TR 85-005

Psychovisual Issues in the Display of Medical Images

Stephen M. Pizer

Department of Computer Science
University of North Carolina
Chapel Hill, North Carolina 27514

March, 1985

To appear in Proc. NATO Advanced Studies Institute on Pictorial Information Systems in Medicine, 1985.

Psychovisual Issues in the Display of Medical Images

Stephen M. Pizer, Ph.D.
Departments of Computer Science and Radiology
University of North Carolina
Chapel Hill, North Carolina, U.S.A.

Abstract

Given a recorded image as a continuous or discrete array of measured or computed intensities, display is the process by which that image is presented to the human viewer as a light image. PACS imposes certain requirements on display such as the use of digital, electronic display devices, the provision of wide-ranging interactions, and the presentation of multiple images simultaneously so that they can be compared. At the same time the digital displays give one considerable flexibility in specifying the display process, providing options that importantly affect the information transmitted from the recorded image to the observer. Essentially one must match the display process to the needs of the observer and capabilities of the display devices. In this paper the parameters of display will be set forth, relevant properties of the human visual system and of display devices will be surveyed, and display processes to provide the required match will be described. In particular, matters related to the size of the display, the number of display pixels, interpolation, the display scale, and the assignment of recorded intensity levels to the display scale will be covered.

1. Introduction

For the purposes of this paper the term display refers to the process by which images recorded in a computer memory are made visible using electronic displays. The concepts presented herein will also be applicable to hard-copy and non-computer based displays. When starting with a computer, the image is originally represented as an array of numbers representing intensity in a picture element (*pixel*). We call these numbers the *recorded intensities* of the image. Examples of recorded intensities are CT numbers, digitized light levels from a radiographic film, and numbers of scintigraphic events in a pixel. In display these recorded intensities are possibly resampled and are transformed to displayed intensities.

The objective of the display process is not to correct the distortions that exist in the recorded image but to transmit most effectively the relevant information in it. Thus noise, blurring, spatial sampling, spatial distortion, intensity distortion, and intensity nonuniformities over space that come from the way the image was measured or processed are not to be corrected at this stage but rather considered part of the information that may be displayed.

Psychovisual issues related to the display of medical images may usefully be divided into those related to spatial parameters and those related to intensity. In space the major parameter of concern is the size of the display pixels. In intensity the major concern is the means by which each recorded intensity is made to correspond with a displayed intensity, commonly an amount of light. It is useful to think of specifying this correspondence in two parts: first one selects a display scale to be used, and second one chooses a scheme of assigning recorded intensities onto the displayed intensities on this scale. In this paper first spatial issues will be discussed, then intensity issues. Finally a short discussion of issues of perception of motion and of the third dimension will be given.

2. Spatial Issues of Display

In the space domain we must be concerned with resolution, distortion, and sampling. These properties of the display process must be distinguished from properties of the recorded image with the same name. As has been indicated above, the latter will be assumed as given, while the former are to be determined. Thus we will assume that the resolution of the imaging device is known. Further, we will assume that the spatial sampling in the recorded image is adequate to either that resolution or a lower resolution suitable for the use to which the image is to be put. For display, resolution has to do with the accuracy with which pixels can be placed on the display screen. Normally this is much higher than the recorded image or the visual system requires, so we will treat display resolution no further. Nor will spatial distortion of display be covered further, as it is normally small and the visual system, sensitive to local context, is quite forgiving in this regard.

2.1. Spatial Sampling

The spatial sampling of display is a crucial matter. The principal objective here is to avoid the observer seeing the individual display pixels while allowing him or her to make all the spatial distinctions that are allowed by the resolution of the recorded image. Due to humans' great sensitivity to texture, visible pixels decrease sensitivity to contrast in the recorded image (this is called the *pixel artifact*). Since acuity increases with luminance in the video range, our constraint is that even at high video intensities such as 30-70 footlamberts, depending on the device, the display pixels should not be visible, while even at relatively low video intensities all resolution provided in the recorded image should be accessible.

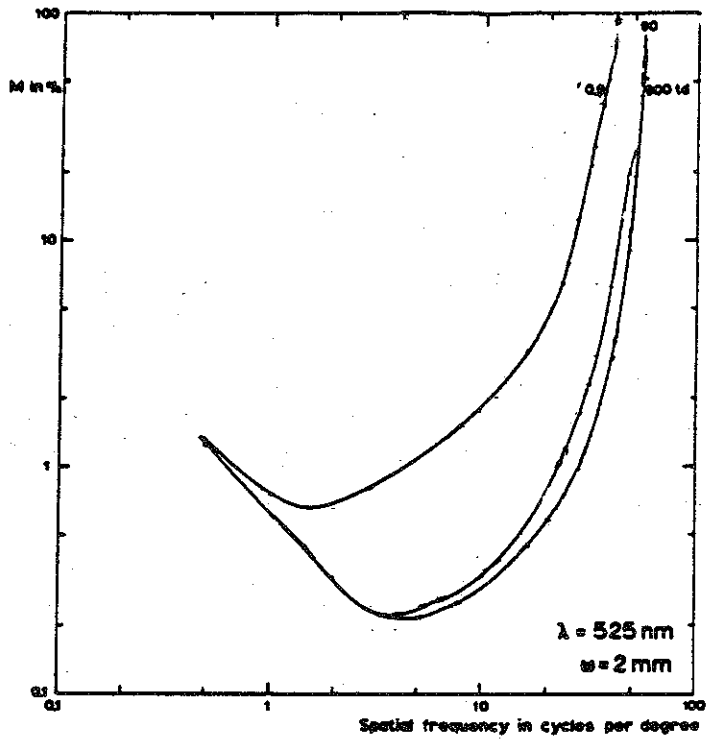

Figure 1. Threshold-modulation curves for green light $\lambda = 525mm$, at three retinal illuminance levels (0.9, 90, and 900 trolands) and a pupil diameter of 2 mm. From [van Nes, 1968].

Curves giving contrast threshold as a function of spatial frequency are given in figure 1 for luminances of 44 footlamberts (high video – about 90 trolands for a 3mm diameter pupil), 440 footlamberts (about half of a high luminance from a light box – about 900 trolands for a 3mm diameter pupil), and 0.44 footlamberts (low video – about 0.9 trolands for a 3mm diameter pupil) for a typical observer [van Nes, 1968]. It must be realized that these curves are dependent on the spatial structure of the pattern being perceived. Nevertheless, from these curves we can conclude roughly that pixels ought to be less than $1'$ of arc in diameter and that the highest resolution to be seen in the image, measured in line pairs, should correspond to no less than $6'$ for video, or $3'$

for a light box. Since the sensitivity to pixels is masked by image noise, pixels of well over $1'$ of arc can be tolerated with noisy images, such as scintigrams [Sharp, 1981].

We conclude that for low-noise images displayed on video we should use approximately 6 display pixels to represent a resolution distance in the recorded image. On the other hand, the sampling theorem together with the fact that summation across a pixel causes some blurring leads to the well known conclusion that to represent the imaged scene without significant loss of information, the recorded image should have approximately 3 pixels per resolution distance. We conclude that normal display sampling should be approximately twice the storage sampling for each spatial dimension. Of course, it is possible to increase the storage sampling, but the gain is only in removing the need to resample at display time, not in image information. Methods of spatial resampling (interpolation) for display will be discussed in section 2.2.

What are the effects of screen size and viewing distance on spatial display requirements? The viewing distance determines the size that corresponds to the numbers of minutes of arc specified above. At a normal viewing distance of 50-60cm $6'$ corresponds to about 1mm. For a lifesize 35cm x-ray displayed in video, the observer can see only 1 lp/mm resolution, only a fifth (or less) of that available in the measured image (for an easily achieved resolution of 5 lp/mm). However, it must be emphasized that only the greater illumination of a light box improves this for hard copy display viewed at the same distance, and then only by about a factor of two. The advantage of hard copy display is that one can move closer if one wants to achieve the full resolution available from the recorded image.

