Journal of Digital Imaging

VOL 2, NO 4 NOVEMBER 1989

Medical Image Work Stations: Functions and Implementation

Stephen M. Pizer and David V. Beard

Electronic work stations are becoming a standard means of presenting medical images for diagnosis and consultation, and they will become more widespread as picture archiving and communication systems (PACS) come into use. These work stations must allow the user both to perceive the patterns necessary for accurate diagnosis and to "navigate" efficiently within large sets of related images, ie, quickly find and compare desired images. The work stations must operate without a feeling of "friction" and have an affordable cost. In this report we survey the tasks and system objectives, first regarding the perceptual needs and second with respect to the navigational needs. We then survey the technology available to satisfy these needs and conclude with a list of needed research and technology that can be expected or should be provided in the future.

KEY WORDS: medical image, display, work station, perception, navigation.

THIS REPORT focuses on the problem of presentation, for diagnosis, of collections of static two-dimensional (2D) gray-scale images that are stored in a quickly accessible archive. Neither cinematic nor three-dimensional (3D) presentations are discussed, except in passing. We assume that the "recorded images" to be displayed have been produced from acquisitions that are adequate in spatial and contrast resolution and in spatial and contrast sampling (number of pixels and gray levels). For a system suitable for displaying images from a wide range of acquisition modalities this will mean that different image sizes, perhaps from 64 × 64 to 4,096 × 4,096, will all need to be displayed and that different recorded intensity ranges, perhaps from 8 to 16 bits, will need to be presented.

The display system must provide two capabilities to the user: the ability to perceive the patterns held in the intensities in the recorded images, and the ability to get around among the images making up a diagnostic set. We will lay out the details of each of these capabilities and then discuss the technological means of providing these capabilities.

DISPLAY SYSTEM FUNCTIONS

Perceptual Functions

Perceptual tasks. The primary uses of medical image work stations are for diagnosis, consultation with referring physicians, and review of images either during image acquisition or after diagnosis. The displayed images, to the greatest degree possible, must allow accurate detection of anatomic or physiological objects, accurate characterization or measurement of their features (such as shape, size, or intensity) and accurate comparison of features between images taken at different times or on different acquisition modalities. Furthermore, effective integration of features across the third spatial dimension or a time sequence must be possible.

System objectives. To achieve these goals, the display system must first have adequate values of various display parameters, such as screen size, number of pixels, gray-scale dynamic range, and number of digital intensity levels.[1] Second, it must adequately assign the digital intensity levels to the intensities in the recorded image so that diagnostically important contrasts in the latter can be perceived on the displayed image.

The screen size should be chosen to allow consultation appropriate to the display station type (Fig 1). The number of pixels and digital

From the Departments of Computer Science and Radiology, University of North Carolina, Chapel Hill.

Supported by National Institutes of Health Grants Nos. RO1 CA44060 and PO1 CA47982.

Address reprint requests to Stephen M. Pizer, PhD, Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599-3175.

0897-1889/89/0204-0001$03.00/0

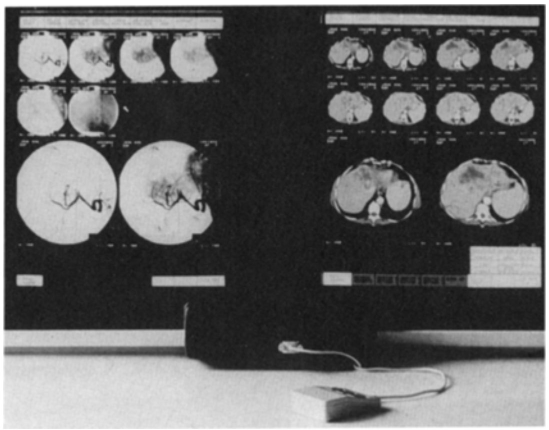

Fig 1. Consultation console. (Courtesy GE Medical Systems, Milwaukee, WI.)

intensity levels chosen needs to be large enough to avoid distinguishable adjacent pixels and adjacent intensity levels, yet small enough to be affordable and allow acceptable image-paint times. Normally 256 intensity levels are adequate, but this is only the case if the levels are approximately evenly spaced in brightness (perceived intensity). This requirement frequently is not met. Regarding the number of pixels, 1,024 × 1,024 is all that normally can be provided today, although 2,048 × 2,048 is becoming affordable. Nevertheless, the resolution necessary for certain diagnoses, such as with radiographs of fine fractures and pneumothorax, may require 4,096 × 4,096 sampling, so there must be some means of presenting the full information in such images on a more coarsely sampled screen. Roaming and zooming within the image, therefore, seems necessary. Means for achieving this function will be discussed later.

