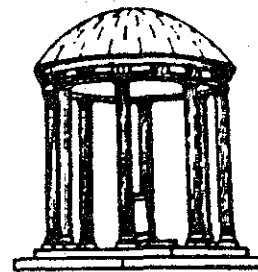


**THE EFFECTIVENESS OF
CONTRAST ENHANCEMENT**

Technical Report 85-029

John Burton Zimmerman

The University of North Carolina at Chapel Hill
Department of Computer Science
New West Hall 035 A
Chapel Hill, N.C. 27514



**THE EFFECTIVENESS OF
ADAPTIVE CONTRAST ENHANCEMENT**

by

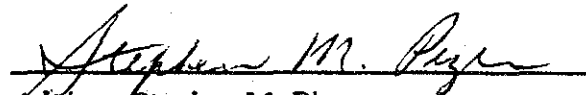
John Burton Zimmerman

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

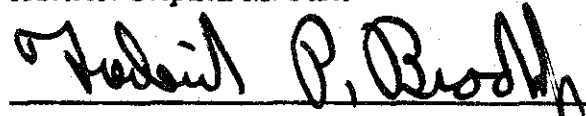
Chapel Hill

1985

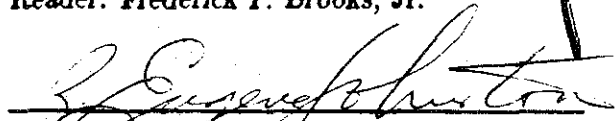
Approved by:



Advisor: Stephen M. Pizer



Reader: Frederick P. Brooks, Jr.



Reader: R. Eugene Johnston

Copyright © 1985
John Burton Zimmerman
ALL RIGHTS RESERVED

Acknowledgements

*To travel hopefully is a better thing than to arrive,
and the true success is to labour.*

—R. L. Stevenson [1850-94]

As my old friend John Lavery used to say "Why do you think they call it a graduate career anyway?" Given the length of my own graduate career, it is not possible to mention everyone who has contributed to the process which has culminated in the production of this dissertation; nevertheless, please accept my humble thanks for your advice and friendship.

My thanks particularly go to my advisor and chairman, Steve Pizer; he has persevered in this endeavor even when my own enthusiasm was flagging. His technical expertise and sound scientific judgement have contributed mightily to this dissertation. Gene Johnston has been of tremendous help in overcoming the technical obstacles involved in the observer experiments and has been a frequent and helpful listener on the occasions when a friendly ear was what I needed most. Jan Koenderink has given valuable advice and encouragement in the production of the quality measure and on the general subject of visual perception. Thanks to my committee for their patience and advice during the production of the written document.

I would like to make special note of three teachers who have played a large role in my development as a scholar. Edward Deeds showed me what science is all about; Wayne Christiansen was able to communicate to me his joy in the scientific process; and Richard Marius encouraged my love for the humanities.

It is not possible to complete an undertaking such as this without the support of family and friends; four friends in particular have meant much to me: Thomas Gary Bishop, Michael Branton, Philip Jewell, and Richard White. My gratitude to all of you both for your technical advice and moral encouragement. I would also like to thank John and Kathy Austin, Vicki Baker, and Ann Brice for their friendship and support during the last stages of the dissertation; of course, thanks are also due to the Basement Crew at the University of Illinois. Susan Kirstein has been a constant source of wisdom and encouragement; she believed I could do it even when the evidence would have suggested otherwise. Thanks, Susan.

Finally, I would like to express my love and gratitude to my parents; they have been supportive and caring throughout the long course of my studies and I know that their pride in my accomplishments equals my own. It is to them that I dedicate this dissertation.

This dissertation was prepared by the author using the \TeX text formatting system and the \BIBLIOTeX reference formatting program. \TeX was written by Donald Knuth and collaborators at Stanford University; \BIBLIOTeX was written at the University of Arizona for use with *troff* and modified by Gary Bishop at UNC to work with \TeX .

JOHN BURTON ZIMMERMAN. The Effectiveness of Adaptive Contrast Enhancement
(Under the direction of STEPHEN M. PIZER.)

Abstract

A significant problem in the display of digital images has been the proper choice of a display mapping which will allow the observer to utilize well the information in the image. Recently, contrast enhancement mappings which adapt to local information in the image have been developed. In this work, the effectiveness of an adaptive contrast enhancement method, Pizer's Adaptive Histogram Equalization (AHE), was compared to global contrast enhancement for a specific task by the use of formal observer studies with medical images. Observers were allowed to choose the parameters for linear intensity windowing of the data to compare to images automatically prepared using Adaptive Histogram Equalization. The conclusions reached in this work were:

- There was no significant difference in diagnostic performance using AHE and interactive windowing.
- Windowing performed relatively better in the mediastinum than in the lungs.
- There was no significant improvement over time in the observers' performance using AHE.
- The performances of the observers were roughly equivalent.

An image quality measure was also developed, based upon models of processing within the visual system that are edge sensitive, to allow the evaluation of contrast enhancement mappings without the use of observer studies. Preliminary tests with this image quality measure showed that it was able to detect appropriate features for contrast changes in obvious targets, but a complete evaluation using subtle contrast changes found that the measure is insufficiently sensitive to allow the comparison of different contrast enhancement modalities. It was concluded that limiting factors affecting the measure's sensitivity included the intrinsic variation in normal image structure and spatial and intensity quantization artifacts. However, humans were able to reliably detect the targets used in the experiment.

This work suggests that 1) the use of adaptive contrast enhancement can be effective for the display of medical images and 2) modeling the eye's detection of contrast as an edge-sensitive process will require further evaluation to determine if such models are useful.

Table Of Contents

Acknowledgements	iii
Table of Contents	vi
1 Introduction: Contrast Enhancement and Image Quality	1
1.1 The First Problem: Display of Nonvisual Images	2
1.2 The Second Problem: Evaluating the Solutions	3
1.3 Overview of the Current Work	4
1.4 Implications of this Work	6
2 Previous Work in Adaptive Contrast Enhancement	7
2.1 Perception and Contrast	7
2.2 Contrast Enhancement	9
2.2.1 Linear Min-max Windowing	9
2.2.2 Histogram Equalization	11
2.3 Adaptive Contrast Enhancement	13
2.3.1 Characterization of ACE Methods	14
2.3.2 Desirable Properties of an ACE Mapping	16
2.4 Examples of ACE Mappings	16
2.4.1 Constant Variance Enhancement	17
2.4.2 Peli-Lim Adaptive Filtering	19
2.4.3 Implementation	20
2.4.4 Results of Application to Medical Images	21
2.5 Local Range Modification	24
2.5.1 Implementation	24
2.5.2 Results of Application to Medical Images	27
2.6 Other Methods of Adaptive Contrast Enhancement	28
3 Properties of Adaptive Histogram Equalization	30
3.1 Theoretical Foundations of AHE	30
3.1.1 Motivation: Information Transfer	30
3.1.2 Previous Work on AHE	34
3.1.3 Pizer's Adaptive Histogram Equalization	36
3.2 Empirical Results to Date	40
3.2.1 Basic Properties of the Method	40
3.2.2 Effect of Interpolation	44
3.2.3 Effects on the Image Histogram	45
3.2.4 Artifact Generation	50
3.3 Possible Modifications of AHE	52
3.3.1 Improved Interpolation Schemes	53
3.3.2 Varying Contextual Regions	54
3.3.3 AHE Combined with Histogram Equalization	54
3.3.4 Summary and Conclusions	55

4	A New Image Quality Measure	57
4.1	What is Quality?	57
4.2	Previous Approaches to Image Quality	58
4.2.1	Physical Models	58
4.2.2	Signal Detection Theory	59
4.2.3	Visual Psychophysics	60
4.3	Overview of the Image Quality Measure	60
4.3.1	General Approach	61
4.4	Definition of the IQM	62
4.4.1	Model of Visual Contrast Detection	64
4.4.2	Self-Similar Receptive Field Arrays	67
4.4.3	Matched Filtering of Response Function	68
4.4.4	Definition of $P(D_i T_i, I_n)$	70
4.4.5	Calculation of the IQM	72
4.4.6	Implementation	72
4.4.7	Algorithm for Calculating the IQM	73
4.4.8	Extensions and Testing of the IQM	74
5	Experimental Results	75
5.1	Purpose and Goals	75
5.1.1	Comparison of Contrast Enhancement Modalities	75
5.1.2	Evaluation of the IQM	76
5.2	Experimental Methodology: Contrast Enhancement Comparison	76
5.2.1	Design Goals	76
5.2.2	Design Decisions	77
5.2.3	Selection of Case Sample	79
5.2.4	Preparation of the Trial Images	84
5.2.5	Selection of Observers	87
5.2.6	Experimental Equipment and Layout	87
5.2.7	Observer Procedure	88
5.3	Results: Comparison of Contrast Enhancement Modalities	91
5.4	Experimental Methodology: IQM Validation	103
5.5	Results: IQM Validation	103
5.5.1	Possible Factors Influencing the IQM Calculation	103
5.5.2	Conclusions	104
6	Summary and Directions for Further Work	106
6.1	Summary of the Current Work	107
6.2	Future Directions	109
	References	112
1	AHE-Windowing Observer Study (ZDOS)	116
1.1	The Experiment in Brief	116
1.2	Description of the Study	117
1.2.1	What the study is designed to show	117
1.2.2	Theoretical basis of the study	118
1.2.3	Limitations	118
1.2.4	Frustration	118

1.2.5	Terminology	119
1.2.6	What you'll see	119
1.3	Experimental Procedure	123
1.3.1	Physical Environment	123
1.3.2	Windowing	123
1.3.3	Choice of Rating Values	124
1.3.4	Conduct of the Experiment	125
1.4	The Observer's Cookbook	125
1.4.1	Getting Started	125
1.4.2	Things You Shouldn't Need to Know	127
1.4.3	Checklist	130
1.4.4	Summary of commands	131

Table of Figures

2.1 Perception of a Digital Image	8
2.2 Contrast of Object on Background B_0	9
2.3 Linear Min-max Windowing	10
2.4 Images Processed by Windowing	12
2.5 Histogram Equalization	13
2.6 Constant Variance Enhancement	19
2.7 Block Diagram of Peli-Lim Adaptive Filtering	21
2.8 Functions Used for Peli-Lim Images	22
2.9 Image Processed with Peli-Lim Filter	23
2.10 Image Division for Local Range Modification	25
2.11 Interpolation for Local Range Modification	27
2.12 Image Processed with Local Range Modification	29
3.1 Transfer of Information in Digital Images	32
3.2 Ketcham's Local Area Histogram Equalization	35
3.3 Sample Points for AHE	38
3.4 Interpolation in AHE	39
3.5 AHE Chest Image	41
3.6 AHE Spine Image	42
3.7 AHE Compared with LAHE	45
3.8 Image Histograms with AHE	47
3.9 Histograms of Medical Images with AHE	48
3.10 Regional Histograms of AHE Images	49
3.11 Artifact Generation with AHE	51
3.12 Digital Chest Radiograph with AHE	52
3.13 Frequency Domain Artifacts with AHE	53
3.14 AHE Followed by Histogram Equalization	55
4.1 Overview of the Self-Similar Sensor Array	65
4.2 Koenderink-van Doorn Sensor Element	65
4.3 Density Function of $F_{\xi}(n)$	69
4.4 Comparison of Density Functions	71
5.1 Minimum Lesion Intensity vs. Average Laplacian	84
5.2 Minimum Lesion Intensity vs. Average Laplacian	84

Chapter 1

Introduction: Contrast Enhancement and Image Quality

*And what is good, Phaedrus?
And what is not good—
Need we ask anyone to tell us these things?
— attributed to Socrates [470-399]*

The development within the last three decades of electronic imaging and digital image processing has provided us with the ability to visualize the world in ways heretofore impossible. It is now common to produce visible images which depict scenes invisible to human senses. Along with these new capabilities have come a number of problems. One of these is the question of how to display images so that the information they contain can best be utilized by the observer of the image. Electronic imaging can capture the properties of scenes which may be remote from the observer both in space and time. Digital image processing then allows the information in the spatial and intensity variations of the image to be transformed to improve the quality of the image so that it may be more easily interpreted by a human being. As part of this processing, an algorithm (*display mapping*) is applied to the recorded image to choose which display levels of a particular display device are to correspond to the recorded values. Choosing an effective display mapping can be very difficult in the case of images which arise from intrinsically nonvisual sources.

Considerable progress has been made in developing good display algorithms for use with the various display devices now available. However, evaluating the effectiveness of these algorithms has proven to be a hard problem as well. Intuitively, the fundamental idea of all such methods is to produce the best possible image. Unfortunately, there is no generally agreed upon standard by which the merit of an image may be evaluated. An image which is satisfactory in the context of some particular task the observer wishes to perform may be inadequate for a different task. Clearly, goodness depends at least in part upon the task of the observer.

The two problems outlined above motivate the research in this dissertation: the need for good display mappings for nonvisual images and the evaluation of the effectiveness of these algorithms. Adaptive contrast enhancement (ACE) has shown potential for displaying nonvisual (and visual) images effectively and automatically. In the current work, two projects have been undertaken. First, an evaluation by formal observer studies of an example from the class of adaptive contrast enhancement mappings, Adaptive Histogram Equalization, has been performed to determine its efficacy as a display mapping.

Second, an objective measure of image quality has been developed to measure the goodness of a particular image relative to a specific task, that of the detection of simulated lesions in medical images. Such a measure allows the evaluation of display techniques without the use of formal observer studies. The developed image quality measure (IQM) has been compared to the performance of real observers. The last section of this chapter considers the implications of this work. It is assumed throughout that the reader has a basic knowledge of contemporary digital image processing. A recommended introduction to image processing is that of Castleman, 1979.

1.1 The First Problem: Display of Nonvisual Images

Often in digital imaging, an image will represent a two-dimensional distribution of some physical parameter such as intensity of a radiation field, temperature, radioactive decay, or X-ray attenuation. Images such as these that do not represent directly perceivable visual scenes occur in medicine, radio astronomy, remote sensing, and other areas of imaging. As the capabilities of imaging devices have improved, the quality of such images has improved dramatically. The ratio of discernible signal to noise has become quite large; there is information present in even small subranges of the intensity data in an image. The range of information collected by an imaging device may far outstrip the capabilities of the display devices which convey this information to the observer.

The display of nonvisual images for visual interpretation involves the choice of some display device such as photographic film or a video screen. A mapping must then be selected relating the intensity values in the recorded image to the available display levels of the device. If this intensity mapping has been well chosen, the information which the observer wishes to see in the image will be perceivable on the device as intensity contrast between the various parts of the displayed image. It will be assumed here that such images are monochromatic and are characterized by a single variable, intensity.

The selection of an appropriate mapping is often not obvious. The most straightforward way of choosing this display mapping is to linearly scale the intensity values in the image into the display range of the device. However, the range of values in the image may be large while the information desired is contained in one or more small intervals of that range. For example, in medical imaging a transverse computed tomography (CT) image of the thorax may contain information about both the heart and lungs. However, the heart information may be in some range of intensity values, the lung information in another, while the entire range of data contains information about the complete image. A linear scaling of the complete data range into the display scale of the device would compress each of the subranges of the data into only a few display levels. The difficulty is in choosing a display mapping which will display all the information in the picture as well as possible.

The typical solution to this problem has been to employ some form of contrast enhancement mapping to emphasize certain parts of the data range. Techniques such as the linear scaling of some intensity subrange (*window*) of the data into the full range of the display device (*linear min-max windowing* or *linear windowing*) and histogram equalization have been widely used. The problem with most such approaches is that they apply a certain mapping function to all of the pixels independent of the local content of the image. This ignores the fact that most of the information of interest in the image is contained in the local distribution of recorded intensities. Recently, considerable progress has been made with methods of contrast enhancement that adapt to the local image information. Chapter 2 discusses some of the various forms of adaptive contrast enhancement which have been proposed. These techniques have shown great promise, but it is not known how these methods compare in effectiveness to the better forms of global contrast enhancement such as linear min-max windowing with the window interactively chosen.

1.2 The Second Problem: Evaluating the Solutions

In choosing an adequate display mapping for images that contain a large range of significant data, many approaches are possible. However, comparing these various methods is hard. One wishes to use the method which will produce the best image, but there is no general agreement on what is meant by *best*. A mapping for a medical image which is best for displaying gross anatomy may be woefully inadequate for detecting subtle pathology. The pragmatic solution is to proclaim that image as best which is most utilitarian: a good image shows what the observer needs to see to do a specified task. A good image will allow the observer to perform the task better than a bad image.

This approach to defining the quality of an image implies that the ultimate arbiter of goodness is a human observer performing a specific task. The evaluation of the effectiveness of an image enhancement method then must be done by using controlled observer studies, yielding a measure of the effectiveness of the observer and hence of the enhancement method. A considerable medical and psychological literature exists on the performance of such observer experiments [Swets and Pickett, 1982; Metz, 1978].

Unfortunately, the performance of formal observer studies for a full evaluation of a particular display mapping technique is both tedious and difficult. An alternative approach is to attempt to define some objective measure of image quality related to the properties of an image and the specific task to be performed using that image. If a quality measure can be found whose value correlates well with the performance of human observers for the same task, it allows the evaluation of display mappings and other imaging and image processing methods to be performed more quickly than by the use of formal observer studies. It also allows the choice of parameters which will optimize a particular mapping relative to the quality measure for a given image. Many measures of image quality have been defined; however, most of these are *ad hoc* and little has been done to compare them experimentally to the performance of human observers.

1.3 Overview of the Current Work

The problems discussed in the two preceding sections are aspects of the larger question of optimal image display. Though one would like the ability to decide which algorithm applied to the image will produce the best possible result in all cases, the current research has restricted itself to two approaches. First, within the problem of developing good mappings for the display of nonvisual images, the use of adaptive contrast enhancement has been evaluated for its effectiveness in displaying images. Second, an image quality measure has been defined for evaluating the contrast enhancement mappings. These two thrusts have been implemented as four projects:

Empirical Investigation of ACE. Several methods of adaptive contrast enhancement have been examined empirically to determine their ability to perform as effective display mappings for the particular problem of the display of medical images. One method, Pizer's Adaptive Histogram Equalization (AHE) [Pizer, 1981a; Pizer, 1981c; Pizer *et al.*, 1984] has been examined in detail for its properties with respect to artifact generation, alteration of its parameters, and performance on medical images of varying modalities.

Observer Evaluation of AHE. The promising technique of AHE was compared to linear min-max windowing using formal observer studies for a specific task, detection of simulated lesions in real clinical images. Gaussian targets of varying intensity and size were introduced into the clinical images, which were then processed with AHE. Experienced radiological observers were then asked to perform a detection task on the processed images. The same task was performed on the unprocessed images with the observers allowed to interactively choose multiple linear min-max windows. The results of these experiments were compared using the techniques of signal detection theory to determine the relative effectiveness of the two methods.

Definition of an Image Quality Measure. An image quality measure (IQM) was developed to allow the *a priori* comparison of contrast enhancement techniques. Since the quality of an image is directly related to the ability of an observer to perform a specified task, the proposed image quality measure is the probability that the observer will perform the task correctly.

To measure this probability quantitatively, a formulation must be developed which allows the calculation of the relevant probabilities. Let T_i be the true situation in some image (for example, that a lesion is or is not present) and D_i be the decision that the observer makes about the image (the decision that the lesion is or is not present). Then the probability that the observer makes a correct decision is given by $P(D_i, T_i)$, the joint probability that given true situation T_i the observer makes the correct corresponding decision D_i . The image quality measure is defined as the total probability that the observer makes a correct decision,

$$IQM = \sum_i P(D_i, T_i). \quad [1.1]$$

In order to calculate the IQM, the assumption is made that the joint probability can be defined in a way which depends principally on the the eye's response to the presence of edges in the image. In this work, a model of part of the human visual system is used to calculate this response. It incorporates the model for visual receptors given in Koenderink and van Doorn, 1978. A fuller discussion of the measure and its calculation in the case of real images containing a known object is given in Chapter 4.

Evaluation of the IQM. The proposed image quality measure was used to predict the performance of the two contrast enhancement methods of the second project on the images used in that study. These predictions were then compared with the results of the observer studies.

1.4 Implications of this Work

The implications of this work are two. First, Adaptive Histogram Equalization was shown to be competitive with interactive linear windowing for a specific detection task. Although it is not possible to generalize this to all detection tasks, this result nonetheless implies that AHE may be used effectively in a clinical setting. If diagnostically significant information is not lost, then AHE has advantages over current techniques in that it allows diagnostic determinations requiring simultaneous viewing of different organs, produces enhanced images without manual intervention, and provides reproducible results. The reported result encourages day-to-day trials in a true clinical situation. That is the next step in this work.

Second, it was shown that while the IQM seems to be sensitive to the contrast features of importance in the image, it is not sufficiently sensitive to subtle contrast changes to allow the reliable comparison of contrast enhancement methods. A number of factors have been identified which may cause this lack of sensitivity; in particular, the normal variation of structure in medical images causes a large variation in the output of the quality measure. Since the targets used in the current work are very small both in spatial extent and intensity variation, the presence or absence of the target is only a small perturbation on the normal structure. The use of measures which are sensitive to the same properties of contrast as the human visual system is a promising approach to defining an objective image quality measure; the next step in this work should be an attempt to account for the structural variation which is present in the images used for the calculation of the IQM.

Chapter 2

Previous Work in Adaptive Contrast Enhancement

Study the past if you would divine the future.

— K'ung Fu-Tse [c.551-479?]

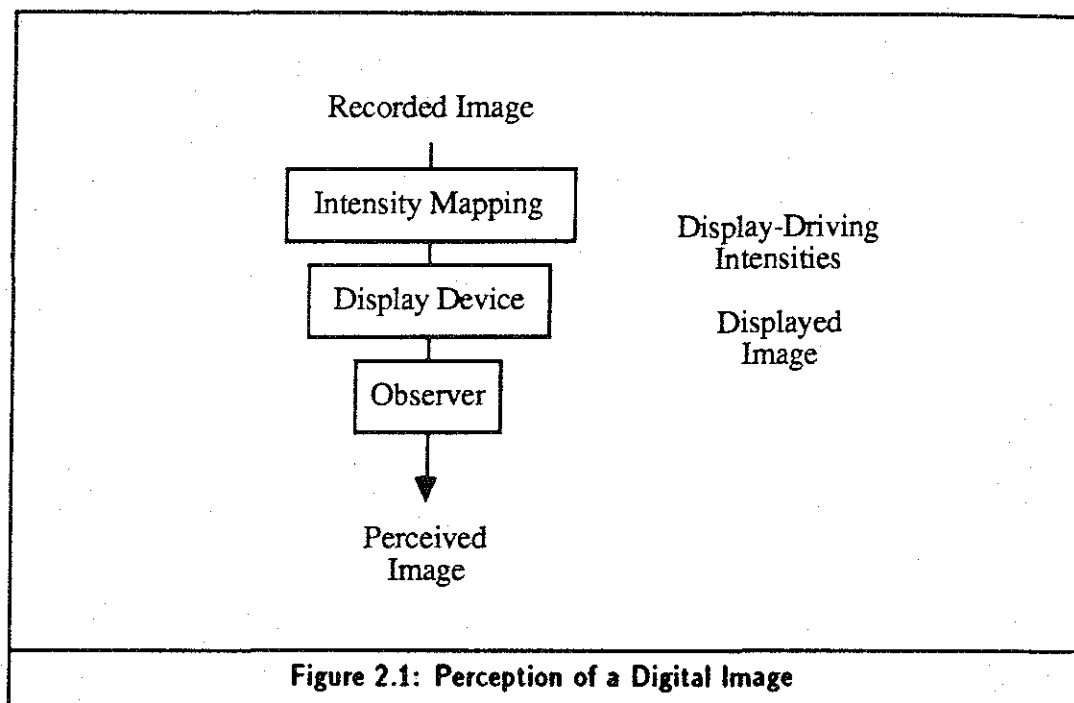
2.1 Perception and Contrast

Assume that we have a recorded digital image, which may have been acquired in a way that has introduced distortion and noise. The image is represented as a two or three dimensional array of numbers, which must now be displayed so a human can observe it.

