

# Opportunities for Visual Computing in Healthcare

Frederick Lee Kitson, Tom Malzbender, and Vasudev Bhaskaran

## Reprint

IEEE MultiMedia

Volume 4, Number 2

April-June 1997



Washington ♦ Los Alamitos ♦ Brussels ♦ Tokyo

PUBLICATIONS OFFICE, 10662 Los Vaqueros Circle, P.O. Box 3014, Los Alamitos, CA 90720-1314 USA

# Opportunities for Visual Computing in Healthcare

Frederick Lee Kitson, Tom Malzbender,  
and Vasudev Bhaskaran  
*Hewlett Packard Laboratories*

Electronically connecting medical experts with patients permits remote care delivery from diagnostics to surgery. Traditional visual computing tasks such as volume rendering or image processing may also become more available when electronically local. Recent innovations in media processing and client-server technology will help process both video and graphics in real time and enable an electronically local environment for healthcare.

**E**volving multimedia technologies portend a convergence of content creation, content processing, and content delivery, as shown in Figure 1. As all forms of media—film, audio, video—become digitized, the communication and processing of such digital data follows. This in itself presents a major shift for content distributors and challenges anyone's ability to control and license such lucrative items as a full-length movie or such private data as their recent X-ray.

The communications industry is anxious to embrace and ultimately capitalize on this trend to convergence. Even as telephone company alliances such as Tele-TV (composed of Nynex, Bell Atlantic, and Pacific Telesis Group) form to control interoperability, so do large cable operators such as TCI and Time Warner. The initial Tele-TV request for 4.5 million set-top boxes (typically connected to a TV for decoding video) demonstrates the scale on which such devices will eventually be deployed. Many computer companies have also shifted computational products toward video servers and on-line interactive devices. Industries on the boundaries of these titan market segments, such as network switching, entertainment software, and satellite broadcast, show startling growth, but the real excitement lies at the intersection of all three industries.

The healthcare industry in the center of Figure 1 leans more toward information processing and

communications than media, with digitized content such as patient records and images. The new applications and capabilities of the Internet best exemplify that intersection. Many have compared the Internet explosion to the California Gold Rush: Often no one knows where to find gold or how long it might take to mine it, but they expect such a broad enterprise to be teeming with wealth. Commentators have also pointed out that most of the business generated with the Gold Rush supported the miners—Levi jeans, Wells Fargo transportation and banking, and mining supplies. We can expect to see the same thing happen with the Internet.

The intersection of media and computing is most evident in the plethora of multimedia products available. In medical imaging, the NeuroNet<sup>1</sup> application illustrates this intersection (see Figure 2). Created at the Center for Clinical Neurophysiology of the University of Pittsburgh Medical Center, the fully integrated system transparently combines the collection, processing, and presentation of real-time data sources. This includes all physiological monitoring functions with non-real-time functions and extensive on-line database information. NeuroNet has yet to reach the potential of visual computing in that it has little support for multimedia in general, such as integrating text, audio, images, graphics, and video.

Medical applications will require and exploit the ability to display 2D images, 1D waveforms, 3D animation, 3D volumetric data, and live video. Because a patient's physiology is dynamic and requires extensive monitoring during surgery, the concept of a closed-loop system seems appropriate. During neurosurgery, for example, measurements such as heart rate can be evaluated against brain-stem stimulation. With video monitoring and appropriate communications, medical professionals can collaborate during evaluations, such as between teams of neurosurgeons, neuroanesthesiologists, and neurophysiologists. As a testimonial to such a system, collateral hearing loss during neurosurgery has all but been eliminated by stimulating the ear and monitoring appropriately.

## Visual computing technology

To support the seamless integration of multimedia, systems must possess the proper computational components (algorithms and architectures) to manipulate data types such as audio, image, video, and graphics. Medical imaging, for example, might require support for image and video compression, volume rendering, modeling, and



high-performance communications.

High-performance graphics and imaging have typically required specialized support. This trend should continue, especially for medical imaging where volumetric data is often of interest. We will briefly introduce typical results and capabilities that make inherently visual applications possible.

#### Lossless compression

Medical imaging poses a great challenge, demanding—for diagnostic and legal reasons—lossless compression algorithms that yet have high compression ratios for reduced storage and transmission time. A novel algorithm we developed for the lossless compression<sup>2</sup> of continuous-tone images combines the simplicity of Huffman (as opposed to arithmetic) coding with the compression potential of context models, enjoying the best of both worlds. This algorithm is based on a simple fixed-context model, inspired by the insight obtained from more complex universal-context modeling techniques. The model is tuned for efficient performance in conjunction with a collection of context-conditioned Huffman codes, realized with adaptive, symbol-wise, Golomb-Rice codes.

The core of the algorithm targets smooth, natural photographic images such as medical X-rays. In addition, an embedded run-length coding enhancement, also implemented through Golomb-Rice codes, yields very good compression performance for non-smooth types of images found in compound documents, such as text, compound documents, and computer-generated graphics.

Our algorithm (originally called Loco-I for Low Complexity for Images) attains, in one pass, compression ratios similar or superior to those obtained with state-of-the-art schemes based on arithmetic coding, at a fraction of the complexity. In particular, it proves superior to the current JPEG<sup>3</sup> lossless compression system in compression

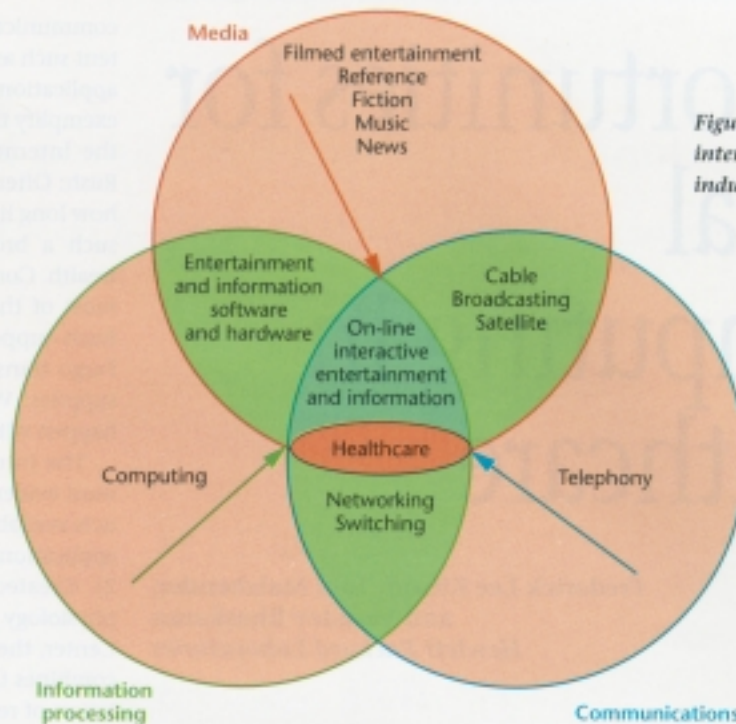


Figure 1. Growth at the intersection of the three industries.