The assignment of digital display levels to the digital intensity levels in the recorded image is an unavoidable step of image display. The most common method is interactive intensity windowing. In this approach the user selects a range of recorded intensities that is to be linearly mapped to the full range of digital display-scale intensities, with recorded intensities below or above the specified recorded intensity range mapped to black and white, respectively.

Recent research and practice has led to automatic means of assignment of display scale levels to the recorded intensities, means that frequently allow the full perception of contrast without the need for intensity windowing. Avoiding intensity windowing saves user time, especially considering that the simultaneous display of many images may require the choice of many different intensity windows. It also can allow the integration into a diagnosis of features that would require different intensity windows. Although all such methods lead to a loss of absolute intensity information, this fact does not normally lead to clinical problems, perhaps because of the human inability to perceive absolute intensity.

Methods of automatic assignment of display scale levels to recorded intensity levels produce an assignment that depends on (ie, adapts to) the local recorded image values. The method of "unsharp masking" is frequently used on chest radiographs (Fig 2A shows a conventional intensity windowed radiograph, Fig 2B shows the same image after application of unsharp masking). Researchers have investigated several versions of unsharp masking in which its parameters adapt across the image. These versions seem somewhat useful but have not come into common use. Another approach, called "contrast-limited adaptive histogram equalization" (CLAHE), is very effective for showing in a single displayed image all diagnostically useful contrast that is contained in a recorded image.[2-5] Its best effect seems to be in computed tomography (CT) (Fig 3A shows a conventional intensity windowed CT image, Fig 3B shows the same image after application of CLAHE), magnetic resonance imaging (MRI), and digital subtraction angiography (DSA) and also in very low contrast radiographs, such as radiotherapy portal films (Fig 4A shows a conventional intensity windowed portal film, Fig 4B shows the same image after application of CLAHE). Although both unsharp masking and CLAHE have been found important in clinical use, there has been some concern with these methods producing shadows at high-contrast edges and showing image noise too well.

Unsharp masking[6] consists of removing some of the local background intensity so as to allow more display levels to be used for portraying local recorded intensity variations. It operates by taking a weighted sum of the recorded image and an edge-enhanced form of that image. The edge-enhanced form is computed by subtracting the local background, that is, a local spatial average.