With the requirement of 1 pixel per minute of arc the viewing distance of 50-60cm implies a display pixel size of 0.15mm or less. Once this pixel size is established, moving towards the screen provides little information increase because of the contrast losses caused by the increasing pixel artifact as you move. Similarly, decreasing the number of display pixels without incurring the pixel artifact is possible only with a smaller displayed image.

In practice it appears [Burgess, 1985] that observers choose a viewing position that varies by task over about a factor of two. Also, performance has a quite flat optimum as a function of viewing distance, falling off by only 5% a factor of two from the optimal distance. However, observers' comfort falls by a considerable amount when they are a factor of two from the distance at which they are most comfortable.

With a display pixel size of 0.15mm an image width of 35cm requires about 2000 display pixels. One can use these pixels to achieve a perceived resolution of approximately 1/350 of the image size if the recorded image is sampled to about 1000 pixels. Thus a CT scan with 1 lp/mm resolution and an (abnormal) sampling of 1000×1000 to support this resolution would require a screen of approximately 35cm with 2000 display pixels. No smaller screen would do unless the viewing distance were decreased and the pixel size changed to accommodate the same number of pixels. And to take advantage of the full resolution of a radiograph (at least 5 lp/mm) while maintaining the 50-60cm viewing distance, one would require a display screen of over 1.75 meters with about 10,000 pixels in each dimension. Since this is clearly undesirable, both economically and ergonomically, an ability to zoom and roam is desired. This would allow part of a recorded image to be viewed at full resolution on a reasonable size screen.

In summary, digital display differs from analog display in the fact that the pixel artifact determines the viewing distance and thus the perceivable resolution in a displayed image. Thus the perceivable resolution can be increased significantly only by zooming (if the recorded image is sampled finely enough to support the increase). The numbers given above can be modified by a small factor depending on the conditions of the image and viewing environment, but the basic limitation of perceivable resolution by the pixel artifact remains.

2.2. Interpolation

The comparison of two images from different imaging modalities is most conveniently done when the displayed images are of the same size. However, the discussion above implies that the displayed size of the image should be in inverse proportion to the resolution of the recorded image, and information losses will be obtained if the image is simply enlarged without resampling. Thus, image comparison across imaging modalities normally implies interpolation. Similarly, the convenience of displaying an image at life size can only be effectively achieved via interpolation. Finally, we have shown above that the benefits of minimizing storage requirements imply that interpolation be part of the display process, at least for relatively noise-free images. Therefore, in this section the methods of interpolating finer sampling from a recorded image are discussed.

Parker et al [1983] give an excellent survey of interpolation concepts and methods. The following is paraphrased from their article.

Generally interpolation is accomplished using a dimensionally separable weighting function $f(x)$ to produce the interpolated image $i'(x', y')$ from the recorded image $i(x, y)$ by the convolution

$$i'(x', y') = \sum_{x,y} f(x' - x)\ f(y' - y)\ i(x, y).$$

Probably the two best methods are bilinear interpolation in which

$f(u) = 1 - |u|$ for $|u| < 1$ and zero otherwise;

and high-resolution cubic spline interpolation in which

$f(u) = u^3 - 2u^2 + 1$, for $|u| < 1$,

$f(u) = -u^3 + 5u^2 - 8u + 4$, for $1 < |u| < 2$,

and $f(u) = 0$ otherwise.

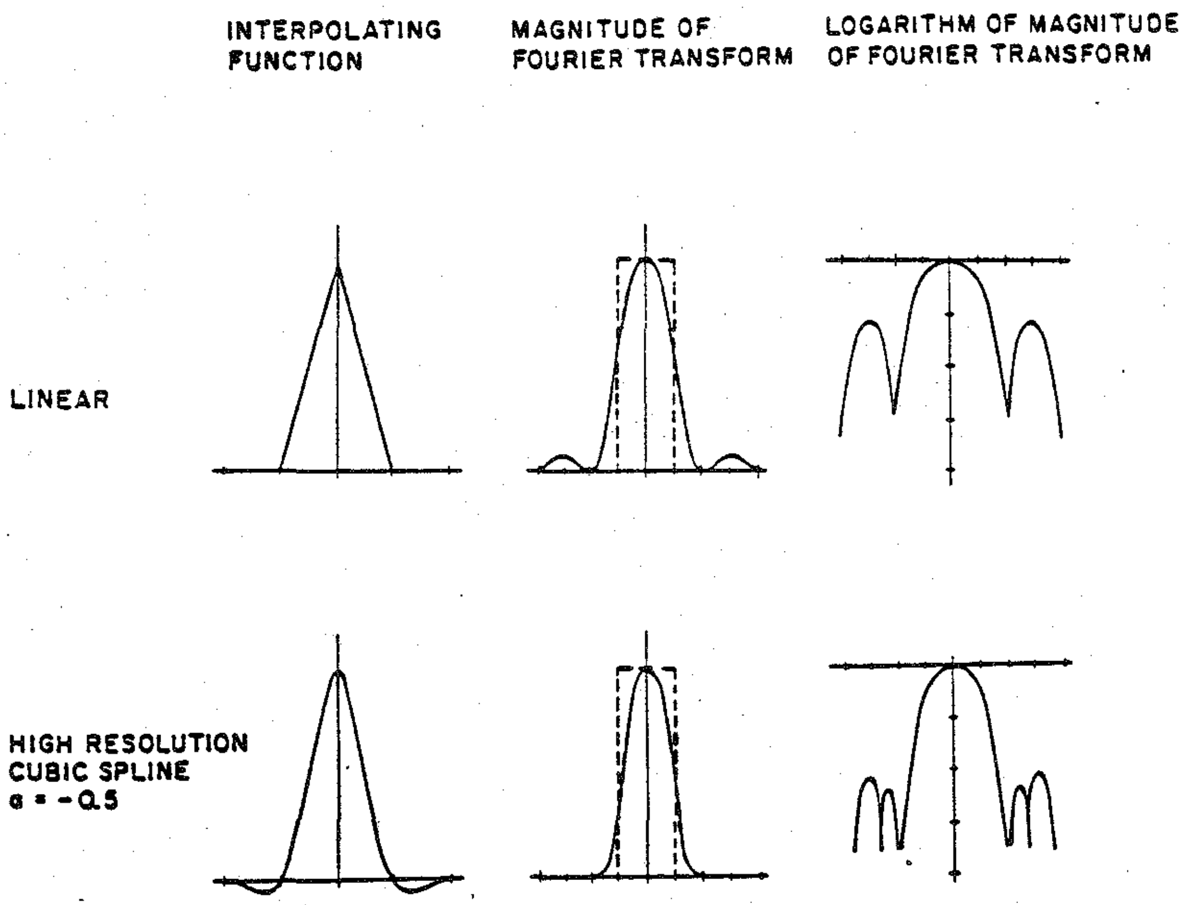

Figure 2. Two interpolating functions. From [Parker, 1983].

For both of these $i'(x, y) = i(x, y)$ at the original sample points, i.e. the result is a true interpolation leaving the image unchanged except for filling in new pixels, a desirable situation. Generally, Parker shows (see figures 2 and 3) that bilinear interpolation causes some resolution loss but is quite efficient, since each new pixel only involves a weighted sum of four original pixel values; whereas the high resolution cubic spline method better transmits the resolution in the recorded image, but it requires sixteen original pixel values per new value and thus is about four times slower.