Figure 2. Medical multimedia workstation running NeuroNet.

with comparable complexity. Consequently, it has become the baseline algorithm known as JPEG-LS and will replace the lossless compression method in the current ISO (International Standards Organization) standard. The basic com-

Figure 3. JPEG-LS coder block diagram.

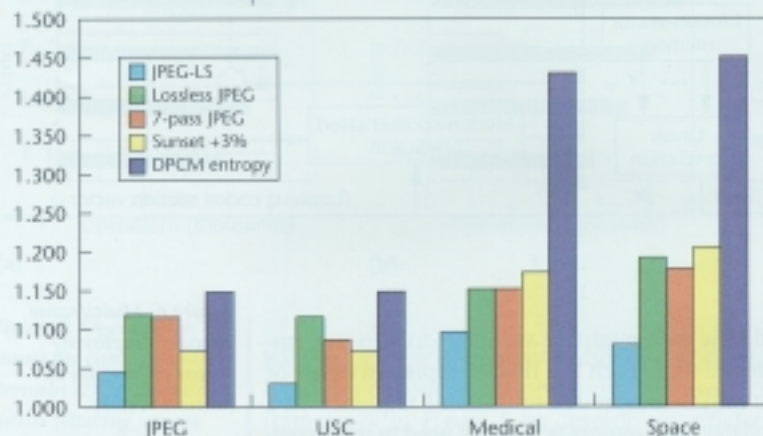
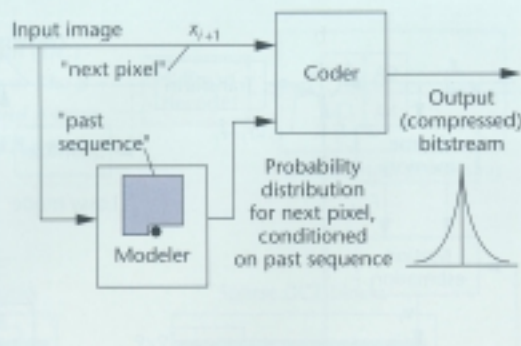


Figure 4. Typical lossless compression results for JPEG-LS and competitive lossless compression methods (a lower number on the ordinate represents more compression). JPEG, USC, Medical, and Space are standard image databases.

ponents of Loco-1 (coder and modeler) appear in Figure 3.

We have evaluated this algorithm on several standard suites of images. The results of JPEG-LS appear in Figure 4 with average compression normalized on the ordinate to the best-known algorithm (a context-based algorithm<sup>2</sup>), which requires almost an order of magnitude more running time. The left-most result is for JPEG-LS (where lower is better), indicating its almost universal advantage.

### Lossy compression

Although preferable in most cases, the lossless mode hinders high compression ratios. For instance, a 4K diagnostic quality chest radiograph of 8 to 10 Mbytes could be reduced by a factor of two to four using lossless compression such as the JPEG-LS method. In a teleradiology application, transmission would take about 10 minutes over a 64 Kbps line, limiting the use of lossless compression for such applications.

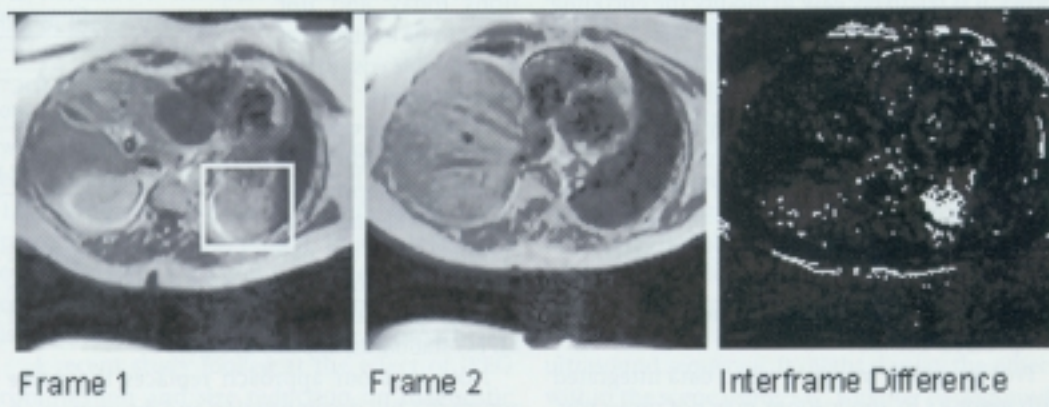
Many recent studies have investigated the usability of JPEG's lossy mode for compressing medical images. By carefully designing the quantizers in the JPEG compression system, studies indicate that echocardiographic images, for instance, retain a very high diagnostic quality<sup>6</sup> for compression ratios as high as 16:1. Similar studies have targeted other compression techniques.<sup>7</sup> The Dicom (Digital Image Communication in Medicine) standard<sup>8</sup> supports all JPEG compression modes (lossless and lossy).

Recently, Aware Inc. developed a commercial image-compression package called Accupress for Radiology, which compresses radiographs (X-rays), computer tomography (CT), magnetic resonance imaging (MRI), and angiograms with up to 16 bits of grayscale. Reviewers reportedly found no clinically relevant degradation up to compression ratios of 30:1.<sup>7</sup>

### Video compression

The emergence of video compression standards has prompted a resurgence in multimedia applications. In medical imaging, compression technologies serve to establish a telepresence in the electronically local environment. The MPEG-2 standard<sup>9</sup> is a good candidate for high-quality broadcast of multiframe sequences.

Figure 5. Motion tracking shows how the region marked in Frame 1 has changed in Frame 2. The interframe difference image shows how much the region has changed.