A simplified representation of the process of image display and perception is shown in Figure 2.1. The recorded image undergoes an intensity mapping (which may be the identity mapping) and is then used as input to a display device. The intensities received by the display device will be referred to as the *display-driving intensities*. The display device produces a visible image which is seen by some observer. The image then enters the observer's perceptual system.

The characteristics of the display device and the human visual system introduce distortions which cause the perceived image to differ from the display driving image. For common display devices such as CRT screens, the display may not have the capability to respond to changes in the display-driving intensities equally well over all of its range; for example, it is well known that most CRTs do not respond well to changes in the low intensity range. The visual system then processes the light entering the eye in complex ways.

If the intensities of the display driving image are to be seen with fidelity, the display device/observer combination should be linear, that is, it should respond equally to a fractional change in recorded intensity no matter in what part of the intensity range the change lies. The introduction of nonlinear distortions by the display device and observer can be approximately corrected by applying a linearization mapping to the display driving



intensities before they enter the display device. This process of *linearization* is discussed by Pizer, 1981b. Thus, this source of distortion in the intensity values of the image can be approximately removed to within the dependence of contrast perception on image structure. A display device whose display-driving intensities have undergone this linearization procedure will be called a *linear display device*, though it is in fact the combination of display device and observer that have been made linear. In the discussion hereafter, it is assumed that we have a linear display device.

In looking at the image, the observer is sensitive to both its spatial and intensity properties. The spatial distribution of intensities forms objects in the image which the observer detects because of *contrast* between the intensities of the object and ground. Contrast may be defined as this difference in intensity. A workable definition is that given by Hall, 1979; assume there is an object of brightness B superimposed on a background of brightness B_0 . Then the contrast, C , is given by Weber's fraction,

$$C = (B - B_0)/B = \Delta B/B. \quad [2.1]$$

Figure 2.2 shows an object of height ΔB on a background B_0 . This definition allows for both positive and negative contrast, negative contrast implying that the object is dimmer than its background. In real images, it is not simple to define the contrast of an object,

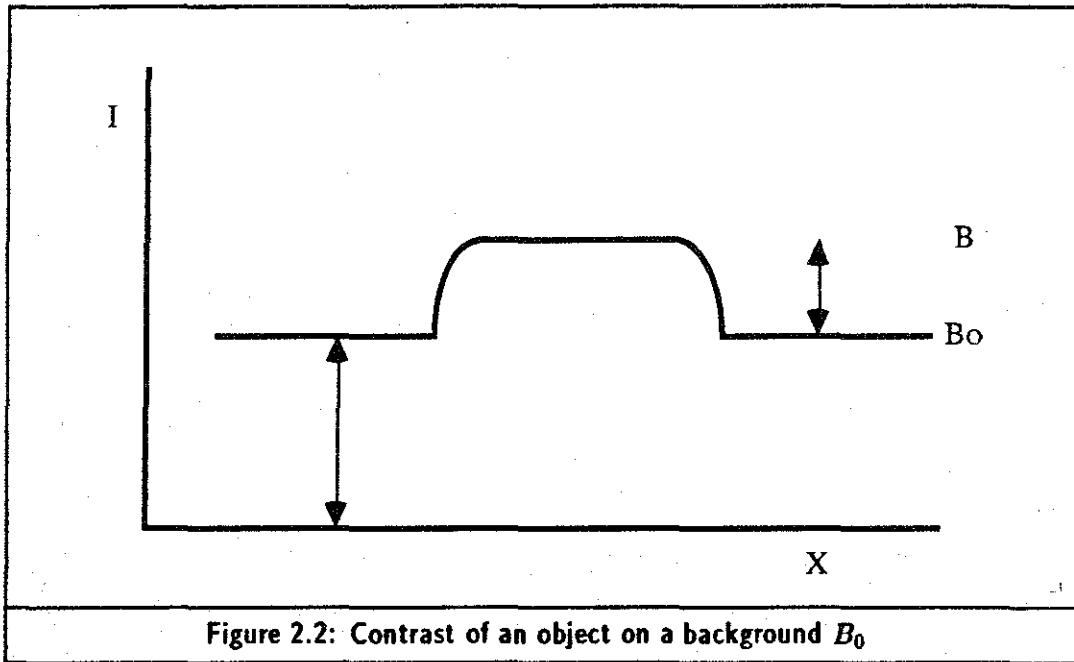


Figure 2.2: Contrast of an object on a background B_0

since there is often noise (both structured and unstructured) which obscures the outline of the object.

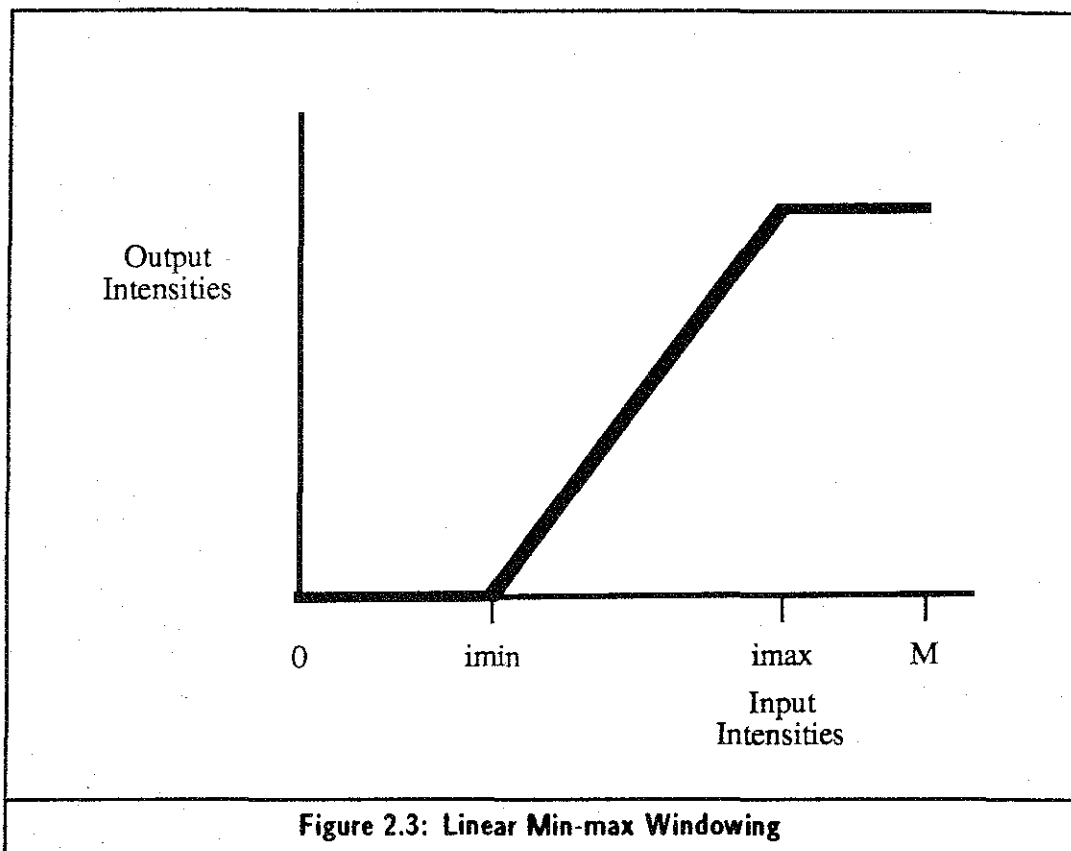
This definition of contrast is a purely physical one, in that it refers only to light intensities. It does not discuss the *perception* of contrast, where there are complex psychophysical effects which must be taken into account; for example, the visual system reacts preferentially to the presence of edges in detecting contrast.

2.2 Contrast Enhancement

Often the range of recorded intensities in an image is sufficiently large that the contrast of a particular object is reduced to only a few of the display levels of the display device. Contrast enhancement is the process of increasing the contrast of an object so that it occupies a larger fraction of the total display range. To increase the contrast, the difference in intensity between the object and ground should increase relative to the ground. We consider as examples two simple ways of doing this: linear min-max windowing and histogram equalization.

2.2.1 Linear Min-max Windowing

In linear min-max windowing, the contrast of a particular object is increased at the expense of the contrast of other objects in the image. A subrange of the recorded intensities (a *window*) is chosen and that subrange is mapped linearly into the total range of display



driving intensities. Figure 2.3 shows this process. More exactly, suppose we choose the intensities lying between (i_{\min}, i_{\max}) as the window which is to be linearly remapped (as in Figure 2.3). Let the desired range of intensities in the image after processing lie in the range $(0, M)$. The transformation which will take (i_{\min}, i_{\max}) into $(0, M)$ is

$$I'(x, y) = \begin{cases} 0, & I(x, y) \leq i_{\min}; \\ \frac{(M-0)}{(i_{\max} - i_{\min})}(I(x, y) - i_{\min}), & i_{\min} < I(x, y) < i_{\max}; \\ M, & I(x, y) \geq i_{\max}, \end{cases} \quad [2.2]$$

where $I(x, y)$ is the intensity of a pixel at location (x, y) . Note that intensities greater than i_{\max} or less than i_{\min} are mapped to the maximum or minimum intensities. This implies that any information in the image intensities outside the chosen window is lost.

As an example of a situation where windowing increases the contrast, consider the following: suppose that $i_{\min} = B_0$, $i_{\max} = B$, where B_0 and B are as in Figure 2.2. Then the contrast before windowing is

$$C = \frac{B - B_0}{B} = \frac{\Delta B}{B_0 + \Delta B},$$

which is less than 1 if $B_0 > 0$. After windowing, $\Delta B = M - 0$ and $B_0 = 0$. The contrast then is

$$C = \frac{M - 0}{0 + (M - 0)} = 1.$$

Thus for this choice of window, the contrast of the image is increased.

This technique is commonly used in medical imaging to increase the contrast in regions of interest to the physician. Typically, several windows are chosen interactively by the physician or technician and the contrast enhancement applied for each window. The disadvantages of this technique are that it requires manual interaction with the image and it does not allow the simultaneous appreciation of different organs in the image. Furthermore, due to the manual interaction either the results are not easily reproducible (another physician might choose a different window for displaying the data), or only a limited selection of standard windows are used. However, linear windowing is easy to apply to the image and is capable of great sensitivity when used by an experienced observer. An example of linear windowing is shown in Figure 2.4. The original image is shown, as well as two windows chosen to render the lungs and mediastinum areas of the chest.

2.2.2 Histogram Equalization

Histogram equalization is an enhancement technique which attempts to use the available display levels as well as possible by distributing the pixels evenly among them. An excellent discussion of histogram equalization is given by Castleman. The cumulative distribution function (CDF) of the image intensities is calculated and used as the enhancement mapping [Figure 2.5]:

$$I'(x, y) = C \times \text{CDF}(I(x, y)). \quad [2.3]$$

The constant C scales the output image into the desired range. The result is an image whose histogram is flat except for effects due to the discrete nature of the recorded intensity values. Histogram equalization uses the statistics of the intensities in the image to effect the enhancement; the result is that the intensity values where there are the most pixels are allocated the most levels for display. Assuming that we have a linear display device and the noise in the image is stationary, histogram equalization has the effect of maximizing

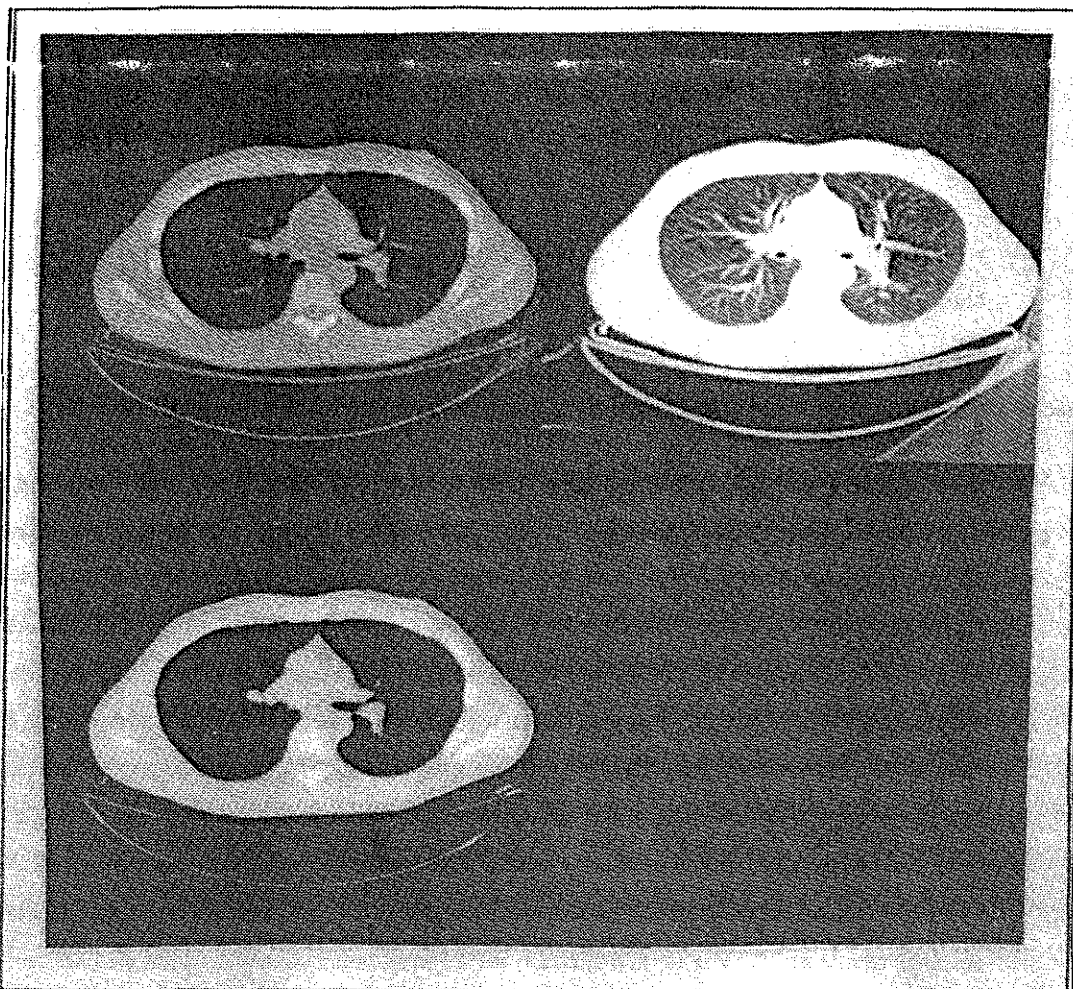
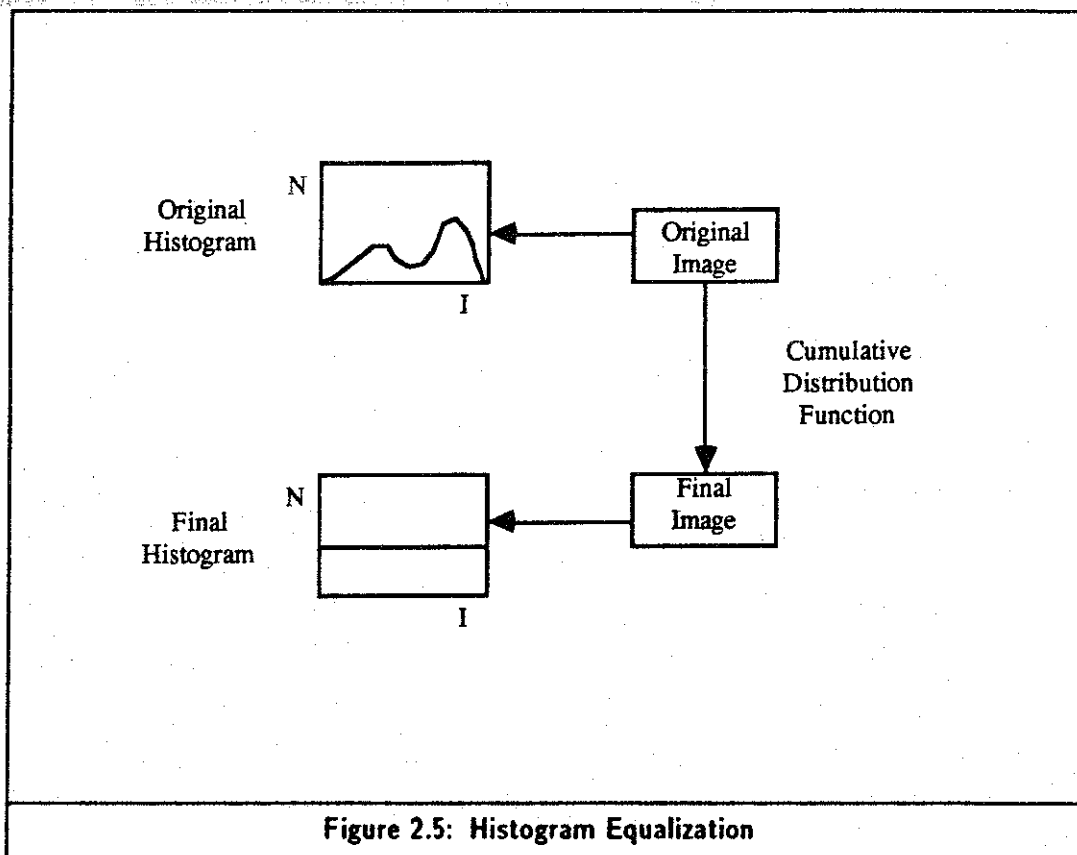


Figure 2.4: Images processed by windowing. The image is a chest CT scan, 512×512 pixels.

Upper left: original image

Upper right: window to show lungs

Lower left: window to show mediastinum



the per pixel transfer of Shannon information, an objective proposed by Cormack and Hutton, 1980. A proof of this assertion is given in Chapter 3. An extended discussion of histogram modification according to several criteria is given in Hummel, 1975 and Hummel, 1977.

2.3 Adaptive Contrast Enhancement

Both of the examples above are instances of *stationary* intensity mappings. In a stationary mapping, the value of the mapping function depends only on the intensity value, I , of the input pixel:

$$I'(x, y) = f(I(x, y)), \quad [2.4]$$

where I' is the new value of the intensity of a pixel at location (x, y) in the image. A stationary mapping is not dependent on the location of the pixel or local properties of the image, such as the intensity distribution in a neighborhood of the pixel. However, the global properties of the image, *e.g.*, the image histogram, are often used to define the

mapping. A *nonstationary* mapping has the property that the mapping varies depending on the location of the pixel in the image:

$$I'(x, y) = f(I(x, y), x, y). \quad [2.5]$$

The choice of f may still depend on the global image properties.

Adaptive Mappings. One class of nonstationary maps are the *adaptive maps*, which use the local intensity properties of the image to define the mapping. For each pixel at location (x, y) in the image, a mapping is chosen which depends on the intensity of the pixel $I(x, y)$ and on the intensity distribution, D_Ω , in some neighborhood Ω about the pixel:

$$I'(x, y) = f(I(x, y), D_\Omega(x, y)). \quad [2.6]$$

The neighborhood Ω is referred to as the *contextual region* of the pixel at (x, y) . It is the set of pixels which are within some distance from the pixel according to a given distance metric. The advantage of such a mapping is that the function f can be chosen to enhance contrast by making the resulting intensity value at each pixel dependent upon the intensity values of its contextual region Ω . However, because the adaptive mapping uses local intensity variations, it is possible that the global relationships in the image will be changed, *e.g.*, two pixels of the same initial intensity $I(x, y)$ may be mapped into different values $I'(x, y)$ if their contextual regions are different.

The function f is frequently defined such that it would achieve some goal if it were applied to every pixel in the contextual region. This goal will be called its *local contrast enhancement goal* (LCEG). An example of such an LCEG is to require that the variance of the pixel values be fixed in the contextual region; other examples of LCEGs will be seen in the next section. A number of ACE techniques can be categorized in this way.

2.3.1 Characterisation of ACE Methods

Most of the proposed ACE methods use local statistics such as the mean, standard deviation, and histogram of the pixel intensities in Ω to calculate the new intensity value. Frequently, the method will manipulate these statistics to achieve some LCEG in the output image.

Many ACE mappings can be seen to be a form of high-pass filtering, removing the low spatial frequency components of the image so that the contrast of the remaining high frequency component can be enhanced. This is similar to the processing of the visual

system, which is insensitive to slowly changing illumination. The low-frequency removal can be approximated by subtracting the local mean from the image; the resulting filter can be cast into a standard form [Fahnestock and Schowengerdt, 1983] that describes many ACE mappings. If \bar{I}_{xy} is the mean value of I_{xy} in some contextual region Ω ,

$$I'_{xy} = f(I_{xy}) \quad [2.7]$$

$$= A [I_{xy} - (1 - B)\bar{I}_{xy}] + C$$

$$I'_{xy} = A [(1 - B)(I_{xy} - \bar{I}_{xy}) + BI_{xy}] + C. \quad [2.8]$$

Here A and C are scaling factors which adjust the final result into the desired range. It is clear that this filter is a weighted sum of a high frequency component ($I_{xy} - \bar{I}_{xy}$) and the original image I_{xy} with the constant B controlling the amount of high frequency present. Since the ACE mapping depends on the intensity values within Ω , pixels outside Ω will have no effect on $I'(x, y)$. Thus, spatial frequency components of the image whose wavelengths are larger than the size of the contextual region will be heavily attenuated.

A more general formulation allows a function of the local statistics of the image, $T_1(W)$, to modify the high frequency component of the output image. Here W is some set of statistics of the intensity distribution, D_Ω . Then

$$I'_{xy} = A [(1 - B)T_1(W)(I_{xy} - \bar{I}_{xy}) + BI_{xy}] + C. \quad [2.9]$$

This filter may be modified yet again by allowing the high and low frequency gains to be controlled independently. We replace the term BI_{xy} with $T_2(I_{xy})$, a function of the local mean, so that

$$I'_{xy} = A [(1 - B)T_1(W)(I_{xy} - \bar{I}_{xy}) + T_2(\bar{I}_{xy})] + C. \quad [2.10]$$

The functions T_1 and T_2 now define the particular adaptive contrast enhancement method and are usually chosen to achieve the LCEG of the method. Note that while T_1 is a general function of the local statistics, T_2 depends on only one of those statistics, the local mean. While other generalizations of this filter are immediately obvious, the current form adequately describes many ACE methods.

2.3.2 Desirable Properties of an ACE Mapping

Before considering specific methods, the desirable properties of an ACE mapping should be examined. Certainly, such methods should enhance the contrast in the image, but they must also be usable in a practical setting. Some desirable characteristics are listed below.

Artifact Generation. Many methods which enhance the contrast in an image also introduce artifacts into the image. High frequency enhancements are particularly susceptible to the introduction of ringing and overshoot. A good ACE method, when used in a reasonable fashion, should introduce a minimum of artifacts.

Fast Implementation. A method may work superbly, but if it requires an excessively large or slow implementation, it may not be usable in practical settings. It must be kept in mind, however, that a method which is intractable in software on a Von Neumann architecture machine may be quite practical if implemented in hardware or on a special-purpose machine.

Stability of Objects. Because of its dependence on the contextual region, an ACE mapping may change the relative intensities of objects in different parts of the image. This can cause similar objects to appear differently or cause a single object to break up into multiple smaller objects, slowing the recognition and interpretation of objects in the image.

Reasonable Parameter Space. If the method has an excessive number of parameters, it may be very difficult to search the parameter space for effective values for different types of images. The size of the parameter space increases exponentially with the number of parameters. The search for optimal parameter values may easily be trapped at a local maximum of the parameter space.

