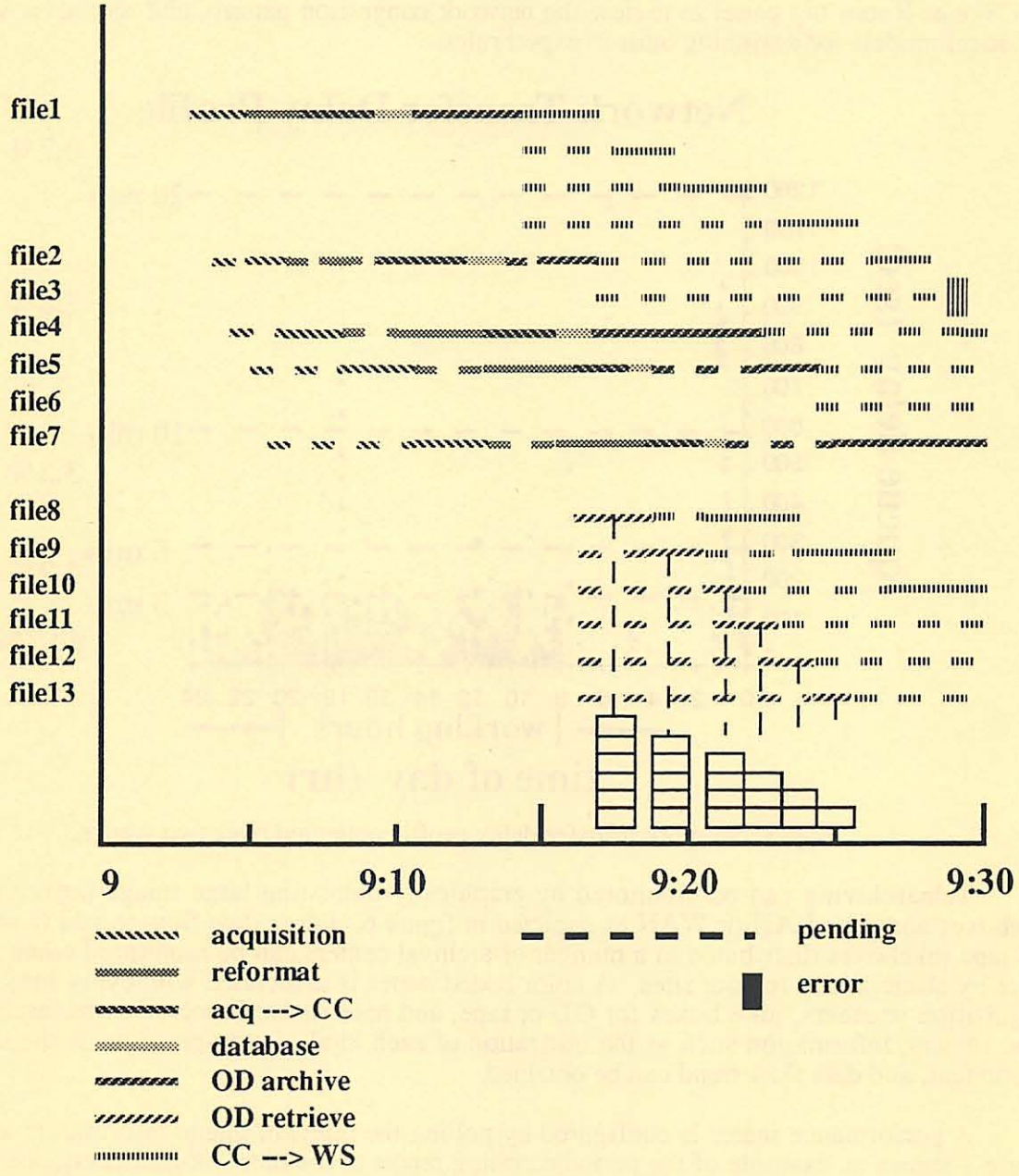


Figure 4. Dynamic image flow monitoring in real time

2.3 Network Activity Trends , Statistics and Queues

Real time monitoring dynamic image flow can help to ensure a smooth operation in the system. However solving or preventing the common errors and the failure conditions still rely on the collection of network activity trends and performance statistics. Next, we demonstrate some War Room applications on system statistics, monitoring data flow in telearchiving, and periodic queue report for cluster controllers.

Figure 5 illustrates network statistics over a period of two weeks. These statistics shows the network congestion pattern, and the percentage of delayed transfers. The detail description and analysis of the statistics is presented in reference 12. This activity trend figure can be shown on PACS War Room big panel to review the network congestion pattern, and moreover it can support statistical models for designing built-in expert rules.

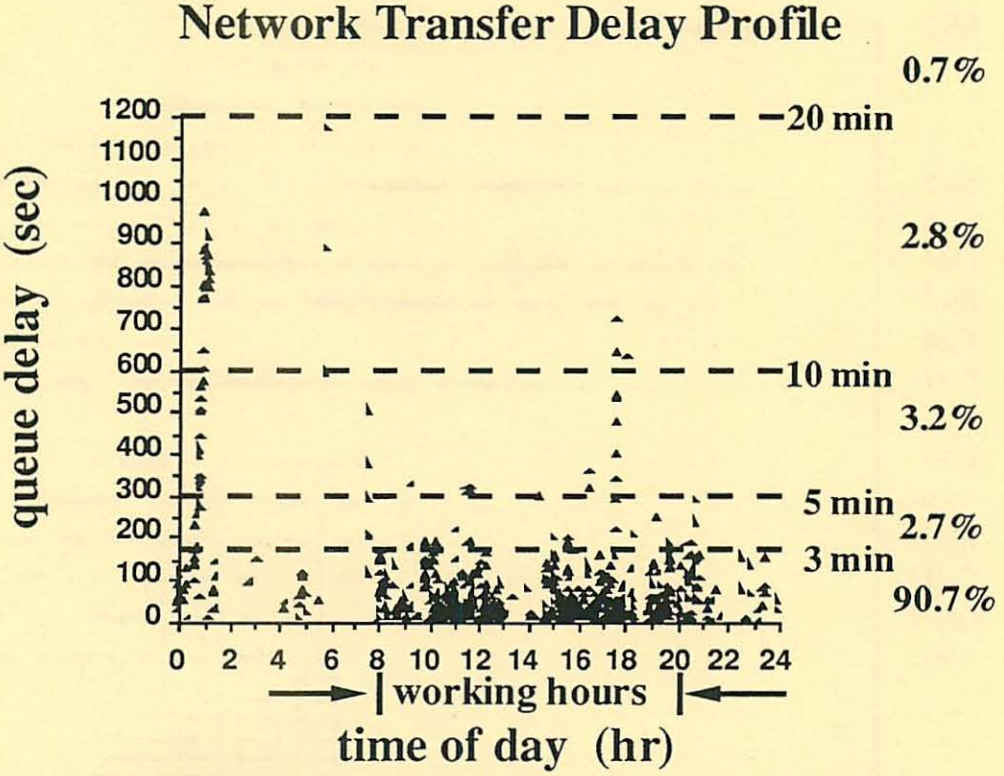


Fig. 5. Network transfer delay profile collected over two weeks.

Telearchiving can be monitored by graphically indicating large image movements between jukeboxes and over LAN or WAN as depicted in figure 6. Large data flow to and from optical disk and tape jukeboxes distributed in a number of archival centers can be monitored when retrievals are made by clinicians in remote sites. A color coded meter is associated with every major component (acquisition scanners, juke boxes for OD or tape, and hard disks) involved in telearchiving. From these meters, information such as the utilization of each kind of storage devices, the status of each component, and data flow trend can be obtained.

A performance meter is configured by polling the relevant queue information as in figure 7. Figure 7 shows an example of the periodic polling report of the network transfer queue in the cluster controller.[3] The report plots the three different job status: succeeded, pending, and error. The scales of severity are indicated by various colors or patterns. The first bar in the graph indicates normal condition. Second bar is the meter for job pending conditions, which generally describes bottleneck condition. An operator should be alert to the cases where the number of pending jobs exceed the normal range in the bottom section (in figure 7), and goes up to the marginal(second and third section from the bottom) and the dangerous range(the top section). If pending bar reaches the dangerous range, the operator should start to page the relevant personnel. The third bar also shows the error condition with different severity. The definition of different severity is similar to the pending bar with different numerical threshold values. This error bar indicates an alarm condition which requires immediately action such as error logging into a tracking system, manual recovery procedures,

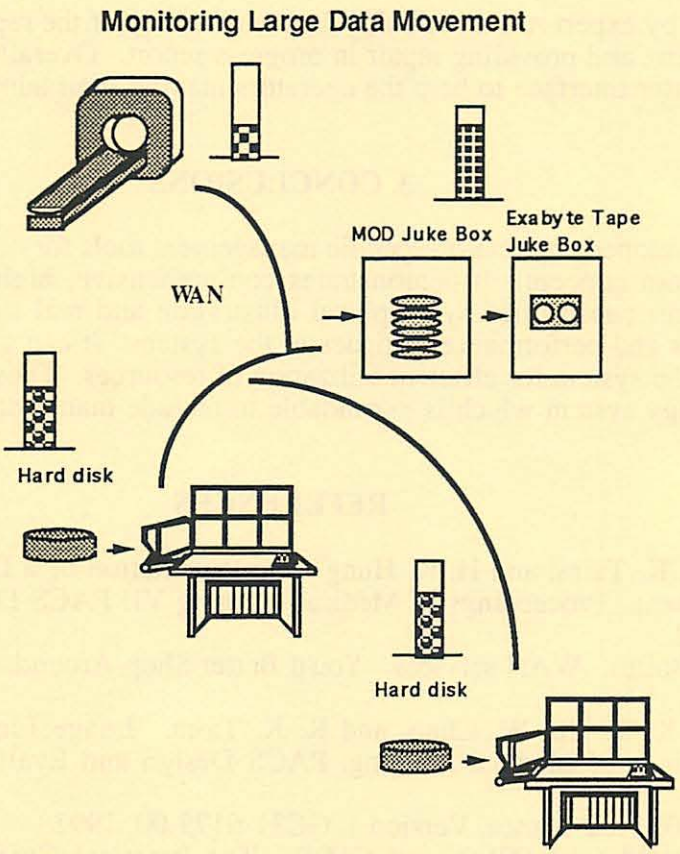


Figure 6. Monitoring large data movement in telearchiving

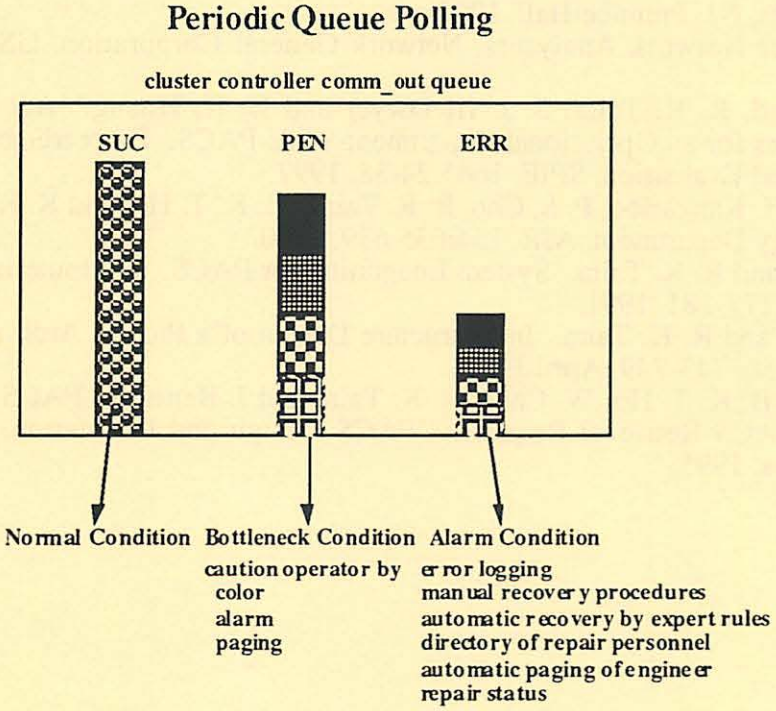


Figure 7. Periodic queue polling report for different job completion status

automatic recovery by expert rules, searching for the directory of the repair personnel and paging the responsible engineers, and providing repair in progress report. Overall, this set of designs provides friendly graphical user interface to help the operators making their administrative decisions quickly and efficiently.

3. CONCLUSIONS

We have developed application-specific management tools for our teleradiology system under the PACS War Room concept. It demonstrates comprehensive, highly automated, and effective network management capabilities by graphical illustration and real time monitoring of the image flow, activity trends and performance statistics of the system. It can give commands to the remote elements and tune the system for efficient utilization of resources. These managerial functions scale with the teleradiology system which is expandable to include many more potential remote imaging centers.

REFERENCES

1. A. W. Wong, R. K. Taira, and H. K. Hung. Implementation of a Digital Archive System for a Radiology Department. Proceedings of Medical Imaging VI: PACS Design and Evaluation, SPIE, 1645:182-190, 1992.
2. B. Quiat, G. A. Bolles. WAN services: You'd Better Shop Around. Network Computing, 4: 68-80, 1993.
3. D. T. Chen, B. K. T. Ho, W. Chao, and R. K. Taira. Image Flow Management in a PACS Network. Proceedings of Medical Imaging: PACS Design and Evaluation, SPIE, 2165:233-240, 1994.
4. AIX NetView/6000 At a Glance, Version 1, GC31-6175-00, 1992.
5. W. Stallings. SNMP, SNMPv2, and CMIP: The Practical Guide to Network-Management Standards. Reading, MA, Addison Wesley, 1993.
6. M. T. Rose . The Simple Book: An Introduction to Management TCP/IP - based internets. Englewood Cliffs, NJ, Prentice-Hall, 1991.
7. Expert Sniffer Network Analyzers, Network General Corporation, GSOOK92AGS6109 PS08, 1994.
8. B. K. Steward, R. K. Taira, S. J. III Dwyer and H. K. Huang. Acquisition and Analysis of Throughput Rates for an Operational Department-Wide PACS. Proceedings of Medical Imaging VI: PACS Design and Evaluation, SPIE, 1645:24-38, 1992.
9. H. K. Hung, H. Kangarloo, P. S. Cho, R. K. Taira, B. K. T. Ho, and K. K. Chan. Planning a Total Digital Radiology Department. AJR, 154:635-639, 1990.
10. K. K. Chan and R. K. Taira. System Integration for PACS. Computerized Medical Imaging and Graphics, 15(3):177-181, 1991.
11. H. K. Hung and R. K. Taira. Infrastructure Design of a Picture Archiving and Communication System. AJR, 158: 743-749, April 1992.
12. D. T. Chen, B. K. T. Ho, W. Chao, R. K. Taira and J. Bentson. PACS Intelligence: Automatic Feedback for PACS Retrieval Requests. PACS Design and Evaluation, SPIE, this volume, San Diego, California, 1995.

Preprint from SPIE: PACS Design and Evaluation 2435-02

Design of a Graphical User Interface for an Intelligent Multimedia Information System for Radiology Research

Ricky K. Taira, Clement Wong, David Johnson, Vikas Bhushan, Monica Rivera†, Lu Huang, Denise Aberle, Alfonso F. Cardenas*, W.W. Chu**

Department of Radiological Sciences
* Computer Science Department
† School of Medicine
University of California, Los Angeles

1.0 Introduction

With the increase in the volume and distribution of images and text available in PACS and medical electronic health-care environments, it becomes increasing important to maintain indexes that summarize the content of these multi-media documents. Such indices are necessary to quickly locate relevant patient cases for research, patient management, and teaching.

The goal of this project is to develop an intelligent document retrieval system that allows researchers to request for patient cases based on document content. Thus we wish to retrieve patient cases from electronic information archives that could include a combined specification of patient demographics, low level radiologic findings (size, shape, number), intermediate-level radiologic findings (e.g., atelectasis, infiltrates, etc.) and/or high-level pathology constraints (e.g., well-differentiated small cell carcinoma) [Cohen90]. The cases could be distributed among multiple heterogeneous databases such as PACS, RIS, and HIS. Content-based retrieval systems go beyond the capabilities of simple key-word or string-based retrieval matching systems. These systems requires a knowledge base to comprehend the generality / specificity of a concept (thus knowing the subclasses or related concepts to a given concept) and knowledge of the various string representations for each concept (i.e., synonyms, lexical variants, etc.).

Figure 1 shows the overall system architecture currently under development at our institution. We have previously reported on a data integration mediation layer that allows transparent access to multiple heterogeneous distributed medical databases (HIS, RIS, and PACS) [Taira94]. The data access layer of our architecture currently has limited query processing capabilities. Given a patient hospital identification number, the Access Mediation Layer collects all documents in RIS and HIS and returns this to the information to a specified workstation location. In this paper we report on our efforts to extend the query processing capabilities of the system by creation of custom query interfaces, an intelligent query processing engine, and a document-content index that can be generated automatically (i.e., no manual authoring or changes to the normal clinical protocols). An example query that the system targets to answer is the following:

"Find cases of Hispanic patient's with non-calcified masses seen on CT greater than 3cm in long axis diameter in the left lower lobe of the lung, that on pathology were small cell carcinoma".

Note that this query involves retrieval of cases whose information content is contained in HIS (patient demographics, and pathology histology reports), RIS (non-calcified masses greater than 3cm in the left lower lobe of the lung), and PACS (CT images).

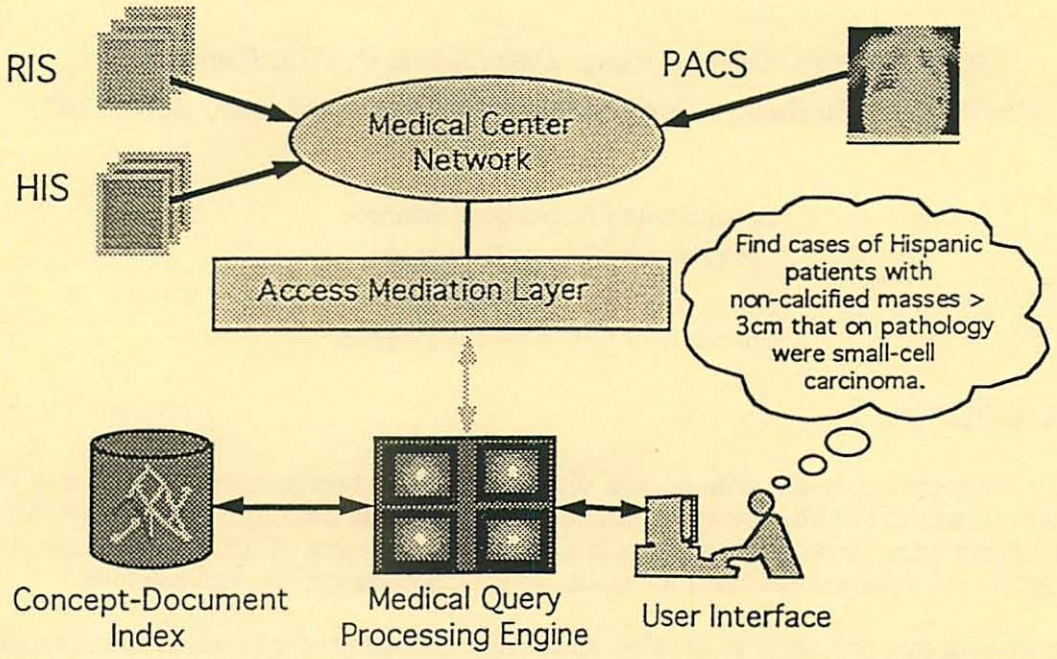


Fig. 1 - System goals are to 1) develop automated methods for document indexing, 2) develop an intelligent query processing engine, and 3) develop a graphical interface for the expression of high-level research queries.

2.0 Materials and Methods

Development occurred in three important areas: 1) automatic indexing of radiology and pathology free text reports, 2) design of the database index and data models, and 3) design of the user query interface. In this paper, we concentrate our discussion on the system's graphical interface.

Our first step was to compile a test database from a specific domain. Our testbed cases were restricted to lung cancer patient's seen at the UCLA Boyer Clinic all of which typically have CT scans and some of which have projectional x-ray images. Most have biopsies for conformation of pathology. Out of the 109 patient's, we found 30 confirmed lung cancer patients. For these 30 patients, there were 153 pathology reports and 178 thoracic radiology reports with references to our target findings of "masses" and "(abnormal) lymph nodes". All pathology, medical, admissions, and laboratory reports for these patients were retrieved using our integrated HIS/RIS/PACS interface [Taira94]. We retrieve the CT and CR images for these patients from the PACS.

2.1 Automatic Indexing of Medical Free Text Reports

Each medical procedure has a text report (e.g., laboratory, radiology, medical, surgical, pathology, etc.) associated with it that is highly information rich. Radiology reports contained detailed descriptions of the objects and findings seen in radiology images. Thus the thoracic radiologist performs the difficult task of image analysis / interpretation. Our system relies on these detailed semantic descriptions of the image findings.

We developed a processing algorithm to perform free text analysis on radiology chest X-ray and CT reports and biopsy pathology reports. Our goal was to provide the user with a comprehensive index summarizing the important findings transcribed in these reports. We imposed the requirement that there would be no departure from the traditional mechanism for radiologic reporting. Specifically, radiologist's would not be required to perform manual indexing or entries of abstract keywords or codes such as the American College of Radiology's diagnostic coding system. A brief description of the algorithm follows. Similar free text analysis systems are described in greater detail in the literature [Huag90, Sager87, Berrut89].

Text analysis occurs in three stages: 1) structural analysis of report text, 2) a lexical analysis involving identification of phrases, and 3) analysis of sentence structure to determine relationship between phrase concepts. We do not discuss resolving phrase ambiguities in this paper. In step 1, the structural analysis involves parsing the radiology (or pathology) report into sections (e.g., study description, clinical history, findings, conclusion). For radiology reports, the "Findings" section of the report is isolated. For pathology reports, the "Final Diagnosis" section is isolated. Within the "Findings" section, paragraphs (finding enumerations) are located and sentences within each paragraph are identified (see Fig. 2).