Analysis of MPEG-2 video compression suggests that the encoding requires 5 GOPS (giga-operations per second), of which 3.6 GOPS go to motion tracking (motion estimation). Compression results from sending only the areas changed from frame to frame. Figure 5 shows an example where the objective is to determine the changed areas in Frame 2 relative to Frame 1. The changed area, which appears in the interframe difference image, comes from a motion-estimation procedure. Various search algorithms—such as log search, hierarchical search, or conjugate direction search<sup>8</sup>—find the pixels in the reference frame for motion estimation. These suboptimal search strategies reduce the complexity of motion-estimation substantially, but they also compromise the overall compression ratio or image fidelity.

We are investigating approaches for motion estimation that do not compromise compressibility. Figure 6 shows one approach using linear prediction of motion vectors. Using correlation in both the spatial and temporal domains reduces the search area by an order of magnitude with comparable coding fidelity.

We also developed a fast algorithm<sup>9</sup> for motion estimation. It yields a 231-times speedup over brute-force motion estimation by exploiting the computer's ability to do Boolean operations on one-bit thresholded images as opposed to 8-bit arithmetic. This algorithm is parallelizable and easily adapts to software-only encoding implementations to exploit the multimedia instruction support in Intel's Pentium processors, including MMX.

The motion-estimation method shown in Figure 6 looks particularly attractive for ultrasound multiframe data, since this motion estimator works very well even with noisy data. Note that we show the encoder supporting a lossy mode as well as the lossless mode. MPEG-2 supports only the lossy mode; however, as Figure 6 shows, it is relatively easy to modify the encoding loop to support the lossless mode as well. The lossless mode might prove useful in interframe compression of ultrasound sequences, as it can improve the compression ratios by a factor of two to three compared with a lossless compression scheme such as JPEG-LS. The latter compresses the multiframe sequence on a frame-by-frame basis without exploiting interframe dependencies.

#### Image and video processing in the compressed domain

Processing the image or video data integrated with graphics in applications may not prove triv-

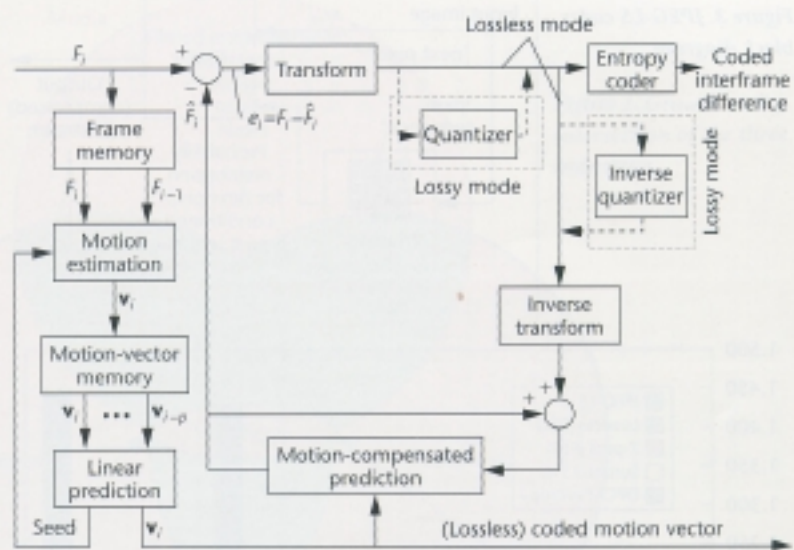


Figure 6. Multiframe encoder employing linear prediction of motion vectors.

ial. The data might be available only in its compressed form, such as a JPEG-compressed image or an MPEG-compressed multiframe sequence. In a collaborative session, you might need to merge several video streams into a single stream for display. This compositing process can get quite expensive if you unwittingly adopt a naïve approach wherein you first decompress each video stream, perhaps scale and decimate each, merge them with the other video streams, and eventually recompress them to form the merged video stream.

The increased complexity of the compositing process arises from the fact that an MPEG video decoding task requires nearly 1 GOPS per video stream and encoding requires up to 5 GOPS. Such computational power is not readily available on workstations today and not likely to be available at a reasonable cost in the next few years.

We have developed an alternate approach to video compositing that is substantially less complex than the naïve approach. The general framework of our approach, which applies to compressed bitstreams conforming to the JPEG and MPEG standards, appears in Figure 7.

Note that our approach replaces the naïve approach's full decompression and recompression

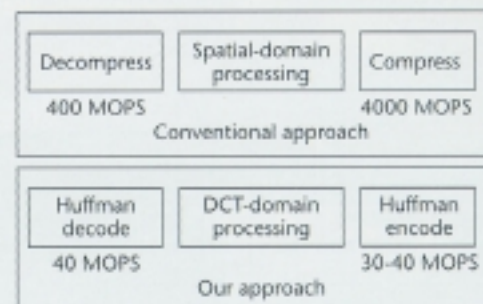


Figure 7. Compressed domain processing versus the conventional spatial domain processing (MOPS is millions of operations per second).



Figure 8. Compressed domain (DCT-domain) downscaling-by-two. The  $8 \times 8$  DCT blocks A1, A2, A3, and A4 are downsampled to yield a single  $8 \times 8$  DCT block Y.

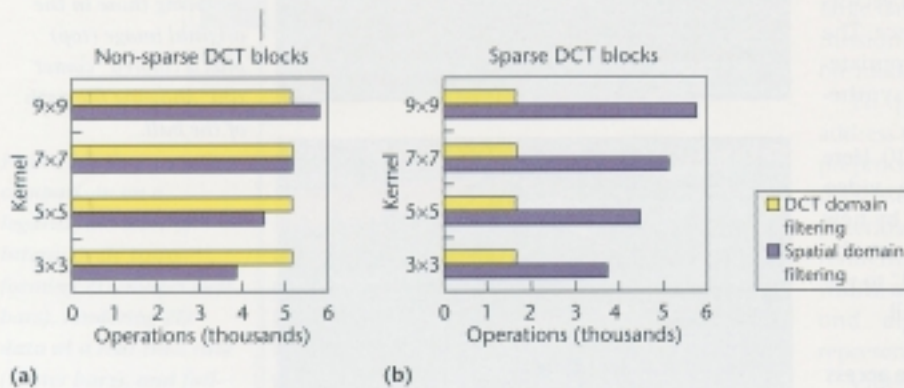
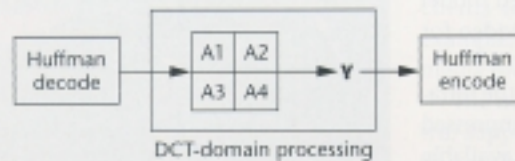