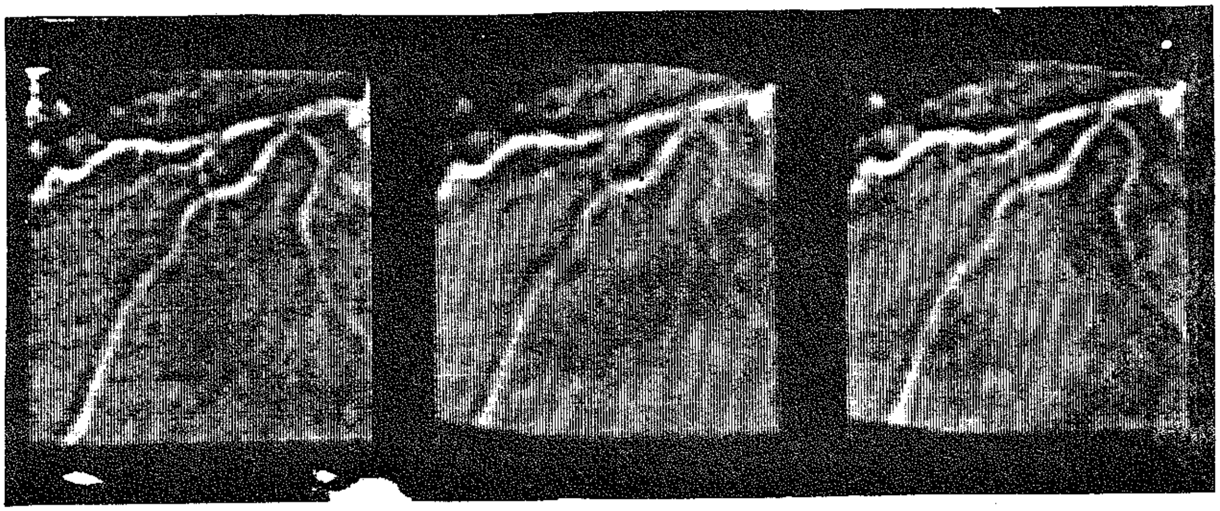

Figure 3. Image resampling. (a) Initial image of a coronary angiogram. The primary data is 64 × 64 with a display dimension of 128 × 128. (b) Resampling using the bilinear interpolating algorithm. Notice the loss of sharpness at the edges of the vessels. (c) Resampling using the high- resolution cubic spline. From [Parker, 1983].

3. Intensity Issues of Display

For display, at each pixel the recorded intensity must be transformed into a displayed intensity. The displayed intensity is normally in the form of luminance or color, but it may sometimes involve some other parameter such as apparent height or motion. This section focuses on matters related to choosing this transformation.

More completely, the transformation that we are concerned with is from recorded intensity to perceived intensity (see figure 4), since it is in the perceived image that the information needs to be optimally available. Thus, we can distinguish three different intensities: the *recorded intensity* that is input to the process, the *displayed intensity* that is produced by the display device, and the *perceived intensity* (i.e. brightness) that can be thought to be generated in the observer's visual system. This perceived intensity will be more carefully defined in section 3.1.

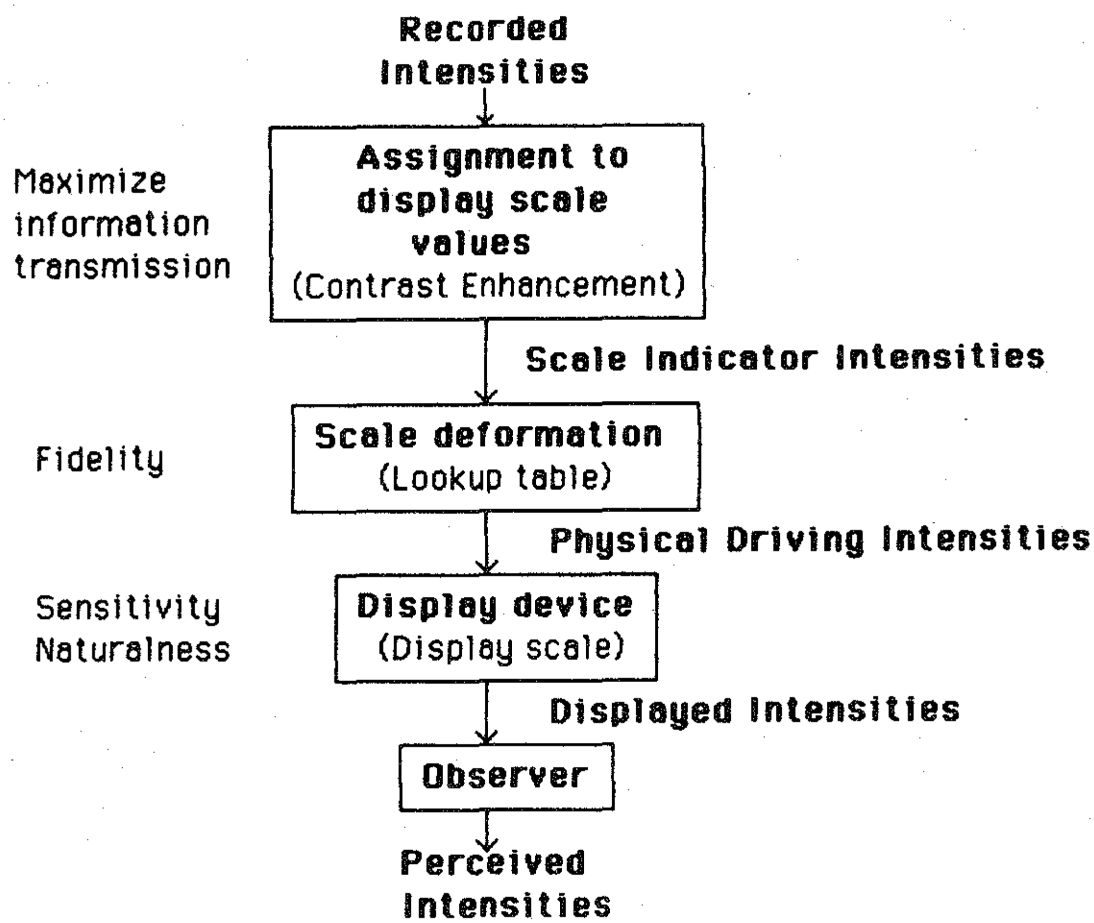

Figure 4. Intensity types and their transformations.

We can control only the transformation from recorded to displayed intensities. It is useful to divide this transformation into three components (see figure 4) in order to separate the concerns of

(1) maximizing information transmission (dependent on the image and the viewing task),

(2) choosing a display scale, i.e. a path through color space or some other space, that provides satisfactory intensity distinctions (dependent on the observer), and

(3) controlling the rate at which we move along the display scale path (dependent on the display device).

In the first component one attempts to achieve the property that intensity differences increase with the importance of seeing the difference and thus information is optimally presented. This step is sometimes called *contrast enhancement*. The result of this step is used to select an intensity on whatever display scale is used; it is the *scale indicator intensity*. The display scale, a path through color or other space, is considered at this point to be controlled in regard to the rate at which the path is traversed. This control over the local stretching or compressing of the display scale is achieved by interposing a transformation between the scale indicator intensity and the *physical driving intensity* (e.g. one or more voltages), which together with the basic display device (e.g. CRT) determine the displayed intensity.

In order to choose the above-mentioned component transformations, we must specify what is meant by optimal information transmission, display scale path, and rate along this path. These issues are treated in reverse order in the next sections.

3.1. Display Scale Linearization

To make it possible to design contrast enhancement so as to maximize information transmission, the succeeding transformations of display scale position assignment, display scale deformation for distortion correction, and perception (see figure 4) must not distort the contrast relationship. That is, they must faithfully transmit intensity ratios: perceived intensity must be linear with the display scale indicator intensity that is the output of the contrast enhancement. Achieving this linearization will have the secondary benefit of standardization across display devices: Any image will appear the same on any display device except for differences due to variations in overall display scale sensitivity. That is, the relative values of perceived contrasts will be independent of the display device.