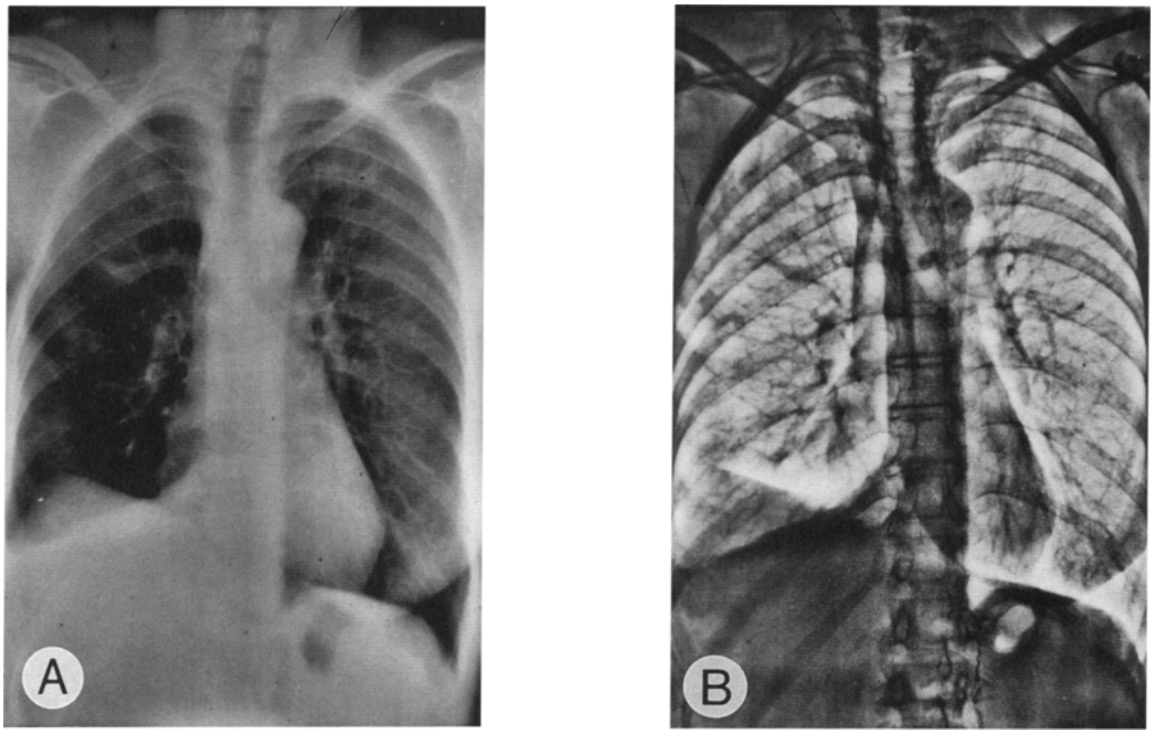

Fig 2. (A) Intensity windowing versus (B) unsharp masking on a chest radiograph (courtesy of J. Sorenson).

CLAHE[3,4] operates by letting each pixel's display-scale level be proportional to its rank in recorded intensity, when compared with the recorded intensities in a region surrounding that pixel (its contextual region); however, contrast enhancements beyond a specified factor are not allowed. The contextual region area is frequently around 1/16 of that of the image but may be changed depending on the detail of interest.

How best to assign display-scale intensities (integer display-driving values) depends on the displayed intensity scale (luminances) used to portray these values.[4] More investigation is needed to find the best displayed intensity scale for each assignment scheme, such as intensity windowing, unsharp masking, or CLAHE. This research must take into account that perception of these displayed intensities depends on the local intensities and their spatial structure.

The information in any single image must frequently be integrated across the third dimension or across time to make the diagnosis. Methods of 3D or cinematic display can be useful for this purpose, but these are beyond the scope of this report. We will now discuss the means of integrating the information when time or the

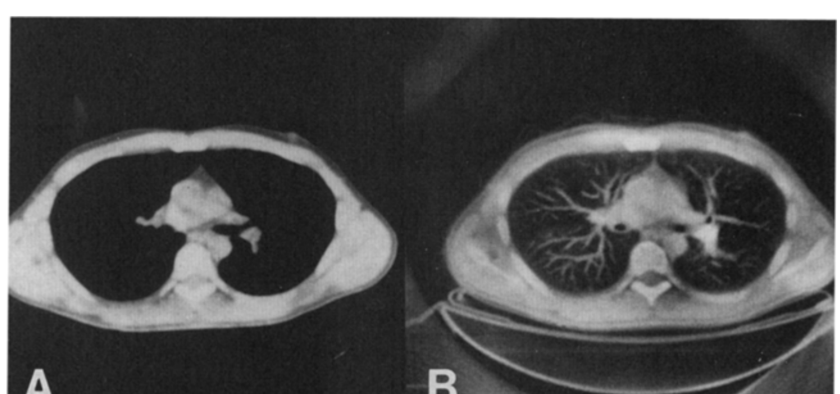

Fig 3. (A) Intensity windowing versus (B) CLAHE on a CT chest.