Figure 9. Operations count for compressed domain (DCT domain) linear filtering versus the naive spatial-domain-based approach for various filter kernel sizes.

with the lower complexity Huffman-decode and Huffman-encode functions only. We used the framework in Figure 7 to develop an efficient method for downscaling an image. In general, any image or video image with a spatial resolution of  $N \times N$ , compressed using the discrete cosine transform (DCT) coding techniques (for example, JPEG, MPEG, H.261, H.263), can have the data processed directly in the compressed domain. Doing so yields a standards-compliant bitstream that when decompressed at the decoder gives an image or video at a spatial resolution of  $(N/k) \times (N/k)$ , where  $k$  is the image decimation factor, typically 2, 3, or 4. Figure 8 depicts the compressed domain approach for downscaling by two.

Note that the basic operation here synthesizes the  $8 \times 8$  DCT block Y representing the downsampled image from the four  $8 \times 8$  DCT blocks A1, A2, A3, and A4. The basic equation for downscaling in the DCT domain is

$$Y = \begin{bmatrix} S & T \end{bmatrix} \begin{bmatrix} A1 & A2 \\ A3 & A4 \end{bmatrix} \begin{bmatrix} S^t \\ T^t \end{bmatrix}$$

where S and T are constant  $8 \times 8$  matrices whose elements take on values in the set  $\{0, 1/8, 1/4, \text{ and } 1/2\}$ . Computing Y requires 880 operations compared with the naive approach, which needs 5,376 operations. For  $k = 3$  and  $k = 4$ , the computation complexity of the compressed domain approach is three times lower than the naive approach.

A recent study looked at the effect of JPEG compression and size reduction on diagnostic

quality of medical images such as chest radiographs.<sup>10</sup> The researchers found that downscaling the image by two and performing JPEG compression to a ratio of at most 25:1 did not reduce the diagnostic quality yet lowered the transmission time significantly.

As an alternative to their method we propose performing the JPEG compression on the original high-resolution image first, then doing the compressed domain downscaling. Our approach should yield better compression ratios, since it compresses a more highly correlated high-resolution image. Furthermore, our compressed domain downscaling is more efficient than their spatial-domain downscaling. Note that our approach can have the same diagnostic quality as well.

Downscaling is a special case of the more general operation of linear filtering, which is usually cast as a convolution operation in the spatial domain. We applied the compressed domain processing method<sup>11</sup> to develop the DCT domain equivalent of the spatial domain convolution operation. As shown in Figure 9, the compressed domain approach requires fewer operations than the conventional spatial domain (naïve) approach. The compressed domain approach also has the added advantage that for sparse DCT blocks, the computation complexity remains constant for a range of convolution kernel sizes.

A filtering operator proves useful in many applications requiring noise reduction, image enhancement, and so forth. In medical imaging, an edge detection operator would be very useful. Since the edge detection operation can be viewed as a convolution operation followed by thresholding, we can employ the compressed-domain-based filtering approach. This yields a 16-times speedup compared to using a Sobel-edge detector,<sup>12</sup> which operates on the spatial domain data.

In some applications, the user can personalize the viewing experience. For instance, while viewing an ultrasound sequence, the user might first select a region of interest on the screen. Then the processor would track the object within this region using any motion information available in the compressed bitstream representation for the ultrasound sequence. It might display the object within the scene using, say, a lighter background.



Other information, such as a synthesized model of the object, could be overlaid on the video for diagnostic purposes.

Compositing a synthesized model on the ultrasound sequence can take place in the compressed domain if the ultrasound sequence were available as, say, an MPEG compressed bitstream. Basically, this process works by tracking the motion vectors of the selected region in the video sequence. The tracked area is then merged with the appropriately transformed bitmap derived from the synthesized model of the object.

A sample application appears in Figure 10. Here several regions of advertising in the video sequence shown on the top are replaced by new advertisements as shown in the picture on the bottom. Also, adding a colored "comet tail" to the lower image highlights the path of the ball.

#### Multimedia content retrieval and database access

Content-based image retrieval systems—such as Query by Image and Video Content (QBIC<sup>®</sup>) at IBM Almaden Research Center or Chabot,<sup>34</sup> a relational database management system for images—can now enhance medical imaging applications. QBIC, for example, supports queries of large image and video databases based on example images, user-constructed sketches and drawings, and selected color and texture patterns. Chabot lets users include image-analysis and information-retrieval tools in the query process. Searches can combine text and image operators such as identifying an object's general shape or location. Many video applications would also benefit from scene transition analysis to aid editing. Current research aims to provide dynamic as well as static analysis of scenes.

Audio is a component of video and a proper multimedia type in its own right. Typical attempts to retrieve audio, even by class such as applause, music, or speech of a particular speaker, can be daunting.<sup>35</sup> Content-based retrieval systems promise to provide access to multimedia information, making possible a richer computer interface that would support speech I/O and avoid a keyboard interface, especially in the home.

Conventional database access has become more feasible, with systems such as the Calgary AgeNet trial sponsored by Calgary Regional Health Authorities and Kasper Medical Laboratories. The



Figure 10. Merging graphics on MPEG-compressed video using compressed domain compositing. The bottom image includes two new advertisements replacing those in the original image (top) plus a colored "comet tail" showing the path of the ball.



Figure 11. Graphics in video and dynamic hot spots. Clicking on the dynamic hot spot region in the original (left) triggers the application to link to further information, in this case outlining the vessel for clearer viewing (right).

trial aims to evaluate a regional healthcare information network with a telecommunications infrastructure. AgeNet focuses on the geriatric community, with data from several hospitals, private labs, and agencies. It includes demographics, medical records, referrals, pharmacy records, admissions, health records, and homecare services.