A method for linearization requires first a knowledge of the perceived intensity corresponding to any physical driving intensity, and this in turn requires a definition of perceived intensity. These can be defined in terms of absolute intensity judgements of observers or of judgements of intensity differences. Since the latter is more relevant to pattern perception in medical images, the definition in terms of intensity differences is more appropriate here. The natural units for this perceived intensity are those in which the observer perceives intensity differences, that is *just noticeable differences (jnd's)*. It is useful to define the perceived intensity of zero as the intensity which is at the bottom of the display scale. Then other scale locations will produce a perceived intensity that is a specified number of jnd's above the minimum intensity.

The jnd is defined as that change in physical driving intensity that results in a just perceivably different displayed intensity, where we must carefully define what we mean by just perceivably different. The jnd is in units of driving intensity, and it is a function of the reference value of driving intensity in which a change is being perceived. This function we call the *jnd curve*, giving jnd vs. reference intensity (see figure 5).

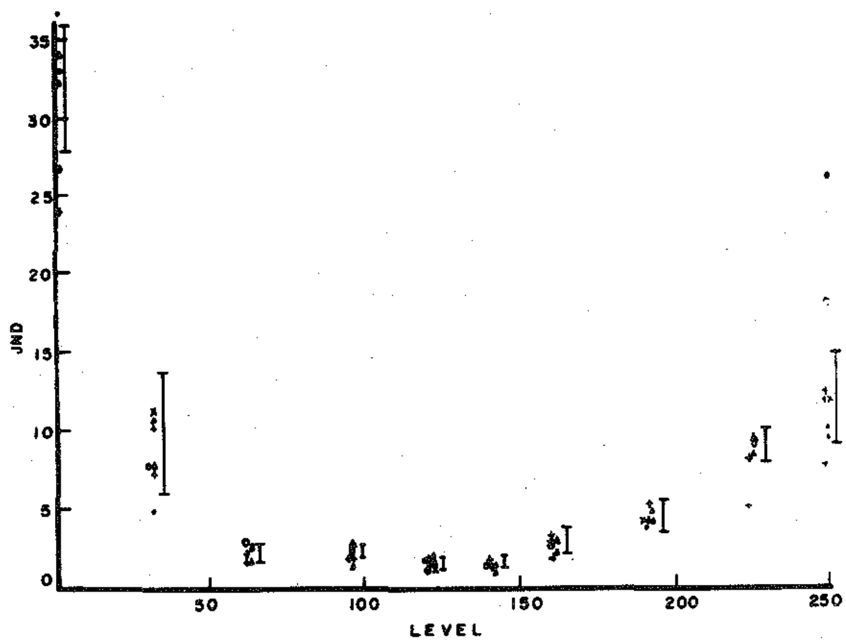

Figure 5. Measured jnd curve for grey scale on a Tektronix 690SR monitor.

A definition for "just noticeably different" must include a full specification of the target, background, and viewing environment together with a probability of correctness (true positive probability) defining the detection of a change and a false positive probability specifying the conservatism of the observer. Many definitions of jnd have failed to include this last factor, the probability that the observer will see a difference when none exists, and since observers can easily vary their conservatism, the resulting jnd can vary significantly in some situations. Pizer and Chan [1980] have suggested using 50% as the true positive probability defining detection and 5% as the false positive probability defining the level of conservatism. Determining the jnd involves in principle carrying out an ROC experiment [Green & Swets, 1974] to determine the change in driving intensity producing an ROC curve passing through this (0.05,0.50) criterion point. Since the two displayed stimuli being distinguished are so close in intensity, it is reasonable to expect, and Chan [1982] showed, that the variances of the two decision variable distributions corresponding to change and no change are the same, that is the ROC curve can be described by the single variable conventionally called d' describing the standard-deviation-normalized difference between the means of the decision variable distributions. The criterion ROC curve can be shown to have a value of d' equal to 1.645, or equivalently correspond to a two-alternative forced choice experiment with a fraction correct of 87.8%.

Johnston [1985], Pizer [1980, 1982, 1983], et al have measured jnd curves of various display scales with ROC rating and two-alternative forced choice experiments using a target consisting of two nearby but separated squares in a background with intensity chosen so that an 8 degree region viewed at 60cm centered at the target has a fixed average adapting intensity. Given such a jnd curve, *jnd(i)*, over a range of reference intensities i_{min} and i_{max} defining the full range of driving intensities for a particular display device, Pizer [1982] showed that perceived intensity, *P(i)*, is given by

$$P(i) = \int_{i_{min}}^{i} \frac{jnd'(i)}{jnd(i)\,\log\,(1 + jnd'(i))} di.$$

Thought of as a function of i, we will call this the perceived intensity function.

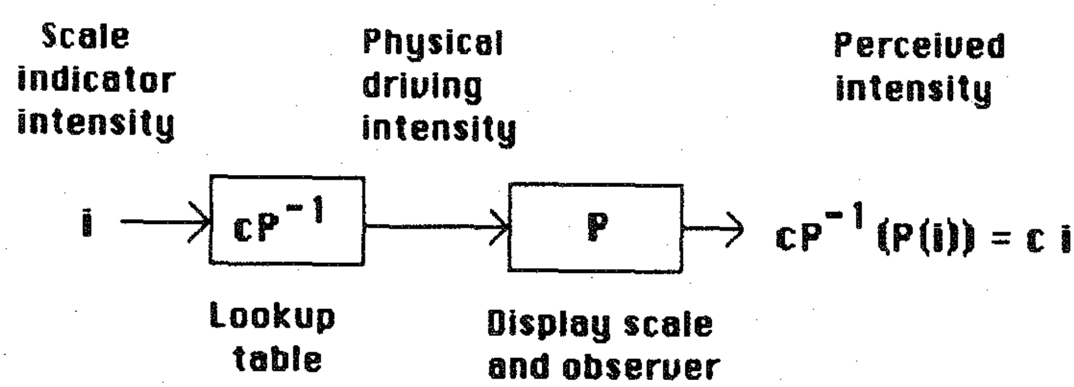

Figure 6. Linearization as the inverse of P(i).

By definition a linearizing function transforms scale indicator intensity to physical driving intensity so that the relation between scale indicator intensity and perceived intensity is linear. Thus (see figure 6) the linearizing function required by a particular device and observer is proportional to the inverse of the perceived intensity function, normalized so that the range of physical driving intensity matches that of the display device. This linearizing function is frequently implemented by insertion in a lookup table of a digital display system. Experience shows that using the physical display scale directly (an identity lookup table) usually results in a strongly different perceived image from one produced with linearization (see figure 7). In particular, the grey scales of video displays normally have an abnormally large jnd (low sensitivity) for small physical driving intensities (see figure 5), and linearization increases sensitivity to contrast in the lower levels of recorded intensity. We will assume in the following sections that all scales being considered are linearized.