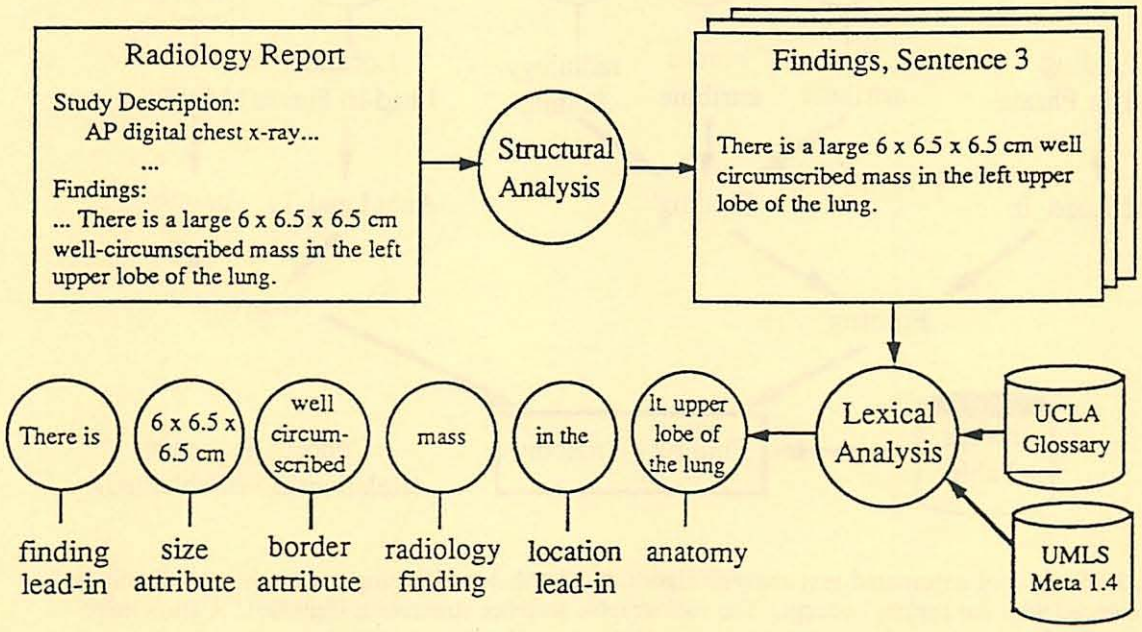


Fig. 2 - Stages 1 and 2 of automated text analysis algorithm. Our approach depends upon comprehensive phrase glossaries for the application domain.

In step 2, a lexical analysis is performed on each sentence. Two glossary sources are used, the National Library of Medicine's Unified Medical Language Sources (UMLS)

metathesaurus and a UCLA-developed thoracic radiology thesaurus. The UMLS metathesaurus (v.1.4) contains over 300,000 medical terms derived from multiple controlled vocabularies, including International Classification of Diseases – 9th Revision (ICD-9), American College of Radiology (ACR) codes, Medical Subject Headings (MeSH), Current Procedural Terminology (CPT) and Systematized Nomenclature of Medicine (SNOMED) [Lindb93]. It serves an excellent electronic source for general medical concepts but is deficient in its entries for radiology because of the lack of spatial, textural, density, and shape-related concepts [Fried92].

The UCLA thoracic glossary supplements UMLS in the following areas: (1) radiology findings related to lung cancer (masses and lymph node involvement), (2) complete thoracic anatomy, (3) geometry, density, texture, temporal, and spatial relationship descriptors, (4) lead-in and lead-out phrases to findings, anatomy, and pathology.

The lexical analysis involves using the two glossary systems to perform term recognition. Instead of extracting every physical findings in the radiology and pathology reports, we attempt only to identify target findings (e.g., masses, lymph node involvement, tumor histology, tumor grade) to limit the size of the task. Each phrase identified is labeled with one of several possible semantic classes seen in medicine including anatomic region, finding, therapeutic method, etc. [Sager87]. The UCLA glossary also include lead-in and lead-out phrases to findings (e.g., "shows", "reveals", etc.), areas of involvement ("extending along the", "in the", etc.), and/or pathology descriptions (e.g., is an indication of, correlates with, etc.). Rules for dealing with negative phrases have also been developed [Bhush94].

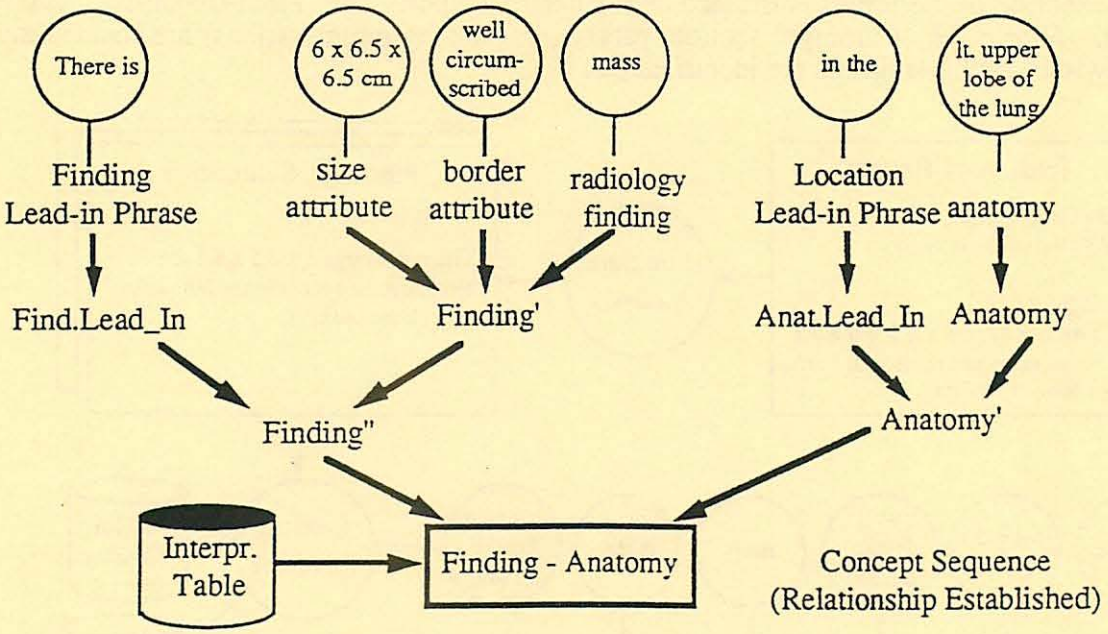


Fig. 3 - Stage 3 of automated text analysis algorithm. Attributes of findings, anatomy, and pathology are merged with the central concept. The radiographic sentence structure is identified. A knowledge-base maintains templates that represent sentence structure and the relationship between the parts.

In step 3, we determine the relationships between the major concepts found in step 2. First, the descriptive attributes of a finding, pathology, and/or anatomy are linked to the central concepts (the finding, pathology, or anatomy themselves). Also, positive lead-in and lead-out phrases are linked to findings, pathology, and/or anatomy. Negative lead-in / lead-out phrases are left unmerged. For each radiology finding, a knowledge frame is maintained which models

the possible attributes associated with the finding. The proximity of the attribute and the knowledge frames together are used to decide which finding/pathology/or anatomy descriptive attributes are in reference to. Thus far, we have modeled attributes of pulmonary neoplasms, and involved lymph nodes (lymphadenopathy). Once all positive lead-in phrases, positive lead-out phrases, and descriptive attributes are merged with the finding, pathology, or anatomy the overall sentence structure is identified. A typical radiographic sentence structure is "Finding-Anatomy". This structure implies that the Finding is located in Anatomy. Figure 3 summarizes the procedure. A knowledge base containing a template of radiographic sentence structure is used to determine the relationship between the concepts (e.g., findings, location, pathology).

2.2 Glossary Generation

A key component in the automated text analysis program and query processing steps is the availability of a comprehensive thoracic radiology thesaurus. Figure 4 shows the structure of the UCLA Thoracic thesaurus. It structure is similar to the structure of Meta-1.4 of the UMLS knowledge sources [Humph93]. It provides a structured method for representing unique concepts by a number of string level representations.

The glossaries were derived from both a "top-down" and "bottom-up" approach. The top-down approach involved going through several classic text books and publications from international conferences on thoracic radiology were used to obtain glossary and index terms [Tudde84, Fraser91, Webb92, WHO81]. This step allowed generally useful terms to be included into the metathesaurus. The bottom-up approach involved scanning several hundred thoracic radiology reports dictated by 4 different staff radiologist and identifying strings used to express various concepts. This step insured that the system worked at a practical level.

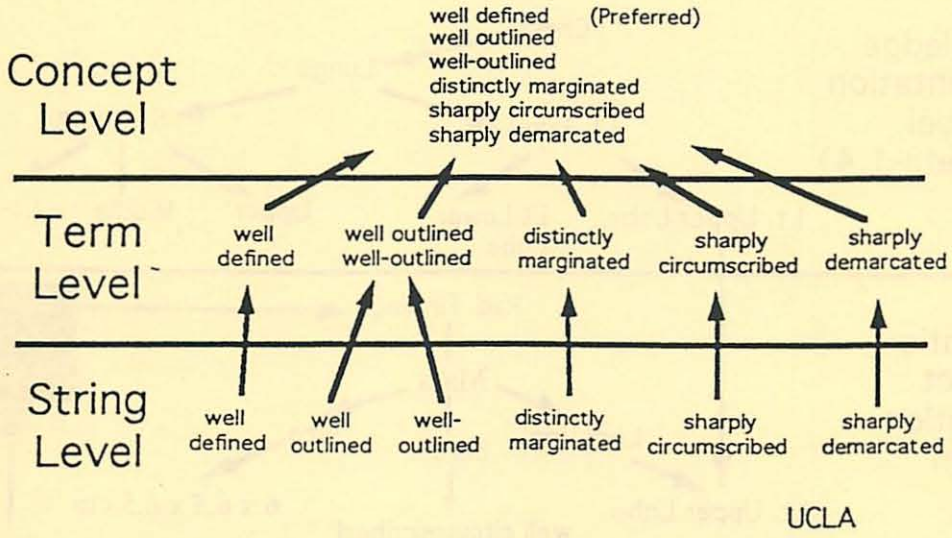


Fig. 4 - Organization of strings, terms, and concepts within the system. There are many string and term representations for a single medical concept.

2.3 Three Level Data Model for Medical Images

Figure 5 shows our three level data model for the abstraction and interpretation of image (and medical report) data. The **Image Stack Level** represents patient images using a dynamic

stacked logical data structure [Josep88]. Stacks consist of two-dimensional variables or pictures registered to the same gridded coordinate system. Multiple stacks of pictures can provide, for example, additional temporal, spectral, and/or spatial information. Stacks may consist of a hierarchy of other stacks.

The **Semantic Object Description Level** provides a description of important findings in radiological images including their location and spatial, textural, and/or density attributes. The methods to extract this semantic information can be from a semantic approach by analysis of the corresponding radiology report or using an image processing approach. At the most abstract level, the description is in the form of the radiology free-text report which is a natural language description of the important findings and observations seen in radiological images. We then use the text analysis routines described above to formalize this description into structured data models (See Fig. 5). Each image analysis and/or text analysis application constructs a different data model dependent upon the target information to be extracted (e.g., masses, lymph nodes, etc.). A object-concept semantic data model [Gupta91, Chu94] approach is used to model object instance descriptions.

The **Knowledge Representation Level** models medical concepts at the class (generic level) as opposed to the instance level. This layer contains data structures that incorporate domain knowledge, and provide a structured means of defining imprecise terms, normal values, and concept relationships [Weila93]. These knowledge structures allow the system to retrieve information based on similar cases of interest and/or expanded scopes of interest. Nodes in this layer can map to instances either to nodes or relationships in the Semantic Object Description Level; hence object instances within an image can be mapped to the larger general data model.

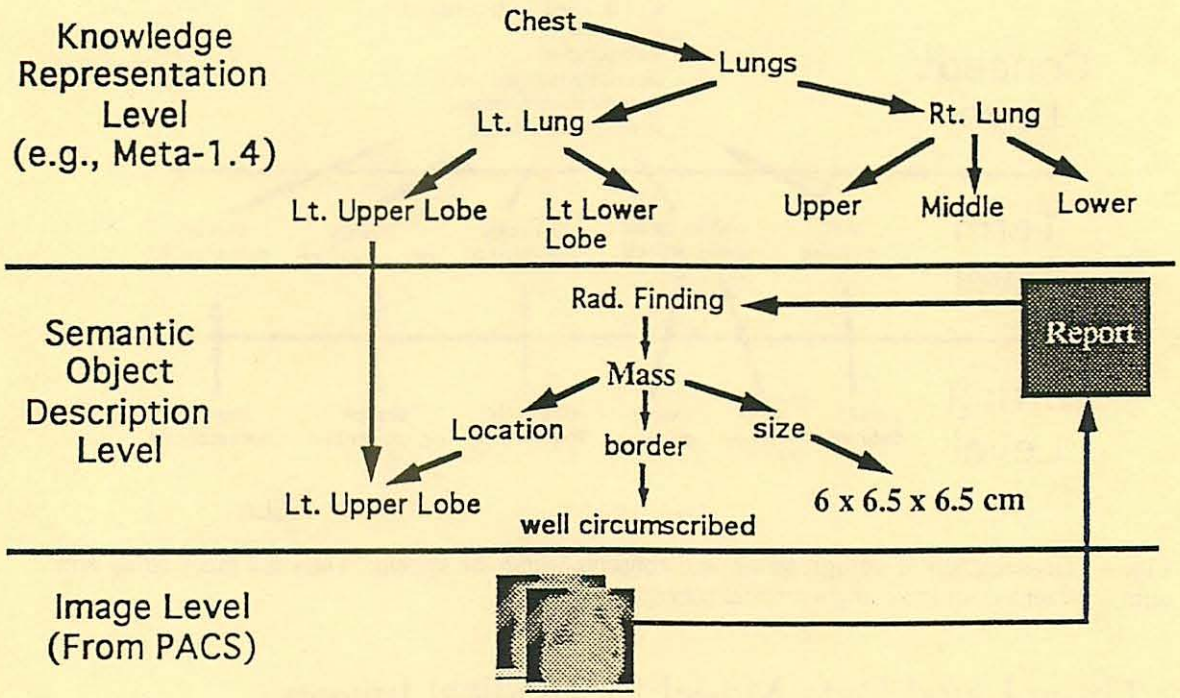


Fig. 5 - Three level data model for medical images.

2.4 User Interface and Query Processing

The user interface and query processing layers allows users to construct research queries and browse the information system. The interface is concerned with the expressibility of English-like queries into a formal query language. We use the PICQUERY+ multimedia interface design proposed by Cardenas et. al. [Carde93], customizing its presentation for thoracic-radiology lung cancer applications. The initial system, allows users to specify three classes of query constraints: Patient demographic profile, patient pathology, and radiological findings. Figure 6 shows the prototype main menu screen.

The patient class filter allows the user to specify the demographic profile of patients whose data is to be retrieved. Icons can be selected to visualize the relationship between values for a particular object attribute. For example, if patient ethnicity icon is selected, a graphical representation showing how different ethnic groups are related is presented as a directed graph (e.g., Orientals include Chinese, Japanese, Korean, etc.). A user may select general class values (e.g., ethnicity = "oriental") or specific values (e.g., ethnicity="Chinese"). Fields requiring temporal values (e.g., age) can be entered using single time points (e.g., 18 months), a time period (e.g., from 18 months to 24 months), or a more abstract value such as a developmental phase (e.g., adolescent). Finally, we will provide not only selectable values for each attribute, but also allow the user to specify different types of correlators including: "equals", "less-than", and "similar-to".

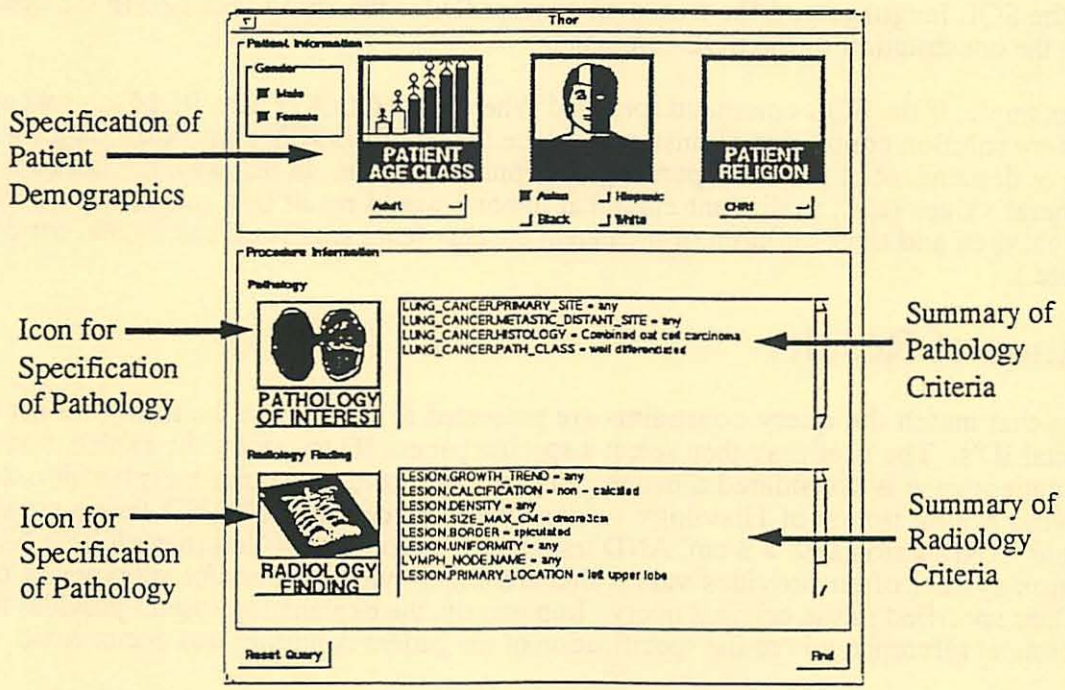


Fig. 6 - Main user window for thoracic radiology research database.

The pathology class filter allows the user to further specify pathology constraints to the query. For each pathology class of interest, a custom form is developed. The lung cancer screen is shown in Figure 7. The upper portion of the screen provides the user with a hierarchical set of lists corresponding to histological classifications of lung cancers. This listing was obtained from data extracted from the ACR Index of Radiologic Diagnosis and the lung cancer histologic

The query processing program is written in the C programming language and uses the Sybase DB-Library C-interface routines. The query processing begins by parsing the input PICQUERY+ table predicate by predicate and constructing an SQL command to query the concept index table created from the results of the text-analysis step (see above).

In the cases of pathology and radiology finding predicates, the system substitutes the attribute string value for its equivalent thesaurus string ID number. The string ID number is unique for each attribute name - attribute value pair. For example, in the UCLA glossary, if the attribute name is "Radiology Finding" and the attribute value is "Mass", then the string ID is 01.1000000.01. A concept ID can be derived from the string ID by truncating the last two digits of the string ID (e.g., 01.000000 is the concept ID for "Mass".). The concept ID identifies all synonyms and lexical variants of the value represented by the string ID.

Using this conceptID number, all synonymous terms, lexical variations, and subclass concepts for the given concept can be identified using the UCLA thoracic thesaurus. The assignment of concept ID's follows a structured approach based on an outline classification paradigm. As an example, the concept "malignant epithelial tumor" has the concept ID in the pathology thesaurus of 01.1400000. A subclass of this concept, "small-cell carcinoma" is assigned the concept ID 01.1410000. A subclass of small-cell carcinoma, "oat-cell carcinoma" is assigned the concept ID 01.1411000. Using this hierarchical coding convention, the zero-truncated portion of the conceptID defines a node within a concept hierarchy (e.g., 01.14 is the truncated substring for the concept "malignant epithelial tumor"). We thus use the "Like" qualifier in the SQL language and the truncated concept ID for the string specified in the query predicates in the construction of the SQL command.

For example, if the SQL command specified Where HISTOLOGY like '01.14', this would result in a query solution containing all instances where the HISTOLOGY class of the tumor was any instance or descendant of the concept malignant epithelial tumor. In this way, predicates that included general values (e.g., malignant epithelial tumors) would result in a solution space that included all children and their children of the parent concept (e.g., small-cell carcinoma, oat-cell carcinoma, etc.).

Presentation of Results

Cases that match the query constraints are presented to the user in the form of a list of patient hospital ID's. The user may then select a specific patient ID to review the explanation as to why this patient case is considered a match. Figure 110 shows an example explanation to a query involving a lung tumor of Histology = "non-small cell carcinoma" AND lesion growth trend = "stable" AND lesion size > 3 cm, AND lesion primary location = "left lower lobe". Note the explanation system often provides values that are either synonyms and/or subtypes of the predicate values specified in the original query. Importantly, the explanation system provides the date and document reference where the specification of the patient condition was documented.

If the user is satisfied with the explanation and/or is curious to view all the HIS, RIS, and PACS records for the patient, the user can select "Full Reports and Images" in the query explanation window to launch a multi-database retrieval of all Radiology, Pathology, Laboratory, Medical reports, Admission records, and select radiology images. The multi-base retrieval at our institution can be performed only using a patient Hospital ID number. See [Taira94] for a screen dump of this interface. If no solutions to a query are found, then the user may which to see a count of the number of cases that match each query predicate separately. This will help assist the user in determining which query constraints should be relaxed. Figure 11 shows the database profiler for the query constraints requested by the user.

The radiology finding filter allows the user to specify cases which demonstrate certain radiographic findings within an image. The radiological findings can refer to several features including: (a) Object composition concepts - gray level, texture, heterogeneity, homogeneity, CT number, proton density; (b) Spatial concepts - size, shape, location, distance; (c) Temporal concepts - changes (e.g., anything new, anything resolved, size, position, etc.), direction of change (e.g., growing, shrinking), rate of change (e.g., fast, slow); (f) Temporal relationships - (e.g., process A "co-occurs with" process B); (g) Spatial relationships - (e.g., mass "adjacent to" aortopulmonary window).

The strategy for specifying image findings is to first ask the user to select a particular type of findings from a well organized menu. A custom menu for each class of findings and or findings common to a disease condition will be displayed. The findings again could include both low level (i.e., morphologic) findings (e.g., circular density, blunt angle, wide mediastinum, etc.,) as well as more abstract radiologic findings (e.g., edema, atelectasis, nodule, consolidation, etc.). Figure 8 demonstrates the common radiographic findings associated with lung cancer patients. Once a finding is selected, attributes for that finding are presented onto a graphical form (see Fig. 8). For example, the finding "lesion" has the attributes "location", "size", "contour", "border characteristics", "internal architecture", "homogeneity", "cavitation", "lobulation", "calcification pattern", and "clinical trend". Under each of these attributes for lesion are possible values from which an investigator may wish to choose.