#### Graphics processing

When using a video sequence or image, you might want to identify or track objects or regions so that you can link other relevant data. Figure 11 illustrates this principle with data from the



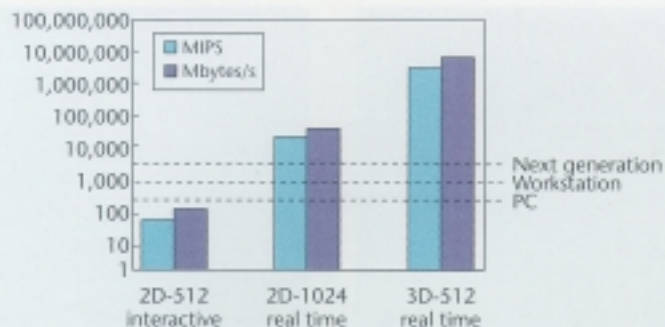


Figure 12. Computation comparison on a logarithmic scale of interactively transforming 2D images (left bars), rendering 2D data at a real-time rate (center bars), and full-motion 3D (right bars).

University of Utah. Here you might interact with a video or image of blood vessels, perhaps establishing a dynamic hot spot on the screen. Then if you click on the region of the screen associated with a particular vessel at a particular time, the application can execute a link, as shown in Figure 11. The region might be identified with an embedded link to further data associated with the particular vessel. We call this notion HyperVideo.

Graphics will provide user interface elements for navigation and presentation, as in a Web application. An application may be downloaded into the user's PC or even a set-top box. Performance requirements of realistic graphics will mandate integration of approximately one million antialiased vectors and texture-mapped polygons per second with video processing. This will enable animated, colorful, and gripping interfaces with a mixture of 3D animation, video, and images.

Architectures to support such visualization or integration of images, graphics, and video range from the PixelFlow technology developed at the University of North Carolina, Chapel Hill to the Talisman architecture<sup>16</sup> from Microsoft. Talisman exploits the coherence in computer graphics to reduce not only the bandwidth requirements for the system but also the frame buffer memory.

Another key technology is the use of image and data compression. Talisman's Image Layer Compositing block combines the image, video, and graphics planes in real time, supporting the type of applications that should be common in medical imaging. Nonetheless, this particular architecture evolved out of a need to support realistic graphics and PC games.

The key computational blocks for multimedia—such as image, graphics, and video support—can be merged into a *media processor*, a specialized processor better suited to acquired data such as multimedia. Several of these high-performance processors are or will soon become available and will enable many low-cost devices or

clients to support interactive multimedia. For example, the Talisman architecture is based on a media processor.

The new digital communications infrastructure will support client-server graphics as well as video and audio, providing a new opportunity for networking. Partitioning the application would let a large database reside on a server and limit transmission and processing to the portion of the visible model needed on the client or home terminal.

To promote this client-server model, we address two algorithmic areas. The first deals with preprocessing and spatially subdividing graphics databases for visibility, then transmitting only the potentially visible subset from server to client for rendering. This saves memory, network bandwidth, and computational requirements. The second algorithm deals with multiresolution representation for graphical databases.

Video may be fed from the server and the graphics composited on the workstation, PC, or set-top with some of the innovations already discussed. Microsoft refers to the next generation of home clients as "communicating PCs," which have the aforementioned video and graphics capability with a broadband interconnection. Although somewhat debatable, many see Asynchronous Transfer Mode (ATM) as a preferred packet-oriented protocol for multimedia traffic because of its notion of quality-of-service (QoS).

Although the PC represents today's information appliance (and game platform), the concept persists of a multimedia communications device attached to the TV set, collectively known as interactive TV or WebTV. The three basic components of an interactive TV architecture are content (information servers), a network, and the set-top box. A form of the set-top box might well appear as a bedside device attached to hospital or home TVs.

#### Volume rendering

Volume rendering converts multidimensional data sets into 2D images without an intermediate geometric primitive such as polygons.<sup>17,18</sup> Often the data is point samples of scalar quantities sampled in three-space, as with density data in medical imaging. Computing and transmitting 3D data, assuming one byte per pixel, poses a further challenge as relatively quantified in Figure 12. This bar chart shows on a log scale the number of instructions in millions per second (MIPS) and the number of Mbytes per second needed to process and transmit 3D volume data for Fourier analysis. The leftmost bars represent interactive transform-



mation of 2D images (10 frames per second). The center bars indicate rendering of 2D data to a resolution of  $1,024 \times 1,024$  at a real-time rate of 30 fps. Finally, the rightmost bars illustrate the 3D case at  $512 \times 512 \times 512$  pixels and full motion (30 fps). Notice that about three orders of magnitude separate the requirements for each case.

Today's workstations do not provide the necessary raw performance for such 3D rendering in real time. Algorithmic innovations such as Fourier volume rendering<sup>10</sup> look promising for this application. Figure 13 shows the spatial data on the left and the 2D frequency transform on the right, from which such a spatial image is computed. The technique exploits the Fourier Projection Slice theorem, which lets you take appropriately filtered slices of a 3D transform volume and perform a 2D inversion instead of the usual 3D data projection. This approach typically reduces the computation for generating a projection image by two or three orders of magnitude. Unlike spatial volume rendering it is limited to linear projections, such as those computed for CT projections.

Typical sources of medical volumetric data include CT, MRI, and positron emission tomography (PET). Recently various researchers have demonstrated the possibility of extracting volumetric data sets from conventional ultrasound imaging hardware. We have built a prototype system for interactive collection and rendering of 3D volumes. With it, the ultrasound technician can acquire a set of slices, preview them spatially registered, and compute surface renderings of relevant structures interactively while the patient waits. If the renderings warrant it, the technician can collect more data immediately.

Our system uses a commercial magnetic orientation and position transducer from Ascension Technologies in Burlington, Vermont along with conventional ultrasound probes. Figure 14 shows the six-degrees-of-freedom magnetic receiver mounted on a 2.5-MHz ultrasound transducer. This arrangement allows us to determine position of the imaged slice in real time. Any video information such as echo intensity, Doppler color, and energy appearing on the ultrasound display is simultaneously read into a Hewlett Packard 735 workstation via a frame grabber.

Next the image slices are rendered in their appropriate locations for immediate display. A user can specify a region of interest, within which the data is uniformly resampled. Resampling from arbitrarily oriented slices at interactive rates proves difficult because of the nonuniform data.