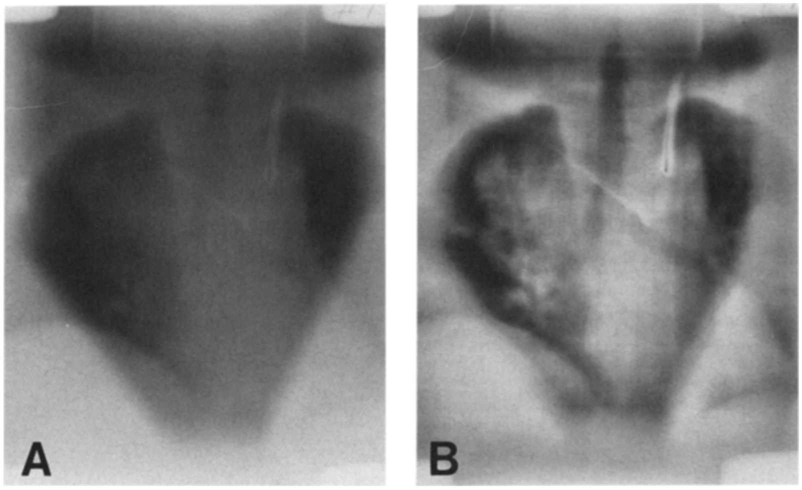

Fig 4. (A) Intensity windowing versus (B) CLAHE on a portal film.

third spatial dimension is captured via a series of 2D gray-scale images.

Navigational Functions

Navigational tasks. The previous section dealt with adequately presenting a single image. But diagnosis requires locating, viewing, and comparing many images associated with a patient. Often, a diagnosis depends on a number of studies, both from the same acquisition modality (taken at different times) and from different acquisition modalities (eg, radiographs, CT, MRI, scintigrams, ultrasonograms). Thus, a diagnosis may be based on many tens of images, easily as many as 100.

Radiologists must be able to use the work station to consult amongst themselves and with referring physicians, and to access patient textual information stored with the image or in the radiology information system. However, the most critical task is an individual diagnosis. This task can be divided into a sequence of steps, each carrying out one of three subtasks: the location and selection of an image to be viewed next, the evaluation of features on that particular image, and the comparison of one image with another. Thus, a means of effortlessly locating and moving to another image must be provided, especially for common movements such as viewing the next image in a CT study. The display of roentgenographic film on a bank of light boxes is effective, not simply because it supports the perception of images, but also because it allows easy movement or navigation among all the images associated with the patient, merely by moving the head and eyes (Fig 5).

Navigation problems also arise when radiologists must view, at full resolution, an image that is sampled at more pixels than can fit on the display screen. In such a case, they need to understand where the zoomed region is in the full image so as to be able conveniently to move their attention to a related place in the image for comparison or simply to look at a new feature.

In a similar way, radiologists must be able to relate image markings of regions of interest and text to an image or group of images. Markings are easily related to an image by switching them on so as to be superimposed on the image or off to remove the distraction. Text, however, can be too voluminous to fit on an image or not be related to a particular image but rather to a set and, if superimposed on the image, it can hide image information. A separate screen for text is therefore frequently provided.

System objectives. The movements among images that must be supported depend upon the sequence in which images are commonly examined. Studies have suggested that there is considerable "locality of reference" among the images.[7] That is, just a few adjacent slices in space or time are frequently the subject of focused inquiry, or one or two adjacent slices from one study may be under comparison with corresponding slices from another study.

To make such movement quick and easy, however, radiologists must have a "mental model" of the image set, that is, a cohesive, consistent, and internalized concept of how the images are organized, located, and displayed. The light box provides such a pictorial mental model, allowing radiologists to take advantage of their strong

Fig 5. Navigation on a lightbox.

spatial memory and thus find desired images simply by moving the eyes or head.

Several work-station designs attempt to provide the radiologist with a mental model of all the images associated with a patient; either as a sequence or, analogously to a lightbox, as a 2D array or as a pile of such arrays, as might appear on a desk. An attractive strategy has been to provide a full array of all the images, in miniature form, as an *aide-mémoire* to reinforce the mental model or metaphor in the radiologist's mind. Pointing on this navigational view yields the selected images in full size, and operations such as moving to the next or previous image in any dimension are supported (Fig 6).

Similar methods are necessary with the partial display of a large image and the user's need to locate a new place on the image. Here radiologists have a natural mental model for the information. Nevertheless, a means of quickly roaming within the image is necessary, and some systems also provide a means of jumping to a new location. Such navigation requires a means of allowing radiologists to comprehend which place in the large image they are presently viewing. For example, reference for the viewed region to a smaller view of the whole image may be provided. Moreover, such an approach provides radiologists with the necessary means for comparison of the zoomed region to other parts of the image, albeit with the latter at lower resolution.