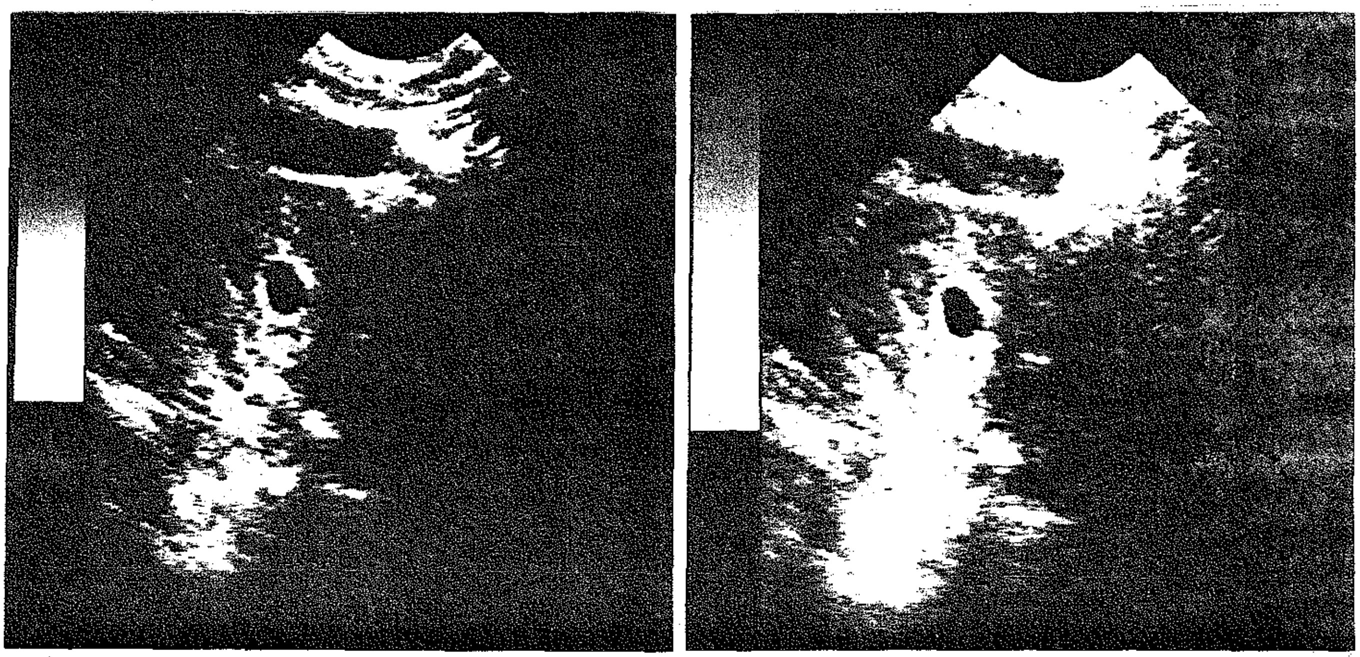

Figure 7. Ultrasound image of the intrauterine sac shown on an original unlinearized (left) and linearized (right) video display.

Since the jnd curve is in principle a function of observer, viewing environment, and image structure, we need to know how sensitive the linearizing function is to these factors. That is, do we need to change linearizing functions as the observer, viewing environment, or image structure changes? Johnston [1985] has demonstrated that inter-observer variations in jnd curve are comparable to intra-observer variations and thus a single linearizing function is satisfactory across observers. Johnston and Pizer respectively have pilot results indicating that although the jnd curve does change with viewing environment and image structure, the linearizing function, determined by the ratio between values along the jnd curve, does not change significantly with these factors. Thus it appears that a single linearizing function suffices for each display device. But more research is necessary to determine the correctness of this result.

The jnd curve is determined by noise in both the observer and the display device. To the extent that the observer noise is the dominant factor, the linearizing function can be determined by measuring photometrically the relation between physical driving intensity and displayed intensity and then using the relation between displayed intensity and jnd derivable from earlier observer experiments. We have developed a program to accomplish this determination of the linearizing function based on photometric measurements for grey scales, and we are happy to distribute copies of this program.

3.2. Display Scale Choice

Assuming linearization, i.e. fidelity of the transformation between scale indicator and perceived intensities, the display scale should be chosen on the basis of overall sensitivity, naturalness, and edge production. Sensitivity indicates the lowest contrast in scale indicator intensity that can be perceived to a criterion degree. Naturalness specifies the ease with which an observer can determine the the relative difference between pairs of scale indicator intensities. Edge production determines informativeness in that false edges limit the comprehension of the recorded image and true edges are the major factor in producing comprehension.

3.2.1. Sensitivity

The overall sensitivity of a display scale is measured by the total number of jnd's across the scale. We call this number the *perceived dynamic range (PDR)* of the scale. Remember that these jnd's were measured in terms of some fixed target, and thus the absolute value of the PDR is target-dependent. But the relative values for different display scales are very informative, and the absolute values reported below are in terms of jnd's that match reasonably well our sense of what is "just noticeable".

The PDR is the same as the perceived intensity corresponding to the top of the display scale:

$$PDR = \int_{i_{min}}^{i_{max}} \frac{jnd'(i)}{jnd(i) \log\left(1 + jnd'(i)\right)} di.$$

Its value for various scales, using a target of separated squares on a uniform background (see section 3.1), is indicated in Table 1. It can be seen that various pseudocolor scales have a greater PDR than the grey scale. But other factors must be taken into account before we can conclude which is the best scale.

Scale	PDR (jnds)
grey	90
heated-object	120
magenta	113
rainbow	200

Table 1. Perceived dynamic range for various display scales. From [Pizer, 1982].

Even for a natural, continuous scale such as the grey scale, it is not clear whether informativeness increases monotonically with the PDR. It is possible that we can be too sensitive to contrast in the recorded image. Increasing the sensitivity, for example by choosing a display scale with a higher PDR, eventually causes the observer to see image differences that are dominated by noise or distracting image structure. Does such an increase of the contrast of these noise differences decrease informativeness even though the contrast of differences due to signal are increased proportionally?

The present scientific evidence [Burgess, 1982], based on artificial images, indicates that informativeness does not fall as sensitivity is increased to show the noise better and better. However, the behavior of radiologists in limiting contrast so that noise is not well seen when reading medical images and one's intuition suggest the opposite. It appears that further study is needed with realistic images before this question will be settled.

3.2.2. Edge Presentation and Artifacts; Number of Discrete Intensity Levels

The informativeness provided by a sensitive scale is also affected by the way in which it causes edges to appear. The human visual system is very sensitive to edges. It both carries out edge enhancing transformations and seems to encode information in terms of edges. Edges determine not only image objects [Hubel & Weisel, 1974], but the way in which object surrounds affect object detection [van der Wildt, 1985]. Therefore, isointensity contours that appear as edges in the display but do not represent edges in the recorded image (the *contour artifact)* must be avoided because they limit what is seen. On the other hand edges should be used to present important information.

Edges may be seen when adjacent groups of pixels differ by a large fraction of a jnd or more (with jnd's defined by the separated-square target mentioned above). Thus the difference in the displayed intensity across real edges should always be more than one jnd. To avoid the contour artifact when the objective is comprehension of image patterns (*qualitative display*), one must assure that adjacent digital display scale intensities differ by well less than a jnd. On the other hand, if one wants to present quantitative information such as absolute recorded intensity values or differences, edges are an important means, and adjacent digital display scale intensities should differ by at least 1.5-2 jnd's. Thus display scales for qualitative and quantitative display should be different.

It follows from the above discussion of false edge avoidance that the number of discrete intensities in linearized digital display scales for qualitative display must be in proportion to the total number of jnd's across the scale, i.e. the PDR. The PDR values given in Table 1 are in jnd's determined for separated target regions. But difference detection of adjacent regions is more sensitive than for separated regions, due to the Mach effect of enhancing contrast at edges. Experience indicates that to make every discrete display level significantly less than one jnd from its neighboring levels, in order to avoid the contour artifact in qualitative display, we should arrange that the number of digital levels in the scale is 1.5-2 times the PDR. Thus, 256 is an appropriate number of levels to avoid false edge artifacts with grey or heated object scales.