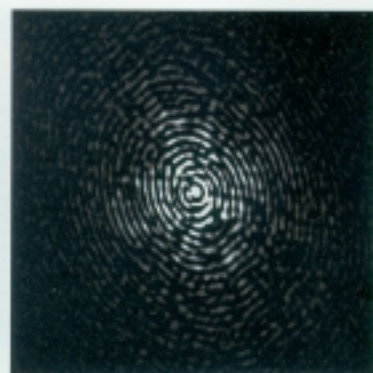
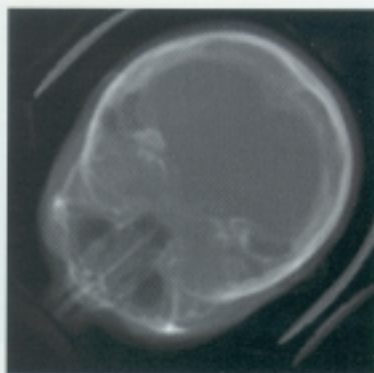


Figure 13. Fourier volume rendering showing the spatial data (left) and the 2D frequency transform from which such a spatial image is computed (right).

Since we do not restrict the ultrasound technician to any specific sweeping pattern, we might get oversampled or undersampled regions of the acquired volume. For this reason our resampling technique varies adaptively between decimation and interpolation to avoid any aliasing artifacts. Once our system has resampled the data into a uniform volume, it can compute surface or volume renderings interactively for visualization of relevant structures. We made electrocardiogram (EKG) gating available during slice acquisition for cardiac applications.

Segmenting and classifying 3D ultrasound data proves especially difficult because the coherent imaging used in ultrasound introduces substantial speckle noise. We have found that the acoustic quantification feature of HP's ultrasound imaging hardware automatically segments arteries, veins, and blood pools for subsequent 3D rendering. Figure 15 (next page) shows an example of a surface rendering based on acoustic quantification for structures within the liver.

Segmentation also proves difficult in modalities such as MRI and CT. We developed a tracking procedure that provides an alternative to segmentation for arterial applications.

Clinicians often want to see cross-sections of arteries to determine the degree of stenosis. Planar cross-sections are inappropriate, since arteries such as the coronaries are tortuous and fall to remain in a single plane. An alternative is to



Figure 14. Orientation and position tracker mounted on an ultrasound probe.



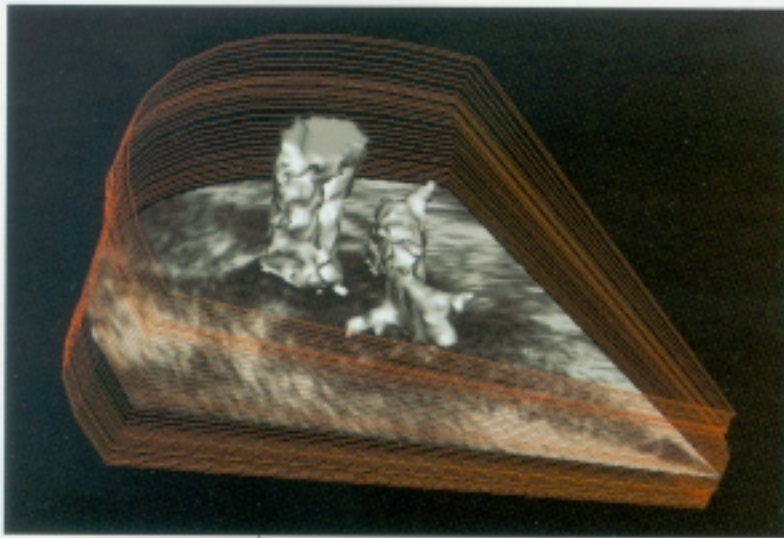


Figure 15. This 3D ultrasound surface rendering based on HP's acoustic quantification highlights veins in the liver.

extract the medial axis of such structures and use it to define a ruled spline resampling surface. This surface stays along the artery, providing a cross-section of a length of the artery in a single view.

We employed finite-impulse response (FIR) filters containing two tiers of concentric rings of sample points, as shown in Figure 16. We sampled the trilinearly reconstructed continuous function and weighted them by either plus or minus one. These *tube filters* are parametrically controlled via two orientation angles, scale, and three translation parameters, and respond maximally when oriented along any bright tubular structure. Additionally, we can use them to find bifurcations along arteries. Note that the evaluation complexity is independent of any parameter settings. This contrasts with the more conventional FIR approach, whose complexity increases with the cube of the scale parameter. Constraining the parameter search lets us use these filters to extract artery paths in a matter of seconds on an HP 735 workstation.

#### Role of the Information Superhighway

The system of interconnected heterogeneous data servers that make up the World Wide Web presents both opportunities and challenges in accessing and processing multimedia. Currently the Web fosters mainly multimedia access and delivery. Its limited bandwidth and almost infinite data have provoked a scramble by network and cable providers to offer Web connectivity with cable bandwidth or telephone company switching. Realizing a distributed multimedia system that connects every home and business is the vision of the Information Superhighway.

The emergence of Web-centric programming languages such as Java has shifted the Information Superhighway paradigm. No longer just a way to access multimedia, the Web can also process multimedia efficiently. Some of the processing functions discussed here, such as compressed domain processing, can be judiciously incorporated within Java scripts to enable a more powerful visual computing environment on the Information Superhighway. Healthcare could benefit from more sophisticated applications that provide for the downloading of "applets" associated with a patient's data, for example. Such mechanisms could enable home monitoring of patients and protect the security of a patient's records.

#### Cable modems

Despite substantial investments, trials, and product announcements, the realization of a true digital video-on-demand (VOD) system has encountered some setbacks. Scaling video servers to the level required for full commercial deployment remains years away. The \$300 fully interactive set-top box, although also under active development,<sup>23</sup> requires state-of-the-art multimedia and communications. It might take a few years to achieve this cost goal. The next real advance in the information superhighway will be deploying a hybrid-fiber coaxial system—cable modems, which are just now being introduced into the consumer market for Web access.

Digital data will link to a distribution hub, where five fibers (needed to offer the necessary aggregate bandwidth for a typical community) will support each local hub. This arrangement allows the management of about 2,000 homes—a number good for network management and amortization of costs. Each home will have a cable modem attached to a PC or other information client such as a set-top box.

This structure offers a major advantage: It makes available tens of megabits per second for delivering data or even MPEG-2 video to the home. This contrasts with today's tens of kilobits per second, which limits applications that use images, audio, or video. With digital quadrature amplitude modulation, a 30-Mbps data stream can exist in a 6-MHz analog channel. Also, a "back channel" handles interactive applications. This shared communications channel manages units of megabits per second.