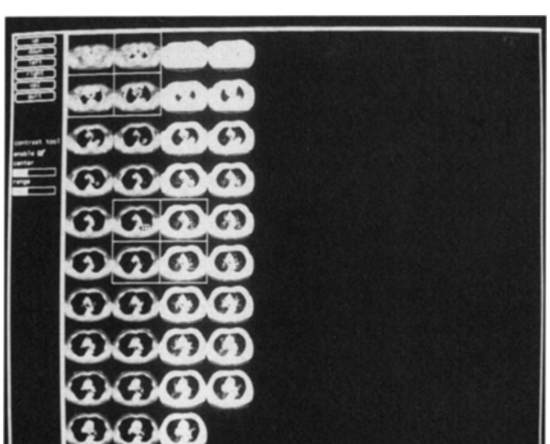

Fig 6. Radiology work station with navigational view.

Even if the navigational and perceptual capabilities of a work station are satisfactory, a work station can be ineffective unless ergonomic issues such as lighting, table height, screen slant, and multiscreen layout, are adequately handled. These issues have been well summarized by Farrell and Booth[8] and Horii.[9]

SOLUTIONS: THE TECHNOLOGY

Available Technology

Perceptual functions. Display transformations intended to present an image with adequate contrast lose quantitative intensity information available in the recorded image. Furthermore, it can be expected that in a few cases it will be desirable to use a nonstandard contrast enhancement, eg, with different parameters of the chosen enhancement method, to match the needs of a particular image. Therefore, one cannot simply compute the display-ready image once and for all and discard the original. On the other hand, one is reluctant to incur the increase in storage costs associated with storing the display-ready image in addition to the recorded image.

The alternative is to reapply the contrast enhancement for each presentation or transmission from the archive to the display station. For the interactive method of intensity windowing, the contrast enhancement must be computed on-line. While this most commonly is done using look-up table techniques, this approach is affordable only when one or at most a few different intensity windows are to be applied for a screenful of images. When many images are to have different windows, the contrast enhancement must be computationally applied.[10]

The adaptive, noninteractive methods must be computationally applied, and the time required is greater than can be allowed for a reapplication for each new presentation. Instead the enhancement is performed for all the images in a study when they are transmitted to the display station. If this is not to impede the diagnosis, this enhancement must take place in a time comparable to transmission times from the archive, ie, approximately 1 second per image for tens of images. Fast contrast enhancement computers are therefore of interest.

A 512 × 512 image unsharp masking can easily require 25 million arithmetic operations. Interpolative CLAHE, an approximate form of CLAHE, can require four times as many. This approximate method produces so many artifacts, as compared with real CLAHE, that the latter method, slowing the speed by perhaps 50 times more, is preferable.

These calculations can take place on a general-purpose computer, on a device oriented to image computing, or on a special-purpose processor designed only to carry out the particular contrast enhancement task. Present general-purpose computers cannot carry out the contrast enhancements at the required speed, because these are more than an order of magnitude too slow for unsharp masking and three orders of magnitude too slow for real CLAHE. However, general image computers using parallel processing are now being provided, either integrated as part of a work or display station or as an optional add-on. These image computers presently add $20,000 to $50,000 to the cost of the display station. For 512 × 512 images they allow unsharp masking in a large fraction of a second and interpolative CLAHE in 2 to 4 seconds. Real CLAHE still requires a few minutes, but the speed and availability of such engines is increasing quickly. Nevertheless, if real CLAHE is required, special-purpose engines seem necessary. One design is under development that will allow real CLAHE of a 512 × 512 image in 4 seconds, with an intermediate result of almost full quality in less than 1 second.[11] Such engines can be produced at a cost comparable to the general-purpose image computing add-ons.

Navigational functions. Using film and a 4 × 2 array of lightboxes, a radiologist can simultaneously view as many as 120 CT images; only eight or 12 can be simultaneously viewed on a two- or three-screen display work station. This lack of display "bandwidth" is a critical work station design problem that must be overcome either with many screens, a few extremely large screens, or a few screens incorporated into an extremely well-designed user interaction.

Several current radiologist work stations use six or eight 1K × 1K display screens. One, in particular, uses the metaphor of a film and light-box alternator (Fig 7). However, the cost of such a system, as well as its physical space requirements, are high.