Query Processing

Once the user has specified the patient demographics, disease pathology, and radiology constraints, a PICQUERY+ table is presented to the user so that he/she may see a summary of the entire query which can span objects stored in multiple databases. The table consists of three columns containing the name of an object attribute, the relational correlator, and the attribute value. For all non-numeric attribute values, the correlator is actually "like" instead of "=" as shown. If the table summary is acceptable, the user selects the "FIND" button to launch the query. The PICQUERY+ table is written to a UNIX file and the query processing routine spawned to retrieve case matches found in the index table populated by the text analysis program.

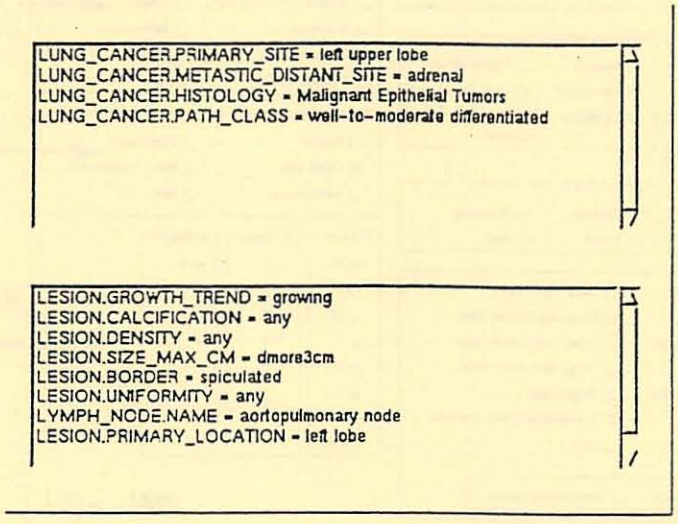


Fig. 9 - PicQuery+ summary table. This table describes the predicates of the query to be launch.

classification from an international conference in Geneva, Switzerland [WHO81]. The histology class filter allows the investigator to specify both very specific histologies (e.g., oat-cell small-cell carcinoma) or general histology values (e.g., malignant epithelial tumors). Radio buttons are provided for the user for the specification of cell differentiation classification for the lung tumor. Additionally, radio buttons are provided to allow the user to specify the location of the primary tumor site and location(s) of metastatic sites.

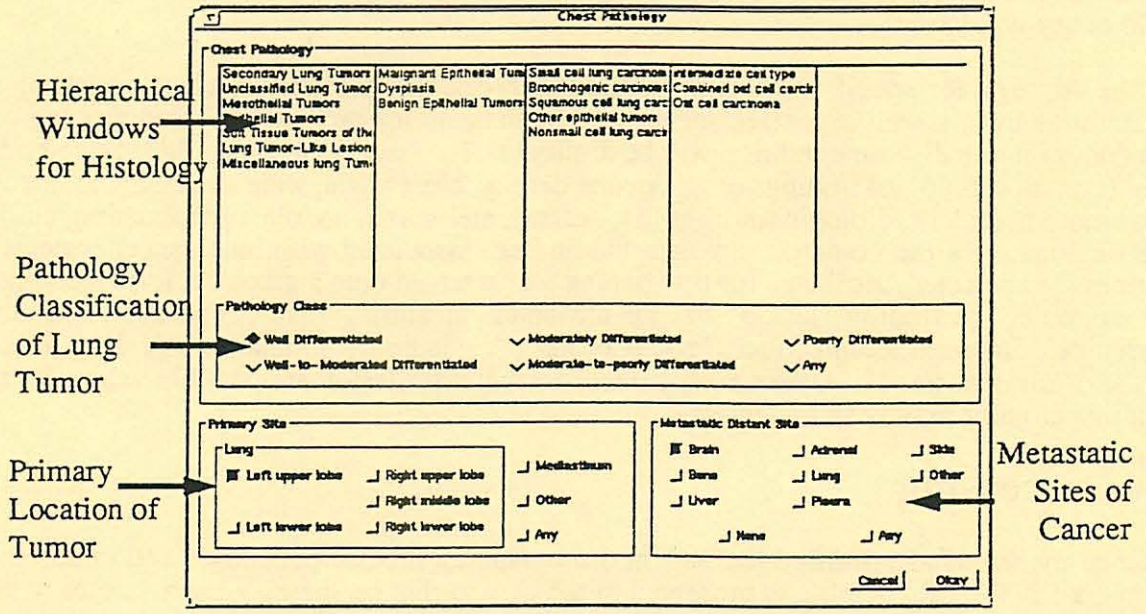


Fig. 7 - Pathology specification window. Customized for patient's with lung cancer.

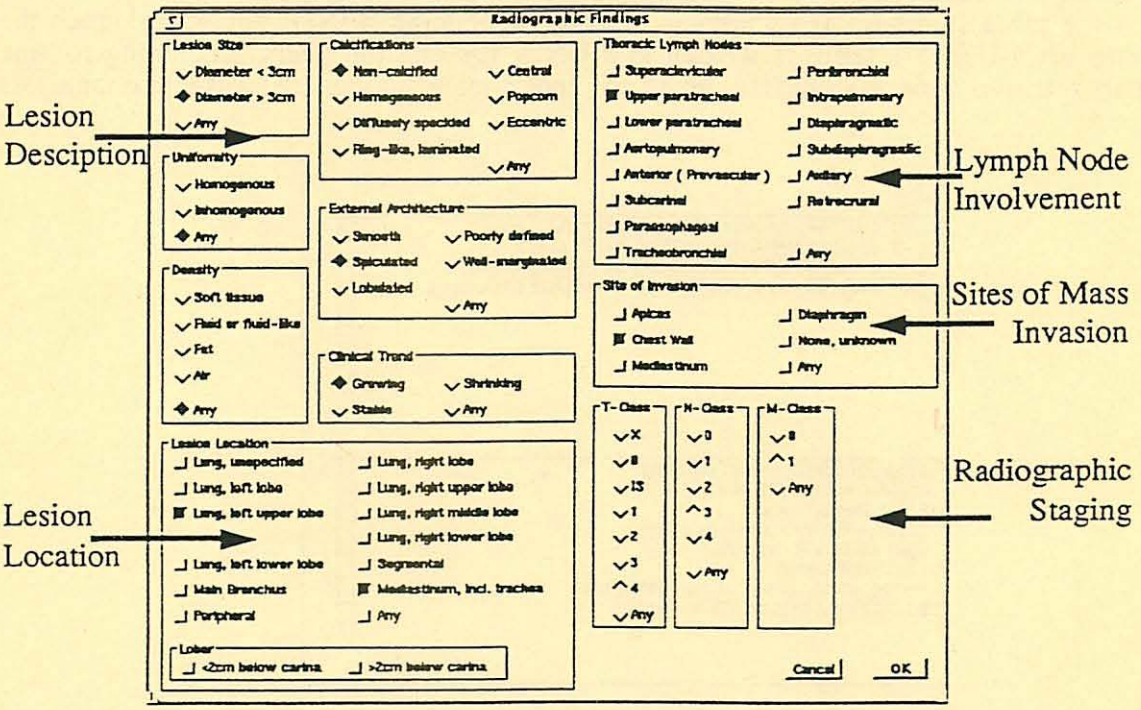


Fig. 8 - Radiology Findings specification window. Customized for patient's with lung cancer.

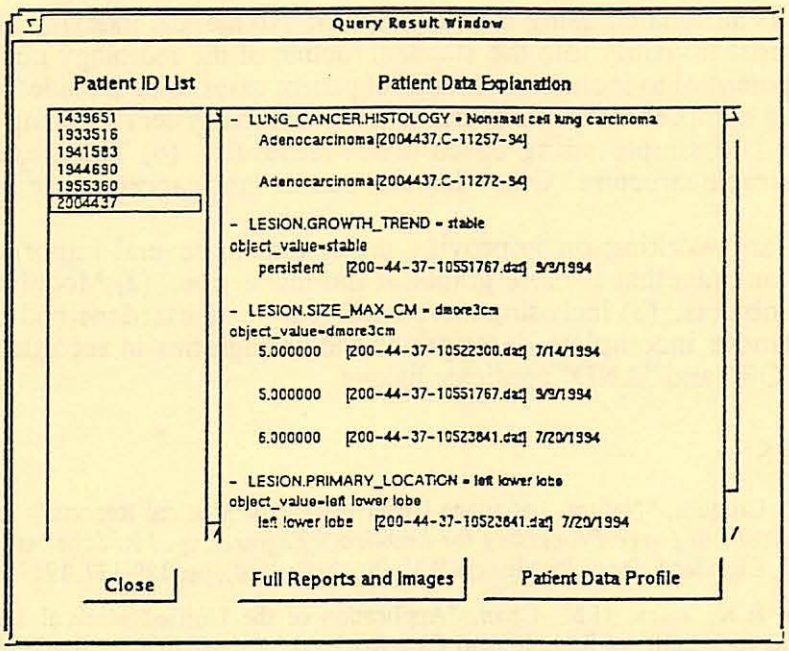


Fig. 10 - User selects a particular case to review the explanation as to why the patient is considered a match. Once the explanation is reviewed, the user may launch a global HIS, RIS, PACS search using the patient's hospital ID to collect all relevant documents.

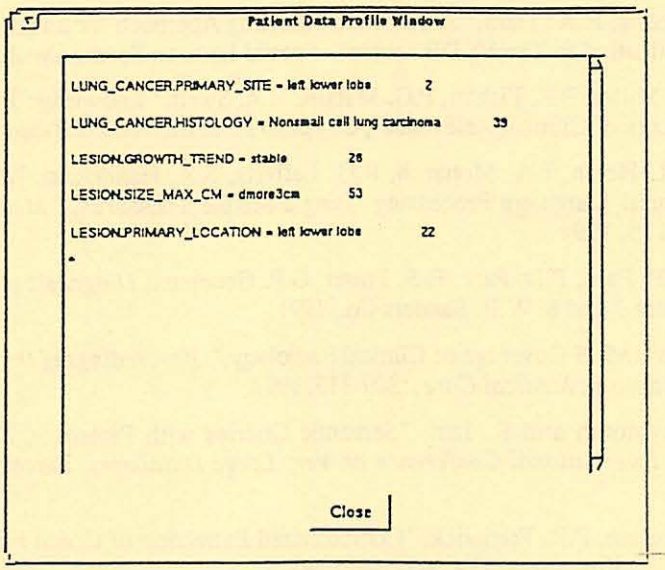


Fig. 11 - Count profile of the number of hits in the database for each query constraint requested by the user independently. The hits to the original query represent patients who match all these constraints (i.e., an AND operation).

3.0 Discussion

We present a prototype system that performs automated indexing of radiology and pathology reports from lung cancer patients. The advantages of this system include: (1)

TAIRA

indexing of concepts is automated using free text reports. No manual indexing is required. (2) Can be integrated almost invisibly into the standard routine of the radiology department. This gives the system the potential to include thousands of patient cases to be included in the system's index. (3) It includes a comprehensive thesaurus and explanation system allowing concept-based matching as compared to simple string based match retrievals. (4) The thesaurus is easily expandable given its simple structure. Other domains outside lung cancer can be included.

Currently we are working on improving the system in several important areas: (1) Developing a Query languagethat is more graphical and interactive. (2) Modeling of temporal and evolutionary data objects. (3) Including more intelligent frame based method of text analysis (e.g., [Haus90]) to handle incomplete descriptions and ambiguities in recorded speech. (4) Providing selectable "OR" and "AND" predicate linking.

4.0 References

[Berrut89] C. Berrut, P. Cinquin, "Natural Language Understanding of Medical Reports," in *Computerized Natural Medical Language Processing for Knowledge Engineering*, J.R. Scherrer, R.A. Cote, S.H. Mandil (eds.), Elsevier Science Pulblishers B.V. (North-Holland), pp. 129-137, 1989.

[Bhush94] V. Bhushan, R.K. Taira, H.M. Chan, "Application of the Unified Medical Language System Metathesaurus to Intelligent Radiological Case Retrieval," *Proceedings of the SPIE*, **2165**:336-347, February 1994.

[Carde93] A.F. Cardenas, I.T. Ieong, R.K. Taira, R. Barker, C.M. Breant, "The Knowledge-Based Object-Oriented PICQUERY$^+$ Language," *IEEE Transactions on Knowledge and Data Engineering*, 5(4):644-657, August 1993.

[Chu94] W.W. Chu, I.T. Ieong, R.K. Taira, "A Semantic Modeling Approach for Image Retrieval by Content", submitted for publication to The VLDB Journal - Special Issue on Spatial Database Systems, 1994.

[Cohen90] A. I. Cohen, P. L. Miller, P.R. Fishler, P.G. Mutalik, H.A. Swett, "Knowledge-Based Radiologic Image Retrieval Using Axes of Clinical Relevance", *Computers and Biomedical Research*, 23:199-221, 1990.

[Evans91] D.A. Evans, W.R. Hersh, I.A. Monarch, R.G. Lefferts, S.K. Handerson, "Automatic Indexing of Abstracts via Natural Language Processing Using a Simple Thesaurus," *Medical Decision Making*, 11(suppl):S108-S115, 1991.

[Fraser91] R.G. Fraser, J.A..P. Pare, P.D. Pare, R.S. Fraser, G.P. Genereux, *Diagnosis of Disease of the Chest*, Vols. I & II, Chapter 1 and 8, W.B. Sanders Co., 1991.

[Fried92] C. Friedman, "The UMLS Coverage of Clinical Radiology," *Proceedings of the Annual symposium on Computer Applications in Medical Care*, :309-313, 1992.

[Gupta91] A. Gupta, T. Weymouth and R. Jain, "Semantic Queries with Pictures: The VIMSYS Model," *Proceedings 17th International Conference on Very Large Databases*, Barcelona, Spain, pp. 69-79, September 1991.

[Haug90] P.J. Haug, D.L. Ranum, P.R. Frederick, "Computerized Extraction of Coded Findings from Free-Text Radiologic Reports," *Radiology*, 174:543-548, 1990.

[Humph93] B.L. Humphreys, D.A. Linberg, "The UMLS Project: Making the Conceptual Connection Between Users and the Information They Need," *Bulletin of the Medical Library Association*, 81(2):170-177, 1993.

[Josep88] T. Joseph and A.F. Cardenas, "PICQUERY: A High Level Query Language for Pictorial Database Management", *IEEE Transactions on Software Engineering*, 14(5):630-638, 1988.

[Lin91] R. Lin, L. Lenert, B. Middleton, S. Shiffman, "A Free-Text Processing System to Capture Physical Findings: Canonical Phrase Identification System (CAPIS)", Proceedings of the

[Lindb93] D.A.B. Lindberg, B.L. Humphreys, A.T. McCray, "The Unified Medical Lanaguage System", in *Yearbook of Medical Informatics* (J.H. van Bemmel, A.T. McCray editors), pp.41-51, 1993.

[Meghi91] C. Meghini, F. Rabitti, C. Thanos, "Conceptual Modeling of Multimedia Documents," *Computer*, :23-30,October 1991.

[Sager87] N. Sager, C. Friedman, M.S. Lyman, *Medical Language Processing. Computer Management of Narrative Data*, Addison-Wesley Publishing Co, 1987.

[Taira94] R.K. Taira, C.M. Breant, F.M Stepczyk, H. Kho, D.J. Valentino, G.Tashima, A.T. Materna, "CASE and Object-oriented Tools for Rapid Integration of HIS, RIS, and PACS," *Proceedings of the SPIE*, **2165**:185-199, February 1994.

[Tudde84] W.J. Tuddenham, "Glossary of Terms for Thoracic Radiology: Recommendations of the Nomenclature Committee of the Fleischner Society," *American Journal of Roentgenology*, **143**:509-517, 1984.

[Webb92] W.R. Webb, N.L, Muller, D.P. Naidich, High Resolution CT of the Lung, Raven Press, 1992, pp. 157-159.

[Weila93] W.J. Weiland, B. Shneiderman, "A Graphical Query Interface Based on Aggregation / Generalization Hierarchies", Information Systems, **18**(4):215-232, 1993.

[WHO81] World Health Organization, "Histological Typing of Lung Tumours, No. 1, 2nd Ed., Geneva, Switz., 1981

PACS Intelligence: Automatic Feedback for PACS Retrieval Requests

Doris T. Chen, Bruce K. T. Ho, Woodrew Chao, Ricky K. Taira, and John Bentson
Department of Radiological Sciences, University of California, Los Angeles, CA 90024

ABSTRACT

Waiting for images from a PACS archive is frustrating for radiologists whenever images are missing from their workstations. We employ system intelligence into our PACS to estimate the image retrieval time and to notify the radiologist upon image availability. When an image retrieval request is made by a radiologist, our PACS manager software is automatically notified. It then collects archival information and system conditions from various PACS nodes (database, file servers, and workstations) through a trap mechanism. A statistical model is developed to estimate the image retrieval delay based on image size, number of prior requests, network topology, and storage device type. A retrieval time table is then graphically displayed on the workstation. Automatic paging notifies the relevant radiologist if the delay is substantial and catastrophic. A radiologist is much more willing to perform softcopy viewing regularly once he is given a good estimation of image retrieval delay, from which he is able to make better decisions in prioritizing his work list. The implementation of PACS intelligence requires enormous information. The PACS manager makes it possible to collect them anywhere in the network. The retrieval time notification is an important example of applying PACS intelligence to assist clinical practice.

Keywords: PACS intelligence, image retrieve, PACS manager, trap, automatic paging

1. INTRODUCTION

Waiting always causes impatience and frustration and this is especially true in daily clinical practice. Almost everyday the radiologists in UCLA need to retrieve images from PACS archive to perform softcopy viewing. Waiting for images to arrive is frustrating for the radiologists whenever images are missing from their workstations. If the time for image retrieve is unpredictable and unacceptable, the radiologists may go back to hardcopy viewing after all. By providing the estimating of the image retrieval time to the radiologists, they would be more willing to perform softcopy viewing regularly.

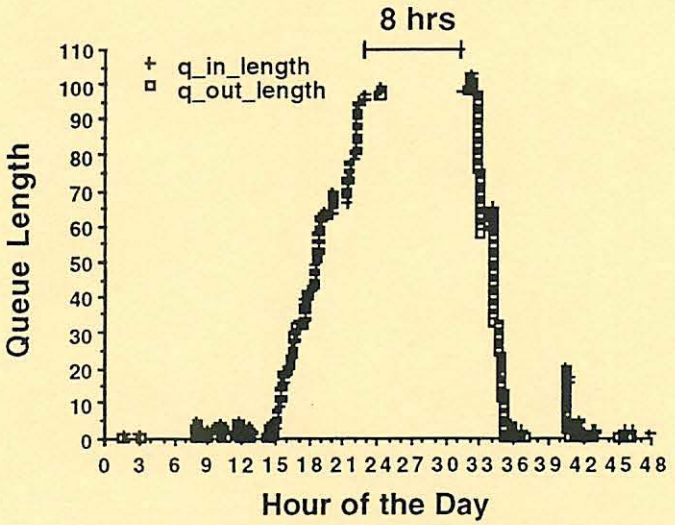


Fig. 1. An example of extremely long queue delay during image retrieve. + represents the number of the jobs enter to the queue, whereas the symbol "empty square" represents the number of the jobs left in the queue .

How long could an image retrieve really take place? A good example of an extensively long image retrieval delay has been demonstrated in figure 1. It indicates when the image retrieval job requests enter the queue, because of the system bottleneck, the jobs are never serviced. The jobs are accumulating within the queue, until they reach at a magnitude of 100 jobs. The bottleneck is not resolved for eight hours, and a situation like this clearly indicates some system failure in the network.

In general, three types of problems could occur in current PACS image retrieve. The first type is catastrophic. The PACS network (Ethernet based) is characterized as heavy image traffic, long image flow path, heterogeneous computer nodes, multiple protocols, and physical media. Within such a complex network, errors regarding computer environment, software and hardware could all contribute to catastrophic system failure in the PACS network.[1] For example, power outage is the most common one to cause hardware failure, and it should be remedied by a graceful power up sequence. Some common errors have been encountered are: 1)communication errors among interprocessors, and errors caused by exceeding resource limitation, 2)file and data errors including non-standard data formatting, data access and protection problem, and data entry errors by the technologists, 3)system call errors, errors in programming logic, and software bugs. All these catastrophic failure will general require the repair from the trained personnel through an automatic paging mechanism,[2] hence they are not the major interest in this project.

In order to study the other two types of PACS problems, a good understanding of the image flow and the associated queues in PACS system would be essential to start with.[3, 4, 5] Figure 2 illustrates all the major components along the imaging chain. The software processes have approximately 80 daemon running at all times. This flow chart demonstrates the complete process steps from acquisition to archive and, finally, to display. Two basic processes: forwarding and retrieval are indicated by solid and dashed lines respectively. An image acquired at the scanner is first reformed by a capture computer, and then transferred to the cluster controller which archives and distributes them to the display workstations in the forwarding process. In the retrieval process, after the user makes a request from the workstation, the workstation daemon first checks if the required images are registered in the database. If they are available, then the daemon starts retrieving. The retrieve process in the cluster controller gets the images from the image archive center(OD and tape juke boxes), and sends the images to the workstations through a communication daemon over the network. In general, there is one queue between every two consecutive processes. For example, The cluster controllers have processes to receive images, to update database, to forward images to display workstations, and to read and write on the optical disks. The way we are making the estimation is by polling the relevant queues in the system, and then check the condition of every single image that is processed.[6]

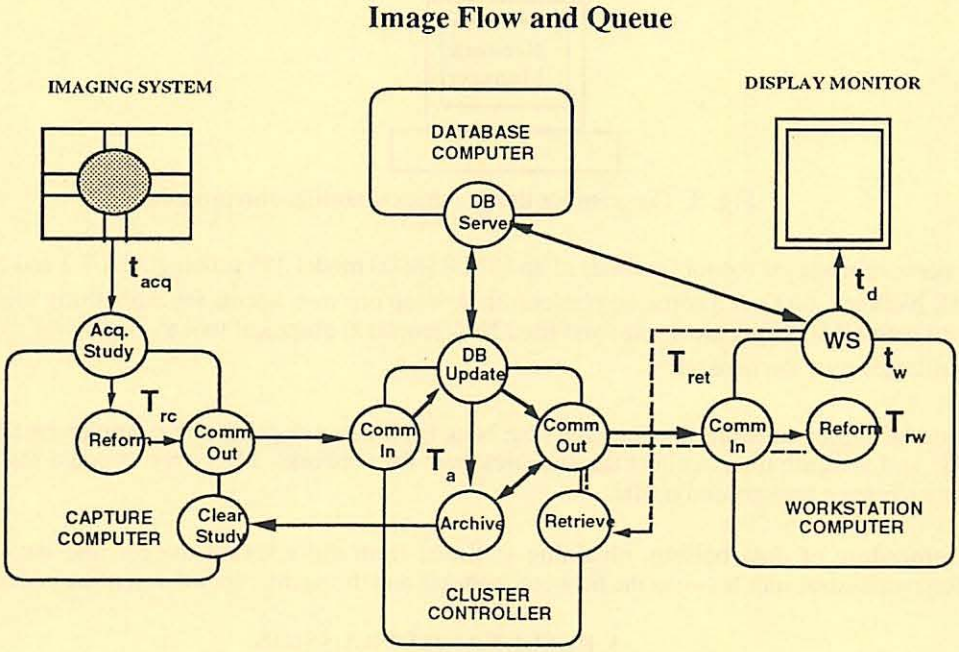


Fig. 2. Acquisition and retrieval of radiological studies require the successful completion of several processing stages.