We developed appropriate communications protocols for such a system and have participated in several cable data consortia, one which



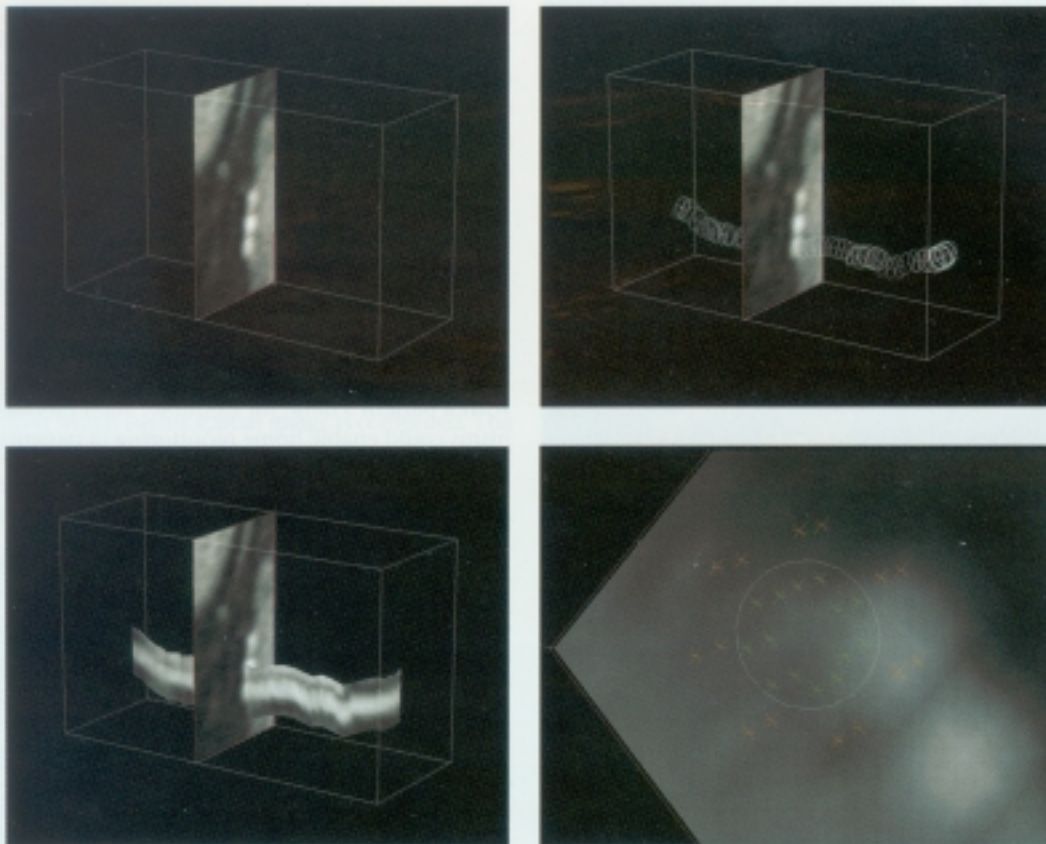


Figure 16. Artery tracking in MRI. Upper left: Ray MRI data showing the cross-section of three arteries. Upper right: Lower artery tracked using parametric FIR tube filter. Lower left: Ruled spline cutting surface extracted from tracked path showing cross-section along the length of the artery. Lower right: Parametric tube filter. Red indicates  $-1$ -weighted resampling points, green indicates  $+1$ -weighted samples.

includes Intel, AT&T, Hybrid Network, and other standards bodies to define international cable data standards. The real advantages of such a system include its ability to operate with the inherent latency, to load balance the multiple upstream channels, and to add such channels incrementally to an existing analog cable infrastructure. This infrastructure might well bring medical care to the home. It provides sufficient bandwidth for measurement data to leave the home and for patient care information such as video to be delivered to the home.

### 3D Web interfaces

The Web spawned many innovations quickly integrated with browsers through mechanisms such as Sun's Java, Microsoft's Blackbird, Macromedia's Director, or Virtual Reality Modeling Language (VRML). As indicated, Java lets you download platform-independent code that enables local execution. Java is built on C and C++, languages optimized for client-server on-line applications. Code is dynamically linked on the fly through an interpreter so that an application runs without precompilation. A Java application

or applet can be sent from the server to the client running HotJava, the player that automatically runs the application. The server can then continue to update the application. Such capability built into Netscape or Microsoft browsers lets you run CD-ROM code and a remote multimedia application through the use of its "socket" technology.

VRML also enables graphics applications that exploit client-server capabilities. For example, imagine a 3D interface metaphor as a 3D house. Navigation in this 3D environment is fairly intuitive, so you might access video and audio information in the family room, use the garage for "trips," seek nutritional information in the kitchen, and consult the medicine cabinet for information on drugs, all while sitting in your own living room.

### Future medical applications

The ultimate application of communications and visual computing might be the facetiously named *long-distance operators*. To explain: The Pentagon funds a telesurgery project named MedFAST, or Medical Forward-Area Surgical Telepresence.<sup>23</sup> The concept is to place the surgical





Figure 17. Robotic manipulator under a remote surgeon's control.

unit in areas of conflict, accidents, or natural disasters. The surgeon, from a safe distance, controls the remote operation through force-feedback devices such as surgical instruments (see Figure 17).

The US Department of Defense Advanced Research Projects Agency (DARPA) also funded the development of the vehicle to transport the MedFAST to the site of the wounded soldier. The entire system could use superhighway information access or specialized aircraft and satellite systems. As a testimonial to the feasibility of this concept, researchers have used a prototype system to operate on a pig at the University of California at San Francisco, with assistance by consultants from SRI's medical electronics laboratory in Menlo Park, California.

Finally, wireless video and data communications offer another opportunity. Many portable devices under development bring wireless video integrated with graphics to potentially provide the medical imaging appliances of the next century. Again, technologies such as client-server graphics and model-based compression like MPEG-4 will help engender these portable clients. Although such devices will play a role in the hospital, a greater opportunity exists in supporting home care with companion home imaging and instrumentation.

We also anticipate other devices for capturing patient data or monitoring patients in the home. We at Hewlett Packard want to effectively combine measurement, communications, and com-

putations in a notion we refer to as MC<sup>2</sup> that will uniquely address the needs of the healthcare industry. The technologies described here indicate the possibilities and how related technologies can combine in a truly synergistic fashion.