Another approach is to use one or two extremely large (2K × 2K) display screens, each with four times the display area of a 1K × 1K screen. Such large homogeneous display areas have the potential to allow an interaction similar to that currently found with film and light box, but with a reasonably sized footprint.

For a work station of one to three screens to be viable, the severely limited display area must be overcome using several methods. First, an image index must be provided to allow the radiologist to quickly locate images and understand their relationship. Second, a minimal effort, in terms of cognitive load and hand motions, should be required to manipulate the work station and select various images for display. Finally, the work station must display the images with enough speed to avoid affecting the radiologist's diagnosis.

Currently, the images arrayed on the light boxes serve first as an image index, allowing the

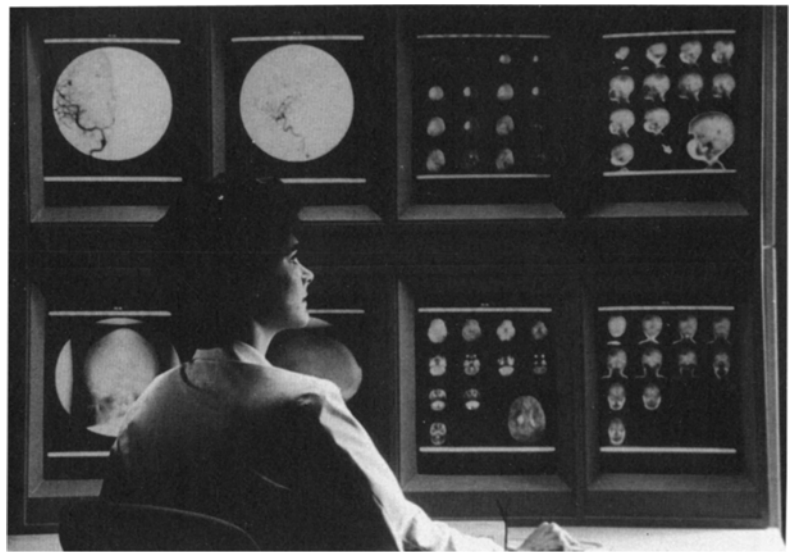

Fig 7. Multiscreen radiology work station. (Courtesy of the Siemens Corporation, Des Plaines, IL)

radiologist to quickly locate a particular image, and second as a means of viewing and understanding the relationships between images. Two methods are currently available to provide this function with a radiologist's work station: the textual and the pictorial indexes (Fig 8). The textual index[10] uses a list of patients, sublists of studies, and sub-sublists of image numbers, to represent all the images available to the radiologist. By pointing to the required patient, study, and image number, the radiologist causes the required image to be displayed. The pictorial index uses greatly reduced CT images to represent all the images for a given patient. These image-icons are arranged either in a plane, analogous to the images on the light-box array, or in a strip. Given the highly spatial nature of the diagnosis task and the need for a clear metaphor that such an index provides, a pictorial index appears preferable.

This index, either pictorial or textual, either can be permanently displayed or only appear on command. The permanent display of the index continually reinforces its metaphor to the user and eliminates the cognitive load and hand motions required to make it appear and disappear. On the other hand, the permanent display of the index takes up valuable screen space that could be used to display more images. The ideal system would allow radiologists to vary whether the index is displayed permanently or only on demand, depending on their current needs.

Besides the layout of the index relative to the images ready for diagnosis, the technology must support a style of layout of those images on the work-station screen(s). Two methods are in vogue, tiled and overlaid. With a tiled layout, images fit next to one another as in a mosaic. The advantage is that images do not occlude one another, but there is a difficulty of fitting different-sized images together and of filling up a screen too quickly. With overlaid images the images can be more flexibly placed, including on top of each other, as on a desk, with only portions of an image appearing. An image is brought to the top of the stack by a command including pointing at the image. The difficulty is the need for the user to continually control the layout, when he or she wants to override the default.

An implementation issue of the overlaid layout provides a further advantage. This layout is normally implemented with an intermediate image buffer between the main memory and the frame buffer that is used to hold the image in a form independent of its screen location.[7] This approach allows operations such as contrast enhancement, roaming, and zooming to be easily and quickly applied.