Our project aims to find out if the image retrieval delay due to the other two types of problems such as the juke box access delay and the network congestion problem. Then a statistical model can be formulated to predict the delay. The medium of the image archive center could be either the optical juke boxes or the tape juke boxes in the near

future. Different types of juke boxes would result in different access rates, hence juke box access delay is definitely the very first element to be considered in our prediction model. Network congestion is again the most common problem in most of today's network. It would be necessary to determine if PACS suffers bottlenecks and if image retrieval delay is contributed by congestion.

2. RETRIEVAL NOTIFICATION METHOD

The brief procedure of image retrieval notification method is demonstrated in figure 3. The network manager station can monitor all the processes involved in image retrieve, and each workstation will have agents which report the status back to the network manager. After the workstation makes an retrieval request, both of these agents will report the data such as image size, prior retrieval queue list, and the network traffic to the network manager, from which the network manager can estimate the image retrieval time. This process is automatic, and the result in a graphical form will be displayed onto the workstations for radiologists' reference.

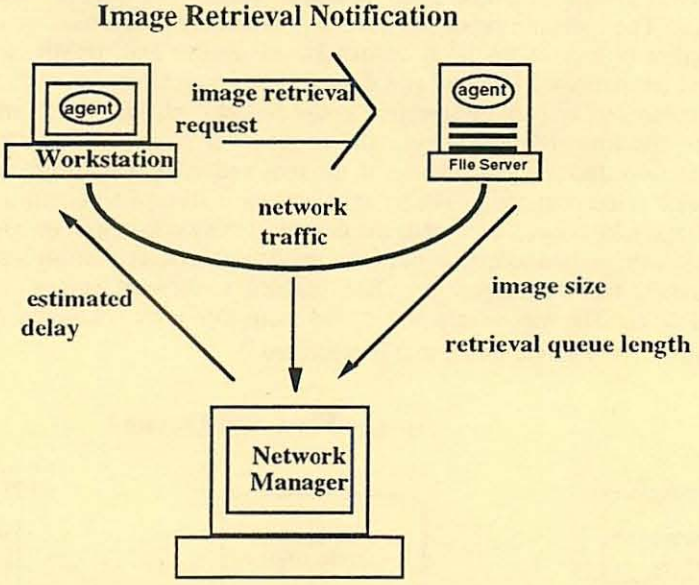


Fig. 3. Diagram for image retrieval notification procedure.

Our network manager console consists of an IBM RS6000 model 355 running AIX 3.2 and NetView/6000.[7] We used IBM AIX NetView/6000 as a software platform to develop our own agents for monitoring image retrieve.[8, 9] We have developed many PACS specific agents, and used NetView/6000 graphical tool to display the statistical performance data that we collected over the network.[2]

A network sniffer software running on a note book from network general[10] is implemented to monitor the total network traffic and utilization and collect the statistics over the network. Moreover, it could also serve as a network traffic generator to create background traffic.

The procedure of data polling, obtaining statistics from the relevant queues and network traffic, and the communication mechanism trap between the network manager and the agents, are all described in detail in reference 2.

3. RESULT AND DISCUSSION

3.1 OD Juke Box Access Delay

The first thing we study is the queue delay for different file sizes. Figure 4 shows the portion of the queue delay due to the OD retrieve. Images sizes range from 8 MB to above 70 MB are tested. Overall, there is no significant unexpected OD delay, and the baseline speed is about 217 KB/sec. Normally, if the images fall on to the baseline, a 50 MB image should be transfer from OD within 4 minutes. There are cases which have excessive delay, and the reasons are still unknown.

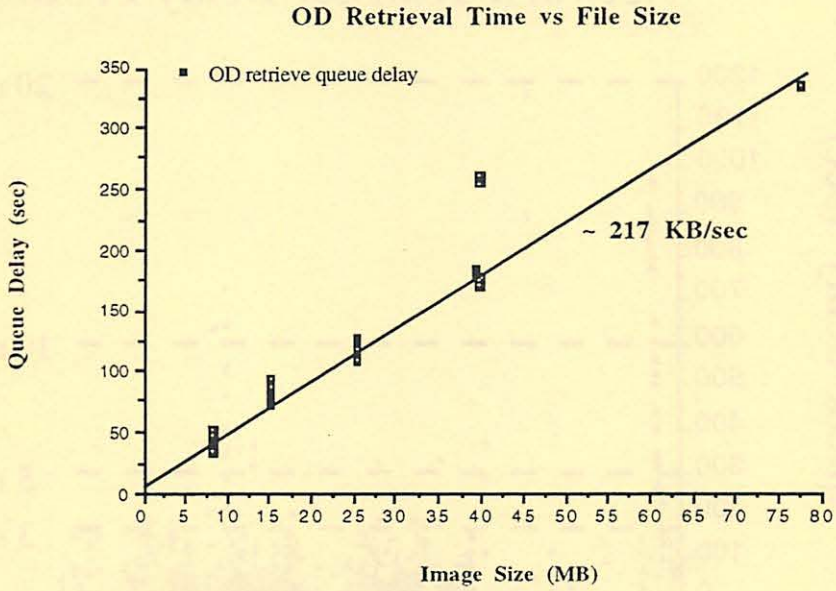


Fig. 4. A range of images with different file sizes are retrieved from the OD juke box.

3.2 Network Congestion

Next, we investigated the effect of the network congestion on image retrieval delay. Many factors may contribute to the delay during the network transfer. Does the image file size effects the delay? If a image retrieve is requested at the different time of the day, is it going to have the same delay? Since the traffic load is closely related to the time of the day, how is traffic load going to affect the delay? In order to simplify these complex problems, we investigate them in three stages.

3.2.1 Image File Size and the Time of Day

First, we look into the delay at different time of the day. Figure 5 shows a network transfer profile over two weeks period of time. It is very prominent that most of the image transfer occurs within the working hours from 8 am to 8 pm. In the predominant cases, 90.7% of the file transfer can be completed within 3 minutes(180 sec), and this is very acceptable. If all the images could be transferred within such a speed, there would not be any problems in the image retrieve. We are now more concern about the rest of 9.3% of the file transfer, and it shows that 2.7% of the file transfer takes 3 to 5 minutes to complete, 3.2% takes 5 to 10 minutes, 2.8% takes 10 to 20 minutes, and 0.7% takes more than 20 minutes. File transfer takes more than 20 minutes definitely shows there is some catastrophic problems occurring in the system, and as we discussed before, would be repaired by the responsible trained personnel. The cases above 10 minutes are always the results of retries, since each retry interval is defined to be 540 sec (9 min) in our PACS. Hence, if there is one retry, it would take at least 10 minutes to successfully complete the transfer by assuming the actual network transfer time is only 1 minute, and if there are two retries, it would take at least 19 minutes to complete by the same assumption as before, and three retries would be 28 minutes and so on. A retry could be caused not only by the system failure which is presented in introduction but also things such as disk space full error, daemons involved in retrieve process down, errors in establishing sockets and acknowledgment message, and the internal UNIX queue problem.[11] Again, all the above problems can not be predicted by our method, and should be notified to the trained personnel. Hence, our investigation is concentrated on the cases within 10 minutes.

The relationship between the network transfer delay and image file size is illustrated in figure 6. The statistics is collected over one month. The size of the images covers a broad range: 0.5 MB to 90 MB. Most of the images are within 30 MB range. All the cases due to the catastrophic problem and the retries are eliminated from the statistics hence in this case all the network transfer delays are under 10 minutes. There is a clear indication that the network transfer delay increases with the file size. A baseline delay is determined to be about 700 KB/sec (within 60% network

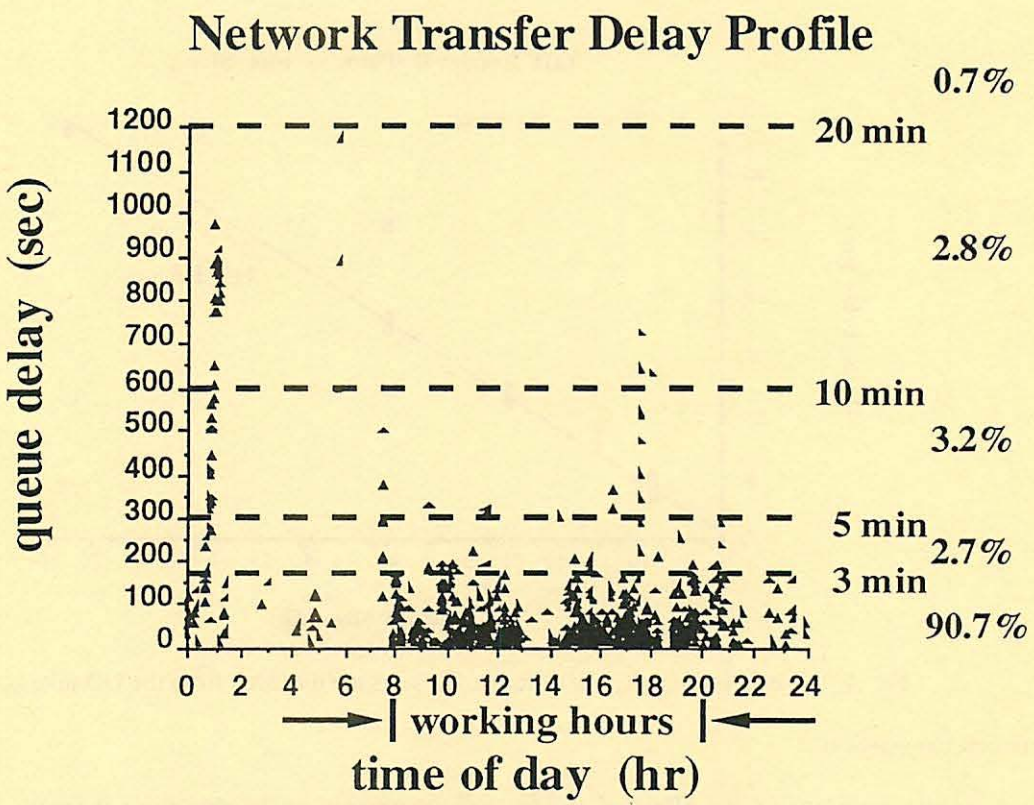


Fig. 5. Network transfer delay profile collected over two weeks.

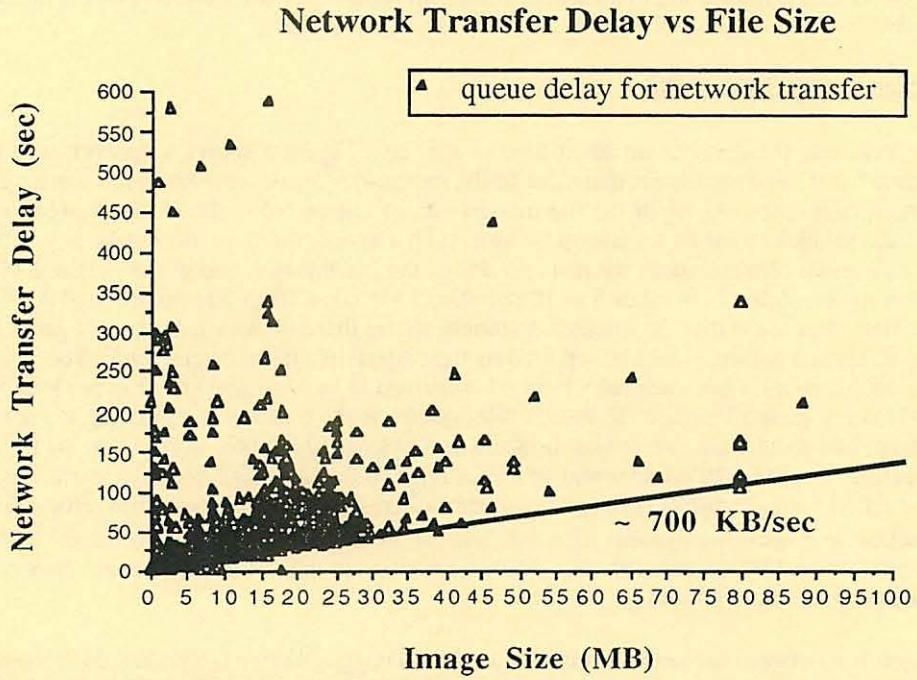


Fig. 6. Network transfer delay vs. the images file size is presented here. The statistics is collected over one month. The network transfer speed is determined by drawing the baseline of the data.

utilization, and it is calculated by using the current transfer rate compare to the standard Ethernet speed), which is very typical for Ethernet(10 Mb/sec or 1250 KB/sec). In this figure, there are cases where the delay is excessive. We need to develop a statistical model to see if excessive delay is a sizable variation in delays for each image size. This variation is expected due to the non-deterministic nature of the Ethernet CSMA/CD(carrier sensing medium access/collision detection) protocol.[12, 13, 14]

3.2.2 Constant Background Traffic Load

An experiment on studying the effect of constant background traffic load on to the transfer delay is performed. The experiment is illustrated in figure 7. In the experiment, the background traffic is generated by a network sniffer with the traffic level at 20%, 40%, and 60% of the total network utilization. A total of 23 images with a magnitude of 550 MB has been selected to be retrieved under the different background traffic level.

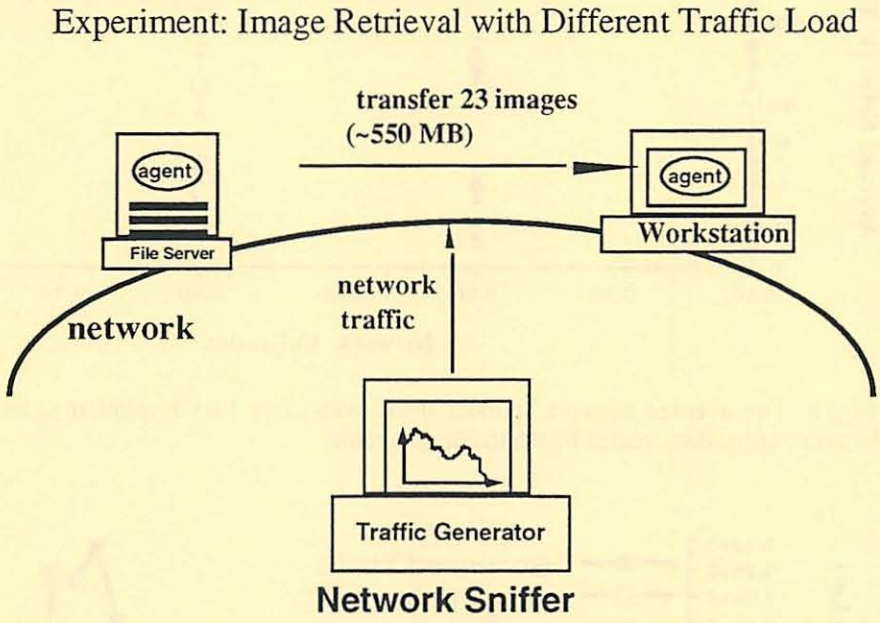


Fig. 7. In this experiment, a set of images has been tested with different traffic load.

The result of the experiment is presented in figure 8 and 9. Figure 8 shows the average network transfer delay at different background traffic load. There is no significant change of the delay for background traffic load of 0, 20%, and 40%. The delay at 40% and 60% shows some increasing tendency especially for the 77 MB large image file. For the images smaller than 77 MB, the effect of network load is not obvious. As we observe in figure 6, most of the images are smaller than 30 MB, and the normal background traffic load is smaller than 60%, therefore the network congestion effect is not expected to be dominant on network transfer delay. Figure 9 quantitatively illustrates the generated background traffic, combined with the traffic due to image transfer, and the network collision during the experiment. The network collision is negligible compare to the background and the image transfer traffic.

3.2.3 Network Transfer Delay vs. Network Traffic

An investigation on network transfer delay vs. the constant background traffic has been performed and the result is presented above. Further study for the actual clinical environment should be carried out to confirm the experimental result. Statistics is collected for two weeks for the network transfer delay and the corresponding network traffic load. The network load is collected for total traffic load, and again all the cases due to the catastrophic delay and retries are eliminated from the data.

Since image size is an important factor which affect the transfer delay, we choose all the images have the same size and focus on network transfer delay vs. network traffic load. In figure 10, all the images are 8 MB CR's, and the network load averaged over 10 minutes is converted to a rate in KB/sec. From the figure, 97.12%(236 out of 243 cases)

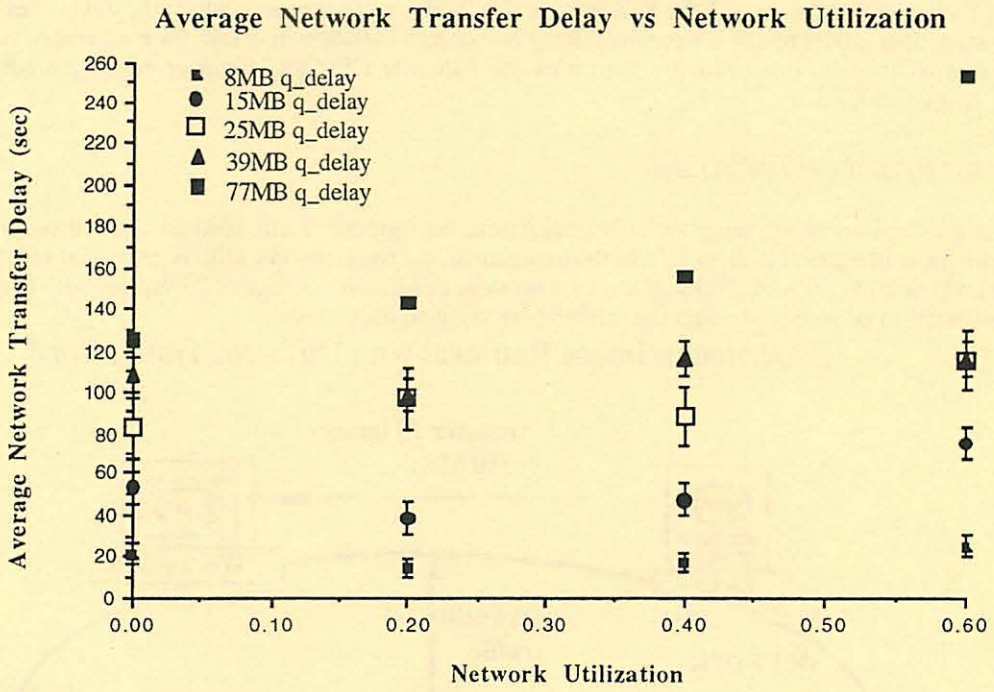


Fig. 8. The average network transfer delay with error bars is plotted against constant network utilization created by the traffic generator.

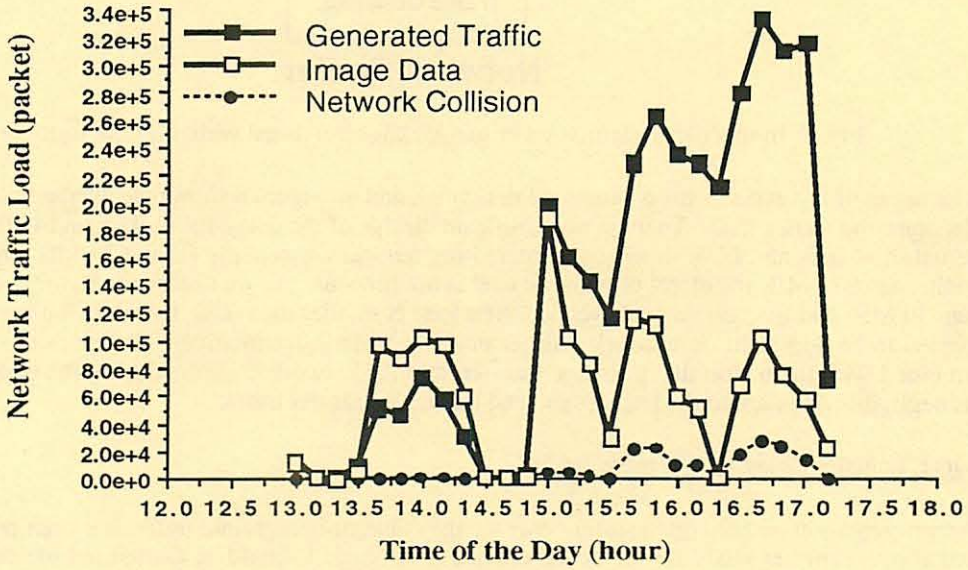


Fig. 9. Traffic load during the experiment.

of the transfer is completed within 60 seconds (1 min), and this is totally acceptable for images as large as 8 MB. The rest of 2.88% transfer is completed within 120 seconds (2 min), and also acceptable. The network traffic load is all under 60% network utilization(750 KB/sec). There is no traffic effect on the transfer delay up to 30% (375 KB/sec) utilization, and after that there is a trend of increasing transfer delay with the network load, however the trend is very small compare to the variation of network delay. Hence, again, for images of 8 MB, the network traffic load is not a dominant factor.