2.4 Examples of ACE Mappings

In this section, we consider some specific ACE mappings. All of these mappings use local image statistics to define the contrast enhancement; in some of them, a local contrast enhancement goal can be formulated. Three methods are considered in detail; as part of this work, they have been implemented and used to process medical images. No attempt is made here to survey completely the ACE mappings which have been proposed in the literature; rather, these techniques have been chosen as examples both because they are representative of a larger body of methods and because they all hold promise as practical methods of Adaptive Contrast Enhancement.

2.4.1 Constant Variance Enhancement

Constant variance enhancement (CVE) is a high frequency enhancement which removes some or all of the low frequency and adjusts the variance within the contextual region to a desired value. The original development was by Harris, 1977. Harris requires that 1) the local mean be removed entirely and 2) the variance in the contextual region be constant for every neighborhood in the output image (hence the name of the method). These two requirements constitute the LCEG of this method. In terms of [2.10], $T_2(I_{xy}) = 0$ and $T_1(W) = 1/S_{xy}$, where S_{xy} is the standard deviation of the pixel intensity values within Ω . The enhancement mapping at (x, y) then is

$$I'_{xy} = A \left[\left(\frac{1}{S_{xy}} \right) (I_{xy} - I_{xy}) \right] + C. \quad [2.11]$$

This simple function gives an image with approximately constant variance. Unfortunately, it suffers in that the contrast gain can grow without limit if $S_{xy} \approx 0$. A more satisfactory solution imposes a limit on the gain caused by the reciprocal of the standard deviation and restores a portion of the low frequency information. Such a filter was proposed independently by Wallis, 1976 and Lee, 1980. Their generalization of the algorithm is

$$I'_{xy} = A \left[\left(\frac{S'_{xy}}{S_{xy} + (S'_{xy}/\text{MAX})} \right) (I_{xy} - I_{xy}) + BI_{xy} \right] + C. \quad [2.12]$$

The term BI_{xy} , for $0 \leq B \leq 1$, restores a fraction of the low frequency component of the image. A and C are as before. Now S'_{xy} represents the desired standard deviation of the output image and the constant MAX imposes a limit on the gain in regions where the standard deviation of the input is small. In this case, $T_1 = ((S'_{xy}/(S_{xy} + (S'_{xy}/\text{MAX})))$ and $T_2 = (1 - B)I_{xy}$. Notice that the exact size and shape of the contextual region are not specified; these choices are parameters of the method which are not explicitly stated.

For CVE, the parameter space is large. While A and C are determined by the requirement that the resulting image must be scaled into the range of the display device, the parameters S'_{xy} , MAX, and B must still be determined.

A variant of CVE called *Local Area Brightness and Gain Control* (LABGC) was proposed by Ketcham *et al.*, 1976. In this mapping, the local variance and mean of the neighborhood of a pixel are again used to implement an adaptive contrast enhancement scheme. A gain factor is used to adjust the local variance. The principal difference between this method and CVE is that the mean is used to ensure that the overall range of the ensuing image fills the display range of the device. This is convenient for the hardware implementation scheme developed by Ketcham and his collaborators.

Implementation. The straightforward implementation of CVE is to calculate at every pixel the mean and standard deviation of its contextual region and then apply the mapping. Unfortunately, this implies a considerable computational burden; if the contextual region is square and k pixels on a side, then the order of this calculation for an $N \times N$ image is $O(k^2 N^2)$. Wallis suggests simplifying the calculation by dividing the image into non-overlapping blocks; the mean and standard deviation are then calculated in each block. These values are assumed to hold at the center of the block and values of μ and σ^2 for other pixels are calculated by two dimensional interpolation, reducing the computational time significantly. A similar technique is used for Local Range Modification and Adaptive Histogram Equalization and will be discussed in more detail in those sections.

A somewhat different implementation was suggested by Narendra and Fitch, 1981 using a recursive filter implementation. This has the advantages that it can be implemented in hardware and allows speeds that would be beyond the capabilities of nonrecursive implementations on general purpose machines.

Results of Application to Medical Images. The constant variance enhancement given in [2.12] was implemented on the VAX11/780 in the Graphics and Image Processing Laboratory at UNC and applied to a variety of medical images. A pixel-by-pixel mapping was used with no attempt to optimize the speed of the implementation. A photograph of a chest CT scan processed with CVE is shown in Figure 2.6. The image is 512×512 pixels and was produced on a Technicare 2060 CT scanner.

The selection of parameters proved to be difficult. For the image shown, there was a tradeoff between the amount of low frequency restored and the size of the desired standard deviation. However, in all cases where a reasonable level of contrast enhancement was achieved, the noise was greatly enhanced and a ringing artifact introduced which is quite pronounced as can be seen in Figure 2.6. The application of CVE also caused objects in many images to break up into subobjects, with the result that it was difficult to compare features in different parts of the image or to recognize common image features. The method was found to be somewhat sensitive to the choice of window size; a small window (≤ 9 pixels on a side) was necessary to give sufficient contrast enhancement, but this choice intensified the ringing artifacts in the image.

A modification of CVE which reduces the ringing artifacts was proposed in Schwartz and Soha, 1977 to limit the range of the grey levels in the contextual region which can influence the calculation of the mean and standard deviation. This suggestion was not implemented in the current project.

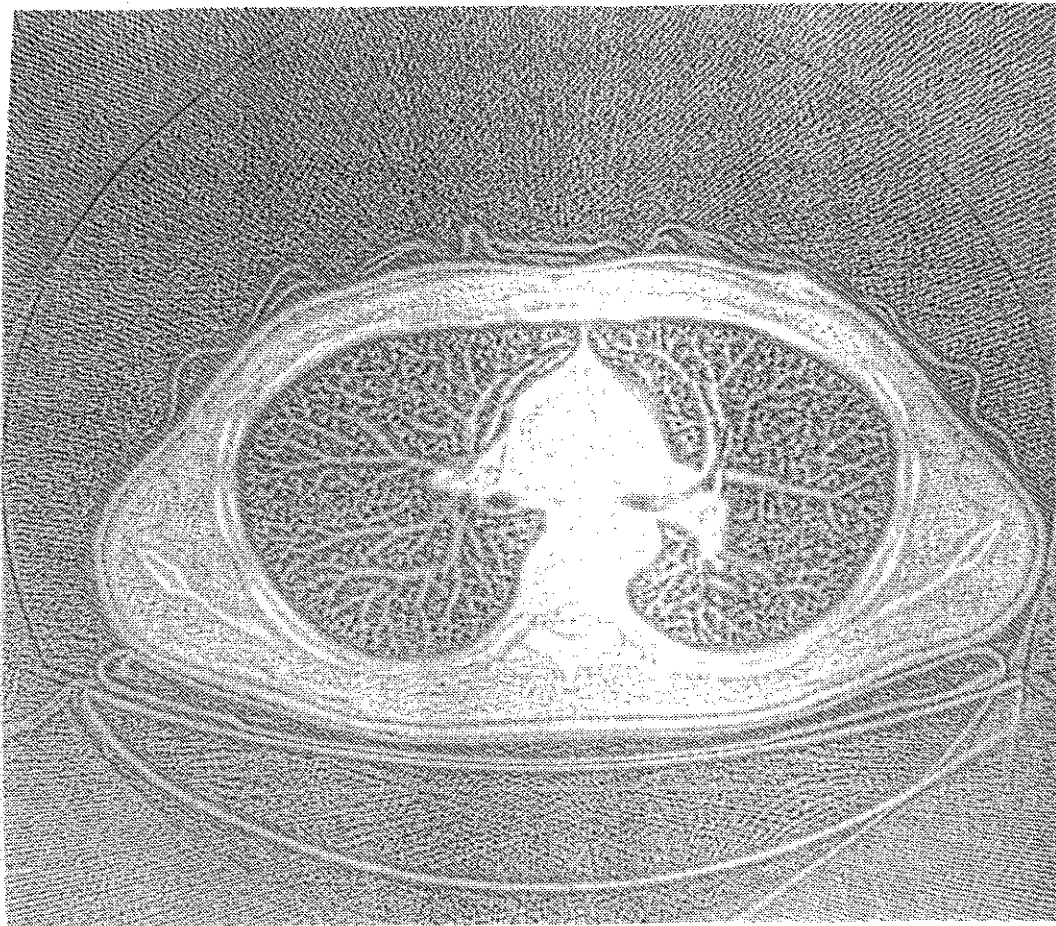


Figure 2.6: A chest CT scan processed with constant variance enhancement. The 512×512 image was processed with a square contextual region 5 pixels on a side with $S'_{xy} = 200$ $B = -117$ $MAX = 500$ and low frequency restoration of .75. The original data range was -1131 to +896.

2.4.2 Peli-Lim Adaptive Filtering

Peli and Lim, 1982, developed an adaptive filter defined by [2.10] with $B = 1$ and T_1 chosen to be a function of \bar{I}_{xy} . They separate the image into high and low frequency components and process each separately before recombination [Figure 2.7]. The resulting formulation is

$$I'_{xy} = A [T_1(\bar{I}_{xy})(I_{xy} - \bar{I}_{xy}) + T_2(\bar{I}_{xy})] + C. \quad [2.13]$$

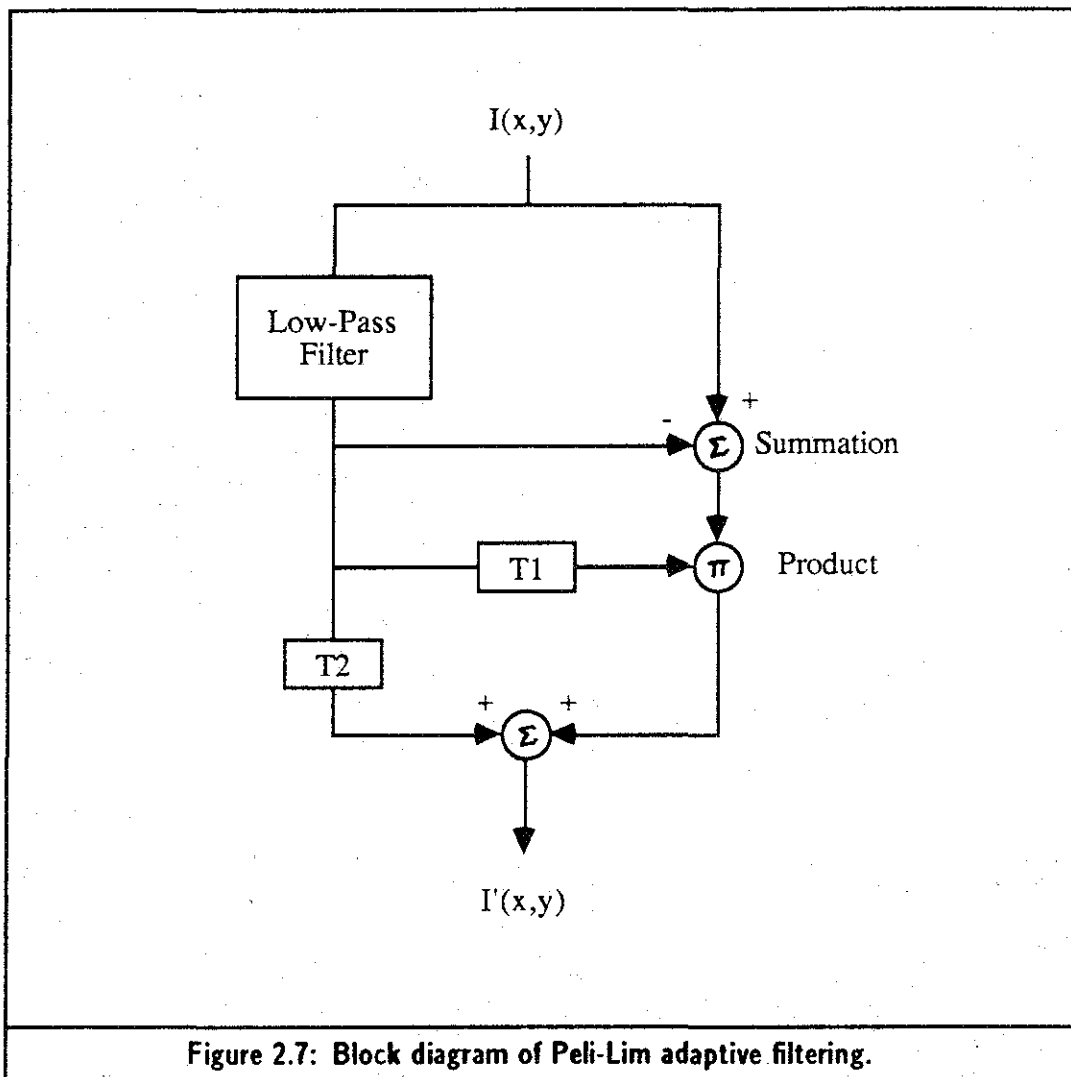
where T_1 and T_2 are chosen to match the properties of a given image or imaging modality. The low frequency component is obtained by averaging over a contextual region that is typically between 5 and 8 pixels on a side. Peli and Lim present three different situations where they have applied their method with good results. In each instance, the functions T_1 and T_2 are carefully chosen to match the problem at hand. For example, in the case of aerial photographs which have been degraded by cloud cover, the detail is in the high frequency variations at middle intensities and the degradation is at low spatial frequencies and high intensities. Thus, T_1 is chosen to stretch the contrast only slightly at low and high mean intensities but to enhance the contrast strongly in the middle intensities. The function T_2 reduces the luminance mean for high mean intensities while leaving low and middle intensities undisturbed.

Unfortunately, this freedom to choose the contrast stretching functions leads to an enormous parameter space that must be searched to find an acceptable filter. There is certainly no hope of finding optimal values for these functions; they must be chosen to match the problem at hand. There is no explicit LCEG for this method.

2.4.3 Implementation

This filter was implemented in much the same way as CVE. A square moving window 9 pixels on a side was used to calculate the local mean and the contrast stretch applied. Again, this implementation is quite costly, though it may be made less so by calculating block means and using an interpolation scheme. It is also reasonable to conceive of hardware implementations.

A difficulty in implementing this method is that the functions T_1 and T_2 often have no simple analytical expression and thus are difficult to express compactly. Several alternative representations are available; one could require them to be piecewise linear or calculate them on the fly according to their desired properties. In the present implementation, a table of function values is supplied to the program, allowing the specification of arbitrary functions.



2.4.4 Results of Application to Medical Images

Using the gain functions shown in Figure 2.8, the image in Figure 2.9 was produced with the filter as described in [2.13]. Aside from the difficulty of determining adequate processing functions and window sizes, it is clear that the Peli-Lim process introduces extensive ringing while not expanding the contrast in large areas of the image. It would appear that this technique is better suited for images which have undergone large scale degradations such as contamination by cloud cover in aerial photographs or underexposure of large portions of the image. For medical images which have considerable detail in all subranges of the data, Peli-Lim filtering has little scope. One possible area of application in medical imaging is that of chest radiographs which like the cloud images are degraded by the projection of large overlying objects. The method could be expanded to allow the

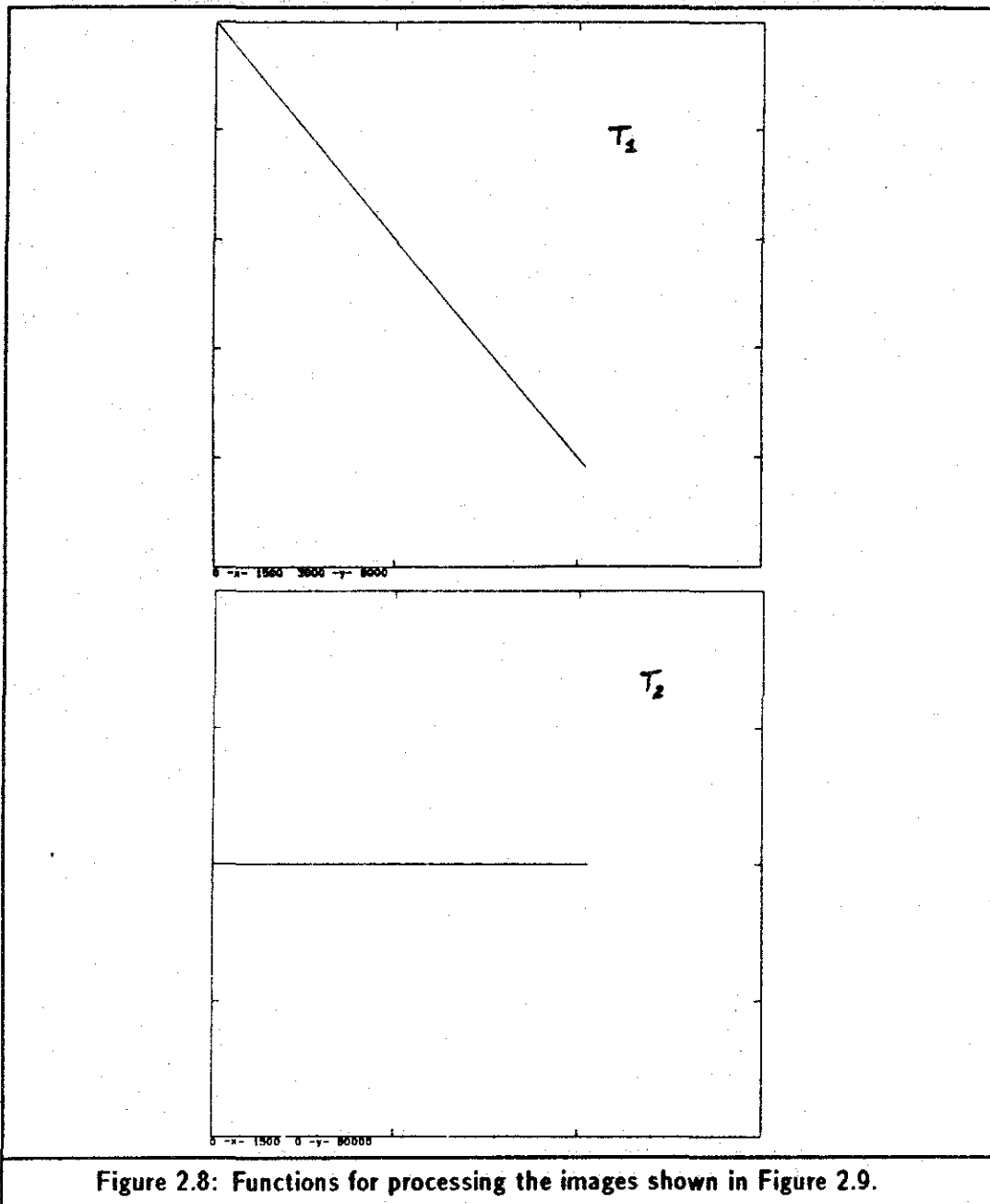


Figure 2.8: Functions for processing the images shown in Figure 2.9.

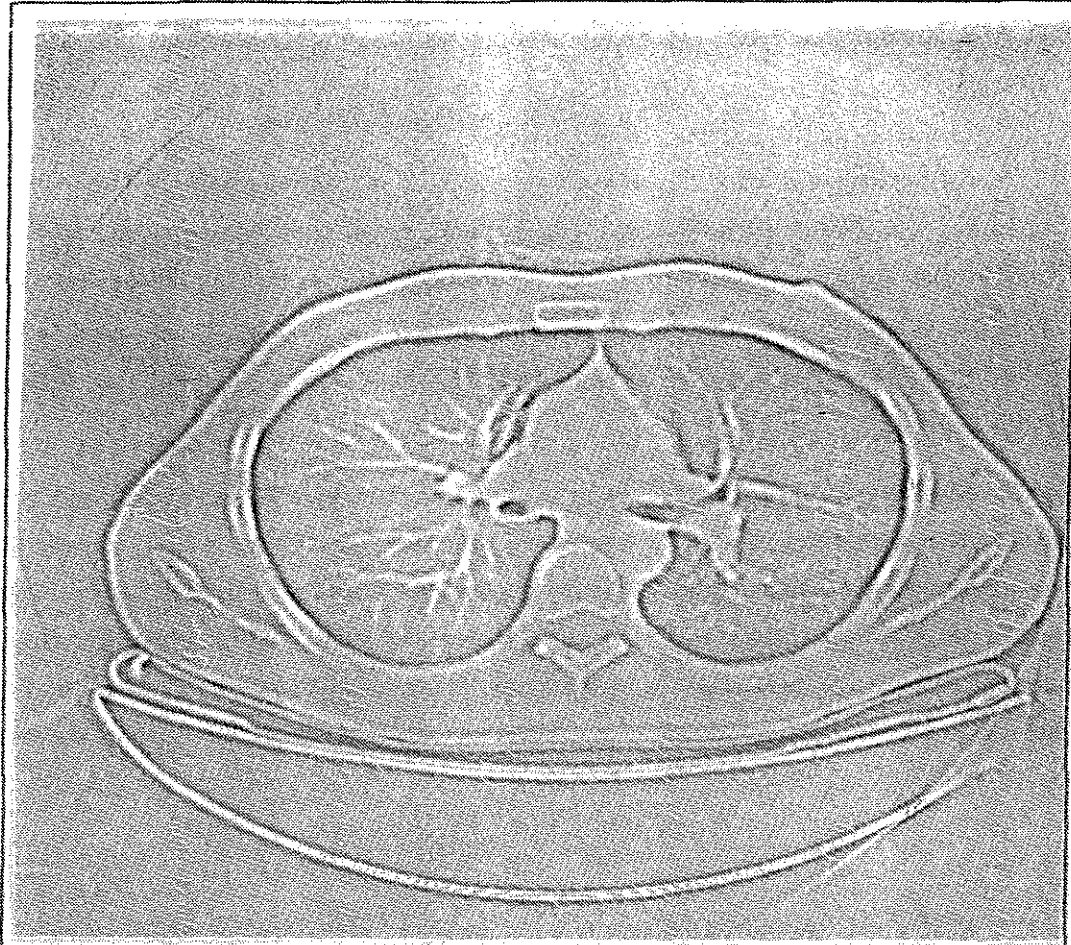


Figure 2.9: A chest CT image processed with Peli-Lim adaptive filtering. The image was processed with a square contextual region 5 pixels on a side and the functions shown in Figure 2.8.

gain functions to vary from point to point, but without a clear guiding theory for the selection of these functions, this seems to compound the problems rather than solve them.

2.5 Local Range Modification

Fahnestock and Schowengert have developed a promising method of adaptive filtering called *Local Range Modification (LRM)* [Fahnestock and Schowengert, 1983]. Their method has only three parameters and is efficiently implementable. Their technique for reducing computational expense is of some interest, as it is applicable to some of the other methods so far examined and is similar to that used in Pizer's Adaptive Histogram Equalization.

LRM is a form of adaptive linear min-max windowing. We first examine a straightforward implementation of adaptive linear min-max windowing before considering the implementation of Fahnestock and Schowengert. Consider a square contextual region Ω of size k pixels on a side centered around a pixel at (x, y) . Then define I_{\min} to be the minimum intensity value and I_{\max} the maximum intensity value in Ω . If C_1 and C_2 are the desired range and mean of the output image, then the intensity mapping for the pixel is

$$I'_{xy} = \frac{C_1}{(I_{\max} - I_{\min})} (I_{xy} - I_{\min}) + C_2. \quad [2.14]$$

Notice that the minimum and maximum within Ω are the only local statistics used. The free parameters here are the scaling constants that determine the output range and the size of the window over which the minimum and maximum are calculated.

2.5.1 Implementation

The problem with the simple adaptive min-max windowing formulation is that for each output pixel in the image, every pixel in Ω must be examined to calculate the minimum and maximum. Although careful development of the algorithm allows this to be done in $O(N^2k)$ for an $N \times N$ image with a $k \times k$ contextual region, it is still costly. One way to reduce the cost is to use an *estimate* of the minimum and maximum at every pixel, rather than the true values. Such methods have been alluded to in previous sections. Here we present a detailed explanation of how this is done in *Local Range Modification*.