It is critical that hand motions not interfere with the diagnosis task. While the use of a mouse, pull-down or pop-up menus, and other techniques of direct manipulation do help, they are not the whole answer. The user must have a clear mental model of how the system works,[12] and the hand motions must follow from it. These motions must be optimized for frequent tasks. For example, requesting the next images in a CT study is extremely common and should require a minimal

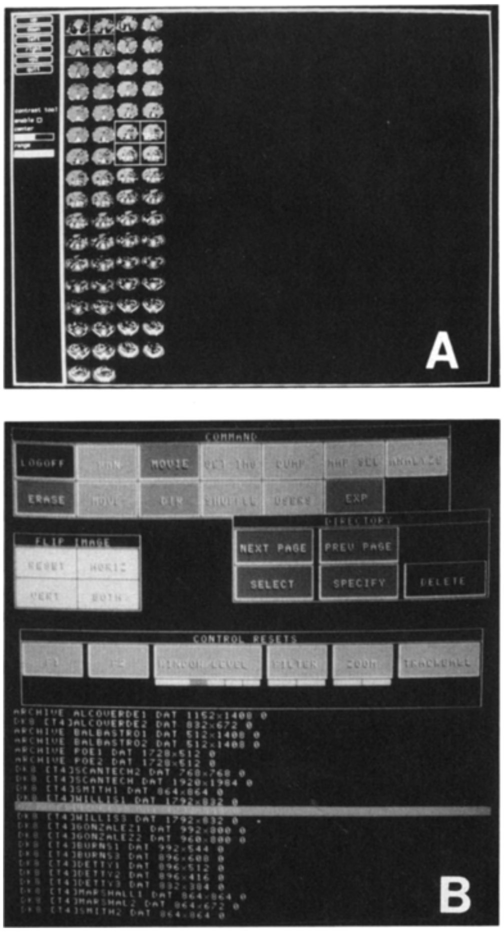

Fig 8. Patient image indexes. (A) Image (pictorial) index. (B) Textual index (courtesy of William Dallas, Radiology Department, University of Arizona).

number of motions. Finally, hand motions should be fine tuned to eliminate interaction errors.

With a small display area, image display speed is critical to avoid disrupting the diagnosis. The required speed depends on the effectiveness of the work station's mental model and on the effectiveness of the image-index; experiments[13] suggest that with a clear metaphor for a mental model and an effective pictorial image-index, a display response of 0.7 seconds for the next two CT images is sufficient for smooth operation.

Needed Research and Technology

Electronic display provides many advantages of image access and image processing, but film on a light box has superior innate spatial and contrast resolution. Indications are that with adequate contrast enhancement and roam and zoom capabilities, together with at least $1,024 \times 1,024$ spatial sampling, diagnosis with electronic displays can match that from film in accuracy. However, research is necessary to establish this fact and also to develop the optimal contrast enhancement and roam and zoom approaches that can achieve this property.

In particular, methods for improved contrast enhancement, avoiding shadows on high contrast edges, need to be developed. Research also must be carried out on the correct display scale to use with such a method, or on the design of a contrast enhancer that builds in the hardware display scale of the device to be used. In addition, faster computing devices that implement adaptive contrast enhancement methods in under 1 second per image need to be made routinely available.

Display systems to be used for consultation must be viewed from a greater distance than single-user stations. They must therefore be considerably larger than many electronic work stations are today. Work station screen space in units of 40×40 cm for a total of 6,400 cm^2 needs to become available.

Comparing the film and lightbox array to any possible work station design shows three critical differences: the considerably reduced image display area, the less familiar hand motions, and, most importantly, the more obscure mental model. One can overcome these difficulties only by constructing the work station to exactly match the radiologist's tasks. Thus, we need to carefully study the tasks radiologists need to perform. For example, what images from what studies need to be examined simultaneously? What patterns of movement of radiologists use to move through a single study or to compare side-by-side the same anatomical features from two different studies? Understanding the diagnostic process will require observing and videotaping the radiologist conducting actual diagnoses with both film and light box and with various radiology work stations. This observation should take advantage of techniques from experimental psychology such as protocol collection and error analysis.