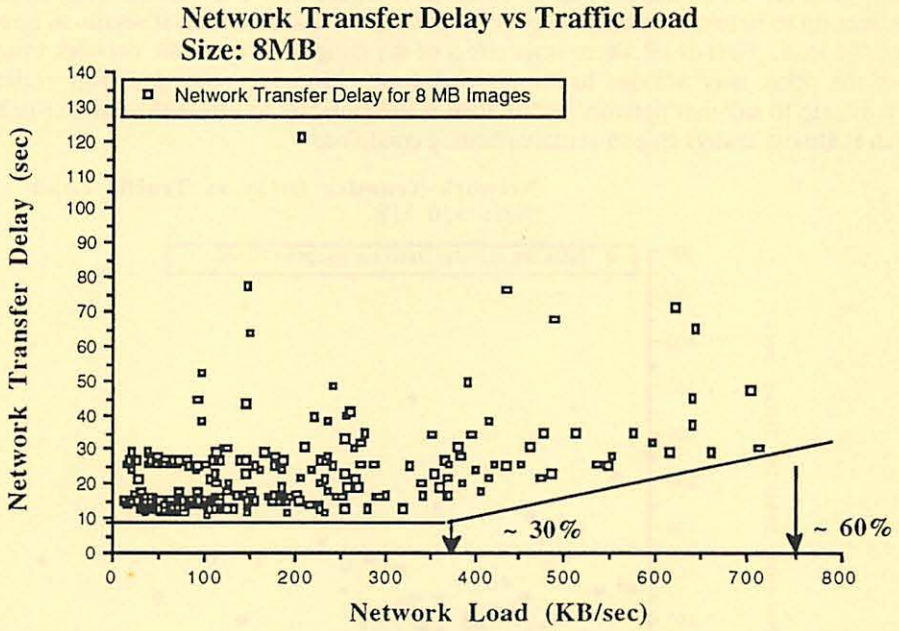


Fig. 10. Statistics of network transfer delay vs. traffic load for CR images at 8 MB. The data is collected over two weeks.

Since our images have a large range in size, it is necessary to investigate files larger than 10 MB. The result is plotted in figures 11 and 12, and both figures represent same piece of data. From the figures, 98.18% (10 out of 354 cases) of the transfer is completed within 3 minutes which agrees with the result we obtained in figure 5, and the rest of 2.82% completes transfer within 7.25 minutes. We are going to discuss about it in more detail later. Figure 11 shows the image size distribution over the network transfer delay. We could see that the images have a broad range starting from 10 MB to almost 90 MB. There is a again a baseline observed with the transfer rate around 700 KB/sec, and it is very clear that transfer delay linearly increases with the images within its statistical variation.

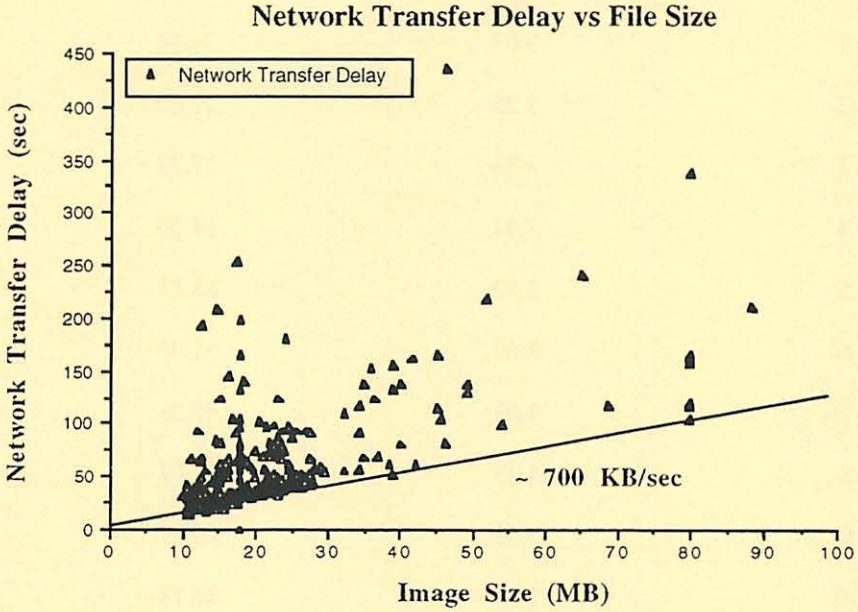


Fig. 11. Effect of file size to the network transfer delay for images larger than 10 MB. Collected over 2 weeks.

In figure 12, we obtain the similar result as in figure 10. The network load never exceeds 60%. There is no network effect up to network utilization 30% (375 KB/sec), and after that it seems to have a trend of increasing delay with the traffic load. First of all, there is an effect of the image size on to the network transfer delay, hence the trend of increase of the delay may not due to the network load. Moreover, the statistical variation is much larger than the increase, it is safe to say that network traffic load is insignificant to affect the transfer delay for the traffic level below 60%, which is almost always true in actual operating conditions.

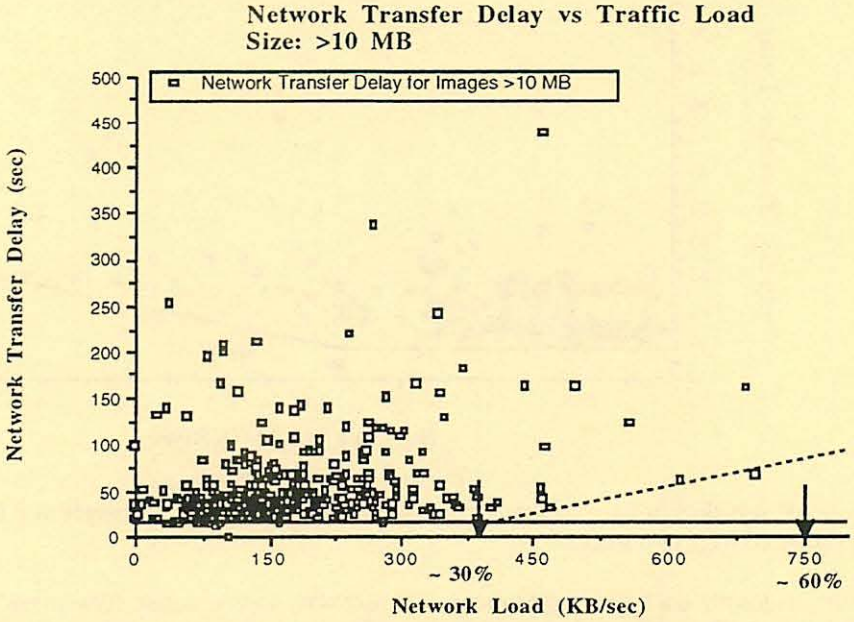


Fig. 12. Network transfer delay vs. the network load for images greater than 10 MB. The file size is illustrated in fig. 11. The statistics is collected over two weeks.

Case Number	Network Transfer Delay (minutes)	Image File Size (MB)	Alias for Destination Computer
1	3.03	24.14	research 1
2	3.25	12.60	faraway 1
3	3.33	17.77	research 1
4	3.47	14.70	faraway 2
5	3.52	88.17	large_file 1
6	3.67	51.96	large_file 1
7	4.03	65.08	large_file 1
8	4.23	17.32	faraway 2
9	5.63	79.77	large_file 1
10	7.30	46.18	faraway 2

Table 1. An illustration of the 10 cases which all exceeds 3 minutes transfer delay

There are 2.82% cases where the transfer delay exceeds 3 minutes in figure 11 and 12, and they are the important information to complete as well as to improve our statistical model. By polling the relevant information from the queue files, we obtain the table above. In table 1, generally speaking, all the above cases are affected by three types of factors. They are 1) large image sizes, 2) complication in network topology and distance, and 3) experimental setup. In cases 5, 6, 7, and 9, their image sizes are all greater than 50 MB, hence their transfer delay is becoming excessive from 3.52 to 5.63 minutes. These cases follow the relationship reported above, the larger the image size, the longer the delay is. For cases 2, 4, 8, and 10, the destination computer for the network transfer is either far away in distance or complex in its topology (multiple bridges and routers connections), so the network transfer delay is above normal.[15] The remaining two cases 1 and 3 are simply because the destination computer "research 1" is a research computer, and it has a less stable environment.

4. CONCLUSIONS

We have developed a methodology to predict the image retrieval delay based on the OD access delay and the network transfer delay. The OD delay is basically follow the linear relationship with the file size. The network transfer delay has many elements to consider. We have determined that image size is the most important factor to affect the delay, and it generally follows a linearly relationship with the delay within numerical variation. The network load is not an important factor to be considered as long as the network utilization does not exceed 60%. The position and its network topology of the destination computers also contribute to the network transfer delay. Normally, for images smaller than 50 MB, the OD retrieve should be done within 4 minutes and the network transfer delay should be completed within 3 minutes.

5. REFERENCES

1. R. K. Taira, K. K. Chan, B. K. Steward, and W. S. Weinberg. Reliability Issues in PACS. Proceedings of Medical Imaging V: PACS Design and Evaluation, SPIE, 1446:451-458, 1991.
2. D. T. Chen, B. K. T. Ho, W. Chao, and R. K. Taira. Image Flow Management in a PACS Network. Proceedings of Medical Imaging: PACS Design and Evaluation, SPIE, 2165:233-240, 1994.
3. K. K. Chan and R. K. Taira. System Integration for PACS. Computerized Medical Imaging and Graphics, 15(3):177-181, 1991.
4. A. W. Wong, R. K. Taira, and H. K. Hung. Implementation of a Digital Archive System for a Radiology Department. Proceedings of Medical Imaging VI: PACS Design and Evaluation, SPIE, 1654:182-190, 1992.
5. H. K. Hung and R. K. Taira. Infrastructure Design of a Picture Archiving and Communication System. AJR, 158: 743-749, April 1992.
6. B. K. Steward, R. K. Taira, S. J. III Dwyer and H. K. Huang. Acquisition and Analysis of Throughput Rates for an Operational Department-Wide PACS. Proceedings of Medical Imaging VI: PACS Design and Evaluation, SPIE, 1645:24-38, 1992.
7. AIX NetView/6000 At a Glance, Version 1, GC31-6175-00, 1992.
8. W. Stallings. SNMP, SNMPv2, and CMIP: The Practical Guide to Network-Management Standards. Reading, MA, Addison Wesley, 1993.
9.M. T. Rose . The Simple Book: An Introduction to Management TCP/IP - based internets. Englewood Cliffs, NJ, Prentice-Hall, 1991.
10. Expert Sniffer Network Analyzers, Network General Corporation, GSOOK92AGS6109 PS08, 1994.
11. W. R. Stevens. Unix Network Programming. Englewood Cliffs, NJ, Prentice-Hall, 1990.
12. Introduction to Local Area Network. Digital, 1982.
13. D. Comer. Internet Working with TCP/IP. Englewood Cliffs, NJ, Prentice-Hall, 1988.
14. B. W. Stuck, and E. Arthurs. A Computer & Communication Network Performance Analysis Primer. Englewood Cliffs, NJ, Prentice-Hall, 1985.
15. Introduction to Network Performance. Digital Nac Publication, 1987.

Preprint from SPIE: PACS Design and Evaluation 2435-51

Clinical Results of HIS, RIS, PACS Integration Using Data Integration CASE Tools

Ricky K. Taira, Hing Ming Chan, †Claudine M. Breant, Lu Huang, Daniel J. Valentino

Department of Radiological Sciences
†Department of Surgery
University of California, Los Angeles

1.0 Introduction

Current infrastructure research in PACS is dominated by the development of communication networks (local area networks, teleradiology, ATM networks, etc.), multimedia display workstations, and hierarchical image storage architectures. However, limited work has been performed on developing flexible, expansible, and intelligent information processing architectures for the vast decentralized image and text data repositories prevalent in healthcare environments. Patient information is often distributed among multiple data management systems. Current large-scale efforts to integrate medical information and knowledge sources have been costly with limited retrieval functionality [Roder92]. Software integration strategies to unify distributed data and knowledge sources is still lacking commercially. Systems heterogeneity (i.e., differences in hardware platforms, communication protocols, database management software, nomenclature, etc.) is at the heart of the problem and is unlikely to be standardized in the near future [Adrio92].

In this paper, we demonstrate the use of newly available CASE (computer-aided software engineering) tools to rapidly integrate HIS, RIS, and PACS information systems. The advantages of these tools include fast development time (low-level code is generated from graphical specifications), and easy system maintenance (excellent documentation, easy to perform changes, and centralized code repository in an object-oriented database). The CASE tools are used to develop and manage the "middle-ware" in our client-mediator-server architecture for systems integration [Wiede92, vanMu92, vanMu93]. Our architecture is scalable and can accommodate heterogeneous database and communication protocols.

Figure 1 shows the environment at UCLA, consisting of the three independent heterogeneous clinical information systems. Our goal was to provide a simple interface to these systems which hides the heterogeneous and distributed nature of the data from the user. Thus, the user need not be burdened with knowledge of network addresses, machine names, passwords, and numerous database navigation paradigms and languages. The following were the clinical requirements of our implementation for an integrated HIS/RIS/PACS system:

1. Simple Interface; Little or No Database Knowledge Required
2. Hide System Heterogeneity and Distributed Nature of Data from User
3. Read-only Access
4. Fast Response to Data Requests
5. No Special Accounts or Privileges Given on server machines

TAIRA

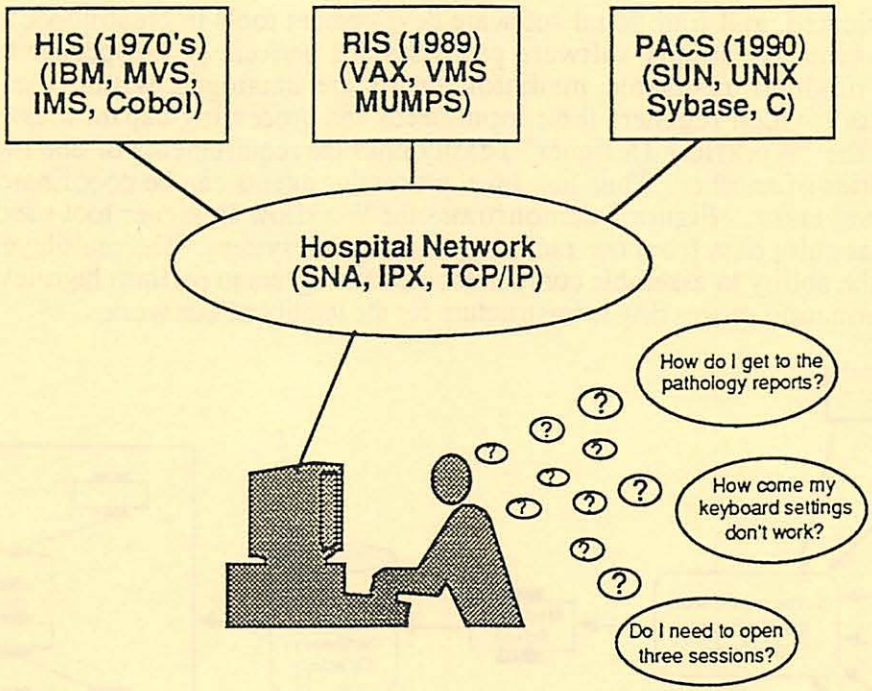


Fig.1 - *UCLA Environment and recurring clinical and research data access problems.*

The following were the technical requirements of our implementation for an integrated HIS/RIS/PACS system:

1. Cannot Affect Operation of HIS, RIS, and/or PACS Databases
2. Must be Scalable and Extensible
3. Must be Easily Adaptable to System and Technology Changes
4. Limited Query Processing (Search by Patient ID and Date)
5. Fast Prototyping

2.0 Materials and Methods

A central issue of our research is the design of an architecture that is open, and that can easily evolve given new data access methods, representations, and storage algorithms. We are currently developing a multi-layered information processing architecture [Carde93, Taira94a] which consists of six layers: (1) user interface, (2) query interpretation, (3) query modification, (4) document / object index, (5) data access / distribution, and (6) autonomous sources. In this paper, we report only on the development of the data access distribution layer. Within each layer exist task-oriented mediator agents that perform such information processing tasks as network communication, database navigation, data storage, data summarization, and data translation.

2.1 System Architecture

Figure 2 summarizes the development of these mediator agents. We use a variety of

TAIRA

CASE, object-oriented, and traditional software development tools in creating various mediator agents. These agents are simply software processes that perform some specific task as stated above. These modular re-usable mediator agents are cataloged within the global data dictionary/directory which registers their input needs and processing capabilities. A graphical interface (called the "Workflow Designer") easily links the requirements of one mediator agent with the capabilities of another. Thus low level processing agents can be coordinated to perform complex high-level tasks. Figure 3 demonstrates the Workflow Designer tool used to link data access agents to acquire data from our radiology information system. The multilayered mediator architecture and the ability to assemble component mediator agents to perform high-level operations serves as the information processing infrastructure for the totality of our work.

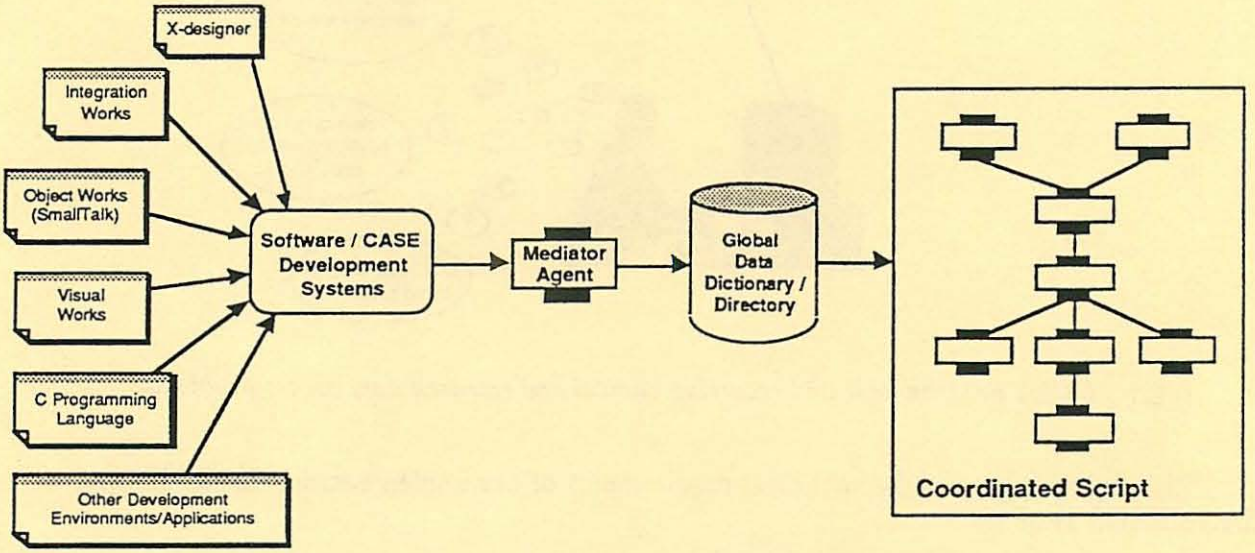


Fig. 2 - Creation, registration, and coordination of information mediator agents within our system. Powerful data integration CASE tools are central to our development philosophy.

2.2 CASE Tools

Previously, few commercial tools or methodologies were available to assist developers integrate database systems. Development time, code complexity, system scalability, and system maintenance are major issues. Recently, the availability of data integration CASE (Computer-Aided Software Engineering) tools has addressed some of these issues [Integ94]. We used the graphical CASE tools Integration Works, (Data Integration Solutions Corporation, Gardena, CA) to serve as our software development platform. Figure 3 shows the four classes of data integration tools we will use: (i) global schema designers, (ii) communication utilities to access data on remote database systems, (iii) data translations, and (iv) jobflow designers. Each tool is designated to allow the user to graphically specify required information to complete the data integration task. These high level specifications are then automatically translated into C++ code. The capabilities of each developed module is registered into an object-oriented data dictionary/directory which manages all available mediator modules (communication scripts, translators, high-level strategies, target and global schemas). With these tools, potentially any database system can be integrated. A brief description of these tools follows. Further details can be found in [Taira94b].

TAIRA

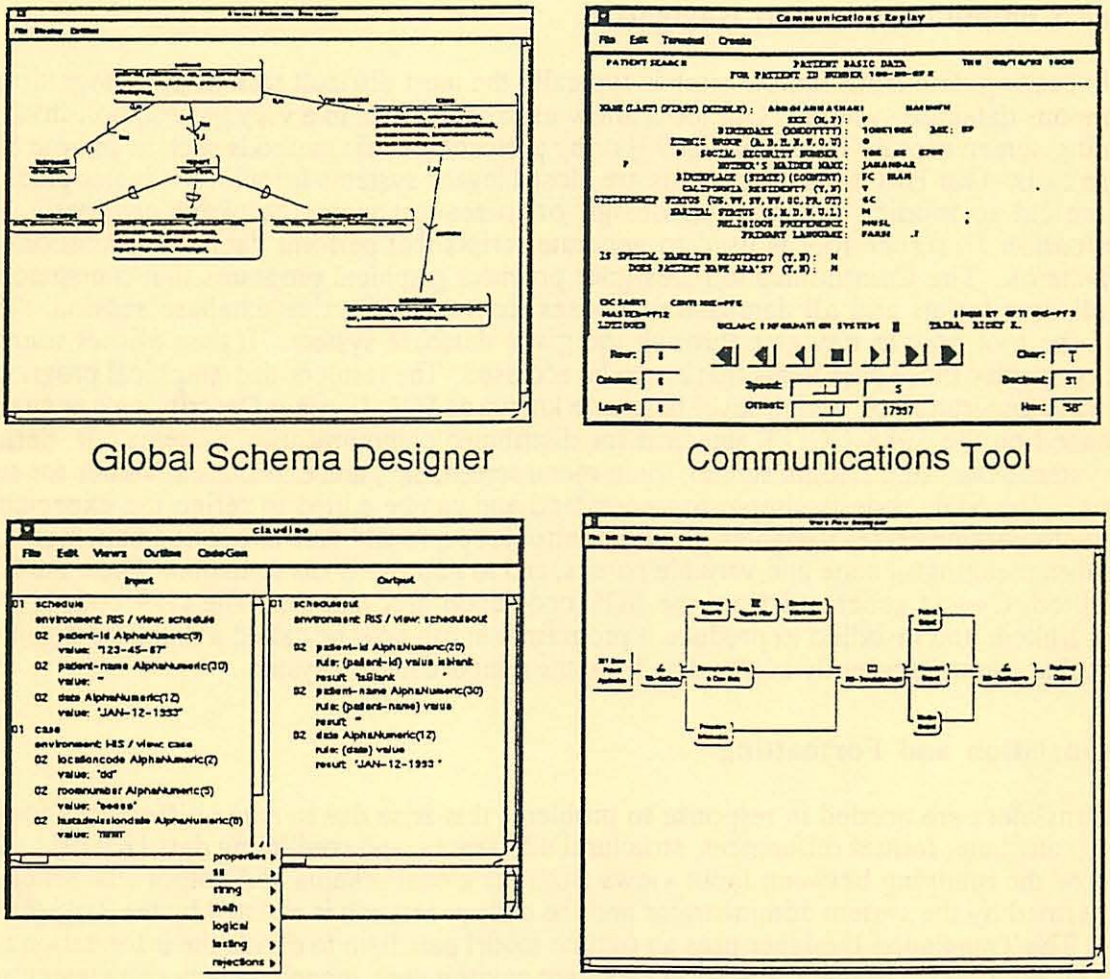


Fig. 3 - *Graphical data integration CASE tools used to rapidly construct C++ code for database integration applications.*

Global Schema Designer

Global schemas can be large and may be too complex to be easily understood without assistance. A graphical representation of the global data aides in the location of distributed information [Sadre93]. The *Schema Designer* is an extended entity relationship (EER) diagramming tool. Local schemata from component databases will be diagrammed using this tool. Input views from the different component databases can then be merged by selecting graphical entities and dragging arrows to interrelate them with other entities. Relationships (aggregation, generalization, and association) and cardinalities can be specified at this stage. Both forward and reverse relationships and cardinalities can be specified. To reduce the complexity of a large ER diagram, entities can be logically clustered into a composite *meta entity*. Key types, the hierarchical level of the data element (for composite attributes), storage length, field type (e.g., alphabetic, numeric, etc.), number of occurrences (for duplicate data items), a sample value, a default value if any, alias names, mappings to semantically equivalent descriptions, constraints as described by integrity and/or conversion rules, and a narrative description are stored.