Consider an $M \times N$ image as shown in Figure 2.10. The image is divided by a rectangular grid into $m \times n$ contextual regions each of size $\Delta x \times \Delta y$ pixels, where $\Delta x = M/m$ and $\Delta y = N/n$. Within each region, the minimum and maximum intensity values of the pixels in the region have been calculated. The regional minimum is shown

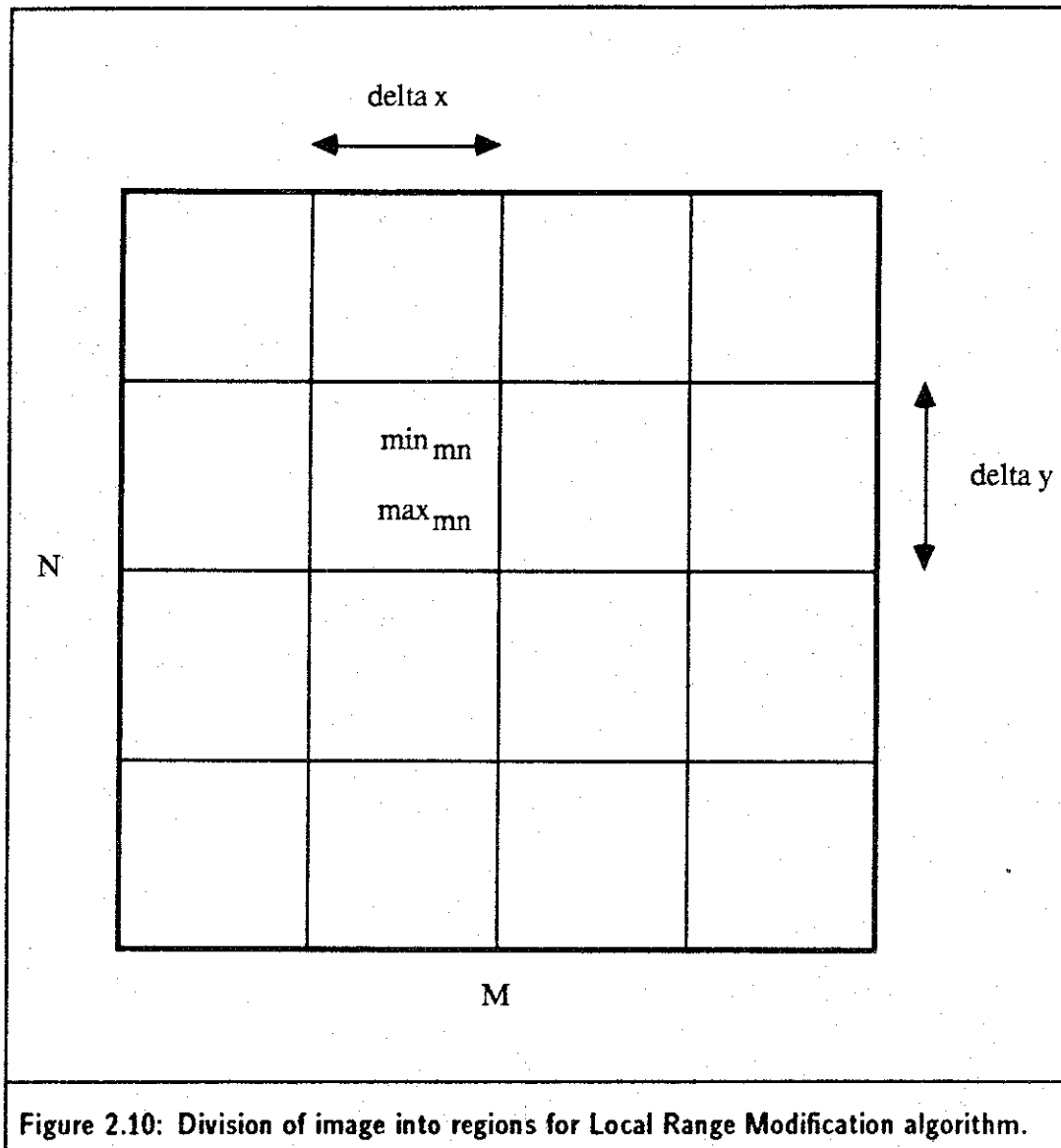


Figure 2.10: Division of image into regions for Local Range Modification algorithm.

in Figure 2.10 as \min_{mn} and the maximum as \max_{mn} . Rather than using these values directly as I_{\min} and I_{\max} in Eq. 2.14 for all pixels in the given region, a scheme is used which allows for the gradual transition from one set of min-max values to another as the pixel being mapped moves from one region to another.

Consider the vertices of the grid which divides the image into regions; one or more regions adjoin at each vertex. Each of these vertices is assigned a pair of values which are the maximum and minimum values for the blocks surrounding the vertex. Let V_{ij} be a vertex; here the indices run over the values $i = 0, 1, \dots, m; j = 0, 1, \dots, n$. Each vertex is assigned a pair of values $V_{ij} = (\text{Max}_{ij}, \text{Min}_{ij})$, such that Max_{ij} is the maximum intensity values in all the blocks that adjoin V_{ij} and Min_{ij} is the minimum intensity value in these same blocks.

Once the vertex values are assigned, an adaptive min-max windowing is applied at every pixel using the minimum and maximum values assigned at nearby vertices to control the amount of contrast stretching. Suppose we have a pixel at location (x', y') in the (i, j) th region of the image. The values I_{\max} and I_{\min} used in Eq. 2.14 are calculated by bilinear interpolation from the values assigned to the four vertices nearest the pixel [Figure 2.11]. Suppose that the coordinates of vertex V_{ij} are (x_i, y_i) . Define

$$\delta x = x' - x_i,$$

$$\delta y = y' - y_i.$$

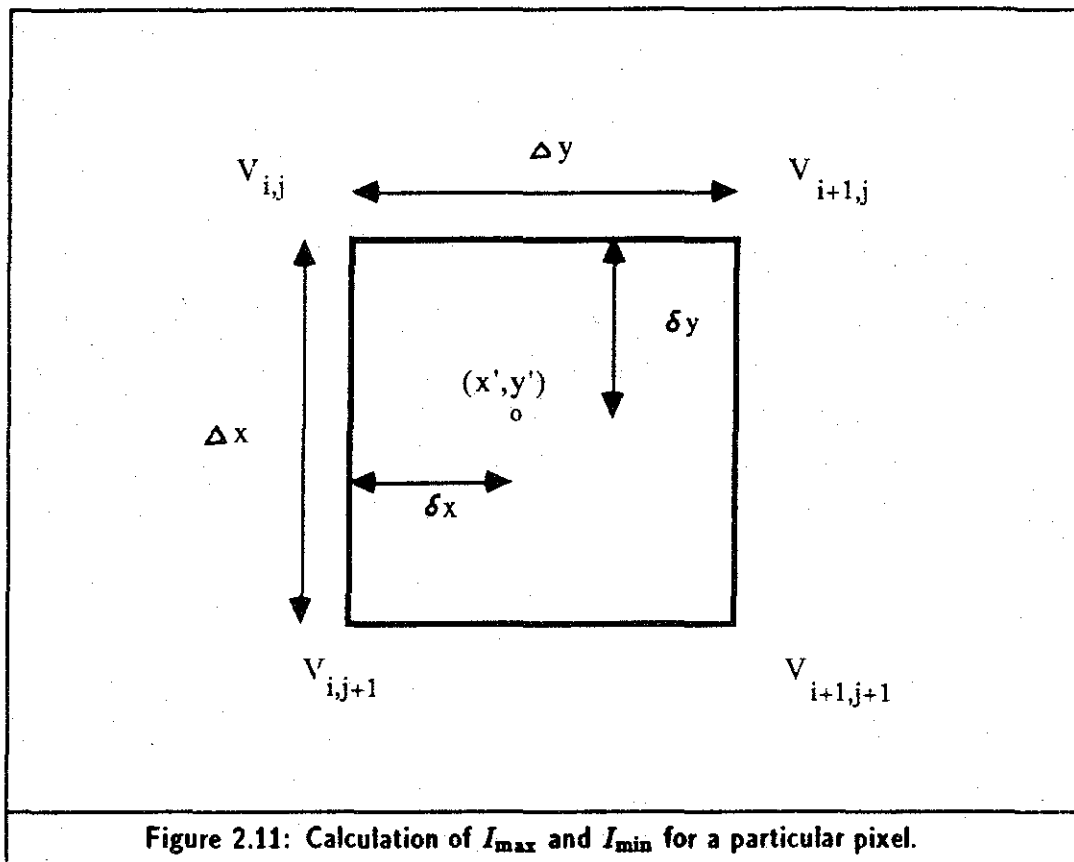
Then the interpolation yields the values

$$I_{\max} = \left[\left(\frac{\delta x}{\Delta x} \right) \text{Max}_{i+1,j} + \left(\frac{\Delta x - \delta x}{\Delta x} \right) \text{Max}_{i,j} \right] \left(\frac{\Delta y - \delta y}{\Delta y} \right) + \left[\left(\frac{\delta x}{\Delta x} \right) \text{Max}_{i+1,j+1} + \left(\frac{\Delta x - \delta x}{\Delta x} \right) \text{Max}_{i,j+1} \right] \left(\frac{\delta y}{\Delta y} \right) \quad [2.16]$$

$$I_{\min} = \left[\left(\frac{\delta x}{\Delta x} \right) \text{Min}_{i+1,j} + \left(\frac{\Delta x - \delta x}{\Delta x} \right) \text{Min}_{i,j} \right] \left(\frac{\Delta y - \delta y}{\Delta y} \right) + \left[\left(\frac{\delta x}{\Delta x} \right) \text{Min}_{i+1,j+1} + \left(\frac{\Delta x - \delta x}{\Delta x} \right) \text{Min}_{i,j+1} \right] \left(\frac{\delta y}{\Delta y} \right) \quad [2.17]$$

The new pixel value is again

$$I'_{xy} = \frac{C_1}{(I_{\max} - I_{\min})} (I_{xy} - I_{\min}) + C_2. \quad [2.18]$$



This method does not fit as easily into the paradigm of high frequency enhancement as the previous filters discussed. It does not separate the image directly into low and high frequency parts, but the size of the contextual regions effectively acts as a high pass filter. The only local statistics which are used are the area minimum and maximum values. The implementation obscures the simplicity of the processing technique.

2.5.2 Results of Application to Medical Images

The LRM algorithm using the approximation of minimum and maximum values was implemented and applied to several medical images. Results for one image are shown in Figure 2.12. The constants C_1 and C_2 are determined by the fact that the image must be scaled into the output range of the display device; for this image, C_1 was set to the range of the data and C_2 was set to the image mean. This leaves only the size of the contextual regions as a free parameter. For the image shown, the contextual regions were respectively $1/32$, $1/64$, and $1/96$ of the image size in both the x and y directions. As can be seen, the larger region sizes provide little enhancement but introduce severe ringing and block artifacts. The smallest region size enhances the contrast well but at the expense of the

stability of objects in the image. The intermediate region sizes enhances the contrast well, but ringing and the breakup of objects are not excessive. The image also has a natural appearance; objects seem to hold together very well. The images shown in the paper by Fahnestock and Schowengerdt are LANDSAT photos; they are enhanced well, but ringing artifacts are present there as well.

2.6 Other Methods of Adaptive Contrast Enhancement

Many other methods of ACE have been proposed in the literature; they will not be considered here. References to a number of papers on ACE and its implementations in hardware and software will be found in the bibliography. In the next chapter, one more ACE method will be examined, Pizer's Adaptive Histogram Equalization. This method has performed quite well in our laboratory and will be examined at length.

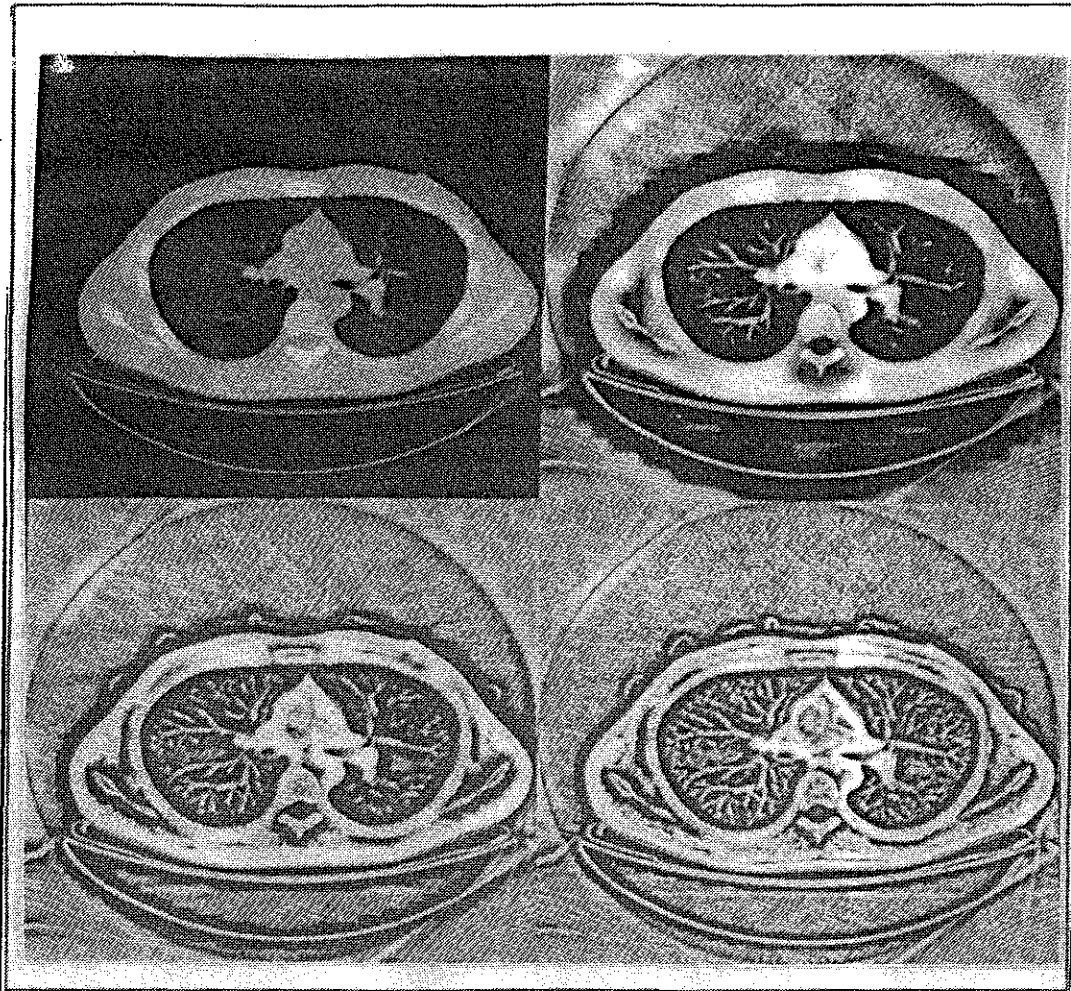


Figure 2.12: A chest CT image processed with the Local Range Modification algorithm.

Upper left: original image.

Upper right: contextual region $1/32$ of the image on a side.

Lower left: contextual region $1/64$ of the image on a side.

Lower right: contextual region $1/96$ of the image on a side.

Chapter 3

Properties of Adaptive Histogram Equalization

*Though this be madness,
yet there is method in't.*

—Hamlet II.ii.203

Adaptive Histogram Equalization (AHE) is a method for adaptive contrast enhancement developed by Pizer [Pizer, 1981a; Pizer, 1981c; Pizer *et al.*, 1984] which is based on information theoretic considerations. In this technique, an attempt is made to minimize the mean pixel uncertainty on a local basis by applying a local histogram equalization mapping at each pixel. This method is fast, automatic, and has produced excellent results for several types of images. A number of authors have suggested mappings similar to AHE. In the first sections of this chapter, those works are summarized and the theoretical development of AHE is given. The remaining sections will describe new developments which yield insight into the effects of AHE on digital images.

3.1 Theoretical Foundations of AHE

The AHE algorithm is similar to the ACE methods already discussed in that it uses the local properties of the image to guide the selection of a contrast enhancement mapping. In AHE, the local property used is the grey-level histogram in the neighborhood of a pixel; the contrast enhancement mapping is chosen to flatten that histogram. The choice of histogram equalization as the LCEG arises from a desire to maximize the information transfer from image to observer.

3.1.1 Motivation: Information Transfer

The original development of AHE was inspired by the work of Cormack and Hutton, 1980, which uses the methods of information theory to consider the process of information transfer from an image to an observer. When the range of data in an image is larger than the available display levels, a data compression takes place. To minimize the

loss of information due to compression, the authors derive an intensity mapping which minimizes the mean uncertainty (entropy) of the image on a pixel-by-pixel basis. This mapping is then applied globally to the image. The development of this *Mean Pixel Uncertainty* (MPU) minimization mapping is sketched below; a full derivation is given in the paper cited and in Cormack and Hutton, 1981. The basic concepts of information theory are well explicated in Abramson, 1963.

Mean Pixel Uncertainty. Consider the process of image acquisition and display shown in Figure 3.1. Let the distribution of intensities in a scene be designated by Λ . When the scene is imaged, a particular source value λ is converted to a recorded intensity n with some probability $P(n|\lambda)$. Thus the imaging process converts the continuous distribution Λ into a discrete distribution $\{n\}$. A given value n is then displayed with some value k and perceived by the observer as l . Conditional probability distributions $P(s|t)$ are associated with each of these transformations. Let $I(\Lambda)_{\lambda \rightarrow l}$ be the average information transferred by a source value in Λ which is perceived as value l by the observer. Then

$$I(\Lambda)_{\lambda \rightarrow l} = H(\Lambda) - H(\Lambda|l). \quad [3.1]$$

Here H is the entropy associated with the given distribution. Thus, the information transferred to the observer by perceiving a value l averaged over the source distribution, Λ , is the difference between $H(\Lambda)$, the *a priori* uncertainty, and $H(\Lambda|l)$, the uncertainty averaged over Λ when l is seen. This conditional entropy is given by

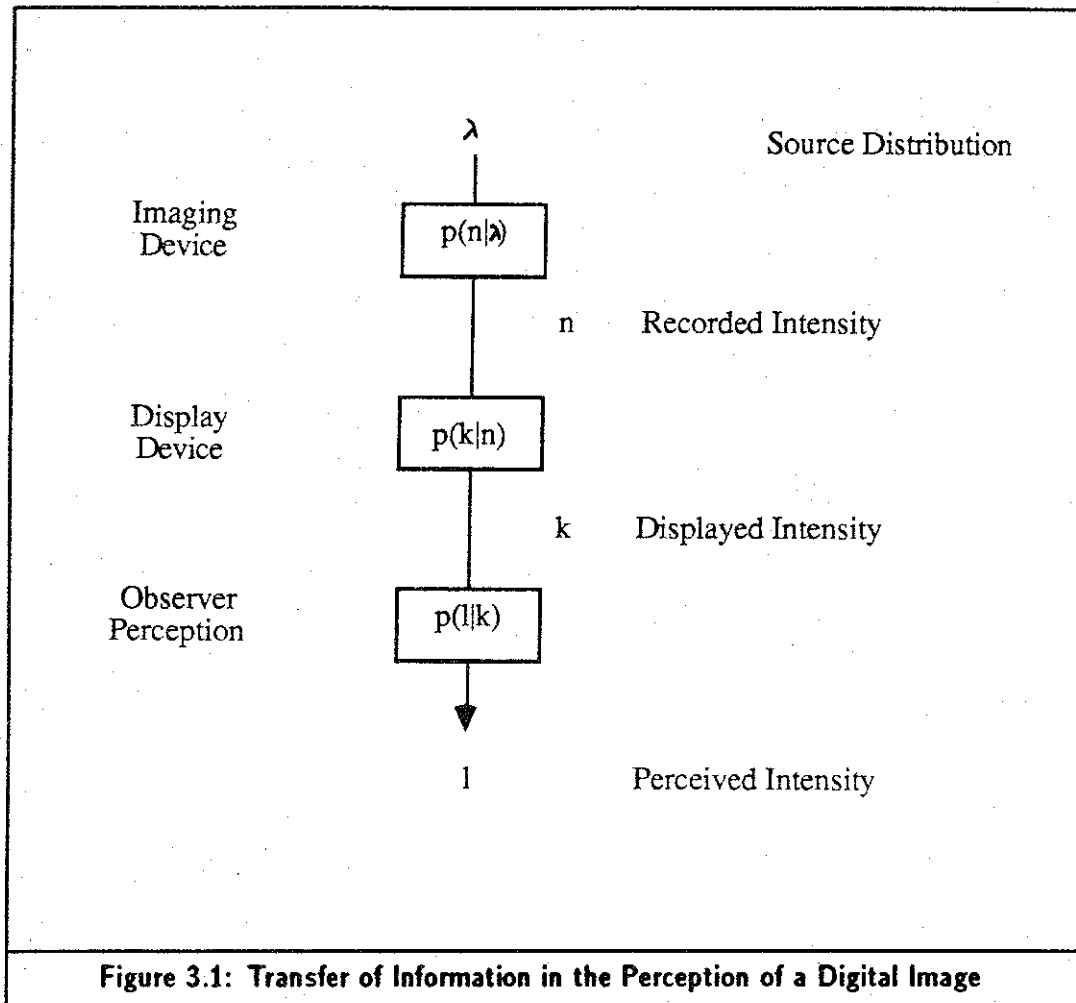
$$H(\Lambda|l) = - \int_0^{\infty} P(\lambda|l) \log[P(\lambda|l)] d\lambda, \quad [3.2]$$

where the logarithm is usually taken to the base 2. The uncertainty then is measured in bits.

Cormack and Hutton evaluate [3.2] in the case where 1) there is no *a priori* knowledge of the source distribution, $P(\lambda)$, and 2) the pixel uncertainty is position independent (*i.e.*, there are no inter-pixel correlations). This expression is in general quite complex, but can be used to calculate the information transferred by observing value l if assumptions are made about the various conditional probabilities.

If we average the uncertainty per pixel over the entire image, we obtain the *mean pixel uncertainty*, \bar{H} , for the image:

$$\bar{H} = \frac{1}{N} \sum_{x,y} H_{xy}(\Lambda|l), \quad [3.4]$$



where N is the total number of pixels in the image and H_{xy} is the pixel uncertainty at (x, y) given that l is perceived. Assuming that the entropy is independent of position in the image,

$$\bar{H} = \sum_{l=0}^M P(l) H(\Lambda|l), \quad [3.5]$$

when $0 \leq l \leq M$ and M is the maximum level perceived. The probability $P(l)$ that a level l will be seen is not known, but must be calculated from the statistics of the given image and the characteristics of the perception mechanism. If a value $0 \leq k \leq M$ is displayed, then

$$\begin{aligned} P(l) &= \sum_{k=0}^M P(l|k) P(k) \\ &= \frac{1}{N} \sum_{k=0}^M h(k) P(l|k), \end{aligned} \quad [3.6]$$

where $h(k)$ is the grey-level histogram of the displayed image.

Minimization of the MPU. Recall from [3.1] that information transferred when a pixel of value l is seen is

$$I(\Lambda)_{\lambda \rightarrow l} = H(\Lambda) - H(\Lambda|l).$$

The mean information transfer per pixel then is

$$\begin{aligned} I(\Lambda)_{\lambda \rightarrow l} &= \sum_{l=0}^M P(l)[H(\Lambda) - H(\Lambda|l)] \\ &= H(\Lambda) - \bar{H}. \end{aligned} \quad [3.7]$$

Notice that to compare the mean information transfer of two images, a knowledge of the source distribution, Λ , is required for each. However, since $H(\Lambda)$ is fixed for a given image, the mean information transfer is maximized for that image when \bar{H} is minimized.