3.2.3. Naturalness; Pseudocolor Scales

Qualitative display is not just used to compare intensities that are nearby on the display scale. Scales that have no natural order, such as the rainbow scale with the large PDR recorded in Table 1, confuse the observer by distorting the relation between scale indicator intensity and perceived intensity for considerably differing intensity values. Although the matter is poorly understood, it appears important for the observer to comprehend immediately which of two intensities corresponds to more, and in fact to have some feeling for how much more. There is considerable literature on the perceived difference between perceivably distinct, possibly colored intensities, that is, on the distance function $d(v_1, v_2)$ between two displayed intensities v_1 and v_2 that an observer effectively imposes on the space of displayed intensities (e.g. colors). We say that a display scale is ***natural*** if perceived differences do not contradict the differences given by integrating differences along the scale. That is, if i_1, i_2, and i_3 are any three scale indicator intensities such that $i_3 > i_2 > i_1$, and v_1, v_2, and v_3 are the corresponding displayed intensities, then $d(v_1, v_3) > d(v_1, v_2)$; an intensity on the display scale does not appear closer to another intensity on the scale than it does to a third that is closer to it on the scale.

It has long been said that pseudocolor scales can increase overall sensitivity, and Table 1 confirms this fact. PDR results for three linearized color scales all monotonically increasing in brightness are given: a heated object scale going from red through orange and yellow to white, a magenta scale going from red through magenta to white, and a rainbow scale approximately going through the hues of the rainbow (while decreasing in saturation so that monotonic increase in brightness can be achieved). But these four scales (see figure 8) are not equally natural; observers informally report [Pizer and Zimmerman, 1983] that the heated object and magenta scales have a natural order, but the rainbow scale does not (we have not yet applied the above process for testing naturalness to these scales). And despite its considerably greater PDR, the rainbow scale appears to give far less information and is preferred by no observers over the grey scale. In contrast, the more natural pseudocolor scales, and especially the heated object scale with the slightly greater PDR seems to give more information than the grey scale. Even this result is disputed by various experiments faulted by the use of nonlinearized scales [Todd-Pokropek, 1983; Burgess, 1985]. Moreover, the apparent advantage of even natural pseudocolor scales is also brought into question by the fact that the visual system has low spatial acuity in distinguishing chromanence changes, as changes in nearby pixels may not be more sensitively perceived as a result of chromanence changes.

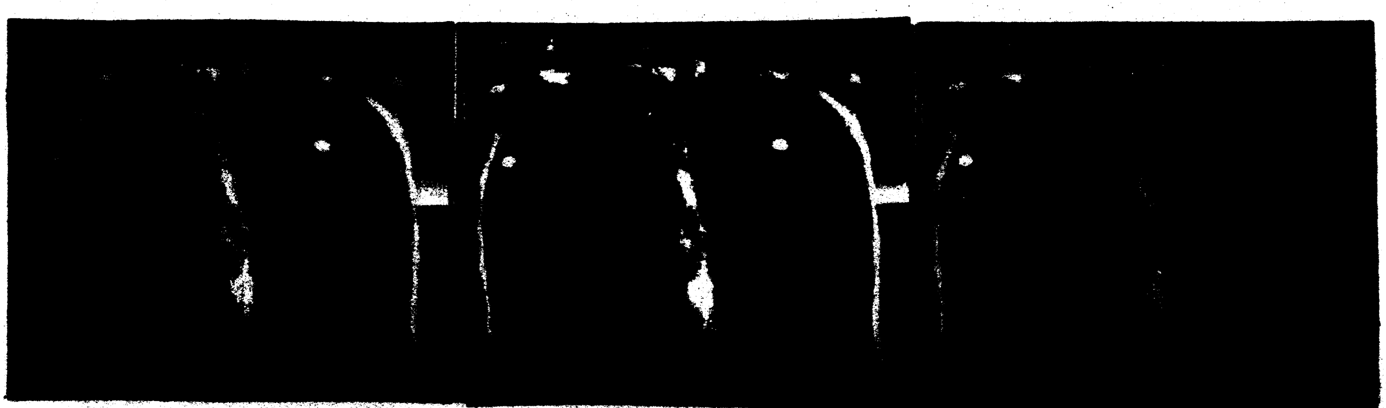

Figure 8. An ahe'd chest radiograph in three linearized scales: grey scale, the heated object scale, and the rainbow scale. The scale appears next to each image.

It is interesting to compare the curves giving red physical driving intensity, green

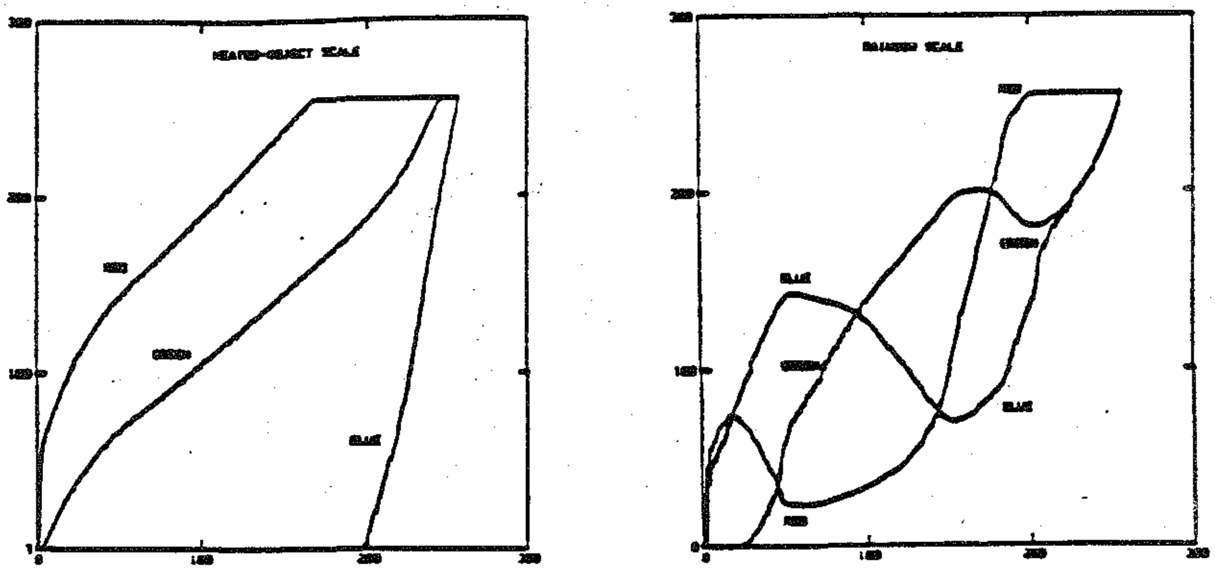

Figure 9. Mappings from scale indicator intensity to color gun-driving intensity for two linearized display scales.

physical driving intensity, and blue physical driving intensity vs. scale indicator intensity for the rainbow and heated object scales (see figure 9). Pizer and Zimmerman report that scales for which these three curves do not cross, as with the heated object and magenta scales, always appear to be natural, while those for which the three curves do cross appear to be unnatural, even when the scales are monotonic in brightness. The heated object scale has in our experience the greatest PDR of the natural scales with the aforementioned non-crossing property. It is thus possible that pseudocolor scales can be used to give increased informativeness over the grey scale, but even if so, it can apparently provide an increase in the PDR of less than 50%.

3.3. Intensity Inhomogeneities

Finally in the area of properties of the display device and the observer, we must face the fact that displayed intensity as a function of physical driving signal is often quite nonuniform across the display area. Cathode ray tubes *(CRT's)* frequently vary by as much as 20% in luminance across the screen for the same driving signal (voltage). But these variations are normally smooth and result in only proportional changes in the brightness scale at each pixel. Therefore they cause quite low frequency variations in the image. The human visual system is principally sensitive to changes in local context; it is quite insensitive to low frequency changes (see figure 1). Therefore correction for intensity inhomogeneity is normally unnecessary even for the rather large inhomogeneities commonly encountered.

ing contrast relative to local context [Cornsweet, 1970]. It is simply unable to make accurate comparisons in either brightness or color between distant locations. Therefore the context in terms of which an information-loss-minimizing assignment should be computed should be quite local; the assignment function should change across the image. It is said that the assignment should be *adaptive*.