Database Communication and Navigation

Accessing data from remote sites is typically the most difficult task in the integration of heterogeneous database systems. Our tools allow us to access data in a very general non-invasive matter using screen capture methods [Lee90] or by protocol specific methods such as remote SQL procedure calls. Our HIS and RIS systems are closed legacy systems for which it is not practical or economical to modify the internal design or to request major software changes. The *Communication Designer* tool is used to generate scripts that perform database operations on remote systems. The Communication Designer provides graphical programs that transparently capture all user inputs and all database responses from an interactive database session. This "teaches" the tool how to navigate through the given database system. It then allows users to graphically specify those data items that are to be accessed. The result of this graphical program is the automatic generation of a high-level language known as *SDL* (System Description Language). SDL is based on the 1988 CCITT standard for distributed communication systems. It defines various system states (e.g., login screen, main menu screen, etc.) and expectation values for state transitions. The SDL code is simple to understand and can be edited to refine the expectation conditions, to parameterize variables, to add control loops, to add data extraction statements, to provide more meaningful state and variable names, and to improve documentation. Once the SDL code is edited, C++ is generated from the SDL code upon user request. The C++ code is then compiled, linked, and installed to produce a program that can now be called a database agent (or mediator), that can transparently access data from the remote database system.

Data Translation and Formatting

Translators are needed in response to problems that arise due to name differences for the same entity attribute, format differences, structural differences, and conflicting data [Kim91]. The definition of the mapping between input views from the global schema and output custom client views is defined by the system administrator and the code generation is assisted by the *Translation Designer*. The Translation Designer uses an outline model paradigm to divide the information into input and output views. In the Translation Designer environment, mappings between elements of input views and output views are defined using rules written in SmallTalk. Many useful translation methods are also available as pull down paste functions, (e.g., centering, formatting, arithmetic operations, string operations) thus not requiring the developer to have a knowledge of SmallTalk in most instances. The translation strategy can be tested by inputting a sample value for a particular input view attribute and seeing directly the translation result that is presented in the output view column. Once the translation is tested, C++ code can be automatically generated that performs the specified translation between the input and output views.

Work Flow Designer

The *Workflow Designer* allows the developer to construct graphical descriptions of how communication scripts, data files used by various tasks, translation programs, and applications programs (i.e., mediators) are assembled, ordered, and interfaced to complete the entire data integration strategy. The user can graphically view different mediator modules (e.g., playback communication scripts, SDL programs, global schemas, translation outlines, etc.) from this tool. The graphical representation allows the developer to visualize the flow of data and control for the distributed set of tasks that may run in parallel. The Workflow Designer is used to generate configuration files that control the run-time environment of the communication scripts, the task specific parameters for each of the various tasks, and the overall workflow specification.

3.0 Results

The execution environment consists of *clients* that are the requesters for information, *mediators* that serve to hide complex protocols, database languages, and navigation schemes from the clients, and *servers* that actually access the target databases (Fig. 4). The global HIS, RIS, PACS mediator module is network accessible and dispatches generic global information requests to the appropriate database servers or mediators in the appropriate language. The layers of mediation encapsulate the complex protocols required to access data from remote sites shielding the user from such burdens.

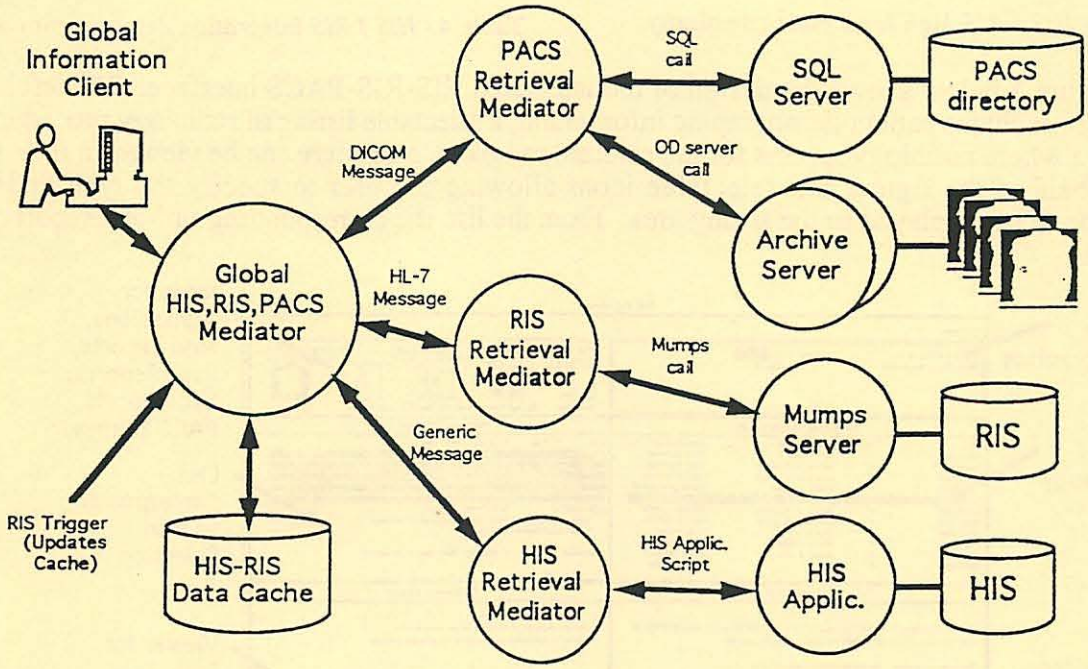


Fig 4. - Mediator agents provide a layer of encapsulation to lower level database access methods which may be difficult for the user to use.

To improve system performance, we developed a global data cache that stores retrieved HIS and RIS records into a networked relational database management system. Performance is further improved by use of RIS triggers including "Patient Arrival" and "Study Complete" to automatically initiate RIS and HIS prefetches to the cache.

Report Type	Avg. Size	Acq Time(Std)
All HIS Data All RIS Data	17.8 KBytes 20.0 KBytes	110s (167s) 45s (37s)

* Per patient, Non-cache retrievals; Cached retrievals < 3.0 s

Table 1* - *Retrieval performance of HIS-RIS integrated system.*

Report Type	Avg. No. (Std)	Avg. Size (Std)
Lab (HIS)	3.7 (8.3)	1.3 KB (2.2 KB)
Path (HIS)	2.0 (3.3)	3.5 KB (4.5 KB)
Medical (HIS)	2.1 (4.2)	2.9 KB (3.2 KB)
Radiology (RIS)	1.8 (11.5)	1.1 KB (0.8 KB)

Table 2 - *Number of documents per patient for various HIS / RIS reports.*

TAIRA 6

Software Module	Lines of Code
HIS Mediator (SDL)	1876 (81 States)
Global Mediator (C)	354
PACS Mediator (C)	455

* List contains only modules that must be managed. Total includes documentation.

Table 3* - *HIS / RIS high level code complexity.*

Software Module	Man-Months
RIS Interface	2.5
HIS Acccess Script	4.2
Global Mediator	1.0
HIS-RIS Cache	2.5
User Interface	1.0

Table 4 - *HIS / RIS integration development time.*

Figure 5 below shows the design of the integrated HIS-RIS-PACS interface. The left half of the figure includes patient demographic information, a selectable listing of radiology procedures, and an area where radiology reports for the selected radiology procedure can be viewed in full. On the right half of the figure, are selectable icons allowing the user to specify the type of HIS information to be displayed in the listing area. From the list, the corresponding full-text report can be viewed.

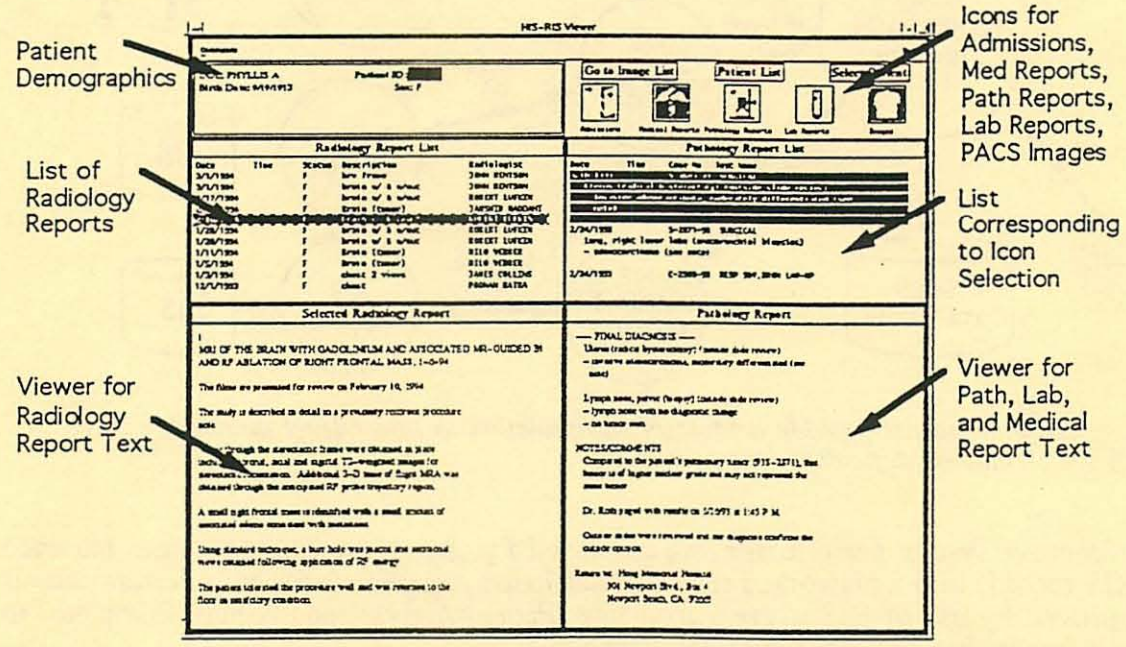


Fig. 5 - Integrated HIS and RIS interface. The right half includes demographics and radiology reports. The left side provides admission records, and medical, pathology, and laboratory reports.

4.0 Discussion

We present a powerful method for developing data integration software for accessing information on various legacy medical information systems. CASE tools were used to efficiently develop code to retrieve information from these systems in a user specified format. The development environment includes graphical tools to specify data integration strategies and to document local and global schemas. Low level code can be directly generated from these high-level communication, navigation, and translation specifications. An object-oriented database

TAIRA 7

provides a disciplined environment for updating and storing the data integration designs.

In conclusion, we were able to develop an integrated HIS-RIS-PACS system rapidly (6 months, with 2 FTE's), with good system response time in most cases (<3s from cache). The major advantages of our approach are its generality (potentially any communication/database protocol can be dealt with), maintainability (changes can be performed at a high level), and it's non-invasive approach (no code had to be installed on the HIS, RIS, or PACS servers).

Currently, the system has limited query processing capabilities. Patient's data can only be accessed via patient hospital identification number and a data range of interest. We are developing an intelligent indexing system to RIS, HIS, and PACS documents using a free-text analysis of radiology and pathology reports. The method uses UMLS (National Library of Medicine's Unified Medical Language System)-like thesauri for radiology [Humph91, Bhush94]. The goal is to retrieve patient documents based on specification of medical concepts.

5.0 References

[Adrio92] W. R. Adrion, E.D. Lazowska, A. van Dam, "From Discipline in Crisis to Mature Science: Evolving Needs for Computing Research Infrastructure," *Computer*, 25(12):18-24, 1992.

[Bhush94] V. Bhushan, R.K. Taira, H.M. Chan, "Application of the Unified Medical Language System Metathesaurus to Intelligent Radiological Case Retrieval", Proceedings of the SPIE, 2165:336-346, 1994.

[Carde93] A.F. Cardenas, R.K. Taira, W.W. Chu, C.M. Breant, "Integration and Interoperability of a Multimedia Medical Distributed Database Systems", *IEEE Data Engineering*, 16(2):43-47, April 1993.

[Humph91] B.L. Humprhreys, D.A.B. Lindberg, W.T. Hole, "Assessing and Enhancing the Value of the UMLS Knowledge Sources," *Proc. of the Computer Applications in Medical Care*, :78-82, 1991.

[Integ94] Integration Works User Guide, Data Integration Solutions Corp., 1994.

[Kim91] W. Kim, J. Seo, "Classifying Schematic and Data Heterogeneity in Multidatabase Systems", *Computer*, December 1991.

[Lee90] W. Lee, A.H. Rowberg, Y.M. O'Leary, Y. Kim, "Integrating a Radiology Information System with a Picture Archiving and Communication System", *Proc. of the SPIE*, 1234:661-669.

[vanMu92] E.M. van Mulligen, T. Timmers, F. van den Heuvel, "A Framework for Uniform Access to Data, Software and Knowledge", *AMIA*, :496-500, 1992.

[vanMu93] E.M. van Mulligen, T. Timmers, J.H. van Bemmel, "A New Architecture for Integration of Heterogeneous Software Components," Methods of Information in Medicine, 32:292-301, 1993.

[Roder92] N.K. Roderer, P.D. Clayton, "IAIMS at Columbia-Presbyterian Medical Center: Accomplishments and Challenges," *Bulletin of Medical Library Association*, 80:253-262, 1992.

[Sadre93] M.H. Sadreddini, D.A. Bell, I.R. Young, "Architectural Considerations for Integrating HIS/RIS/PACS in a Distributed and Heterogeneous Environment", *Proceedings of the SPIE*, 1899:518-527, 1993.

[Taira94a] R.K. Taira, A.F. Cardenas, W.W. Chu, C.M. Breant, J.D.N. Dionisio, C-C Hsu, I.T. Ieong, "An Object-Oriented Data Model for Skeletal Development", Workshop on Applications of Object-Oriented Modeling, *Proceedings of the SPIE*, February 1994.

[Taira94b] R.K. Taira, C.M. Breant, F.M Stepczyk, H. Kho, D.J. Valentino, G.Tashima, A.T. Materna, "CASE and Object-oriented Tools for Rapid Integration of HIS, RIS, and PACS", *Proceedings of the SPIE*, 2165:185-199, February 1994.

[Wiede92] G. Wiederhold, "Mediators in the Architecture of Future Information Systems," *Computer*, March:38-49, 1992.

TAIRA

Image Navigation for PACS workstations

Kelvin T. Leung, Bruce K.T. Ho,
Woodrew Chao, Johnny Chao, Ramesh K. Panwar[†],
Viskas Bhushan, Zoran Barbaric, Barbara M. Kadell-Wootton,
John R. Bentson, Leanne L. Seegar, Hooshang Kangarloo

Department of Radiological Sciences, Medical Imaging Division,
University of California, Los Angeles, CA 90024-1721

[†]Computer Science Department,
University of California, Los Angeles, CA 90024

ABSTRACT

We present a novel image navigation methodology for PACS viewing stations to handle very high volume studies efficiently. This methodology is based on the "customizable folder" concept in which a scaleable electronic worklist is formulated according to the prefetching algorithm based on radiologist's preference and the "virtual panel" concept which provides an virtually unlimited screen space on the station. Furthermore, the "moving strips paradigm" is presented to serve as an efficient "browsing" method to allow rapid and random access to a selection of studies and a "pre-hanging procedure" during the formation of an electronic work list. Finally, the advantages and applications of such methodology are presented.

Key Words: PACS Diagnostic Viewing Station, Image Navigation, Virtual Panel, Moving Strips Paradigm

1. INTRODUCTION

One of the crucial tasks in the development of PACS viewing stations is to design a user-friendly interface which can allow a radiologist go through an electronic work list at the same speed as a film based system without impeding his or her workflow. The ultimate challenge of PACS viewing stations is to replace the conventional lightbox which is commonly used by physicians to review films [1]. Since a display viewing station cannot ever match the film alternator in screen space, it is necessary to implement a image navigation methodology which provides a more effective reading environment. This paper describes a novel image navigation methodology which is primarily based on 1) the "customizable folder" concept, a design concept of formulating the electronic work list according to the prefetching algorithm along with the radiologist's preference, and 2) "virtual panel" concept, a design concept of navigating the work list efficiently with a foot pedal to combat the limitation in screen space that the current viewing stations cannot provide. Furthermore, the "moving strips paradigm" features an efficient "browsing" method to handle very high volume studies by allowing rapid and random access to a selection of images. In case of the default patient folder composition does not match with the needs of radiologists, the "moving strips paradigm" will serve as a image "pre-hanging procedure" to allow radiologists to rearrange the order of medical images in virtual panels.

2. IMAGE NAVIGATION METHODOLOGY

2.1 CUSTOMIZABLE FOLDER CONCEPT

The "customizable folder" concept borrows the idea from the traditional patient folder in which a patient's image films are spooled into a folder by a film librarian and then delivered to the reading room upon a request made by a radiologist.

Before the reading session processes, the librarian/radiologist will take the image films out of the patient's folder and affix those image films onto the conventional lightbox panels in a preferred order. In order to imitate this film-delivery and film-hanging procedure in our PACS environment, a knowledge-based prefetching algorithm and an automatic image-spooling mechanism must be developed and integrated into the system so that digitized image files stored in the global image archives will be delivered to the designated viewing station through the high-speed communication network. Upon the arrival of image files at the viewing station, the image files will be automatically sorted and spooled accordingly onto "virtual panels" (VP) to act as a prelude to "ready-to-fly" image navigation. A "worklist" of the day will consist of a number of "virtual panels" each associated with one, two or more image files. The preparation of the worklist is done before the radiologist or physician steps into his or her office in the early morning. Figure 1 shows the customizable folder concept. First, new images of a patient arrive at the display viewing station; in order to formulate the worklist, the new images of this patient is subdivided into different folders according to its modality. In each modality folder, images will be spooled into "virtual panels" for single-study or multi-studies comparison mode according to the default viewing sequence or the sequence that the radiologist rearranged with the aid of "moving strips paradigm". Rules for such folder composition can be specific to each radiologist or radiology section. Once the order of the virtual panels (folders) is defined (either done by the default sequence or the manual sorting of the librarian), the worklist of the day is then formulated.

The image navigation is controlled by a foot pedal modeled after the motorized alternator. This "customizable folder concept" design will evolve and expand based on an initial set of simple rules and continual feedback from radiologists and physicians. A general rule would be to identify the current study by date, prefetch the most recent prior study, arrange them in the same virtual panel position and have them advance and retreat together according to the foot pedal input. In section 3, we will discuss more about the applications of this folder concept and its examples.

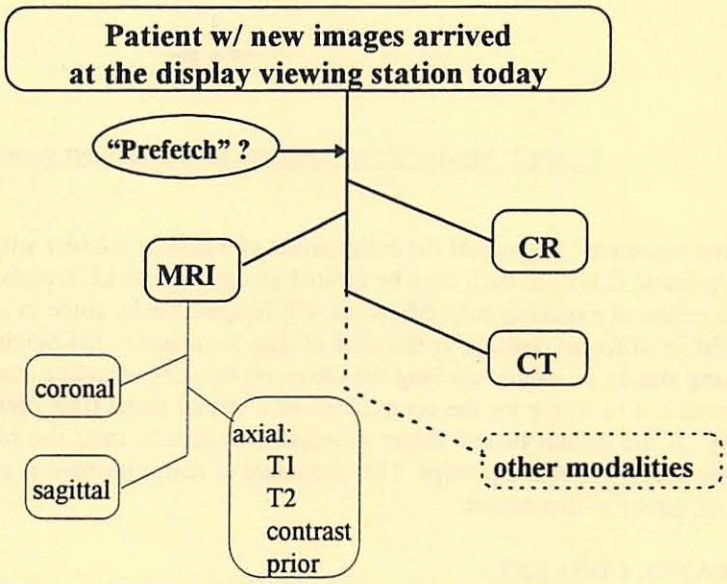


Figure 1: Customizable folder concept for radiologist

2.2 MOVING STRIPS PARADIGM

The "moving strips paradigm" features two different properties in our image navigation methodology. First, it serves as an efficient "browsing" method, especially for very high volume cases acquired by helical CT scanners, to allow rapid and random access to a selection of studies without going through the entire worklist; and second, it provides a mean of facilitating the image "pre-hanging" procedure to allow radiologist to rearrange the images onto different virtual panels if the default composition of a patient's folder fails to meet the radiologist's needs. Figure 2 shows the layout of the "moving strips paradigm" for a 3-monitor display viewing station. In this configuration, the right monitor is used to display the moving strips and the left and middle monitor are reserved for the "virtual panels". For the case of 1K display viewing

station, a moving strip may contain a series of 128x128 pixel icons which are subsampled from the original image file and a screen can display up to eight moving strips per screen page. The screen page can be scrolled up or down by the foot pedal or arrow keys on the keyboard. For MRI and CT images, a single study is represented by a single moving strip, whereas for CR images, a single strip will consist of two CR icons (i.e. two CR image studies). A single strips can be navigated horizontally by turning a dial on the dialbox. For the rapid and random access of selection of studies, this browsing feature allows radiologists to navigate through the entire subsampled images in a fast and rapid manner by turning a dial on the dialbox or stepping on the foot pedal. In addition, the order of image accessing is no longer required to be sequential. If a radiologist wants to exam a particular subsampled image from a moving strip in detail, he or she can click on an icon of that strip on the screen. A full version image associated with that icon will be displayed on the monitors designated for "virtual panels". Moreover, the bracketed icon associated with that moving strip will be synchronized in motion with the full version image. This feature not only provides a random access of full version image based on fast browsing moving strips on screen, but also provides a mean of comparison between synchronized image slices (both in full version and bracketed icon on the moving strip) of similar studies by selecting appropriated moving strips from the right monitor.