If the noise is assumed to be stationary across the image, it can be shown that the use of histogram equalization as the transformation which maps $n \rightarrow k$ results in no information loss on the average, that is

$$\bar{H}_{\lambda \rightarrow n} = \bar{H}_{\lambda \rightarrow k}.$$

Since no information can be gained in the transformation $n \rightarrow k$, this is the best that can be done; if furthermore the display device is linear, the information in the values $\{k\}$ will be transmitted as accurately as possible to the observer. If the function $f: n \rightarrow k$, then it is known from elementary probability that

$$p(k|\lambda) = p(n|\lambda) / |f'(n)|;$$

for histogram equalization, $f'(n) = p(n)$, from which it follows that the transformation $n \rightarrow k$ results in no information loss in the mean.

Critique of MPU Minimization. The foregoing development has suggested the following simple result: if the display device is linear and the noise is stationary, the mean pixel uncertainty can be minimized by histogram equalization. This result is in correspondence with the intuitive feeling that images which have been histogram equalized are better than their untreated originals; of global contrast enhancement mappings, histogram equalization has proven quite durable. However, the assumptions which have been made in the preceding derivations are worth closer examination.

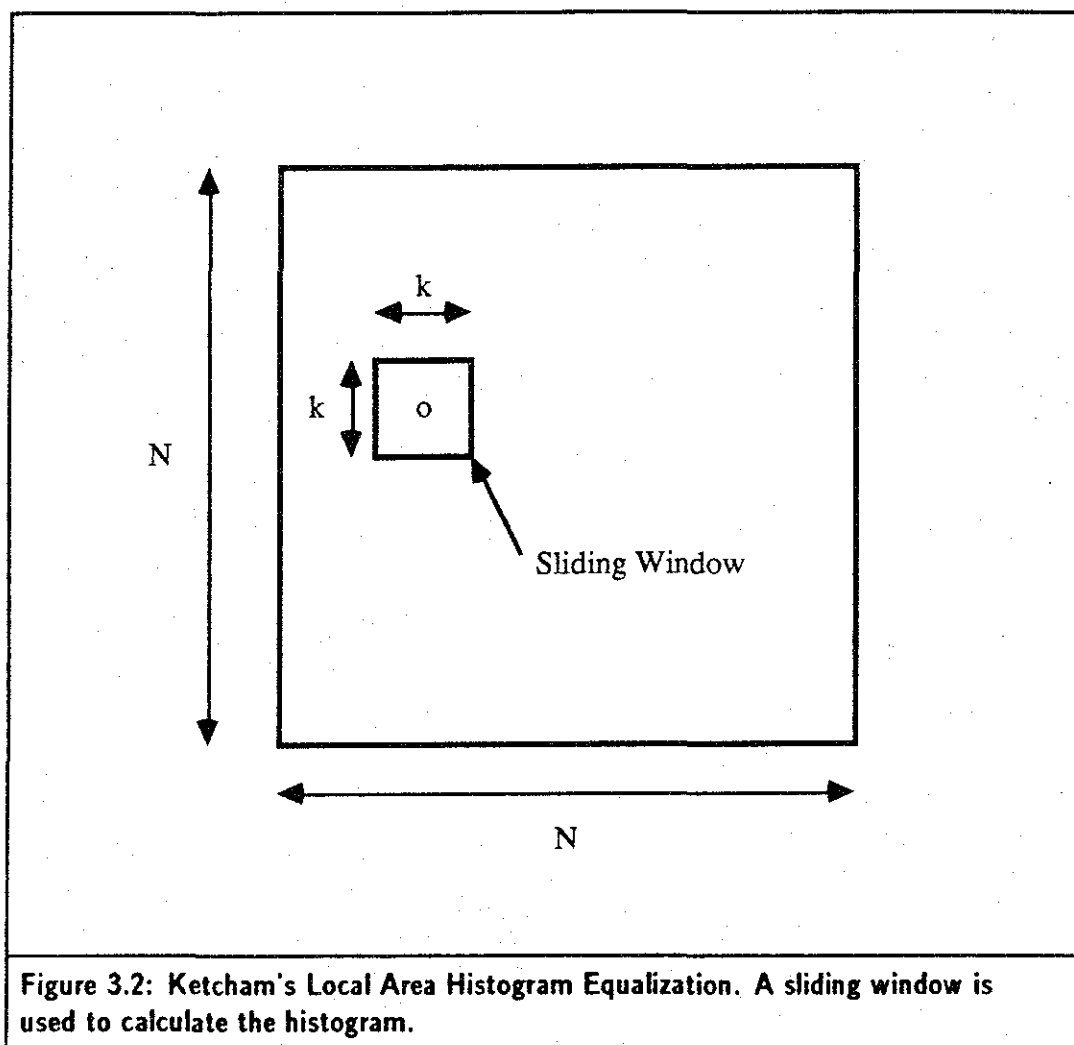
First, it is assumed that there is no *a priori* knowledge of the source distribution. This assumption, while somewhat questionable, allows a development independent of the exact characteristics of the source. Second, it is assumed that the display device is linear, *i.e.*, that $P(l|k)$ is independent of the absolute values of k and l . This is necessary if the fidelity of the information transmission to the observer is to be preserved. Pizer, 1981b has shown how to achieve this property approximately .

The greatest problem is with the assumption that the MPU minimization can be done on a pixel by pixel basis; that is, the values derived from the source at different pixels in the image are uncorrelated. There are two aspects to this assumption. First, it says that the pixel values are independent even if statistical noise is disregarded. This is normally not true; the source consists of objects, thus neighboring pixel values are often highly correlated. Second, it is implied that the statistical noise added by the imaging and perception processes is uncorrelated. Unfortunately, the power spectrum of the noise in many common imaging modalities such as CT is nonwhite. Thus the assumption of pixel independence which allows entropy minimization on a pixel by pixel basis is often not justified. However, the goal of maximizing the information transfer from image to observer is an appealing one; in this light, the pixel by pixel minimization of mean pixel uncertainty using the well-understood, easily implementable technique of histogram equalization can be justified as a first attempt to develop a methodology which will achieve this goal.

3.1.2 Previous Work on AHE

The results of the preceding section imply strongly that an adaptive contrast enhancement mapping based upon histogram equalization is worth investigation. Various such mappings have been proposed previously in the literature. It should be noted that while some of the papers mentioned below predate the work of Pizer, 1981a, that work was developed independently of the research discussed in this section.

Ketcham's Local Area Histogram Equalization. Ketcham *et al.*, 1976. have developed and implemented a straightforward extension of histogram equalization for use as an adaptive contrast enhancement mapping [Figure 3.2]. In this scheme, a sliding window is used to calculate the histogram and histogram equalization mapping in the neighborhood of each pixel. Termed *Local Area Histogram Equalization (LAHE)* by the authors, this technique was implemented by using special-purpose hardware so that images could be



enhanced at video frame rates. They present several examples of the application of this method to images.

Hummel. Hummel, 1977, in a general review of histogram modification techniques, mentions the possibility of local histogram modification. He notes that the image which ensues will not necessarily have a flat histogram even though each pixel has been processed by local histogram equalization. He recommends a post-processing phase in which global histogram equalization is performed after the local histogram equalization mapping. The utility of this method in conjunction with AHE is examined later in this chapter.

Driscoll and Walker. Driscoll and Walker, 1983 show an implementation of local histogram equalization for a particular commercial frame buffer which utilizes special-purpose

hardware to perform the necessary histogram calculations swiftly. The details of their implementation will not be discussed here; a similar algorithm has been developed by Pizer that is applicable to many commercially available frame buffers. The method of Driscoll and Walker shares the common flaw of the straightforward extensions of histogram equalization to adaptive contrast enhancement: special-purpose, parallel, pipelined hardware must be used, since the method is unacceptably slow when implemented on general-purpose computers. Pizer's Adaptive Histogram Equalization algorithm avoids this problem, allowing implementation on ordinary minicomputers, while producing results that are visually equivalent to the direct application of local histogram equalization.

3.1.3 Pizer's Adaptive Histogram Equalization

The Adaptive Histogram Equalization algorithm of Pizer was motivated by the desire to extend the information-theoretic ideas of Cormack and Hutton's Mean Pixel Uncertainty method to the realm of adaptive contrast enhancement. It also attempts to meet the criteria for an ACE method given in Chapter 2: it generates few artifacts, is sufficiently fast for general use, does not cause an objectionable break-up of objects in the image, and has a small parameter space.

Extension of HE to an Adaptive Mapping. The straightforward extension of histogram equalization is essentially that of Ketcham [Figure 3.2]. For each pixel, a neighborhood of some size is chosen about the pixel; the histogram of this region is determined and a histogram equalization mapping calculated. The resulting function is then applied to the pixel; this process is then repeated for the next pixel.

The principal drawback of this method is that it is computationally expensive. For an $N \times N$ image, the calculation is $O(N^2(1 + k + L))$, where L is the number of intensity levels in the image and the contextual region is $k \times k$. The size of the contextual region, k , enters linearly if the mapping for each pixel utilizes the information from the calculation for the previous pixel optimally. Pizer's algorithm reduces the computation time by calculating the histogram equalization mapping only at selected sample pixels; the mapping for all other pixels is then determined by a bilinear interpolation of the mappings for nearby sample points. If there are S sample points, the computation is then $O(N^2 + S(k^2 + L))$. If $S \ll N^2$, this implies a considerable savings. Experience has shown that for many medical images $S = 64$ is ample.

Algorithm and Implementation. Pizer's algorithm is similar to that of Fahnestock and Schowengerdt's Local Range Modification; however, where LRM is conceptualized as a division of the image into blocks, it is easier to consider Pizer's algorithm in the following way. A set of sample points in the image is chosen; at each of these points, the histogram equalization mapping is calculated precisely within a contextual region about the point [Figure 3.3]. The size of the contextual region around each sample point is chosen independently of the number of sample points. This implies that the union of the contextual regions may not cover the image, or that the regions may overlap. Most commonly, and for use in the research here reported, the region size is such that the regions exactly cover the image with the sample points on a regular rectangular grid.

Assume that the grid of sample points is $m \times n$. If the image is $M \times N$ pixels, then the spacing between grid points in x and y is [Figure 3.3]

$$\begin{aligned}\Delta x &= M/m, \\ \Delta y &= N/n.\end{aligned}\tag{3.12}$$

Let the sample points be labeled S_{ij} , $0 \leq i \leq N-1$, $0 \leq j \leq M-1$. A contextual region $\Delta x \times \Delta y$ pixels on a side is used to calculate the local histogram and the cumulative distribution function, CDF_{ij} , within the region. Over all sample contextual regions, this requires $O(S(\Delta x \Delta y + L))$ operations, where L is the number of intensity levels in the image and S is the total number of sample points. If $S \times \Delta x \Delta y = N^2$ (as it will if the contextual regions exactly cover the image), then the calculations are $O(N^2 + SL)$.

Now for each pixel at a location (x, y) in the image, the new image value is calculated by bilinear interpolation of the mappings that apply at the four nearest surrounding sample points [Figure 3.4]. Let the coordinate of S_{ij} be (x', y') . Then define

$$\begin{aligned}\delta x &= x - x' \\ \delta y &= y - y'\end{aligned}$$

and the new value of the pixel shown in Figure 3.4 is

$$\begin{aligned}I'_{xy} &= C \times \left[\frac{\delta x}{\Delta x} CDF_{i+1,j}(I_{xy}) + \frac{(\Delta x - \delta x)}{\Delta x} CDF_{i,j}(I_{xy}) \right] \left(\frac{\Delta y - \delta y}{\Delta y} \right) \\ &+ C \times \left[\frac{\delta x}{\Delta x} CDF_{i+1,j+1}(I_{xy}) + \frac{(\Delta x - \delta x)}{\Delta x} CDF_{i,j+1}(I_{xy}) \right] \left(\frac{\delta y}{\Delta y} \right).\end{aligned}\tag{3.13}$$

The constant C is used to scale the output into the desired range (usually the original range of the image). For pixels which do not have four surrounding sample points

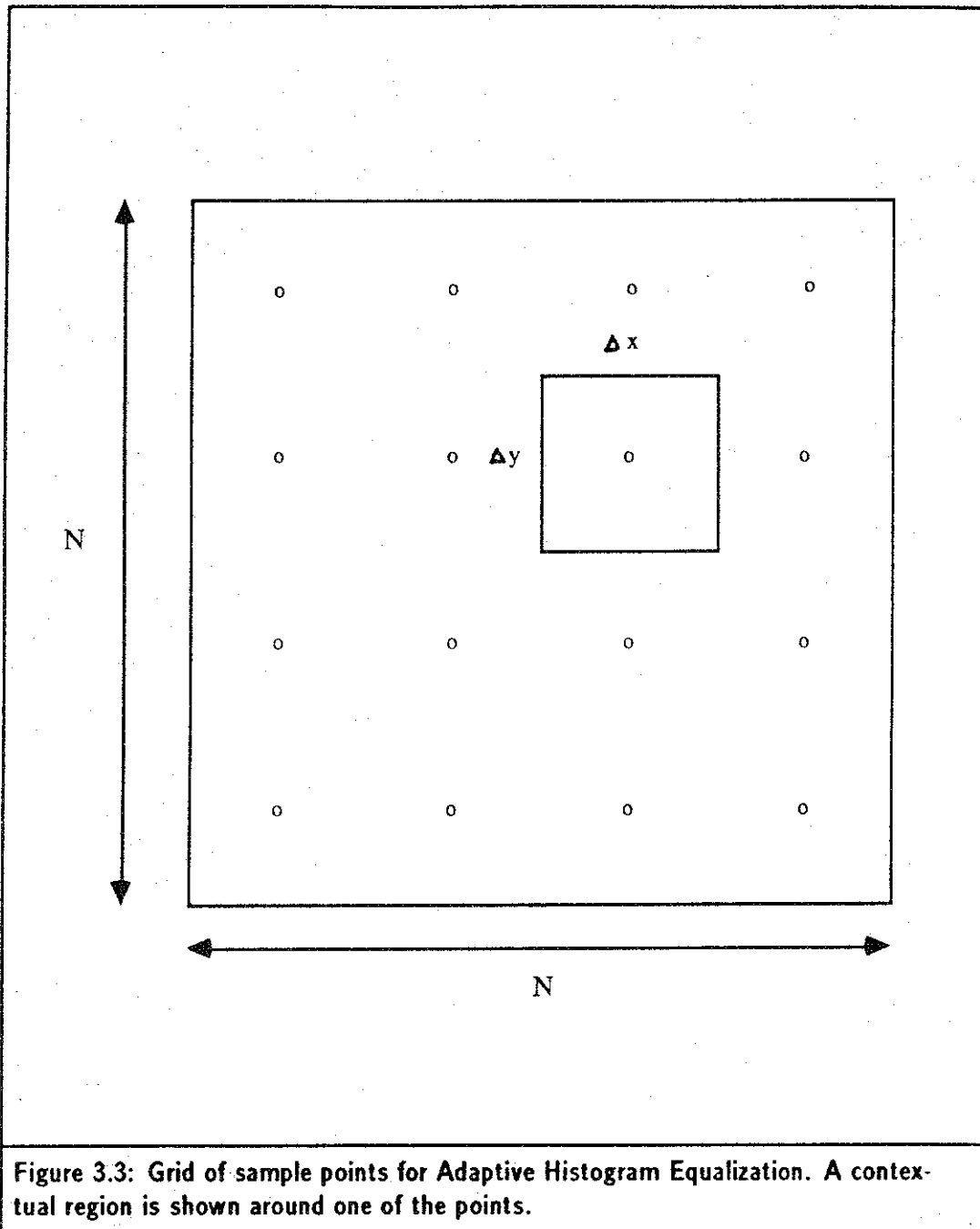
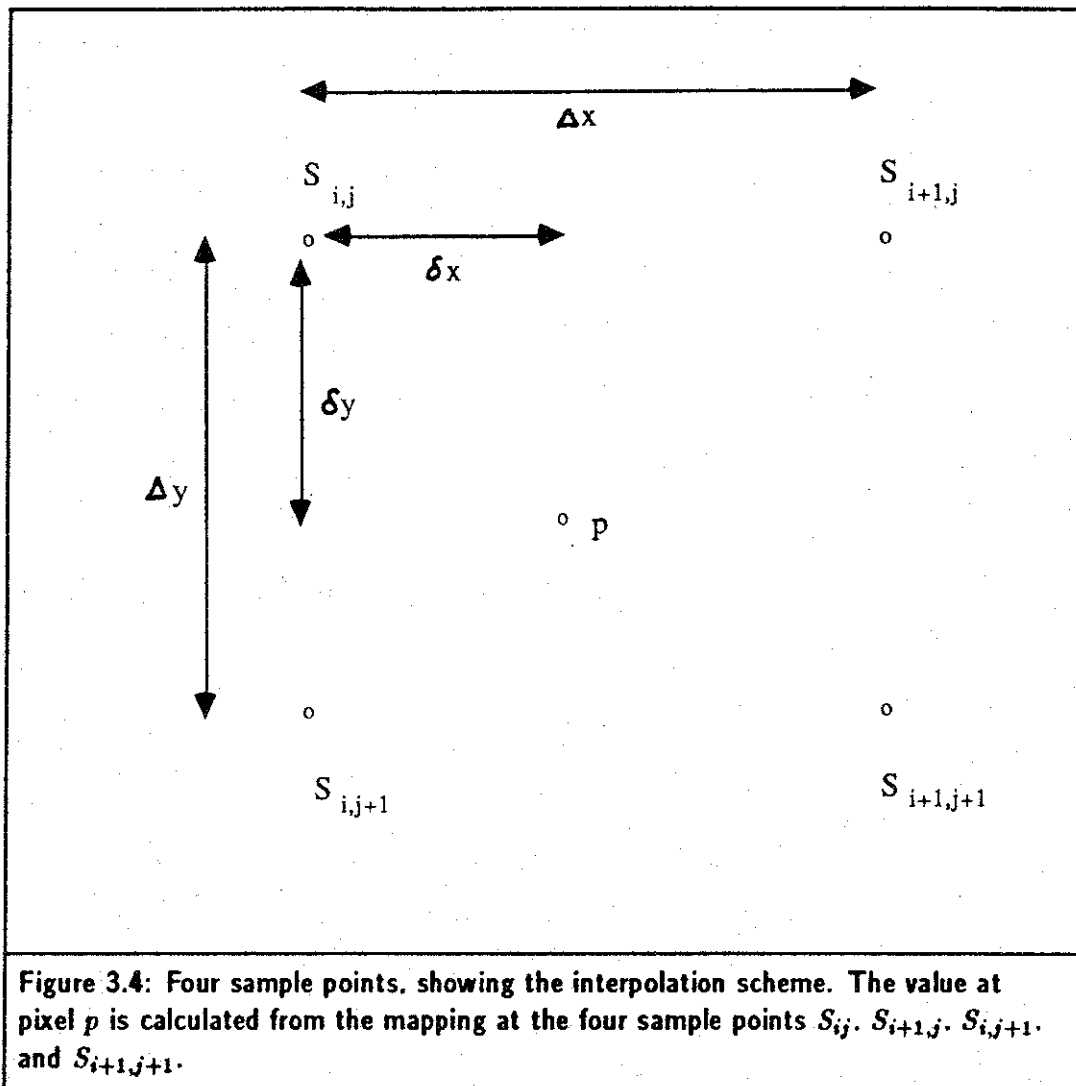


Figure 3.3: Grid of sample points for Adaptive Histogram Equalization. A contextual region is shown around one of the points.



(those along the edges of the image and in the corners), a similar formulation is used with fewer sample points. For edge pixels, a two point interpolation is used; corner pixels use the mapping for the single nearest sample point directly.

The algorithm described above will be referred to as Adaptive Histogram Equalization (AHE). The straightforward extension of histogram equalization using the sliding window algorithm of Ketcham will be referred to as Local Area Histogram Equalization (LAHE).

AHE as a High-Pass Filter. As in the LRM algorithm, AHE does not fit well into the high-pass filter paradigm described in Chapter 2. Nevertheless, AHE does work as a high-pass filter; the mapping at each pixel is influenced only by image values within the

four adjacent contextual regions. Thus, spatial frequencies in the image whose wavelength is larger than the size of the contextual region will be heavily attenuated. The parameter space for AHE is quite small; the number of sample points in each direction and the size of the contextual region are the only free parameters.

Results of Application to Images. Adaptive Histogram Equalization has been implemented on the VAX11/780 with floating point accelerator at the Computer Graphics and Image Processing Laboratory at UNC and used in a variety of applications. The vast majority of images processed with AHE have been CT scans, though it has also been used with NMR (nuclear magnetic resonance) images, digital radiographs and digital subtraction angiograms. AHE has been found to give remarkably good results in exhibiting diagnostic information; in some hundreds of images that have been examined informally by radiologists, very few images exhibited diagnostic information when windowed that was not seen using AHE. No artifacts were apparent when the method was applied using reasonable parameters. It is quite fast, allowing the processing of a 512×512 pixel image in about 1 minute of CPU time. Shown in Figure 3.5 and Figure 3.6 are examples of the application of AHE to CT scans, compared with windowed versions of the image. In these images, sixty-four sample points in an 8×8 grid were used, with contextual regions around the sample points chosen to exactly cover the image. The lower right image has been processed by AHE; the upper right and lower left images are windowed to show particular structures in the image. In each case, the reader should note the ability of AHE to show the same structures as in the windowed versions simultaneously and with no perceptible loss of information.

3.2 Empirical Results to Date

3.2.1 Basic Properties of the Method

In this section, some of the basic properties of AHE are reviewed. AHE has two parameters, the size of the contextual regions and the number of sample points. Throughout this chapter it will be assumed that the contextual regions are chosen to exactly cover the image, leaving the selection of the number of sample points in the x and y directions as the only remaining free parameters. In this case, the larger the number of sample points, the smaller the contextual regions. The choice of contextual regions so as to exactly cover the image plane eliminates both the loss of information, as when the contextual regions do not completely cover the plane and some pixels are excluded from consideration, and

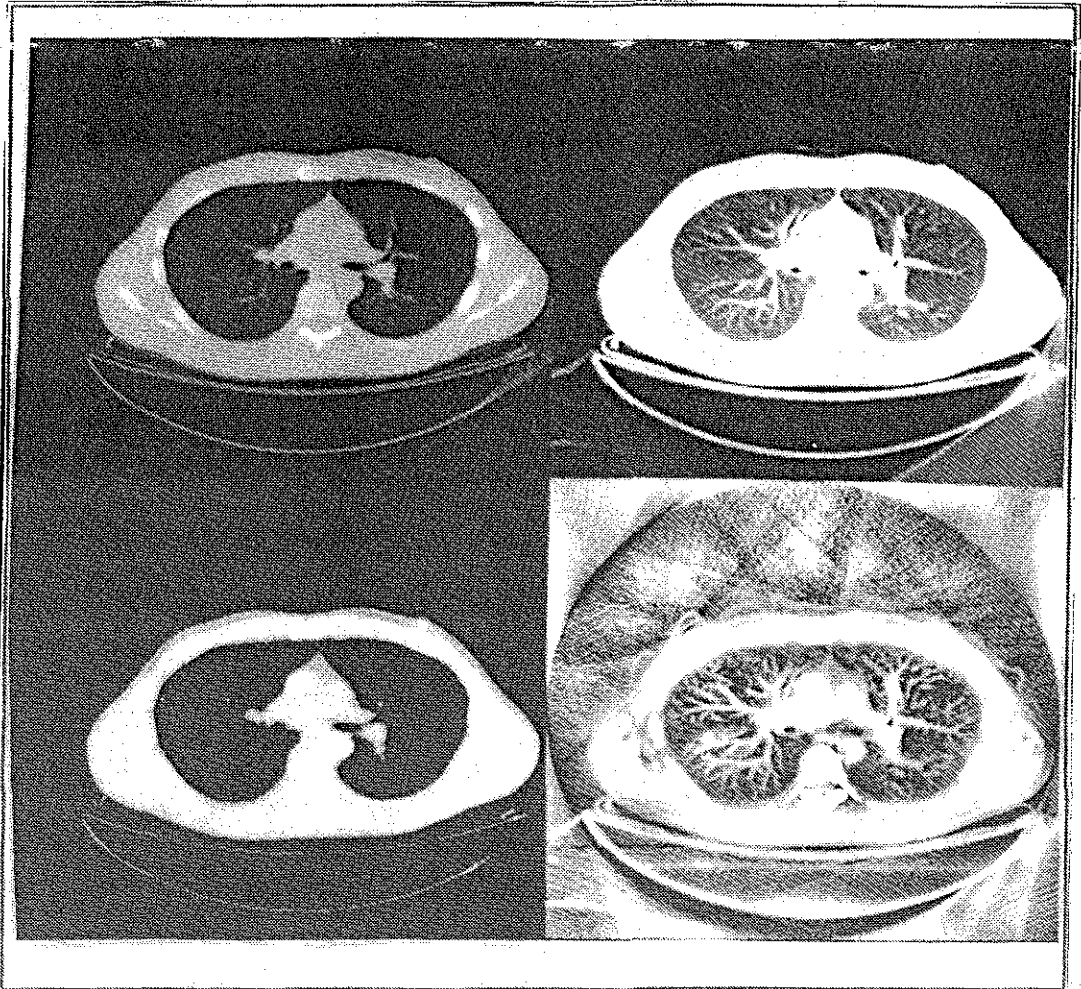


Figure 3.5: CT scan of the chest, 512×512 pixels.