If histogram equalization is the basis of this adaptive approach, the method called *adaptive histogram equalization (ahe)* [Pizer, 1984] is obtained. The method attempts to optimize contrast enhancement everywhere in the image relative to local context, and as a result it provides a single displayed image in which contrasts in widely varying recorded intensities in different image regions, and thus organs, can be easily perceived. A comparison to windowing and to global (nonadaptive) histogram equalization is given in figure 10. Zimmerman [1985] has compared its results to many other methods of adaptive display scale assignment and found it distinctly superior in both producing high contrast and avoiding artifacts. Research is now proceeding to evaluate its effectiveness in a controlled experiment on simulated clinical images (inserted lesions in clinical normals). If it is shown as effective as present anecdotal experience indicates, it is a candidate for a standard assignment method that will avoid the need for interactive contrast enhancement with the large majority of images. We have therefore been facing questions of developing software and hardware that will make its implementation fast. But in this paper we address none of these matters of making the method speedy.

Besides its clinical effectiveness, the major questions with regard to *ahe* are the following.

(1) What should the contextual region size be? Our work with the method to date indicates that when the method operates as described above with each pixel having its own contextual region, a square region that is between 1/4 and 1/8 of the image width on a side is appropriate, with little sensitivity of the result to changes of region size between these extremes.

(2) Should the contextual region be related to the boundaries of nearby objects? The visual system most likely works in just this way [van der Wildt, 1983], but it is hoped and present experience suggests that pattern recognition will not be a necessary part of *ahe* .

(3) Should nearby points in the contextual region be given more effect by some kind of weighting scheme? Locality plays an important role in the operation of the visual

3.4. Assignment of Display Scale Levels to Recorded Intensities

The most common method for assigning display scale indicator levels to recorded intensities is windowing. In this method the user interactively selects a range over which recorded intensities are mapped linearly to the full range of the display scale. While this method has the advantage of giving control to the user, it limits him to a particular type of assignment function and requires that this function be the same everywhere in the image. The following discussion suggests that neither of these choices is close to optimal.

Cormack [1980] has suggested that the assignment function be chosen to minimize information loss in the transformation from recorded to scale indicator intensities (and thus to perceived intensities if the display scale has been linearized). Here information is used in the sense of information theory. While it can be argued that information should be defined in a more task-related way, we will accept Cormack's suggestion as a good starting point.

Cormack goes on to suggest that information be measured on an average per pixel basis, assuming independent pixels. This is clearly a poor approximation to reality, since pixel intensities are heavily correlated, but it does simplify the mathematics. Given this approximation, it can be shown that if the probability distribution describing the noise at a pixel is independent of the intensity of the pixel, the method called histogram equalization [e.g. Castleman, 1979] provides the assignment function. In histogram equalization the assignment function is the cumulative recorded intensity histogram, normalized so that the full range of the display scale is used. With this assignment function if a pixel has an intensity at the p^{th} percentile of recorded image intensities, it is displayed at the p^{th} percentile along the display scale. Thus the display is sensitive to changes in popular ranges of recorded intensities, while few displayed intensity levels are wasted in recorded intensity ranges in which there are few pixels with intensities to distinguish.

If we drop Cormack's approximations but accept his approach of minimizing information loss in the information theoretic sense, we will obtain some assignment function, perhaps approximately that provided by histogram equalization. But in any case this will be a single assignment function designed to optimize information transfer across the whole image. But the human visual system is not equipped to receive information in one part of the image relative to the whole image. It preprocesses the image by record-

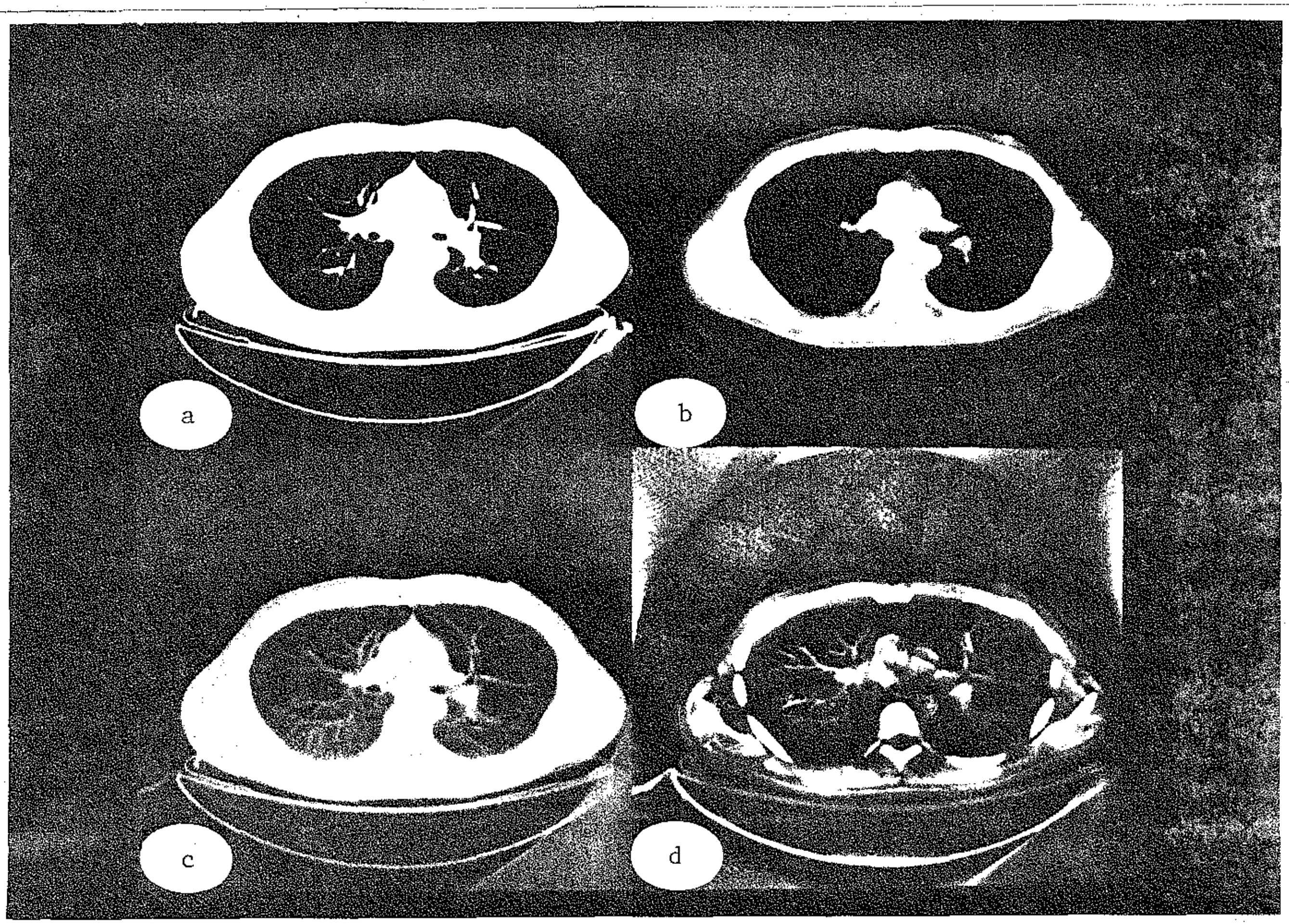

Figure 10. The effect of (a,b) windowing, (c) global histogram equalization and (d) adaptive histogram equalization on a CT scan of the chest.

system.