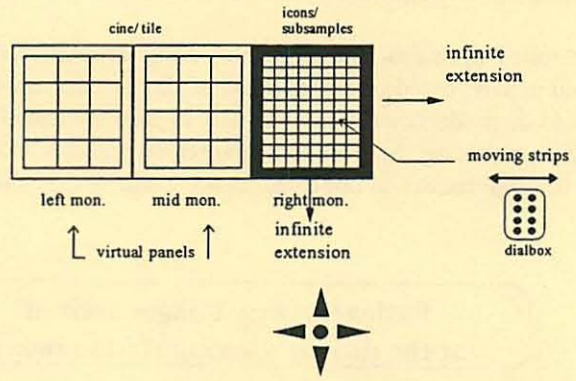


Figure 2: Moving strips paradigm for a 3-monitor viewing station

This "moving strips paradigm" will reveal the entire jacket of a patient's folder with moving strips. Radiologists can easily adjust the Look-Up-Table (LUT) of each strip by turning an appropriate LUT dials (mean or window dials) on the dialbox while pointing the mouse at a moving strip. Moreover, CR images usually come in at different orientations due to the image platter was oriented in different position at the time of data acquisition, radiologists can readjust the orientation of a CR image at the viewing station by simply clicking the mouse on the corresponding icon image. In addition, moving strips can be paired up or remained in single for the composition of a virtual panel (this option is extremely useful for multi-studies comparison mode). If the default patient folder's composition fails to meet the radiologist's needs, it can be changed by reordering the sequence of the moving strips. This procedure of rearrangement is simply accomplished by "drag-and-drop" of a moving strip on screen with a mouse.

2.3. VIRTUAL PANEL CONCEPT

As we mention before, a display viewing station cannot ever match the alternator in screen space. Therefore, we propose the use of "virtual panel" concept to combat this limitation of space. This concept provides a mean of moving a window of visible area (which depends on the viewing station configuration) over the entire panel area (theoretically, it can extend to infinite in size) where images are sitting on it. This concept is most easily illustrated by a 4-monitor display viewing station which has about the same screen space as one panel of the conventional lightbox. The same concept can be extrapolated to the most commonly seen two-monitor workstations [2]. Figure 3a shows a layout of the virtual panels for a 4-monitor display viewing station.

After the "worklist" is generated, each "virtual panel" will be placed in order and have a user-interface button with the patient's name associated with it. The radiologists can navigate through these virtual panels sequentially or randomly. If the system has configured with four monitors, the visible window size would be four-monitor screen size per advancement and

the images displayed on this panel would be advanced together. To further illustrate how the virtual panel concept works, we assume that a system configured with four display monitors (not including the monitor that is used for moving strips paradigm) is used and the layout of the virtual panel is shown in Figure 3a. In this setting, a virtual panel has a window of visible screen of four monitors per advancement [3]. Each monitor on this panel can display the same image files (single-study mode) or four different image files (multi-studies mode). We will discuss single-study mode and multi-studies mode in the following section. The numbers shown on each virtual panel in figure 3a indicate the image filenames that are stored in our electronic "worklist" for each panel and will be used to display on the corresponding monitor at the time of advancement. Referring to the figure 3a, for the sequential image navigation, the virtual panel #1 (in multi-studies comparison mode) containing files 1, 2, 3, and 4 will be displayed on the monitors and the virtual panel will be advanced to the next/previous one upon receiving the "UP" or "DOWN" instruction from the foot pedal. If the instruction is "UP", all the monitors will show the images of file #5 with different slices in single-study mode which is the only image file referenced by the virtual panel #2. Similarly, a particular virtual panel can be accessed by clicking the appropriate button on the user-interface associating with that panel. The image files referenced by that virtual panel will be displayed on the appropriate display monitors accordingly. Furthermore, in the multi-studies comparison mode, images files in each virtual panel may contain a stack of image slices as shown in figure 3b. By the same token as virtual panel, the images shown on these monitors will be advanced and retreated together according to the foot pedal input.

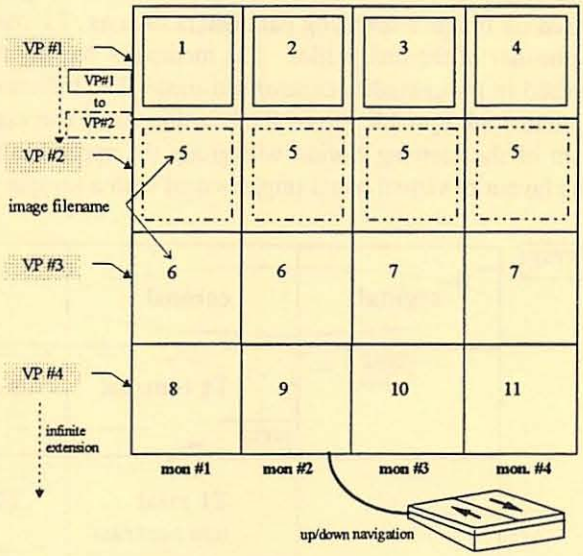


Figure 3a: The layout of virtual panels for a 4-monitor display viewing station.

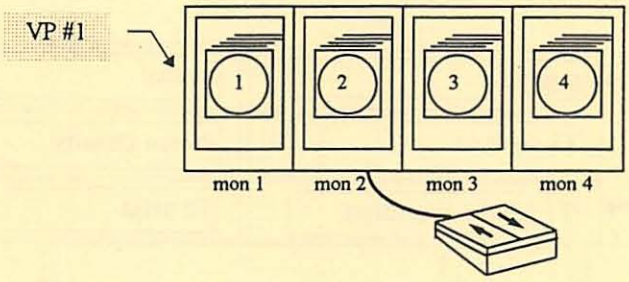


Figure 3b: The close-look of VP#1 where files 1,2,3 & 4 have stacks of image slices

3. DESIGNS OF VIRTUAL PANEL LAYOUT

The virtual panel layout can be designed based on rules for a given composition of a patient's folder. This "virtual panel" concept basically provides two different kinds of viewing mode. They are 1) a single-study mode and 2) multi-studies mode. The single-study mode is a mode that a virtual panel contains only one study. Contrarily, the multi-studies mode allows multiple studies residing on a single virtual panel for comparison. The maximum number of studies that allows in multi-studies mode depends on the configuration of the display viewing station. If the viewing station is configured with two monitors, the maximum number of studies that allows in a virtual panel will be two, i.e. one study per monitor. It is quite obvious that we do not need a very sophisticated rules for single-study mode; on the other hand, more sophisticated or custom-made rules must be used for the multi-studies mode. The formulation of such rules may depend on the radiologist's preferences , the nature of the viewing station (a station used for Neuro-MRI images vs. thoracic CR images etc.), the modality of the image files stored in a virtual panel and so forth. In the following section, we present some general idea of grouping images based on scanning parameters. We show some examples of "virtual panel layout" in conjunction with the three most commonly used modalities of digitized image: MRI, CT, and CR.

3.1 NEURO-MRI VIRTUAL PANEL LAYOUT

An examples of a personal preference for Neuro-MRI images [4] is shown in figure 4 and figure 5. Basically, the images is spooled into virtual panels based on image's scanning parameters such as T1 contrast, T2 contrast and proton density that are usually obtained from the header of the image files. The method of pairing up the images is done by a sample matrix depicted in figure 4 which is used in multi-studies comparison mode. The information stored in the elements of this matrix can be implemented with a small look-up-table. Once the coordinates of the matrix for a set of images is supplied, the automatic spooling mechanism of the viewing station will group the appropriate images into a specific virtual panel for navigation. Figure 5 shows the layout of virtual panel implemented with a sample matrix.

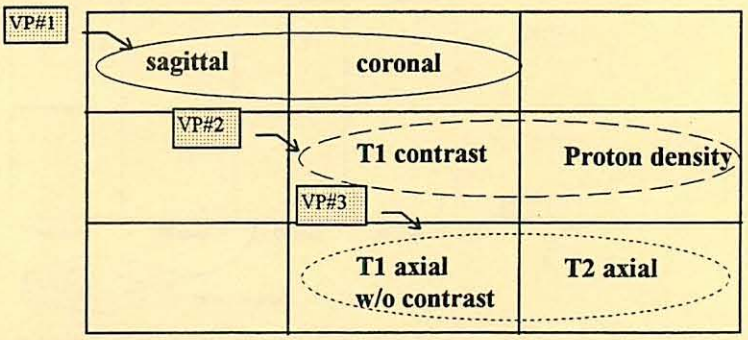


Figure 4: Sample matrix for Neuro MRI images based on scanning parameters

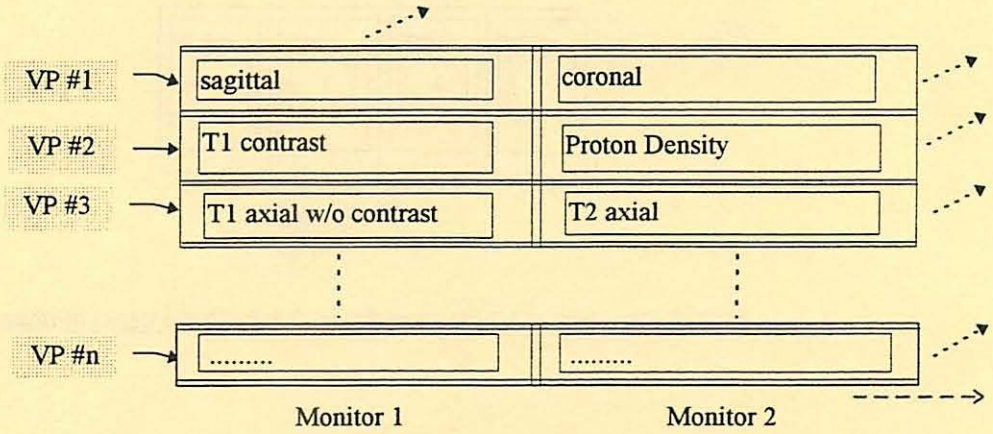


Figure 4: The layout of virtual panels for comparison of multiple MRI images

For single-study mode of MRI images, images can be spooled into virtual panels according to a series of preset sequence. For example, the order of viewing sequence could be sagittal, coronal, proton density, T1 contract, T2 axial, T1 axial w/o contrast and so forth. Figure 5 shows the layout of the virtual panel for single-study mode of MRI images.

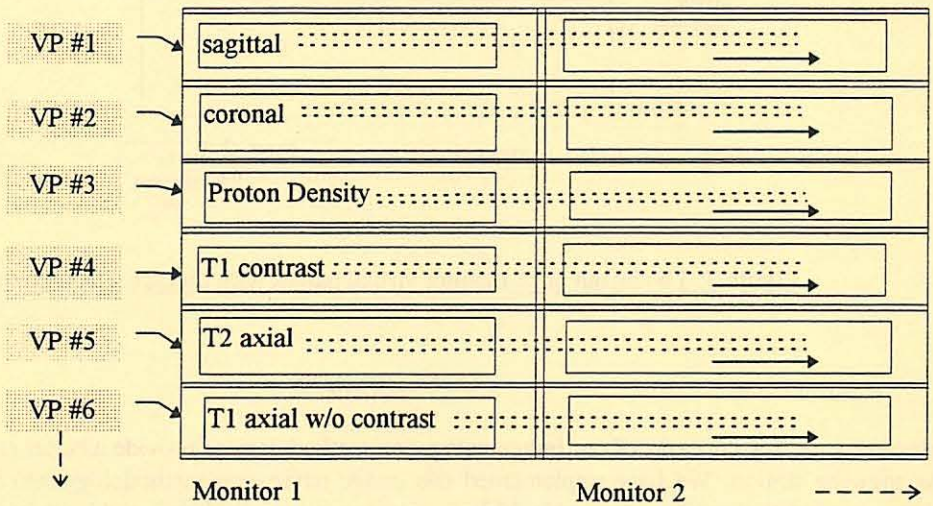


Figure 5: The layout of a single MRI study in each virtual panel

3.2 THORACIC CT AND CR VIRTUAL PANEL LAYOUT

In many cases, the composition of the customizable folder like thoracic CT and CR images can be simpler than that of Neuro-MRI images mentioned above. Basically, we divide the visible window of the virtual panel into two regions. The first half region of the visible window is used to display the current study and the second half region of the visible window is used to display the prior equivalent study for comparison. A layout of virtual panels for CT and CR thoracic images is shown in figure 6 and figure 7 respectively. Notice that in both cases, the current image will be displayed on monitor #1, whereas the prior image will be displayed on monitor #2. Images are grouped into the virtual panels based on the date, time, and scanning parameters like lateral, PA (posterior/anterior), with or without contrast and so forth.

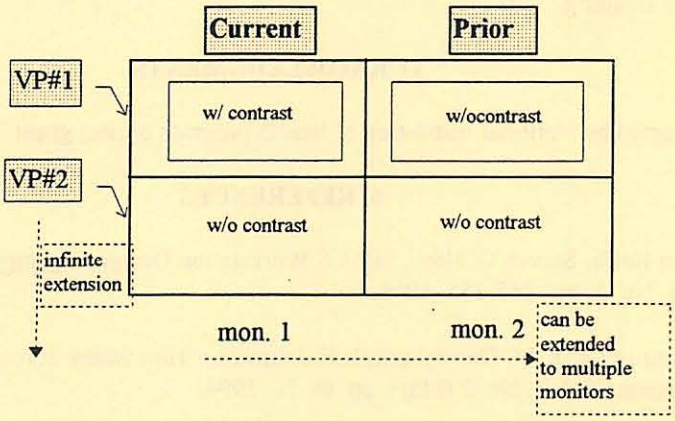


Figure 6: The layout of 2-monitor virtual panels with thoracic CT images

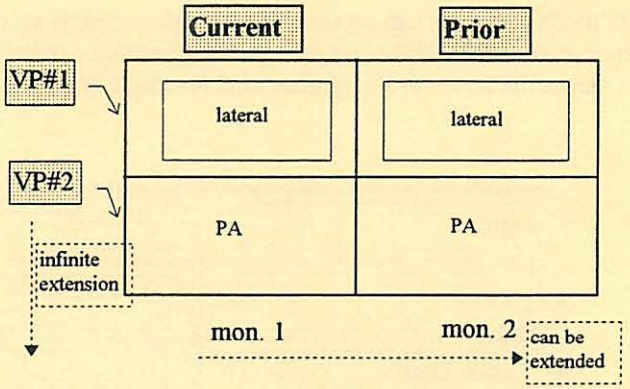


Figure 7: The layout of 2-monitor virtual panels with thoracic CR images

CONCLUSIONS

We have presented some key concepts of our image navigation methodology to provide a better reading environment with a PACS display viewing station. We have implemented this image navigation methodology into the UCLA-PACS viewing station and demonstrated that the "customizable folder", "moving strips paradigm" and "virtual panel" concepts should be the fundamental components in the design of a user interface for PACS viewing station. It is proven that the effects that brought out by this novel image navigation methodology will certainly enhance the existing reading environment in our PACS viewing station. It is achieved by 1) providing an virtually unlimited screen space on the station and a scaleable electronic worklist, 2) eliminating the time consuming process of film hanging on the panels of the conventional lightbox, 3) providing efficient single-study and multi-studies comparison viewing methods, and 4) allowing rapid random or sequential accesses to the virtual panels with a simple navigation tool (i.e. foot pedal or mouse) so that the change in radiologist's practice habit is minimal.

We are able to implement this image navigation features into our UCLA-PACS viewing station because of the capabilities of our hardware. Without a proper hardware design, the use of the display viewing station will become a burden on radiologists and, most importantly, the overall reading efficiency of the display viewing station will not be improved. In order to further increase the reading efficiency, some other factors and research (such as image analysis tool of active contour edge detection and image registration) must be considered and carried out to continually support the trend of prospective digital medical imaging.

ACKNOWLEDGMENTS

This project is partly supported by National Institution of Health program project grant PO1-CA51198

6. REFERENCES

[1] Bruce K.T. Ho, Osman Ratib, Steven C. Horii, "PACS Workstation Design," Computerized Medical Imaging and Graphics, Vol. 15, No. 3, pp. 147-155, 1991.

[2] D.V. Beard, B.M. Hemminger, K.M. Denelsbeck, R.E. Johnston, "How Many Screens Does a CT Workstation Need?" Journal of Digital Imaging, Vol. 7, No. 2 (May), pp. 69-76, 1994.

[3] David V. Beard, Bradley M. Hemminger, J. Randolph Perry, et al, "Interpretation of CT Studies: Single-Screen Workstation versus Film Alternator," Radiology - Computer Applications, No. 187, pp.565-569, 1993.

[4] Dr. John R. Bentson, Private Communication

Implementation of System Intelligence in a 3-tier Tele-Medicine/PACS Hierarchical Storage Management System

Woodrew Chao, Bruce K. T. Ho, John T. Chao,
Reza Sadri, Lu Huang, Ricky Taira

University of California Los Angeles, Department of Radiological Sciences
Medical Imaging Division, Los Angeles, CA 90024

ABSTRACT

Our Tele-medicine/PACS Archive system is based on a 3-tier distributed hierarchical architecture, including Magnetic disk farms, Optical jukebox, and Tape jukebox sub-systems[1]. The Hierarchical Storage Management (HSM) architecture, built around a low cost high performance platform (Personal Computers (PC) and Microsoft Windows NT), presents a very scaleable and distributed solution ideal for meeting the needs of client/server environments such as Tele-medicine, Tele-radiology, and PACS[2]. These image based systems typically require storage capacities mirroring those of film based technology (multi-terabyte with 10+ years storage) and patient data retrieval times at near on-line performance as demanded by radiologists. With the scaleable architecture, storage requirements can be easily configured to meet the needs of the small clinic (multi-gigabyte) to those of a major hospital (multi-terabyte). The patient data retrieval performance requirement was achieved by employing System Intelligence to manage migration and caching of archived data. Relevant information from HIS/RIS triggers prefetching of data whenever possible based on simple rules. System Intelligence embedded in the Migration manager allows the clustering of patient data onto a single tape during data migration from optical to tape medium. Clustering of patient data on the same tape eliminates multiple tape loading and associated seek time during patient data retrieval. Optimal tape performance can then be achieved by utilizing the tape drives high performance data streaming capabilities thereby reducing typical data retrieval delays associated with streaming tape devices.

1. INTRODUCTION

With the introduction of digital radiology we can see a very large increase in the amount of storage required to store patient data. A typical 2Kx2K CR image occupies 8,388,608 Bytes of storage space, and a typical CT scan with 20 slices can occupy up to 10,485,760 Bytes of storage space. With the obvious advantages of digital radiology (on-line access to patient data, reduction in film processing/storage cost), we see more hardware acquisition vendors moving towards a digital interface for image capture and the need to develop a storage hierarchy capable of storing this new generation of multi-media data. The architecture discussed here has the capacity of storing 10+ years of data with near-line performance. Typical HSM features include automatic management of data storage, migration (MD->OD->Tape), and caching (MD<-Tape) of data

(Figure 1). The core concept behind HSM is a migration/caching scheme where new data resides in fast, more expensive media, and migrate towards slower and cheaper media as the data ages and the probability of access decreases[3].

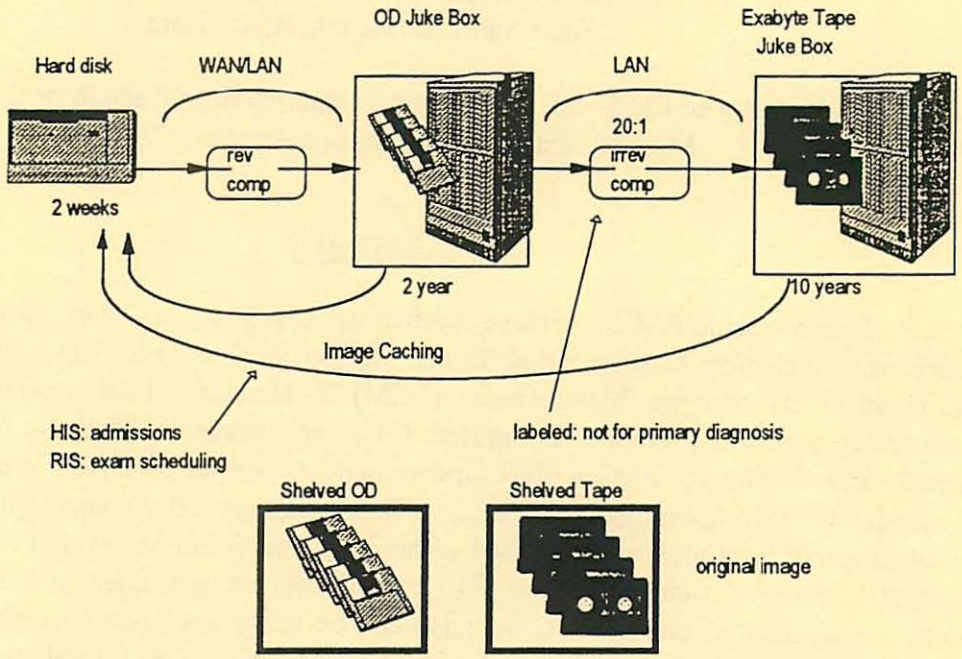


Figure 1

2. Architectures

2.1 HSM Hardware Architectures

There are multiple HSM architectures available today, ranging from 2-Tier HSM - Magnetic Disk and Optical media, or Magnetic Disk and Tape media, and 3-Tier HSM - Magnetic Disk, Optical, and Tape media. Although many existing PACS rely on optical disk, the user must incur the same medium cost for ODs even though the OD may stay on-line for a short part of the archive period[4]. The choice of hardware architecture relies heavily upon long term storage and performance requirements. A cost analysis of media and jukeboxes shows that tape is more cost-effective than ODs. The 3-tier HSM allows for minimal cost, yet provides an almost infinite data storage repository because of its use of tape archival (see 4.3 COST). Implementation of System Intelligence such as image prefetching and patient data clustering provides the HSM with optimizations which yields on-line performance during patient image retrieval.