As our understanding of the diagnostic task improves, we can begin to consider how various navigation strategies and work station designs will aid or hinder radiologists. The best approach is to develop prototypes and evaluate them using controlled experiments and field studies.[12] This can be time consuming, and a faster approach is to construct a model[12,14] of the radiologist's actions using a particular work station design. A keystroke model[14] is an analytic tool that not only considers the times for radiologists' various hand motions but also the time to pause and think about an operation before its execution.

The radiologist's work station memory must be rather large and fast if access speed adequate to effective navigation is to take place. It can be thought of as having three levels: work station disk, work station main memory, and work station frame buffer. Current work stations cannot even move images from main memory to the work station's frame buffer with sufficient speed to fully support the required interaction; moving images from the work station disk to the frame buffer is still far too slow. Main memories that are large enough to hold 100 images and be used directly as a frame buffer seem likely to be useful and are beginning to appear.

Alternatively, we may be able to take advantage of the fact that radiologists' image-access patterns are highly predictable, primarily consisting of moving to the next or previous images in a study or moving within a radiograph. This locality of reference should allow the work station to prefetch images accurately from disk into main memory, allowing sufficient performance with 10 megabytes of displayable main memory per display screen.

Finally, networks with transmissions speeds in the 100 megabit to 1 gigabit per second range are beginning to be discussed in the telecommunication industry. These rates are far faster than the backplane rates of most current radiology work stations. At such network data rates, a viable work station might simply consist of a monitor, a human interface device, a large framebuffer, and a network connection to a very high-speed archive.

ACKNOWLEDGMENT

We gratefully acknowledge our colleagues R. Eugene Johnston, Diane C. Rogers, and Victor Klymenko both for work on which this report is based and for useful conversations. Our radiologist colleagues, J. Randolph Perry and Jeffrey Creasy, have provided a critical contribution of the clinical issues of display. We thank Sharon Laney and Carolyn Din for help with the manuscript.

REFERENCES

1. ACR Report: Engineering Research in Visual Perception, sections III. 1.3-5, IV, and V, Reston, VA, American College of Radiology, November 1986

2. Chang C, Martin N, Templeton A, et al: An adaptive histogram for contrast enhancement of digitized mammographic films: An initial study. Kansas City, KS, University of Kansas technical report, 1988

3. Pizer S, Amburn E, Austin J, et al: Adaptive histogram equalization and its variations. Comput Vision Graphics Image Proc 39:355-368, 1987

4. Pizer S, Johnston R, Rogers D, et al: Effective presentation of medical images on an electronic display station. Radiographics 7:1267-1274, 1987

5. Zimmerman J, Pizer S, Staab E, et al: An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. IEEE Trans Med Imag 7:304-312, 1988

6. Sorenson J, Mitchell C, Armstrong J, et al: Effects of improved contrast on lung-nodule detection: A clinical ROC study. Invest Radiol 22:772-780, 1987

7. Beard D, Pizer S, Rogers D, et al: A prototype single-screen PACs console development using human-computer interaction techniques. Society of Photo-Optical Instrumentation Engineers, Medical Imaging, Vol 767. 1987, pp 646-653

8. Farrell R, Booth J: Design Handbook for Imagery Interpretation Equipment. Boeing Aerospace Company, 1975

9. Horii S, Horii H, Mun S, et al: Environmental designs for reading from imaging work stations: Ergonomic and architectural features. J Digit Imag 2:156-162, 1989

10. McNeill K, Fisher H: A model for radiologic workstation-user interface design. Society of Photo-Optical Instrumentation Engineers, Medical Imaging, Vol 767. 1987, pp 713-716

11. Austin J, Pizer S: A multiprocessor adaptive histogram equalization machine, in Information Processing in Medical Imaging, X, New York, NY, Plenum, 1988

12. Wendler T: Human-computer interaction in radiology, in Computer Assisted Radiology. Berlin, Springer-Verlag, 1985, p 669-678

13. Beard D, Cromartie R, Creasy J, et al: Experiment comparing image-locating on film and the FILM-PLANE radiology workstation. Society of Photo-Optical Instrumentation Engineers Medical Imaging II Vol 914. 1988, pp 933-937

14. Card S, Moran T, Newell A: The Psychology of Human-Computer Interaction. Hillsdale, NJ, Lawrence Erlbaum Associates, 1983, pp 259-312