Upper left: original image

Upper right: window to show lungs

Lower left: window to show mediastinum

Lower right: AHE image, 8×8 sample points.

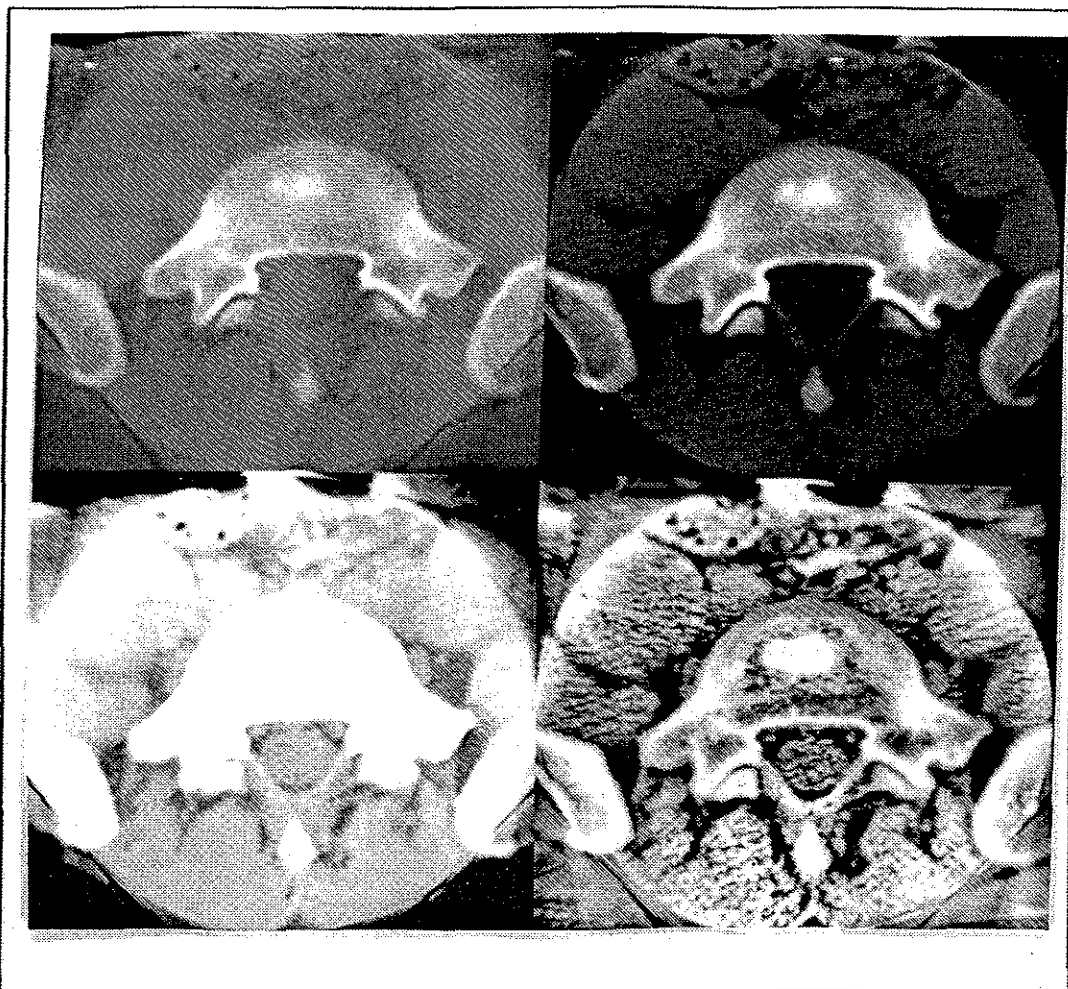


Figure 3.6: CT scan of the spine, 512×512 pixels.

Upper left: original image

Upper right: window to show bone

Lower left: window to show soft tissue

Lower right: AHE image, 8×8 sample points.

redundancy of information, as when the contextual regions overlap and some pixels are included multiple times. The amount of contrast enhancement performed tends to vary inversely with the region size, since smaller regions usually encompass less variation in intensity. Thus, any desired amount of contrast enhancement can be obtained by controlling the number of sample points used.

The results of AHE have proven not to be extremely sensitive to the number of sample points; the number must be effectively quadrupled (the spacing between points halved in each direction) to produce an effective difference in the amount of contrast enhancement. This has the result that a single choice of sample point spacing, an 8×8 grid, seems to be optimal for a wide variety of medical images. If fewer points are chosen (e.g., 4×4), the amount of contrast enhancement is not sufficient to show the fine detail in the image, whereas if a larger number is used (16×16), there is sometimes an objectionable breakup of image features. The fortunate result that a single choice of 8×8 sample points is effective for almost all images implies that in most cases, no manual intervention is needed in applying the AHE mapping. The relationship between the number of sample points used and the generation of artifacts will be discussed in a later section.

The effect of AHE on the contrast of the image is pronounced. In areas where there is effectively no signal (as in the background of CT scans), the noise is readily visible (see Figure 3.5, for example). This is not unexpected, since if there is no signal, the only contrast is that of the noise, and it is enhanced very well. However, in regions when there is signal present, then the signal and noise will both be enhanced.

In addition to the background noise in Figure 3.5, noise is also apparent in the body. A reasonable question is whether this increase in the noise visibility is detrimental to the observer's performance or is merely aesthetically unappealing. Burgess *et al.*, 1982 have suggested that a contrast enhancement which renders the noise more visible will not decrease the observer's performance provided that the contrast of the signal is also enhanced proportionately. They make two claims: first, the image should at a minimum be enhanced until the image noise is just visible. Second, further enhancement beyond this point does not harm and may improve the observer's performance, even though the noise may be quite apparent.

Their first claim is based on the argument that there are two sources of noise involved in the detection of features in an image. First, there is the statistical noise in the image; as the contrast of the image is enhanced, this will become more visible. Second, there

is the internal noise of the observer's visual system. If the external (image) noise is not apparent to the observer, then the internal noise will dominate the perception process. Thus, they argue, the observer's performance will certainly continue to improve if the contrast enhancement is increased until the external noise is greater than the internal system noise.

Their second claim uses results from signal detection theory, supported by observer experiments with simple targets, to suggest that the observer's performance will not be degraded even if the enhancement is further increased, so long as the noise is not enhanced more than the signal to be detected. These arguments suggest that the increased noise visibility in images processed with AHE is not detrimental, since both signal and noise are enhanced; however, there is some doubt whether the results of Burgess and his collaborators can be extended from the simple targets used in their experiments to the complex situations present in real images.

Adaptive Histogram Equalization has also been found to function effectively as a preprocessor for edge-enhancement algorithms such as the Sobel method. This result implies that AHE increases the contrast of edges relative to the remainder of the image.

3.2.2 Effect of Interpolation

The straightforward extension of histogram equalization uses a sliding window to calculate the values of the enhanced image. AHE instead uses the bilinear interpolation of a selected set of mappings to produce the new image values. A reasonable question then is whether AHE produces an inferior result due to this approximation and, if so, to what extent the interpolated method (AHE) is different from the sliding window version (LAHE). Experiments have shown that there is no qualitative difference in the two methods in almost all cases; the results of applying AHE are virtually identical to those of LAHE. However, the contextual region size necessary to produce a given result is different for the two methods. The region size in AHE is effectively four times greater than that of LAHE, since the mapping at a given pixel is influenced by the contextual regions of the four surrounding sample points. When the contextual region size in AHE is one-fourth that of LAHE, the resulting images are very similar, as is shown in Figure 3.7. This figure also shows the histograms of the two images; the result of applying AHE is again essentially identical to that of applying LAHE. Considerable experience has shown these results to be typical.

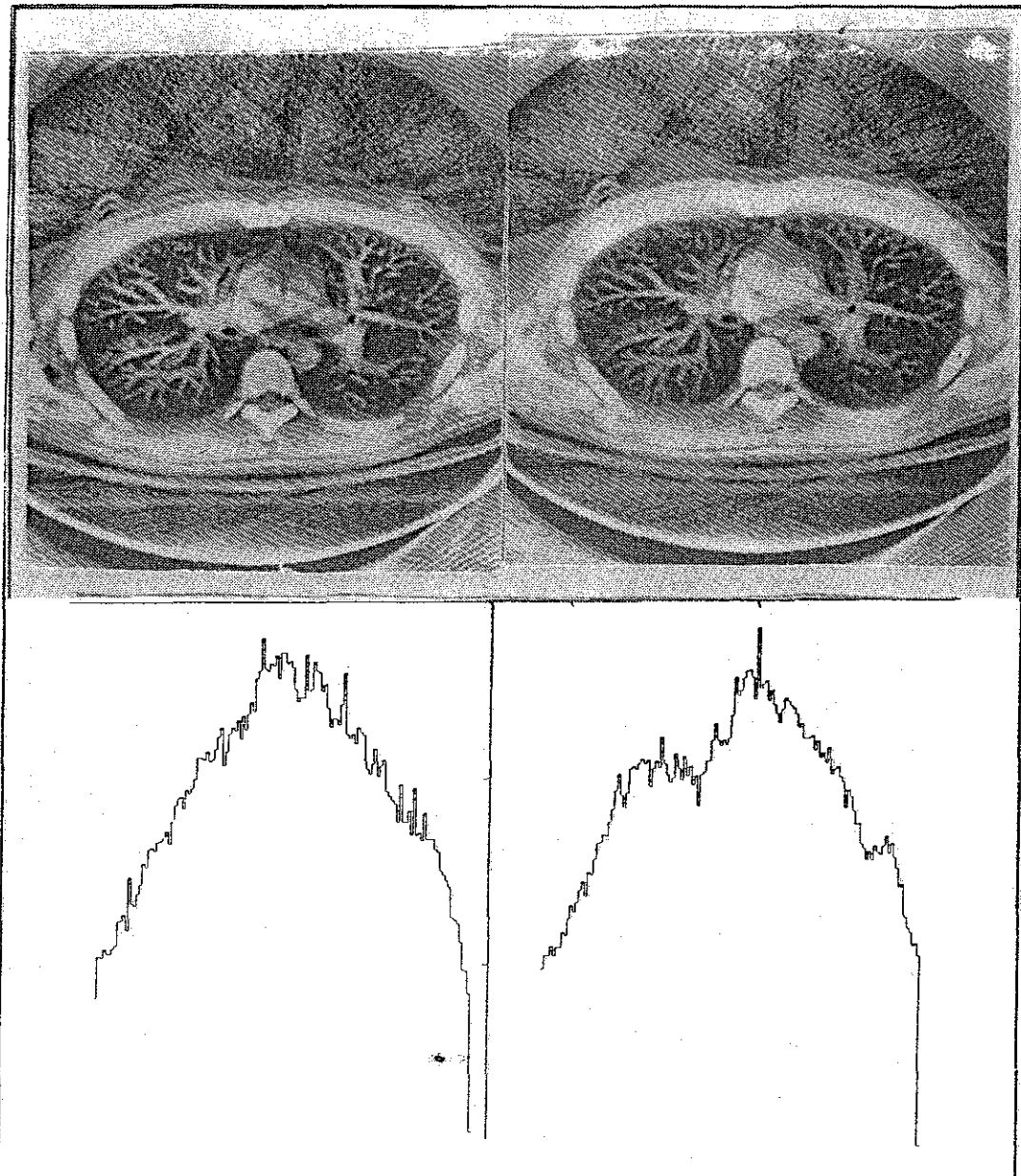


Figure 3.7: AHE compared with LAHE. The image is a chest CT, 512×512 pixels.

Upper left: AHE, 8×8 sample points

Upper right: LAHE, contextual region equivalent to 4×4 AHE

Lower left: Histogram of AHE image

Lower right: Histogram of LAHE image

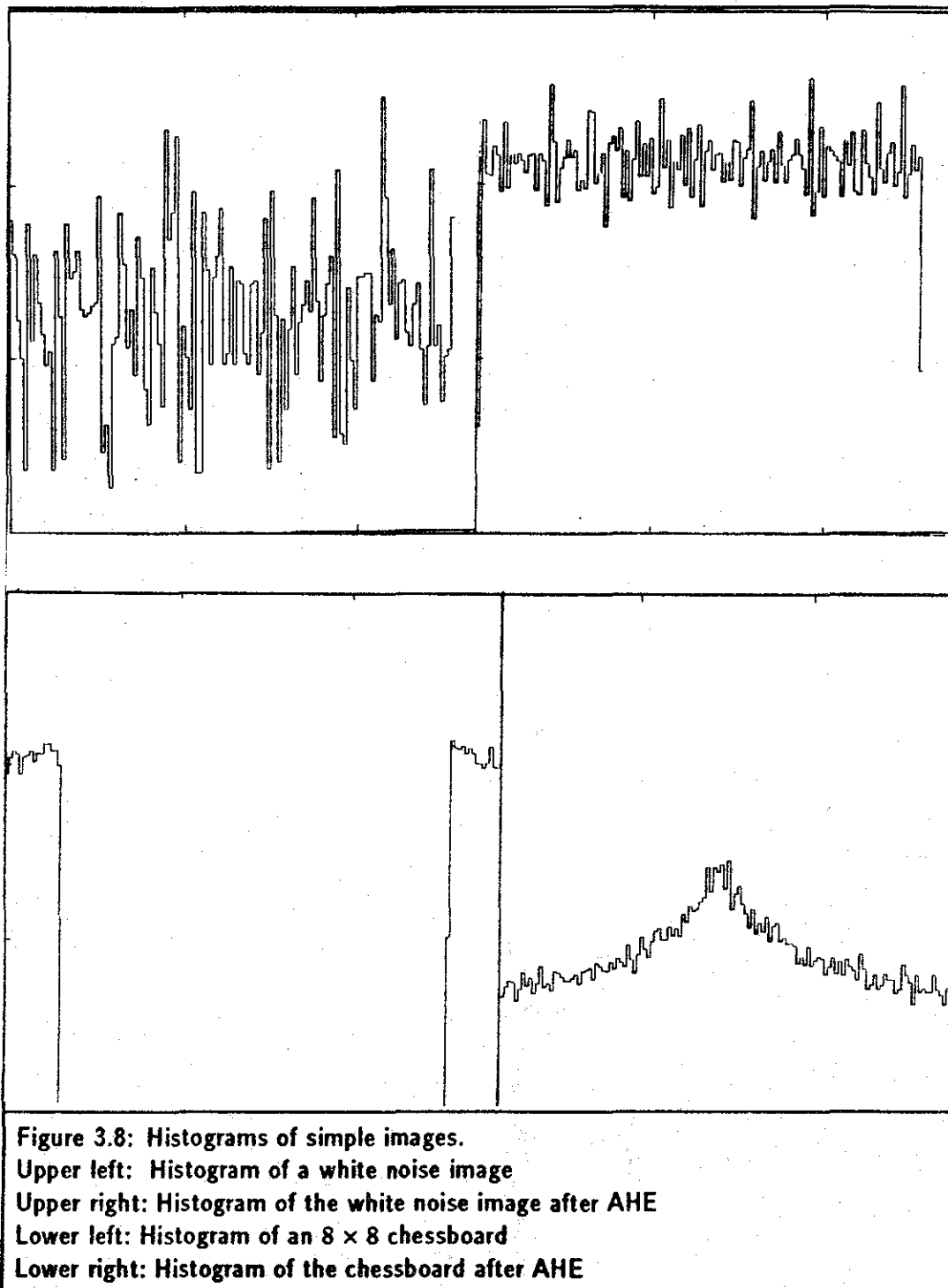
3.2.3 *Effects on the Image Histogram*

Global histogram equalization, when applied to an image, yields an image whose grey-scale histogram is approximately flat. However, for an image processed with AHE, the histogram will not necessarily be flat, since the equalization mappings are applied locally rather than globally.

Figure 3.7 shows the histogram of a typical AHE image. The histogram is peaked rather strongly about the middle intensities. This behavior is common in images treated with AHE, though it varies with the image type. Since a similar behavior obtains with LAHE, it can be assumed that this is not an artifact of the bilinear interpolation, but rather a consequence of the local nature of the histogram equalization and the structure in the image.

For images consisting of uniformly distributed random noise [Figure 3.8], it is found that the histogram of the image is essentially flat both before and after AHE. For a somewhat more complex image, a chessboard with random noise added, the peaking of the histogram is present, though not pronounced (again, Figure 3.8). For real medical images, the histogram of the processed image is peaked; however, these images can be divided into two classes. For images such as CT scans of the body and digital radiographs, the peaking was quite pronounced and increased as the contextual region size was made smaller [Figure 3.9]. For spine images, the peaking was less pronounced for all regions sizes, though it was still present. These results can be explained by looking at the histograms of the individual contextual regions.

Figure 3.10 shows histograms taken from single contextual regions in the images of Figure 3.5 and Figure 3.6. These regions were chosen to be typical of their respective images. For the chest CT, the regional histogram is unimodal with a sharp peak, whereas for the spine, it is much broader and bimodal. The resulting regional histograms after AHE retain these characteristics, with the locations of the peaks shifted and the histograms broadened. This is due to the discrete nature of the histogram equalization operation; the output intensities are not in fact evenly distributed. Since all pixels with a given input intensity are mapped to the same output intensity, the output histogram can have some intensities heavily populated, while others have no pixels assigned to them. This effect can be seen in Figure 3.10. The relation to global histogram equalization is clear; it is not possible to completely flatten the histogram of a digital image using the common method of global histogram equalization. This effect becomes more evident as the number of pixels in the region becomes smaller.



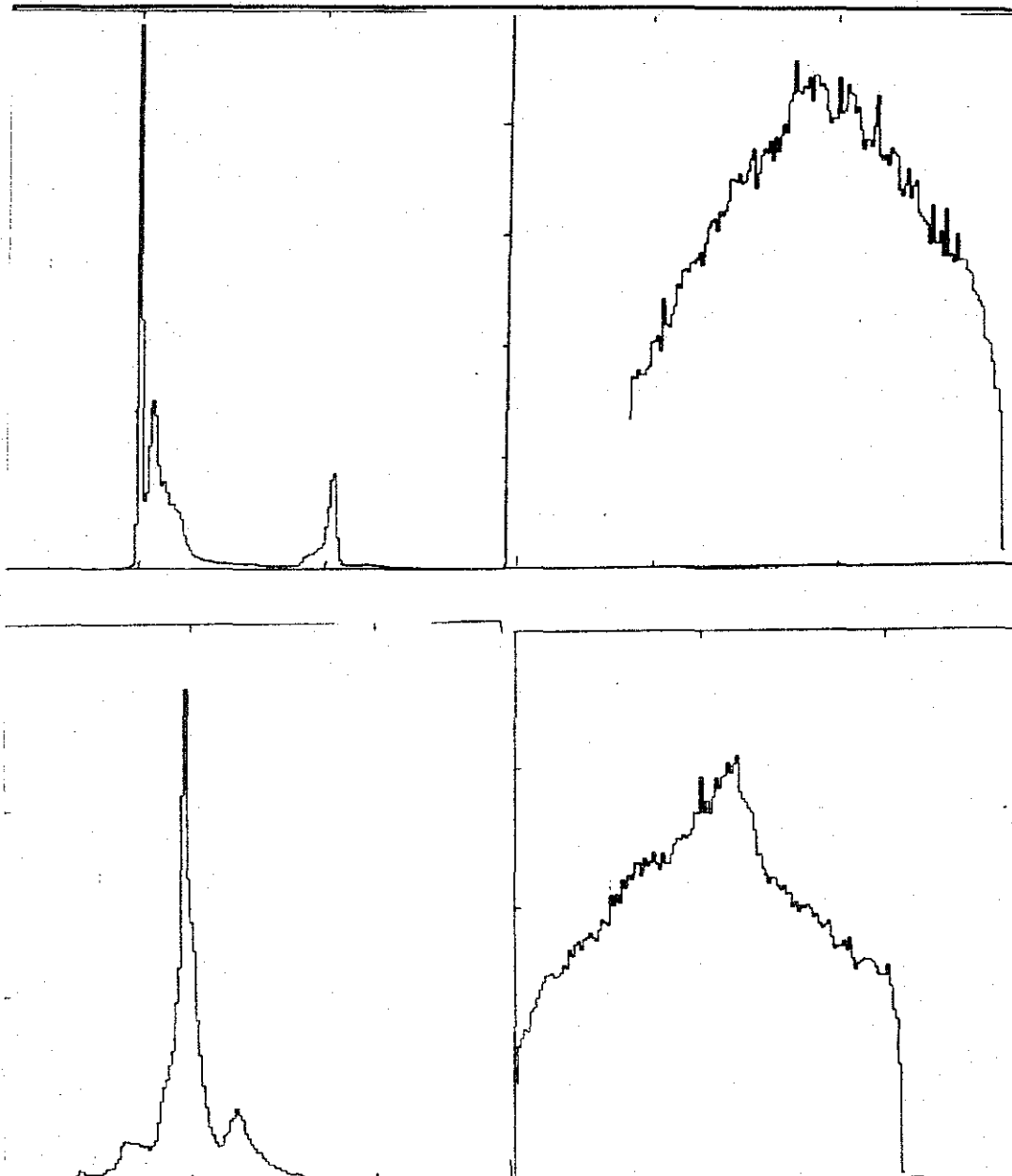


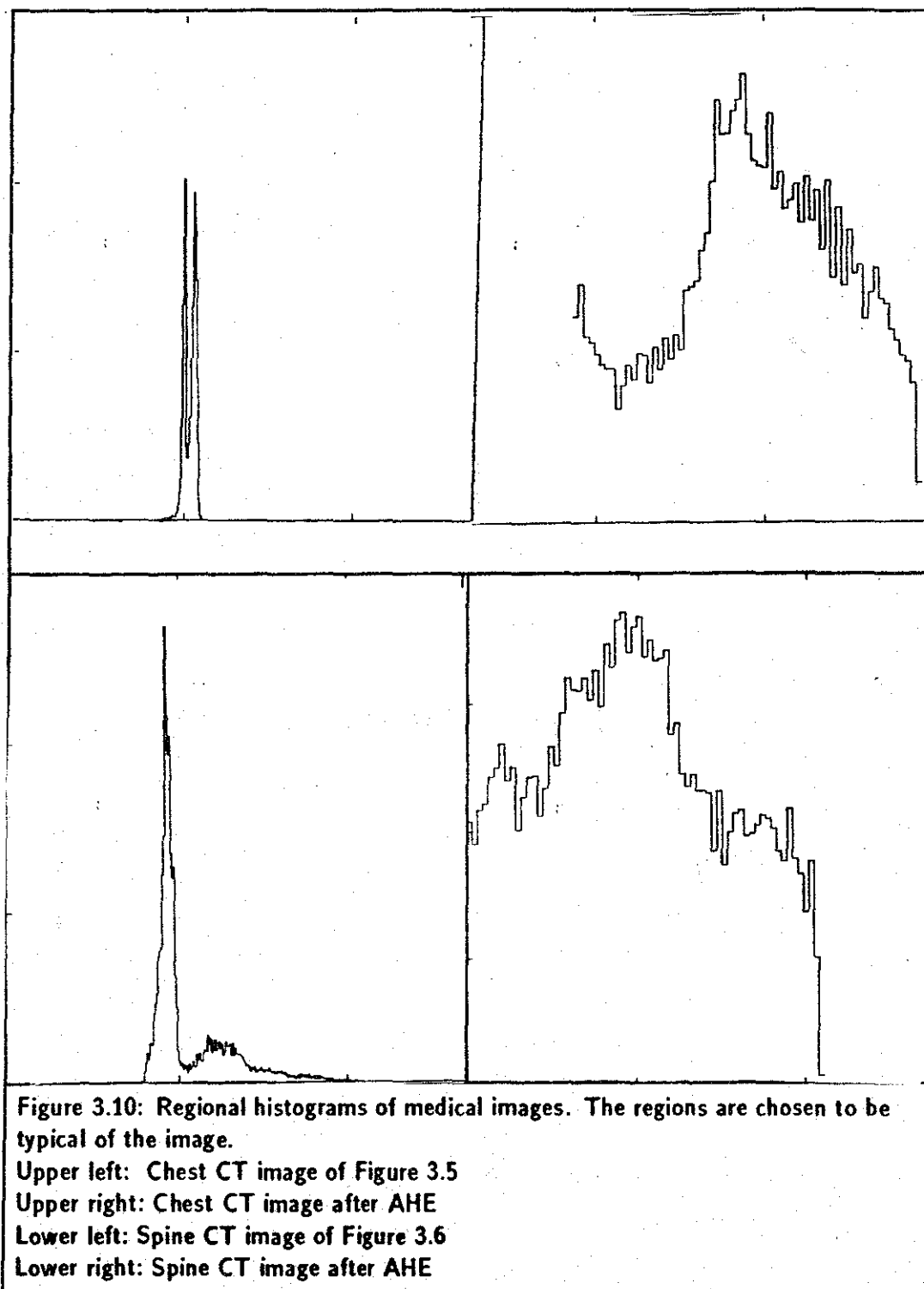
Figure 3.9: Histograms of medical images.