(4) Should the degree of contrast enhancement be limited, e.g. by limiting the height of the histogram? The method sometimes shows image noise disagreeably well, while also showing signal contrast. This returns us to the previously mentioned issue of whether sensitivity to contrast in the recorded image can be too high. A positive answer to that question would indicate that we must investigate ways of limiting contrast enhancement when it varies locally.

4. Other Visual Dimensions: Motion and 3D

It is now not uncommon for medical image display to show objects in motion by cinematic approaches or to show the third dimension. It also seems worth considering the use of motion and the third dimension to enhance the display of one parameter varying in two dimensions or allow the simultaneous display of additional parameters.

A few comments on visual aspects in motion and 3D are thus in order.

The visual properties of motion detection are well surveyed by Nakayama [1985]. Humans have low sensitivity to low velocities. For moderate velocities the sensitivity to change in velocity is described by a Weber's law: we are equally sensitive to fractional changes in velocity, with a threshold near 5% of the present velocity. It is interesting to note (see figure 11) that the spatial frequency to which we are maximally sensitive decreases with velocity. Our resolution for matching directions of motion is about 1 degree.

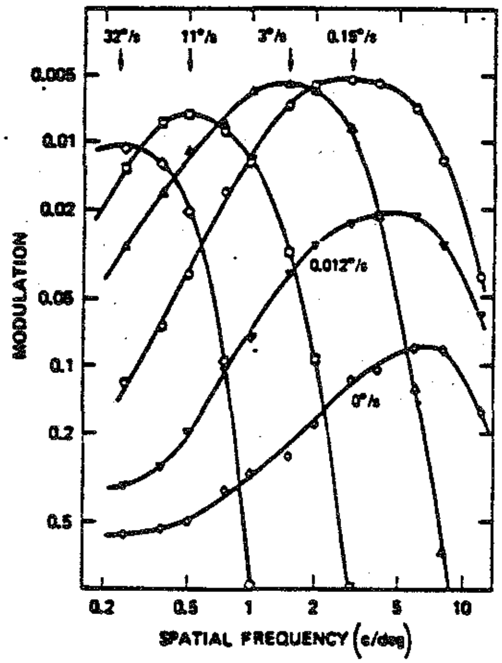

Figure 11. Stabilized contrast-sensitivity curves measured at constant velocity. Data are shown for six different velocities, ranging from 32 deg/s down to zero. From [Kelly, 1979].

In the depth dimension visual resolution in terms of visual angle is about the same as for the other two spatial dimensions. However, when this translates trigonometrically into absolute spatial units, for normal viewing distances such as 1 meter our resolution is on the order of a few mm, far poorer than that in the other two dimensions. The cues to the third dimension are many, including stereopsis (horizontal disparity of the images seen by the two eyes), vergence (the relative angle at which the two eyes are facing, operating up to about 6 meters), accommodation (the force of the muscles focusing the lens, important only for near objects), linear perspective (only important when there are long, straight, parallel edges), head motion parallax, interposition, surface texture, and surface shading. While head motion parallax and interposition appear to be among the most important cues, the stereoscopic effect has received by far the most research attention, and there has been little research in the combination of all of these cues into a single percept. Until such research is done, 3D display will have to be intuitively based.

5. Summary

As can be seen from the preceding, studies of visual perception have much to tell us about the important medical image display issues of spatial sampling, display scale choice and linearization, and assignment of display scale levels to recorded intensities, as well as the use of cinematic and three-dimensional displays. But much research remains to be done, both in basic studies of visual perception and applied studies in regard to the display of medical images. Especially, more understanding of the effect of the display task and of image structure in images of clinical complexity is required.

6. Acknowledgements

I thank Diane Rogers, Arthur Burgess, R. Eugene Johnston, John Zimmerman, and Jan Koenderink for useful discussions, J. Anthony Parker for providing figures, and Richard Bader, Scott Fields, Ching-Man Kwan, and Edward Staab for providing digitized images. I am indebted to Joan Savrock for manuscript preparation and to Bo Strain and Karen Curran for photography. This paper was prepared partially with the support of National Institutes of Health grant number 1-R01-CA39060.

7. References

A. E. Burgess, R. F. Wagner, R. J. Jennings, "Human Signal Detection Performance for Noisy, Medical Images", *Proceedings of the International Workshop on Physics and Engineering in Medical Imaging,* IEEE Computer Society No. 82CH1751–7:99-105, 1982.

A. E. Burgess, Personal Communication, 1985.

F. H. Chan, *Evaluating the Perceived Dynamic Range of A Display Device Using Pseudocolor*, Ph.D. Dissertation, University of North Carolina, Curriculum in Biomathematics and Medical Engineering, 1982.

J. Cormack and B. F. Hutton, "Minimisation of Data Transfer Losses in the Display of Digitised Scintigraphic Images", *Phys. Med. Biol., 25:*271-82, 1980.

T. N. Cornsweet, *Visual Perception*, 1970.

D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, 1974.

D. H. Hubel and T. N. Weisel, "Sequence Regularity and Geometry of Orientation Columns in the Monkey Striate Cortex", *J. Comp. Neur., 158*(3):267-294, 1974.

R. E. Johnston, J. B. Zimmerman, D. C. Rogers, S. M. Pizer, "Perceptual Standardization", *SPIE Conference on PACS*, 1985.

D. H. Kelly, "Motion and Vision, II, Stabilized Spatio-temporal Threshold Surface", *Journal of Optical Society of America, 69*(10) : 1340-1349, 1979.

K. Nakayama, "Biological Image Motion Processing: A Review", to be published in *Vision Research*, 1985.

J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of Interpolating Methods for Image Resampling", *IEEE Transactions on Medical Imaging, MI-2,* (1): 31-39, 1983.

S. M. Pizer and F. H. Chan, "Evaluation of the Number of Discernible Levels Produced by a Display", *Traitment Des Information en Imagerie Medicale* (Information Processing in Medical Imaging), R. DiPaola and E. Kahn, eds., Editions INSERM, Paris: 561-580, 1980.

S. M. Pizer and J. B. Zimmerman, "Color Display in Ultrasonography", *Ultrasound in Medicine and Biology*, 9 (4): 331-345, 1983.

S. M. Pizer, J. B. Zimmerman, and E. V. Staab, "Adaptive Grey Level Assignment in CT Scan Display," *J. Comp. Ass. Tomo.* 8 (2): 300-305, 1984.

S. M. Pizer, J. B. Zimmerman, and R. E. Johnston, "Contrast Transmission in Medical Image Display", *Proceedings of ISMIII* '82 (International Symposium on Medical Imaging and Image Interpretation), IEEE Computer Society (IEEE Catalog No. 82CH1804-4): 2-9, 1982.

P. F. Sharp and R. B. Chesser, "The Influence of Data Interpolation on Image Quality", *Proc. VIIth Int. Conf. on Info. Proc. in Med. Imaging*, Division of Nuclear Medicine, Stanford University School of Medicine, Stanford, CA: 225-239, 1981.

A. E. Todd-Pokropek, "The Intercomparison of a Black and White and a Color Display: An Example of the Use of Receiver Operating Characteristic Curves", *IEEE Transactions on Medical Imaging, MI-2* (1): 19-23, 1983.

G. J. van der Wildt and R. G. Waarts, "Contrast Detection and Its Dependence on the Presence of Edges and Lines in the Stimulus Field", *Vision Res., 23* (8): 821-830, 1983.

F. L. van Nes, *Experimental Studies in Spatiotemporal Contrast Transfer by the Human Eye*, Ph.D. dissertation, Rijksuniversiteit Utrecht, 1968.

J. B. Zimmerman, *Effectiveness of Adaptive Contrast Enhancement*, Ph.D. dissertation, University of North Carolina, Department of Computer Science, to appear in 1985.