2.2 HSM Software Architectures

The software architecture of and HSM can be configured for manual, operator assisted, or automatic migration/caching of patient data. Migration/caching schemes such as Least Recently Used (LRU), watermarks, RIS/HIS triggers can be used. Images can be retrieved via (1) Rule Based Image Routing, (2) Prefetching Triggered by Admission and Exam Scheduling, and / or (3) Ad-Hoc Retrieval of Prior Images From Central Archive by technologist (Figure 2.)

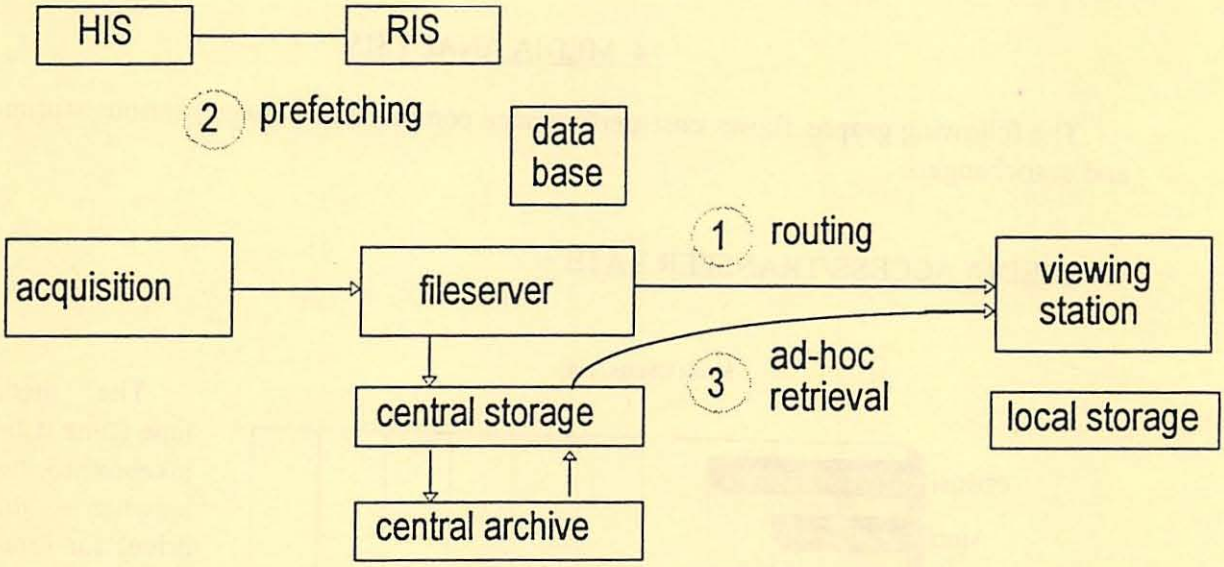


Figure 2 Distributed Image Storage

3. SYSTEM INTELLIGENCE

The key in developing a HSM for Tele-medicine/PACS is to incorporate a 3-Tier HSM utilizing a Disk Farm (2 week capacity), Re-writable MO (2 year capacity), and Tape (10 year capacity) hardware with software enhancements such as Patient Data Clustering and Intelligent Data Prefetching. With this combination of software and hardware integration, we can achieve >90% cache hit ratio on patient data retrieval, minimal media change request, and maximum tape drive performance, thereby eliminating the effects of retrieval delays due to tape streaming devices.

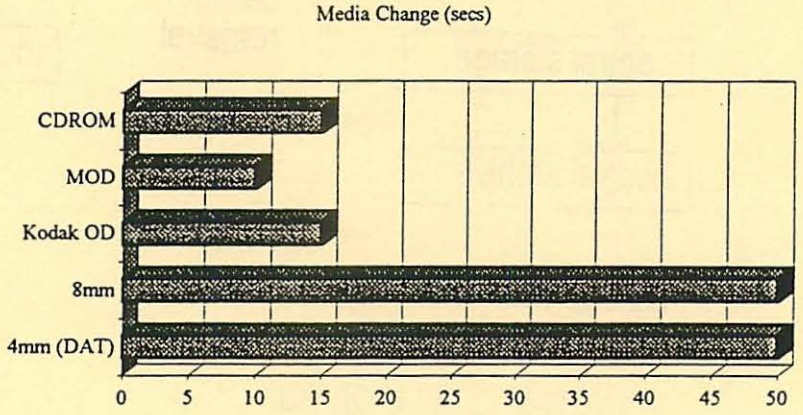
Patient data clustering involves the clustering of data as information is migrated from MO to tape. This strategy allows for the clustering of an individual patient's data to reside "clustered" together on a single tape which will allow for maximum tape throughput during data retrieval. Clustered data on tape allows the tape device to stream data off the medium at maximum transfer rates. In 8mm tape devices, this data can be "burst" to the host at 4MB/sec or 0.5MB/sec sustained [Exabyte Corp., Boulder CO].

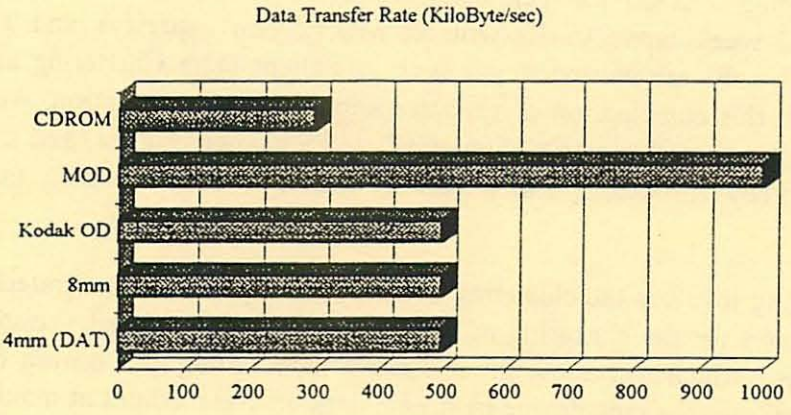
Patient data prefetching involves receiving "hints" from HIS/RIS to establish "triggers" for retrieving data prior to actual reading. Hint from HIS are typically events such as patient admission, discharge, and transfer, and RIS events such as procedure completion and report approval[5]. The lead time for these triggers is usually days or hours which is sufficient for prefetching of prior images.

4. MEDIA ANALYSIS

The following graphs shows cost-performance comparisons between various storage mediums and autochangers.

4.1. MEDIA ACCESS/TRANSFER RATE

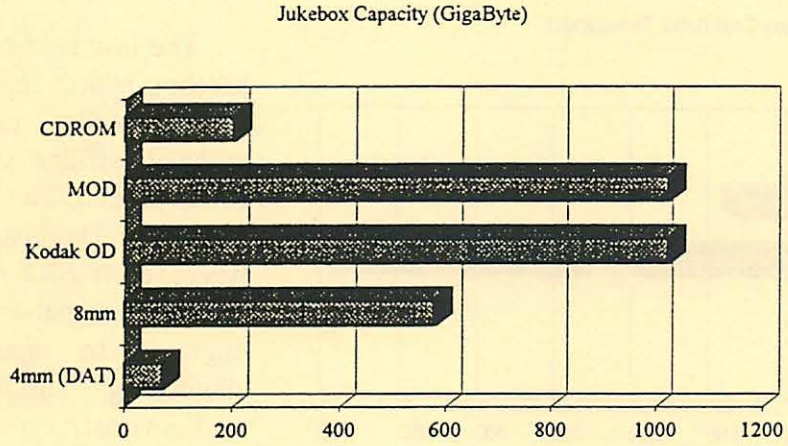


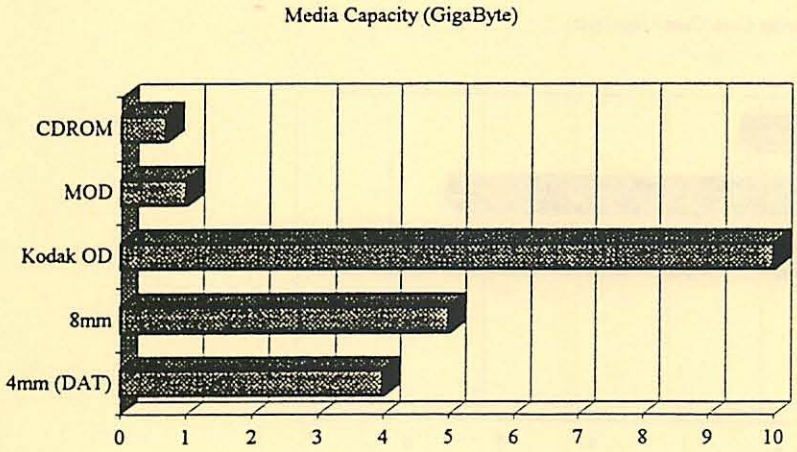


The media access time (time it takes for the jukebox to move media between holder and drive) for random access devices are typically much faster because data is immediately available once the medium is inserted.

The media transfer rate displays sustained data transfer rate for each media type. Because of the continued enhancements to MOD technology, we can see sustained transfer rate for MOD achieve greater than 1MB/sec (transfer rate close to that of Magnetic Disk).

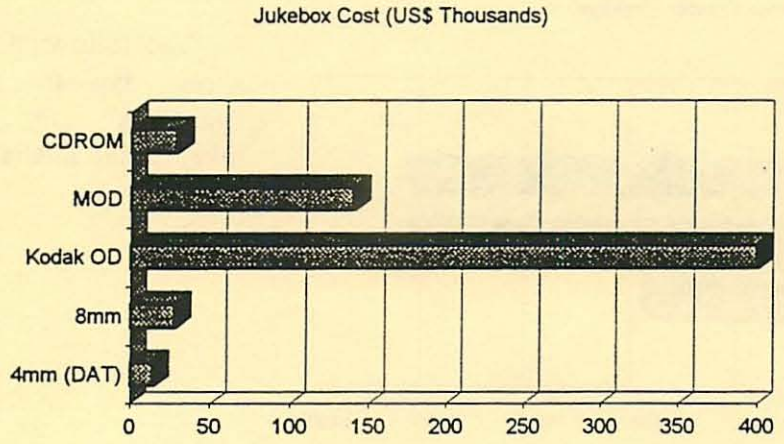
4.2. STORAGE CAPACITY



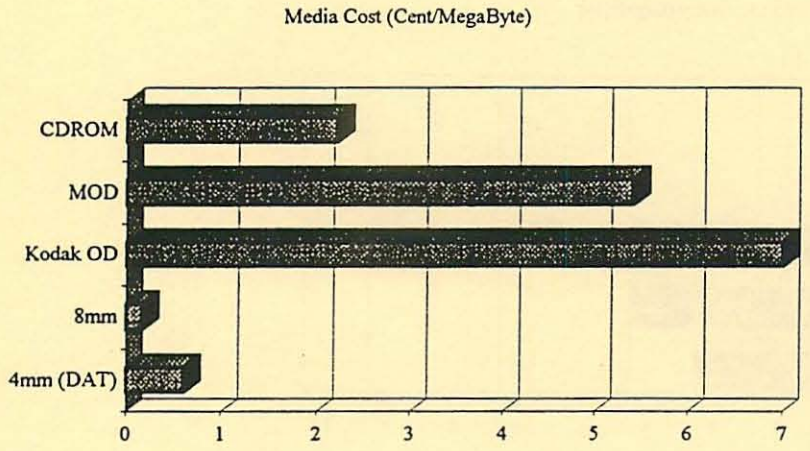


The following graphs show typical storage capacities for both jukebox and media.

4.3. COST



The low cost of tape and jukebox makes tape the ideal medium for long term archival/ off-line storage. A MOD jukebox can be used as secondary archive medium where each MO can be re-used when patient data is migrated to tape, thereby eliminating sustained OD medium cost.



5. CONCLUSION

In conclusion, a Hierarchical Storage Management system incorporating System Intelligence, based upon low cost high performance personal computers (PC) and common-of-the-shelf (COTS) Optical/Tape Jukeboxes, can be implemented at a very minimal cost with performance and storage capacity equivalent to or exceeding that of more traditional multi-million dollar systems. Use of patient information clustering on sequential access devices and intelligent image prefetching is critical to successful implementation of a Hierarchical Storage Management System.

6. REFERENCES

[1] Ho BKT, Sadre R, Chao W, Huang L, Taira RK. Strategy for long term PACS archive using hierarchical storage with image compression and HIS/RIS triggers. *J. Digital Imaging*, 1995

[2] Ho BKT, Taira RK, Steckel R, Kangarloo H. Technical considerations in planning a distributed teleradiology system. *Telemedicine Journal*, March 1995

[3] Miller, SW:Review of IEEE-CS MSS reference model. Digest of Papers, *Tenth IEEE Symposium on Mass Storage Systems*. Crisis in Mass Storage (Cat. No. 90CH2844-9), IEEE Comput. Soc. Press, Monterey, 1990, p3-5.

[4] Tighe LC:Digital archiving. *Medical Imaging* 9(2):52-57, 1994.

[5] Huang HK, Taira RK:Infrastructure design of a Picture Archiving and Communication System. *Am J Rad* 158:743-749, 1992

Automatic Segmentation of Bones from Digital Hand Radiographs

Brent J. Liu, Taira, Shim, Keaton

Medical Imaging Division, UCLA Department of Radiological Sciences
Los Angeles, CA 90024-1721

The purpose of this paper is to develop a robust and accurate method that automatically segments phalangeal and epiphyseal bones from digital pediatric hand radiographs exhibiting various stages of growth.

The method is tested on 1,200 phalangeal and epiphyseal objects to be segmented from 50 cases of pediatric hand images obtained from our clinical PACS. Patient age ranges from 2-16 years. The algorithm uses an object-oriented approach comprising of several stages beginning with the most general objects to be segmented, such as the outline of the hand from background, and proceeding in a succession of stages to the most specific object, such as a specific phalangeal bone from a digit of the hand. Each stage carries custom operators unique to the needs of that specific stage which will aid in more accurate results.

The method is further aided by a knowledge base where all model contours and other information such as age, race, and sex, are stored. Shape models, 1-D wrist profiles, as well as an interpretation tree are used to map model and data contour segments. Shape analysis is performed using an arc-length orientation transform.

A pediatric radiologist qualitatively assessed the results of the object contours and were found to be accurate to within 90% for cases with non-fused bones and to within 80% for cases with fused bones. With accurate and robust results, the method can be applied toward areas such as the determination of bone age, the development of a normal hand atlas, and the characterization of many congenital and acquired growth diseases.

Key Words: Object-Oriented Method, Knowledge-Base System, Shape Analysis, Phalanges, Epiphysis, Arc-Length Transform

Submit for **Oral Presentation** at the **MI95 Image Processing Conference**(Murray H. Loew, Conf. Chair).

Fault Tolerance Techniques to Assure Data Integrity in High-Volume PACS Image Archives

Yutao He, Lu J. Huang, Daniel J. Valentino and Algirdas Avižienis (Univ. of California, Los Angeles, CA, 90024)

Picture Archiving and Communication Systems(PACS) perform the systematic acquisition, archiving and presentation of large quantities of radiological image and text data. In the UCLA clinical PACS, for example, the volume of image data archived currently exceeds 2500 gigabytes. Furthermore, the distributed heterogeneous PACS system is expected to have near real-time response, be continuously available, and assure the integrity and privacy of patient data. The off-the-shelf subystems that compose the current PACS cannot meet these expectations; therefore fault tolerance techniques had to be incorporated into the system. As the first step, we developed system auditing and accounting techniques. Currently the PACS has a HL-7 interface to the departmental Radiology Information System(RIS) which can be utilized to enhance the image data integrity.

Data integrity in the context of PACS means that images newly generated by radiolographic scanners can be reliably archived and old images can be correctly retrieved. Operational experience shows that faults that damage data integrity in PACS fall into two categories: (1) System faults, which include both hardware and software faults; (2) Operator faults, which are human-induced. The former disrupt data integrity by interrupting normal workflow and thus causing data inconsistency, while the latter have more extensive effects.

To detect inconsistency errors in newly generated image data, we developed self-auditing and cross-auditing schemes. Self-auditing makes use of the continuity of study number and its relations with series count and image count for radiologic studies. It checks if these data are consistent between the acquisition site and the archives. Cross-auditing introduces additional data information from the external RIS system which tracks patients and examinations and shares them with PACS. It checks data against RIS at acquisition, archiving and display stages in terms of patient id, procedure id and time. This is intended to detect those errors that occured at an early stage and remained latent with respect to the self–auditing method. To detect inconsistency in old data, we perform random retrievability checks which are scheduled during the light-load period of PACS.

A total of 62.4 gigabytes of MR and CT image data were archived in a three-month period of 1994. A total of 128 errors due to operator and system faults were found from 3724 exams by our error detection technique and subsequently corrected. This result shows that fault tolerance techniques are essential to achieve data integrity in PACS and that our error detection methods are effective.

Next we will investigate the following issues: (1) Introduce data privacy and security techniques for patient data in PACS; (2) Develop automatic error recovery methods; (3) Use a temporal profiling technique for dynamic data redundancy management in the distributed environment of PACS to achieve both high availability and high performance.

Software and System Engineering for a Large-Scale PACS

Algirdas Avižienis, Lu J. Huang, Yutao He and Daniel J. Valentino(Univ. of California, Los Angeles, CA 90024)

A large-scale Picture Archiving and Communications System (PACS) is characterized by distributed processing in an heterogeneous computing environment. This complex system deals with a very large quantity of data and demands near real-time response with guaranteed data integrity. The infrastructure of the UCLA PACS incorporates: digital imaging modalities and image digitizers; local and wide-area communication networks of varying bandwidth, general- and special-purpose computers for system management; storage devices for image data archiving; image display workstations; and image printing devices. The software components include: network communication software; data archiving and retrieval software; image database management and distribution software; image display software; and image print utilities. Dependability is supported by a PACS process management subsystem that collects error reports and initiates recovery processes.

The need to support continuous evolution of an already clinically operational PACS as described above requires good system engineering and maintenance practices. Systematic documentation and specification of a PACS significantly aids the development and operation of a clinical PACS. In this paper, we describe our efforts in establishing system documentation and specification. Our current effort is to standardize the documentation of hardware, network, software and operation configurations for the current clinically operational PACS. There are two aspects of documenting the system: one is to document the physical configuration and the other is to generate a set of logic process and data flow diagrams for the system. For the physical configuration, computer hardware and network connections are drawn with clear indication of different network types and segments. For the logic process and data flow diagrams, we document process resources and logs, process interactions, and data storage for various components of the distributed PACS. Supported by such documentation, a PACS operator can more easily monitor the system and identify system errors. System faults, errors and down-time are systematically recorded for future improvements.

Our next effort is the design of a second-generation PACS that will employ extensive fault tolerance techniques to attain very high availability and to assure data integrity, privacy and security of patient records (See: D. J. Valentino, et. al, "Building a Dependable Next-Generation PACS", SPIE Medical Imaging 1995: PACS Design and Evaluation). In this effort, we promote the use of computer-assisted-software-engineering (CASE) tools to establish and document the system specification and to support system analysis, design, verification and documentation. Established information system design practices are emphasized, which include workflow optimization, object-oriented analysis and design, and adaption of industrial standards. We employ CASE tools that support Object-Oriented Analysis and Design (OOAD, Shlaer and Mellor), Object Modeling Technique (OMT, Rumbaugh), and Requirements Driven Development (RDD, Alford). A major challenge in the design of the fault-tolerant PACS is the specification and dynamic verification of fault tolerance attributes under simulated fault conditions. Advanced CASE tools are essential for this effort.

Virtual Reality in Radiology: Virtual Intervention

Michael Harreld Daniel J. Valentino Gary Duckwiler
Robert Lufkin Walter Karplus

Abstract

Our first goal was to develop a VR-based system for visualization of computer simulated blood flow and stress fields. To develop and validate the blood flow model and simulator, we used three-dimensional computational fluid dynamics (CFD) software to simulate blood flow in an animal model. The 3D geometry for the simulations was constructed using biplanar DSA images of swine aneurysm. A commercial CFD package (Nekton) was used to generate the complex, 3D, pulsatile velocity, pressure, and stress profiles throughout the aneurysm and surrounding artery. These results were stored in an object-oriented data model. The model allows optimized precomputations on the data to provide rapid indexing and interpolation of flow parameters within the computational mesh at multiple time points. This data model is a key component of our interactive virtual reality visualization tools. Starting with the data model, we developed a tool that simulates the measurements made by the Doppler velocity probe use in the animal validations. Using this tool we are able to adjust the model and simulation parameers so as to validate the simulation.

We developed a VR-based workstation as part of our PACS (see: D. J. Valentino, et. al., "Building a Dependable Next-Generation PACS," SPIE Medical Imaging 1995: PACS Design and Evaluation). Distributed file servers are used to acquire, compress, decompress, and archive angiographic sequences. A SPARC-based server with four processing modules and two ZX high-performance graphics boards is used to calculate and store simulation results and to generate stereoscopic 3D images. The VR subsystem includes a stereoscopic display for three-dimensional image viewing, a head tracker for viewpoint selection, and a 3D mouse for manipulation of aneurysm models and virtual tools. The virtual environment allows a physicians to become immersed in the simulated blood flow structure and interactively navigate to explore pressure, stress and velocity within a particular aneurysm. Pressure or stress values are mapped to a color-table and displayed on the wall of the aneurysm.

We are currently developing more sophisticated blood flow visualization tools for the virtual environment, includeing virtual tools to allow direct manipulation of streamlines, particle traces, isosurfaces and other graphical structures. Our next goal is to develop a system that can accurately visualize the results in a virtual environment, predict its future behavior, and give insight into the type of treatment necessary. We have found that VR offers novel solutions for planning and performing, via simulation and visualization, procedures that use therapeutic angiography.