Upper left: Histogram of the chest CT image of Figure 3.5

Upper right: Histogram of the chest CT image after AHE

Lower left: Histogram of the spine CT image of Figure 3.6

Lower right: Histogram of the spine CT image after AHE



Thus, the output histogram of an AHE image can be predicted to some extent by the structure of the input image. For object with fairly uniform backgrounds (such as CT scans of the body) or large areas with low contrast (as in digital radiographs), the sharply peaked histograms of the regions will be retained in the final image. This effect becomes more pronounced as the contextual regions are made smaller, since the regional histograms become more strongly peaked. For images which do not have a uniform background or which have within a contextual region a considerable range of intensities (such as a CT scan of the spine), the peaking phenomenon will not be as pronounced, though it will be present and will increase as the contextual regions grow smaller, in accordance with observation.

Even though the histogram of the output image is not flat, it is not necessarily true that the intensity levels of the output device are not being well used. The peaking of the histogram is the result of processing in areas of the image where there is little signal initially, such as the background of a CT image; here, enhancement into too broad a range of the output device is of no benefit. The result would be a strong enhancement of the noise and an objectionable breakup of objects in the image. This can be seen to happen as the contextual regions grow small and very narrow histograms are mapped into broader peaks. The question of whether the histogram of the image should be adjusted again after applying AHE will be discussed later in this chapter.

3.2.4 Artifact Generation

The final question to be considered in evaluating AHE is whether it introduces artifacts into the image. Most of the methods reviewed in Chapter 2 have serious problems with ringing or overshoot artifacts when used on images with strong edges. AHE has proven to generate few artifacts, either in the spatial and the frequency domains.

Spatial Artifacts. Figure 3.11 shows a 6×6 chessboard to which uniformly distributed random noise has been added. The image was then enhanced using AHE with 6×6 sample points arranged so that the contextual regions in the image exactly overlay the squares of the chessboard. An overshoot effect is immediately apparent. This artifact is due to the interaction of the features (squares) in the image and the contextual region. This is not an artifact of the interpolation process, but rather of the histogram of the area surrounding each point. The effect is identical when LAHE is used, and it is perhaps easier to think of this artifact in that context. In LAHE, when a contextual region does not lie exactly on a square, the regional histogram is weighted toward those values which predominate in the region. Thus if the pixel is in a dark region, there are more dark pixels than light ones

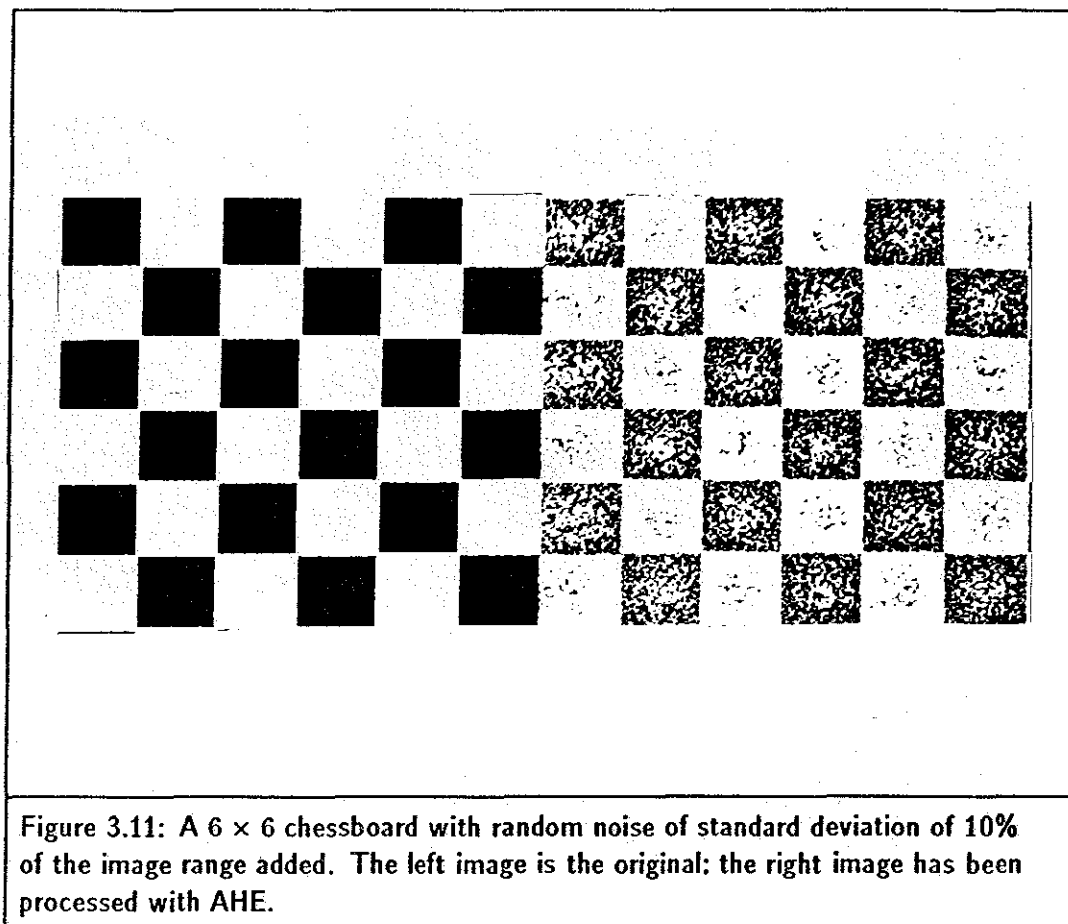
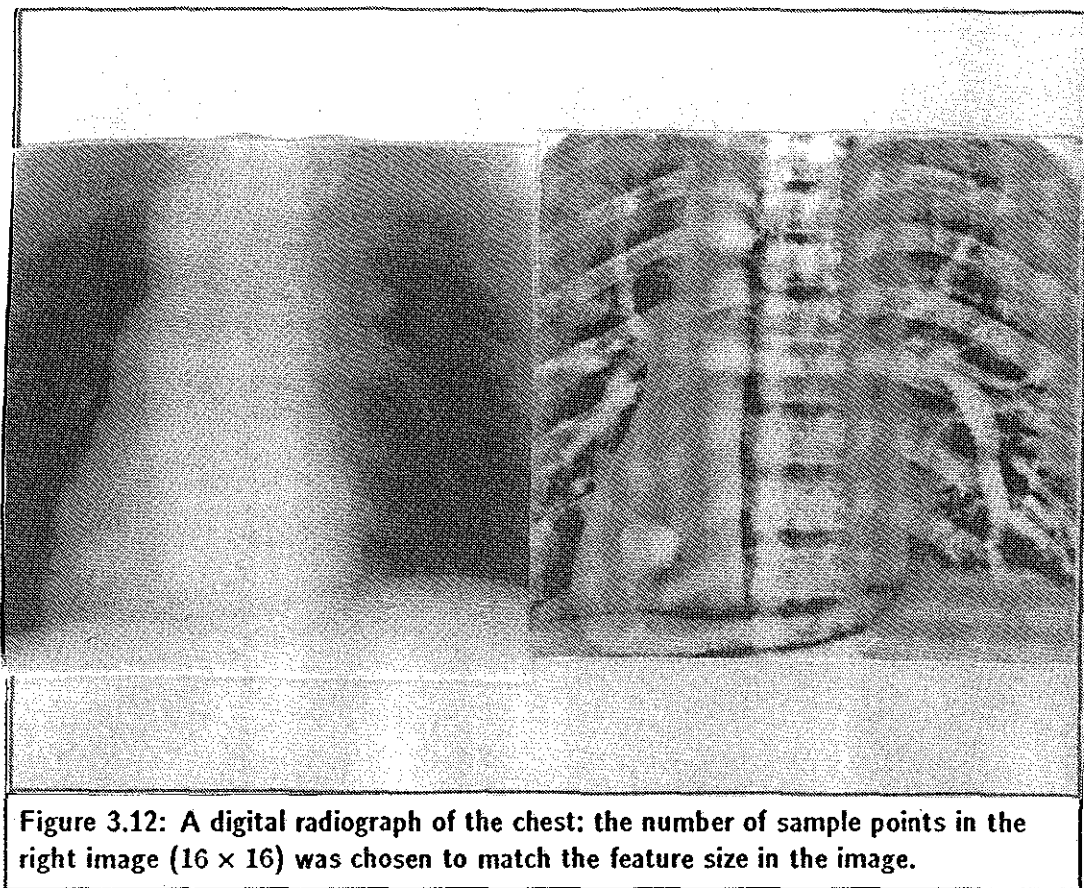


Figure 3.11: A 6×6 chessboard with random noise of standard deviation of 10% of the image range added. The left image is the original; the right image has been processed with AHE.

within the contextual region. This causes the value at the pixel to be pushed to the light end of the intensity scale by the histogram equalization. The same effect holds when the pixel is in the light region. This is similar to the well-known Mach band phenomenon in visual perception, though the mechanism is entirely different.

This artifact is only pronounced when the size of the contextual region is nearly that of the features in the images. If the contextual region is significantly larger than the features, the pixel value is less likely to lie at one of the extremes of the histogram. This overshoot artifact was first noticed in digital radiographs of the chest [Figure 3.12] when the contextual region sizes were chosen to match features in the image such as the size of the rib cage and heart. It can be seen in this figure that there are vertical striations in the image whose widths are approximately that of the contextual regions. Fortunately, this situation rarely occurs in practice. The size of objects of interest in CT scans is usually much smaller than the size of the contextual region necessary to produce adequate contrast enhancement; when the region size is chosen sufficiently small to produce ringing

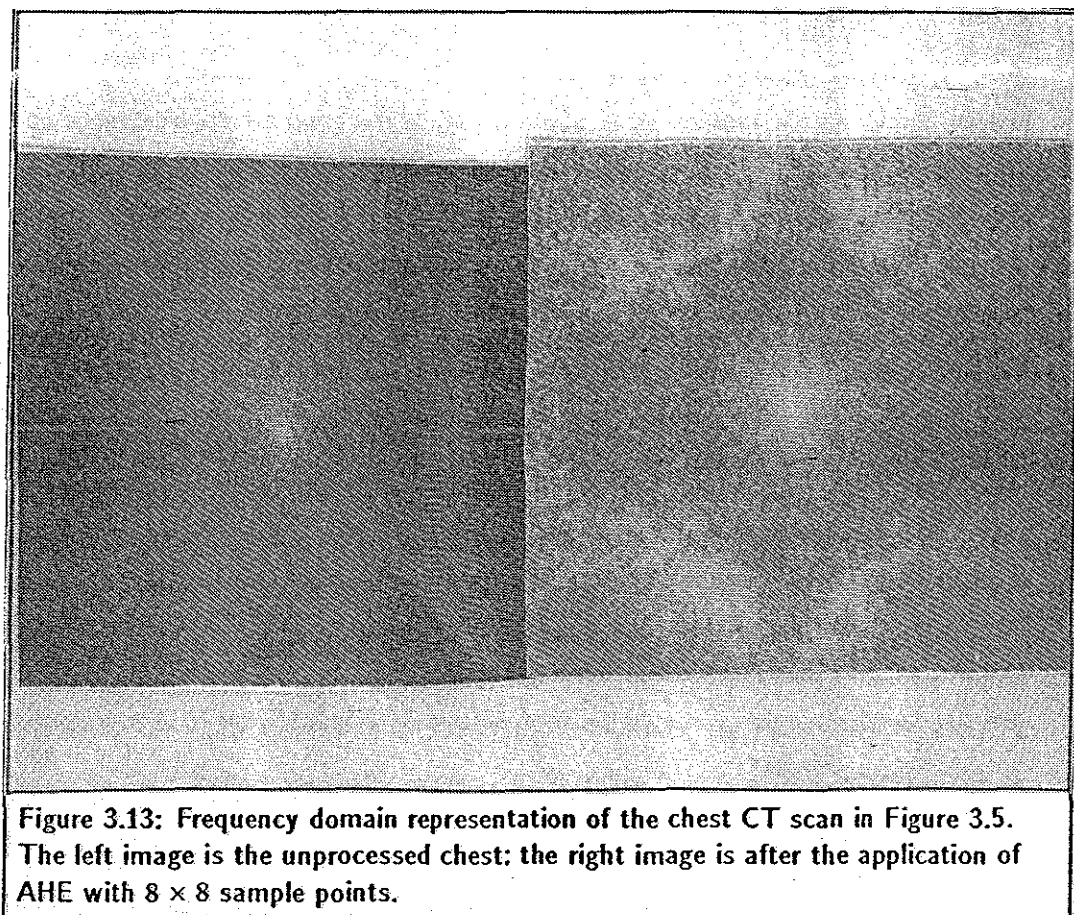


artifacts, the breakup of objects in the image is quite pronounced. The severity of this artifact has been found to be worse using AHE rather than LAHE in some cases, probably due to the effects of interpolation. No other spatial artifacts are known at this time.

Frequency Domain Artifacts. Figure 3.13 shows the frequency domain representation of the chest CT scan in Figure 3.5. The effects of applying AHE seem to be minimal in the frequency domain. There is a redistribution of energy, with the low frequency component attenuated. However, there is little evidence for the introduction of artifacts in the frequency image. There seems to be a smoothing effect in the high frequencies; the regular patterns characteristic of the correlated noise seem to be reduced.

3.3 Possible Modifications of AHE

There are a number of possible modifications of AHE which might improve the performance of the method. Some of these are examined below; for several, their effectiveness is still an open question.



3.3.1 Improved Interpolation Schemes

Bilinear interpolation is the simplest possible scheme for a smooth mapping between adjacent contextual regions. It might seem that a reasonable improvement of AHE would be to adopt a more elaborate interpolation scheme. However, as has been seen in the preceding sections, cases where the production of artifacts differ between AHE and LAHE are rare and occur only in circumstances where the contextual regions are quite small. Thus, AHE seem to perform almost identically to LAHE, which uses no interpolation. The attractiveness of other interpolation schemes, then, lies in the possibility of better performance as a contrast enhancement mapping rather than in the correction of artifacts introduced by the present method. There is little reason to believe that more elaborate interpolation schemes would provide improved performance, while the computational expense is likely to be greater.

3.3.2 Varying Contextual Regions

A related question is whether a different choice of contextual regions about the sample points would provide an increase in either the contrast enhancement or the computational performance of the algorithm. Two possibilities are to 1) have overlapping regions which cover some parts of the image more than once or 2) use regions which do not completely cover the image. The first alternative might provide a smoother transition between contextual regions, since the histogram mappings would not be independent. However, there are few cases where the present scheme is inadequate, and it is not clear that any improvement would warrant the additional computational burden of processing the redundant information.

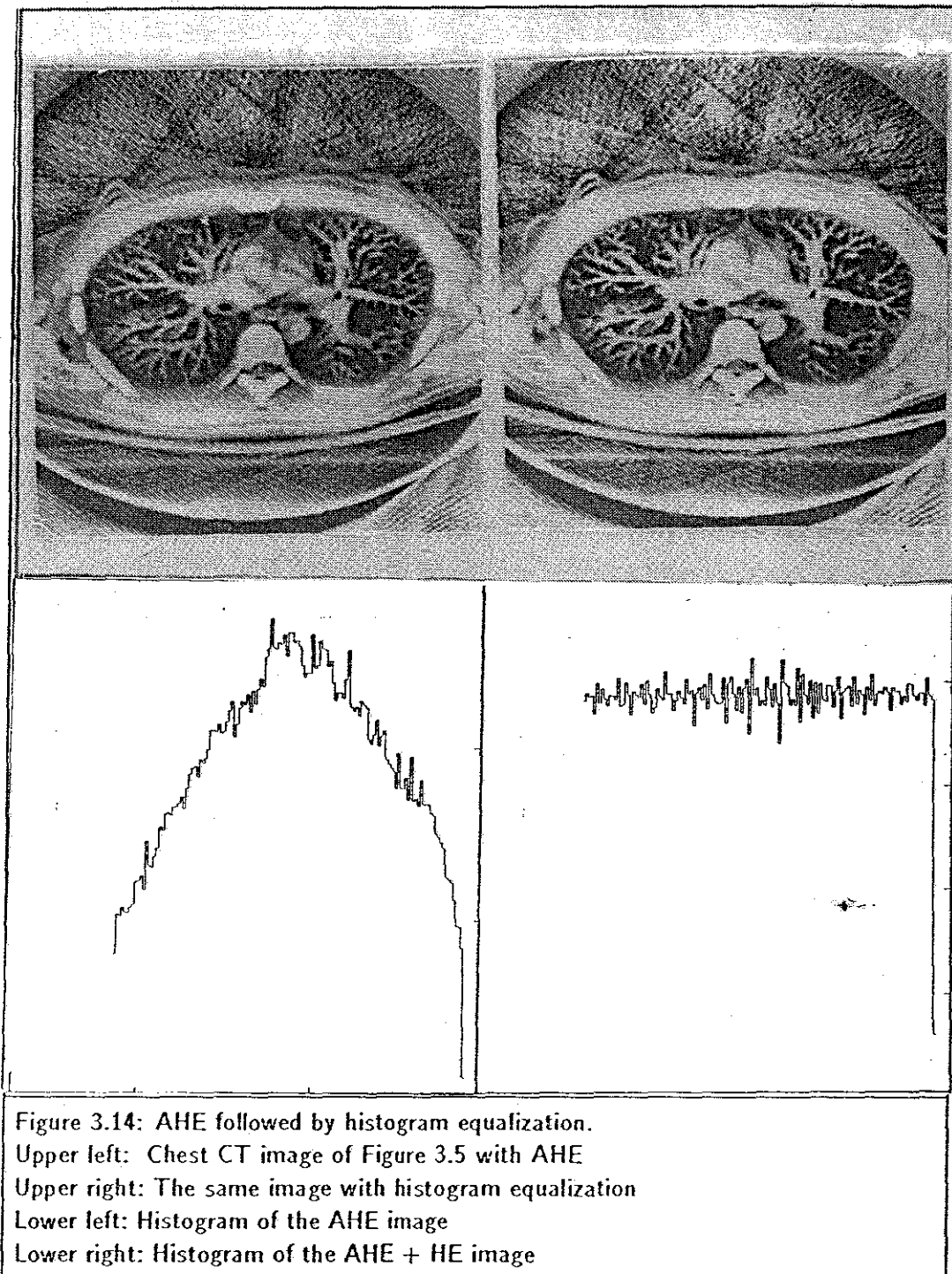
The second alternative offers the possibility of doing almost as well as the current scheme at a reduction in computation. While this method has not been investigated in this work, it seems probable that there is some trade-off point where the degradation in image quality would overwhelm the computational savings. Nevertheless, there is some probability that a savings could be effected at only a minor cost in image quality.

3.3.3 AHE Combined with Histogram Equalization

Another possible modification to the AHE algorithm is to combine it with the use of standard histogram equalization as a pre- or post-processor. As noted above, the histogram which results from the application of AHE is not flat; histogram equalization following AHE yields an image with an approximately flat histogram.

In the current investigation, it was found that using histogram equalization as a pre-processing step has little effect on the final image. This is not surprising, since the application of the global mapping gives no substantive change in the relative pixel intensities in small regions. The histogram resulting from the application of global histogram equalization followed by AHE was the same as that of AHE alone.

Figure 3.14 shows the results of applying histogram equalization after AHE. The resulting histogram is flat; however it is arguable whether the resulting image is improved. The overall contrast is somewhat greater, but there is a loss of detail in the lower intensities. This is to be expected from the earlier discussion; areas where there is signal are already well equalized. The flattening of the peaked histogram resulting from areas of low signal is done at their expense. Further evaluation, both theoretical and observational is in order



3.3.4 Summary and Conclusions

In this chapter, Pizer's Adaptive Histogram Equalization mapping has been examined in detail. The motivation for this method has its origins in information theory; its implementation is intended to make it practical for general use. It has been seen to be fast, automatic and reproducible, effective in producing contrast enhancement for all image objects, and generally free of artifacts. The problem which remains is to determine its relative effectiveness in presenting local contrast compared to standard global contrast enhancement techniques. If it can be shown that there is little or no loss of diagnostic information when AHE is used instead of more conventional techniques, the advantages of AHE outlined in this chapter suggest that it is a preferable technique for use in clinical situations. The determination of the relative diagnostic efficiency of AHE requires careful observer studies; these will be developed in Chapter 5. The problem with such studies are that they are extremely time-consuming and difficult. To evaluate the question at hand as well as to evaluate the possible modifications to AHE described above is a substantial task. A better alternative would be to develop an Image Quality Measure which would allow the evaluation of ACE methods without observer studies. The development of such a measure is the subject of the next chapter.

Chapter 4

A New Image Quality Measure

Quality ... you know what it is, yet you don't know what it is. But that's self-contradictory. But some things are better than others, that is, they have more quality. But when you try to say what quality is, apart from the things that have it, it all goes poof! There's nothing to talk about. But if you can't say what Quality is, how do you know what it is, or how do you know that it even exists? If no one knows what it is, then for all practical purposes it doesn't exist at all. But for all practical purposes it really does exist. What else are the grades based on? Why else would people pay fortunes for some things and throw others in the trash pile? Obviously some things are better than others ... but what is "betterness"? ... So round and round you go, spinning mental wheels and nowhere finding any place to get traction. What the hell is Quality? What is it?

—Robert M. Pirsig [1929—]

In this chapter, a new measure of image quality is proposed which is based on the properties of human visual perception. In the first sections, the problems of deciding what is meant by quality in a digital image and of developing an approach for the measurement of quality are considered. The remainder of the chapter describes the new image quality measure.