### Summary

As we move into the next millennium, we anticipate a rapid deployment of visual computing technology applied to the healthcare industry. In the hospital and clinic, we expect new modalities of diagnosis through volume rendering and image processing. Appropriate media processing support in the clients (PCs or home appliances) and communications connections between healthcare providers and patients will generate many opportunities.

Current methods will become more efficient and ultimately provide better health maintenance through more timely and comprehensive access to information.

Like personal portable distributed computing, healthcare will migrate to reach people in their homes or clinics in a more personal distributed manner. Technology such as videoconferencing or 3D graphics can provide the interface between provider and patient, realizing the electronically local environment needed in a world where people and data are most natural and efficient when local.

MM

### References

1. R.J. Scabassi et al., "NeuroNet: Collaborative Intraoperative Guidance and Control," *IEEE Computer Graphics and Applications*, Vol. 16, No. 1, Jan. 1996, pp. 39-45.
2. M.J. Weinberger, G. Seroussi, and G. Sapiro, "Loco-I: A Low Complexity, Context Based, Lossless Image Compression Algorithm," *Proc. Data Compression Conf. 96*, IEEE Computer Soc. Press, Los Alamitos, Calif., Apr. 1996, pp. 140-149.
3. B. Vasudev and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, Kluwer Academic Publishers, Boston, Mass., Sept. 1995.
4. T.H. Karson et al., "JPEG Compression of Digital Echocardiographic Images: Impact on Image Quality," *J. American Soc. of Echocardiography*, Vol. 8, No. 3, June 1995, pp. 306-318.



5. P.C. Cosman, R.M. Gray, and R.A. Olshen, "Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy," *Proc. IEEE*, Vol. 82, June 1994, pp. 919-932.
6. J.D. Thomas, "The Dicom Image Formatting Standard: What It Means for Echocardiographers," *J. American Soc. for Echocardiography*, 1995, pp. 319-327.
7. Aware Inc., "Application of Wavelet Compression to Digitized Radiographs," *American J. of Roentgenology*, Vol. 163, 1994, pp. 463-468.
8. R. Armitano et al., "Motion Vector Estimation Using Spatio-Temporal Prediction and Its Application to Video Coding," *SPIE Conf., Digital Video Compression: Algorithms and Tech.*, SPIE, Bellingham, Wash., Jan. 1996.
9. B. Natarajan and B. Vasudev, "A Fast Approximate Algorithm for Scaling Down Images in the DCT Domain," *IEEE Int'l Conf. on Image Processing*, IEEE Press, Piscataway, N.J., Oct. 1995, pp. 241-243.
10. B.J. Sullivan and R. Ansari, "Effects of Image Preprocessing/Resizing on Diagnostic Quality of Compressed Images," *IEEE Int'l Conf. on Image Processing*, IEEE Press, Piscataway, N.J., Oct. 1995, pp. 13-16.
11. N. Merhav and B. Vasudev, "A Fast Algorithm for DCT-Domain Image Downsampling," *IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4, IEEE Press, Piscataway, N.J., May 1996, pp. 2309-2312.
12. W.K. Pratt, *Digital Image Processing*, John Wiley and Sons, New York, 1991.
13. M. Flickner et al., "Query by Image and Video Retrieval from a Relational Database of Images," *Computer*, Vol. 28, No. 9, Sept. 1995, pp. 23-32.
14. V. Ogle and M. Stonebraker, "Chabot: Retrieval from a Relational Database of Images," *Computer*, Vol. 28, No. 9, Sept. 1995, pp. 40-48.
15. E. Wold et al., "Content-Based Classification, Search, and Retrieval of Audio," *IEEE Multimedia*, Vol. 2, No. 3, Fall 1996, pp. 27-36.
16. J. Torborg and J. Kajiya, "Talisman: Commodity Realtime 3D Graphics for the PC," *Proc. Siggraph 96*, ACM Press, New York, 1996, pp. 353-364.
17. A. Kaufman, *Volume Visualization*, IEEE Computer Society Press, Los Alamitos, Calif., 1991.
18. M. Bentum, B. Lichtenbelt, and T. Malzbender, "Frequency Analysis of Gradient Estimators in Volume Rendering," *IEEE Trans. Visualization and Computer Graphics*, Vol. 2, No. 3, Sept. 1996, pp. 242-254.
19. T. Malzbender, "Fourier Volume Rendering," *ACM Trans. Graphics*, Vol. 12, No. 3, Jul. 1993, pp. 233-250.
20. B. Furht et al., "Design Issues for Interactive

Television Systems," *Computer*, Vol. 28, No. 9, May 1995, pp. 25-38.

21. P.S. Green et al., "Telepresence Surgery," *IEEE Eng. in Medicine and Biology*, May/June 1995, pp. 324-329.



**Frederick L. Kitson** is the manager of visual computing research at Hewlett Packard Laboratories in Palo Alto, California, where he works on graphics and visualization architectures, multimedia data processing, and communications systems. He is also an adjunct faculty member at the Georgia Institute of Technology. Kitson received his BS with honors in electrical engineering from the University of Delaware in 1974, an MS from the Georgia Institute of Technology in 1975, and a PhD from the University of Colorado in 1981.



**Tom Malzbender** heads the Graphics Algorithms and Architectures group at Hewlett Packard Laboratories in Palo Alto, California. He developed the capacitive sensing technology that forms the basis of HP's line of graphics tablets. His research interests include the role of signal processing in volume rendering, 3D ultrasound, statistical texture modeling, and neural and genetic optimization techniques. Malzbender holds a BS in electrical engineering from Cornell University.



**Vasudev Bhaskaran** is in the Video Algorithms and Architectures group at Hewlett Packard Laboratories in Palo Alto, California. His research interests include image and video compression techniques, image and video processing, and multimedia information transmission. He co-authored *Image and Video Compression Standards: Algorithms and Architectures* from Kluwer Academic Publishers (1995). Bhaskaran received his BTech from the Indian Institute of Technology, Madras, and did his MS and PhD work at Wichita State University in Wichita, Kansas, and Rennselaer Polytechnic Institute in Troy, New York.

Readers may contact Kitson at Hewlett Packard Laboratories, 1501 Page Mill Rd., Bldg. 3U, Palo Alto, CA 94304, e-mail kitson@hpl.hp.com.