4.1 What is Quality?

As suggested by Pirsig, the concept of quality is a slippery one. In the current work, a pragmatic approach to the definition of quality in a digital image has been taken. As stated in Chapter 1, we define that image as best which is most utilitarian. It is assumed that the observer has a definite task to perform and the use of an image of higher quality will improve the observer's performance. While this definition is intuitively satisfying, it nevertheless has some difficulties when used to evaluate specific image improvement techniques. It requires that the task at hand be narrowly defined, precluding a statement about the relative values of processing techniques across a general range of tasks. It takes

no account of the costs related to an operation performed on an image, and it ignores the aesthetic issues of quality which, while difficult to quantify, are nevertheless of importance to those who must deal with the images presented.

In spite of these drawbacks, there are considerable benefits to be gained by attempting to formulate task-oriented measures of image quality, particularly in applications such as medical imaging where the quality of diagnosis is paramount. Thus, the current work focuses on the question of observer performance within a specific context, that of detection of contrast in medical images.

In evaluating both the previous approaches to the definition of image quality and a new definition, it is helpful to formulate the general criteria which would ideally be met by an IQM. First, the measure should be generally applicable; it should apply for a wide variety of realistic tasks. Second, it should be analytic rather than observational in nature in order that it be usable without the necessity of involving human observers. It is not currently possible to satisfy both of these criteria completely; our understanding of the nature of the decision making process of the human observer is insufficient. In the next section, three approaches to the definition of image quality are examined; the ways in which each of these falls short of the ideal will be shown.

4.2 Previous Approaches to Image Quality

Numerous investigators have attempted the definition of measures of image quality. No attempt has been made here to exhaustively survey this field; rather, three general approaches to the topic will be examined. An interesting overview of the subject is given by Todd-Pokropek, 1976; he considers quality measures in two categories, those based on physical models of processing and those based on signal detection theory. A third category is added here for those measures based on visual psychophysics.

4.2.1 Physical Models

The first approach to defining image quality measures is to define a physical model of the imaging process. A simplified representation of the process of image generation or improvement is constructed and then the various parameters of the model used as an index of quality. Applicable parameters might be the spatial or intensity resolution of the imaging process being modeled, the signal to noise ratio, or the degree of contrast enhancement. Since the value of any one parameter will not reflect the total performance of the processing techniques, several parameters are often combined additively or multiplicatively with appropriate weights to form *figures of merit* whose value presumably

reflects the quality of the image. This approach originally arose from a desire to evaluate the quality of different imaging devices, but has been extended to the evaluation of processing techniques as well. Often, the figure of merit will be determined through the use of simulations or phantom studies.

This model has yielded good results in many instances where it has been applied and has the advantage that it can be applied analytically. However, it has a number of drawbacks. First, because the images on which the parameters are based are often very simple, it is difficult to generalize the results. Second, the formulation of the figure of merit requires an evaluation of the relative importance of the various physical parameters of the image in order that proper weights be assigned. Finally, it is not obvious that the parameters measured are the important ones involved in performing a task on a given image. For example, the signal to noise ratio may be evaluated in areas of the image where there is no information of importance.

4.2.2 Signal Detection Theory

There are two principal ways in which signal detection theory has contributed to the understanding of image quality and the formulation of image quality measures. First, signal detection theory has been used to examine the fundamental limits of the performance of the human visual system. Second, it has been used to allow the analysis of psychophysical experiments done using human observers.

The first approach considers the theoretical limits of object detection in situations where there is random noise. For certain simple situations, such as the embedding of bars or disks in noise of varying statistics the optimal strategy for deciding whether an object is present or not can be determined. An "ideal observer" operating with this strategy would use all the information at his disposal as well as is theoretically possible in reaching a decision. This work is reviewed in the paper by Burgess, 1983.

Burgess has compared the results that would be obtained by the ideal observer with the performance of human observers. He finds that there is little difference in the case of uncorrelated noise (the human requires a factor of about $\sqrt{2}$ larger signal to noise ratio for the same performance), though the ideal observer is able to deal better with the presence of correlated noise by using a re-whitening procedure of which the human is evidently incapable. His conclusion is that many of the global approaches which have been used in the past to improve images, such as sharpening or smoothing of the image, are of little efficacy, since the human observer's performance is very close to the ideal in any case. If

it is true that the human performs nearly as well as the ideal observer in many cases, an image quality measure could be formulated by calculating the performance of the ideal observer for sets of images processed in different ways and using this performance as the quality measure. Such a measure would be both general and analytical.

The difficulty in this scheme is that it is not clear that these results can be generalized beyond the simple situations used in Burgess's experiments to images which have objects embedded in complex backgrounds. In the simple situation of a known signal embedded in a uniform background of random noise, the optimal strategy of the ideal observer is to do a matched filtering operation. However, for real images with complex targets embedded in structured backgrounds, the optimal strategy is not always obvious.

A second use of statistical decision theory in determining image quality is to analyze the results of psychophysical observer experiments. The use of ROC (receiver operating characteristic) curves allows the discrimination of different processing methods for a given task; this approach is very powerful and is the basis for the experiment to be described in Chapter 5. Observer experiments have the advantage that they measure directly the quality of images produced by different processing techniques for a given task; however, it is very difficult to generalize the results to arbitrary tasks, and the results are not analytic.

4.2.3 Visual Psychophysics

Since the ultimate arbiter of image quality is the human observer, a third way to approach the definition of a quality measure is to determine, either analytically or through the use of experiments, how the human visual system will react to a given image. If the quality measure is to be analytic, then some model of the visual system must be constructed. A few attempts have been made to construct such a model; the difficulty is that current knowledge does not permit a complete description of the way the visual system will react to a given stimulus. Nevertheless, this approach is attractive as our knowledge of the visual system improves; it is the basis of the IQM which will be described in the remainder of this chapter.

4.3 Overview of the Image Quality Measure

This section describes the new Image Quality Measure; it first gives an overview of the approach which was taken and then describes in some detail each piece of the measure. It concludes with a discussion of some of the strengths and weaknesses of the current approach.

4.3.1 General Approach

The current IQM is based on two premises: first, the IQM must be probabilistic in nature and second, it should be sensitive to the same properties of contrast as the human visual system.

Probabilistic Requirements. There are two ways in which the IQM must be probabilistic. First, it should average the quality measure over an ensemble of images. It is unlikely that a single measurement of the IQM will give a satisfactory evaluation of a contrast enhancement technique. Second, since the detection of contrast in an image is a stochastic process, it should consider and weight the probability of different possible answers which the observer might give.

Model of the Human Visual System. An approach based on a model of the visual system has the following advantages:

- 1) If the model is good, it is likely to be well-correlated with the performance of real observers, unlike approaches which consider only the physical parameters of the image.
- 2) It avoids the problem of weighting too heavily information in the image which is unrelated to the detection problem; for example, an often used technique is to compare the sum of squares for the difference of the original and processed image. This approach ignores the fact that much of the contribution to this quality measure comes from areas of the image where there is no significant information. On the other hand, the current approach allows the quality measure to be tailored to the features in the image to which the eye is sensitive, for example the presence of edges.
- 3) The quality measure is likely to improve as our knowledge of the visual system becomes greater. If it is well designed, the quality measure can be modified easily to incorporate this new knowledge.
- 4) It allows the testing of our understanding of our knowledge of the visual system. If the quality measure proves not to correlate well with observer performance for a given model, it is likely that the model emphasizes features in the image which are unimportant to the visual task at hand.

Assumptions and Restrictions. In the current work, it is assumed that the task which the observer is asked to perform is restricted to the detection of luminance contrast; thus, the IQM considers models of the visual system which are sensitive to such contrast, in particular the presence of edges in the image. Further, it is assumed that the signal to be detected is known exactly by the observer. This condition is satisfied by the observer experiments to be described in Chapter 5; in that chapter the nature of such signal known exactly (SKE) experiments is further described.

Problems With This Approach. The bases on which the current IQM is constructed have a number of drawbacks. First, because the assumed task is so strictly defined, it may be difficult to generalize this IQM to other tasks. As will be explained in Chapter 5, there is evidence to suggest that the performance of an observer in a detection experiment is well correlated with that in other types of experiments such as localization and recognition experiments, so this drawback is not as serious as it might appear. However, it is clear that the signal known exactly experimental paradigm represents only a small part of the general task required of human observers.

Second, it is clear that our current knowledge of the visual system is inadequate to completely model the detection of contrast. We are able only to model in any detail the early stages of the contrast detection ability of the eye; therefore, it is probable that important visual processes will be neglected by the IQM. This is related to the problem of cognition of the "inner screen" in visual perception; the IQM produces a number as a measure of quality; even if this number reflects well the processing of the early visual system, it may not reflect the information used by the later visual system in forming decisions.

The inadequacy of our current knowledge is a serious drawback in that it is possible that the features in the image that seem to correlate with contrast detection ability may be badly selected, thus implying that the current approach is inadequate when in fact it may be that only a better selection of features is needed for the IQM to be of utility.

4.4 Definition of the IQM

The proposed IQM is defined in the following way. Let I be an image which may contain some target of interest, ξ . The joint probability $P(D_i, T_j)$ is the probability that the situation T_j holds in the image and the decision by the observer is D_i . The set $\{T_j\}$ is defined to be the set of all possible true situations in the image; for example, in a detection experiment, $\{T_j\} = \{\text{signal present, signal absent}\}$. The set $\{D_i\}$ corresponds to

the possible decisions that the observer might make about the image given the possible true situations; again in a detection experiment, $\{D_i\} = \{\text{signal present, signal absent}\}$. Thus, for each T_i there is a corresponding D_i . The IQM is then defined to be the probability that the observer will make a correct decision, that is,

$$IQM = \sum_i P(D_i, T_i). \quad [4.1]$$

Let $\{I_n\}$ be some set of images such that in each image, I_n , exactly one of the T_i holds; then using the rules of elementary probability, the expression for the joint probability can be rewritten in the following way:

$$\begin{aligned} P(D_i, T_i) &= \sum_n P(D_i, T_i, I_n) \\ &= \sum_n P(D_i|T_i, I_n) P(I_n|T_i) P(T_i) \end{aligned} \quad [4.2]$$

If it is assumed that each image I_n in a set of N images, $\{I_n\}$, from situation T_i is equally likely, then $P(I_n|T_i) = 1/N$; if there are Q possible true situations and if each situation is presumed to be equally likely in a given image, then $P(T_i) = 1/Q$. The expression for the IQM then simplifies to

$$IQM = \frac{1}{NQ} \sum_{i,n} P(D_i|T_i, I_n). \quad [4.3]$$

The conditional probability $P(D_i|T_i, I_n)$ is the probability that the observer makes the decision D_i is the situation where he is shown an image I_n in which the situation T_i holds. This conditional probability will be based on the properties of the human visual system.

Advantages and Disadvantages of This Formulation. This definition has a number of good features. It isolates the objective portion of the model from that dependent on the model of the visual system, allowing easy modification of the various model components as better formulations are developed. It is possible to calculate the IQM without the use of observer studies. However, the requirement that a *priori* knowledge be used is a disadvantage in that it appears to be impossible to apply this measure when the actual characteristics of the target are unknown. This difficulty is not too severe, since there is no requirement that the target be of any particular form, merely that the form be known in advance. Furthermore, there is some evidence to suggest that the process of contrast detection is iterative; feedback in the later visual system causes the early visual system to look for objects with the same form as something which has been tentatively located. Therefore, the IQM is similar to the known behavior of the visual system in that it attempts to measure the detectability of a target which it expects to see.

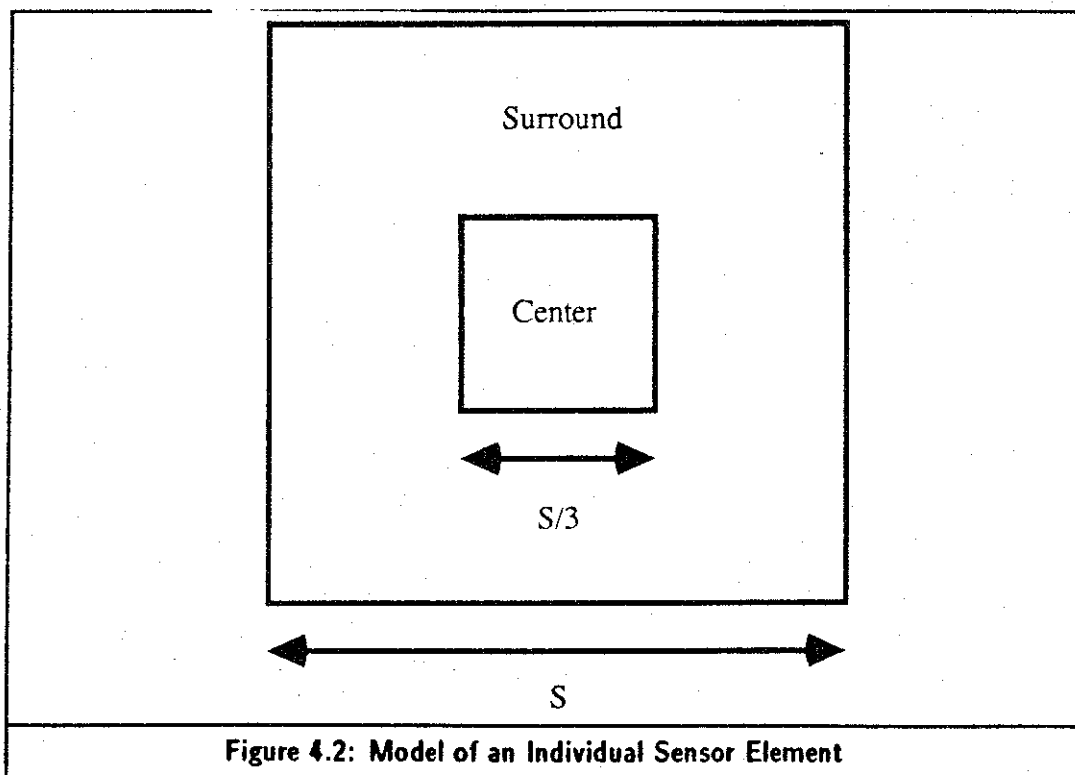
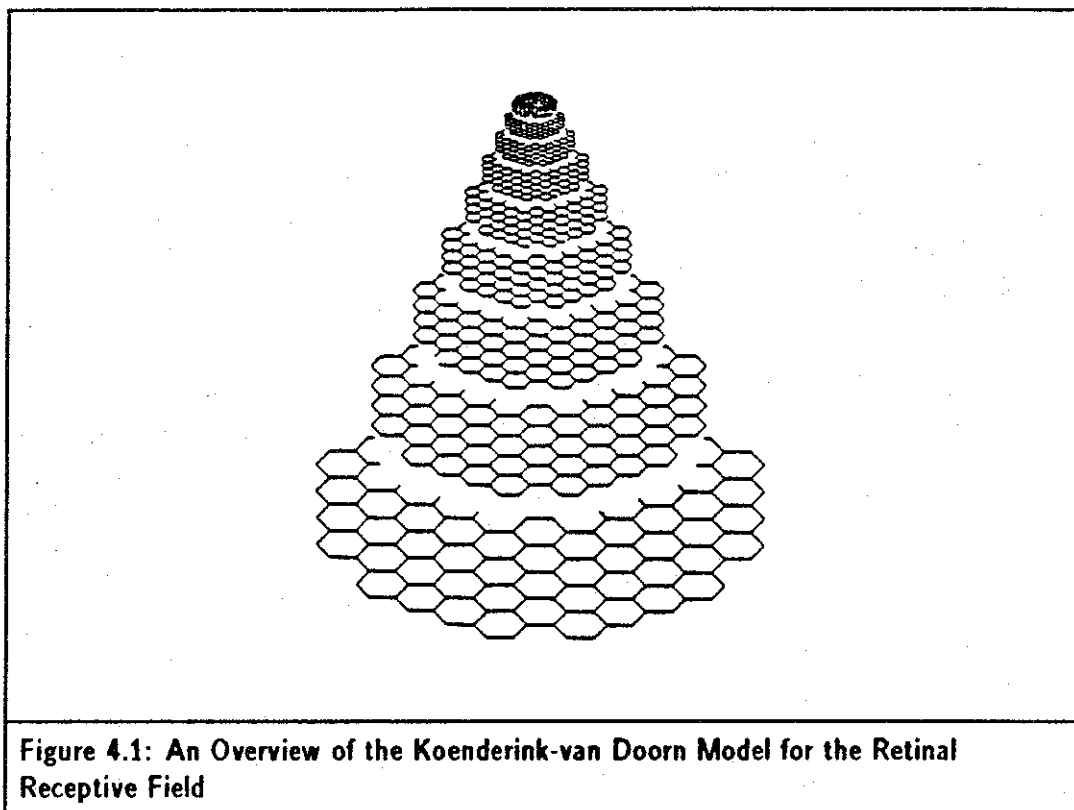
4.4.1 Model of Visual Contrast Detection

The conditional probability $P(D_i|T_i, I_n)$ is the portion of the measure which is dependent upon modeling the early processing of spatial contrast information within the visual system. The approach has been to look at those elements of contrast detection which are indicated by psychophysical and neurophysiological experimentation to be most important in contrast perception. In particular, two important components of spatial contrast detection have been incorporated into the model. First, contrast detection is known to operate at multiple scales of spatial resolution. For example, the modulation threshold for contrast detection seems to be approximately independent of the spatial extent of a given object over a considerable range of sizes; thus a model of contrast detection ability should have a similar scale invariance. Second, the retina's detection of spatial intensity appears to have a bandpass characteristic which causes it to be sensitive to the presence of edges in the intensity field.

A promising model for explaining contrast detection ability in the early visual system is the self-similar sensor array approach of Koenderink and van Doorn, 1978. In their theory, the retina is modeled as an array of sensory receptive fields which have the two desired characteristics: first, the model is scale invariant due to the use of arrays of sensor elements of different sizes which respond to features at many scales of resolution. Second, each individual sensor element has the required bandpass characteristic making it sensitive to the presence of edges at a given scale of resolution.

An overview of the model is shown in Figure 4.1. The retina can be considered as a series of overlying planes of sensor elements. The sensor elements of each plane have a characteristic size such that each succeeding plane can be considered as a simple magnification of the preceding plane. This structure allows each plane to respond to the features in the retinal illumination pattern whose sizes are similar to the sizes of the sensor elements in that plane.

Description of the Sensor Elements. In describing the Koenderink-van Doorn model, the characteristics of the individual sensor elements are considered first, followed by an explanation of the overall structure of the sensor array. In Figure 4.2, the sensor can be seen to consist of a central region with an outer surround. The sensor is characterized by its linear size, S . The sensor is presumed to be sensitive only to the total flux of the illuminance falling on it, with the sensor response determined by the ratio of the flux on the center region of the sensor to the total flux falling on the entire sensor (both the center and the surround). This organization is similar to the well-known on-center, off-surround



receptive field which is commonly attributed to the presence of lateral inhibition in the early visual system.

Let φ_c be the flux through the center of the sensor and φ_t be the total flux through the center and surround. Then in the presence of an illumination pattern, the sensor produces an output of the form

$$R_S = \begin{cases} 0, & \bar{\varphi}_t \leq \bar{\varphi}_z; \\ \frac{\varphi_c}{k\sqrt{\varphi_w\varphi_t}} - 1, & \bar{\varphi}_z < \bar{\varphi}_t < \bar{\varphi}_w; \\ \frac{\varphi_c}{k\varphi_t} - 1, & \bar{\varphi}_t \geq \bar{\varphi}_w. \end{cases} \quad [4.4]$$

There are three regimes of operation for the sensor. When the flux per unit area over the entire sensor ($\bar{\varphi}_t$) is less than some threshold amount $\bar{\varphi}_z$, the sensor does not respond at all. When the flux is greater than this but less than some threshold $\bar{\varphi}_w$, the response is proportional to the inverse of the square root of the total flux (de Vries-Rose behavior). Finally, when the average flux exceeds the Weber-Fechner threshold, $\bar{\varphi}_w$, the sensor responds with a Weber-Fechner behavior. The constant k scales the total flux according to the relative area of the center and total area of the sensor; if A_c is the area of the center of the sensor and $A_t (= S^2)$ is the total sensor area, $k = A_c/A_t$.

Given this definition of the response function, it is possible to predict the sensor response to various illumination patterns. For a flat field or a linear ramp function, φ_c will be equal to $k\varphi_t$, so the sensor output will be zero in the Weber-Fechner regime, in accordance with expectation. For an edge or local intensity extremum of about the same size as the sensor, there will be a response whose magnitude is dependent on the intensity of the edge and the positioning of the feature relative to the sensor center. This behavior implies that there is a response to features whose size is similar to that of the sensor, but that the sensor is relatively insensitive to features much larger than itself. The sensor is also insensitive to features much smaller than itself since small features will not cause an appreciable change in the ratio of the fluxes in the center and surround of the sensor.

The choice of the threshold $\bar{\varphi}_w$ for the transition of the sensor behavior is made according to the criteria given by Zuidema, 1984. Results cited in his paper indicate that the threshold for the change from the de Vries-Rose behavior to the Weber-Fechner behavior occurs for spatial frequencies of about 4 cycles per degree when the background and ambient illumination of the image exceeds 65 foot-lamberts. In the current work, the spatial frequency of the targets varies between about 3 and 8 cycles per degree; the

background illumination does not exceed 30 foot-lamberts and is on average about 15 foot-lamberts, so the response function operates entirely within the De Vries-Rose regime. The threshold for zero response was set at about 0.25 foot-lamberts to avoid artifacts due to the large relative error at low intensities.

4.4.2 Self-Similar Receptive Field Arrays

In the Koenderink-van Doorn model, each sensor is characterized by its linear size, S ; the retina is assumed to be covered by overlapping mosaics of sensors of varying sizes. The requirement we wish to place on the distribution of sensor sizes is that the response of the model should be scale invariant over a wide range of sizes; within this range an arbitrary spatial rescaling of the retinal illumination pattern should not change the total output of the retinal sensors. As explained in Koenderink and van Doorn, 1978, this can be accomplished in the case of a continuous distribution of sensor sizes by requiring that the number of sensors of size S per unit area of the retina be given by the density function $\mu(S) = K/S^3$, where K is some arbitrary constant. It can be shown that a group of sensors obeying this density function with sizes lying between $(S, S + \Delta S)$ will tile the plane without gaps for $(\Delta S/S) = (1/K)$. Thus, at any point in the retinal plane, there are sensors present at many different sizes; the characteristic size of the features at that point will cause sensors of the appropriate size to respond. In the Koenderink-van Doorn model, there is a restriction on the distribution of the sensors of a given size; the smallest sensors are restricted to the center of the retinal plane. The size of the smallest sensor at a given retinal eccentricity, ϵ , increases with increasing ϵ . This restriction will not be considered further here because it will be assumed that the targets to be detected are in or near the fovea; in this region it is assumed that the retina is homogeneous. Physiological considerations also demand that there be a smallest and a largest possible sensor (S_{\min} and S_{\max}).

For the case of a discrete distribution of sensor sizes, scale invariance can be achieved in the following way: the retina is considered to consist of multiple overlying planes of sensor elements; in each plane, the size of the sensor elements is constant. Assume that the planes are in order of increasing sensor size. Then for scale invariance to hold, the size of the sensors in plane $i + 1$ must be some constant multiple of the sensor size in plane i :

$$S_{i+1} = M \times S_i. \quad [4.5]$$

The choice of M and the maximum and minimum sensor sizes allowed determines the number of different planes of sensors. Within each plane, the sensors of size S_i are assumed

to be arranged in a regular array such that their center regions exactly tile the plane, with their surround regions allowed to overlap.

Given these sensor arrays, a two-dimensional response function for the retina can be calculated for an arbitrary retinal illumination pattern. The response of each sensor is assumed to be as in Eq. 4.3; the response is distributed evenly in the plane over the area of the center region of the sensor, A_c . Define $\bar{R}_S(x, y) = R_S/A_c$ for all (x, y) within A_c ; then the total response function at a point (x, y) is defined by

$$R(x, y) = \sum_{\{S_i\}} \bar{R}_{S_i}^2(x, y) \Delta S_i \quad [4.6]$$

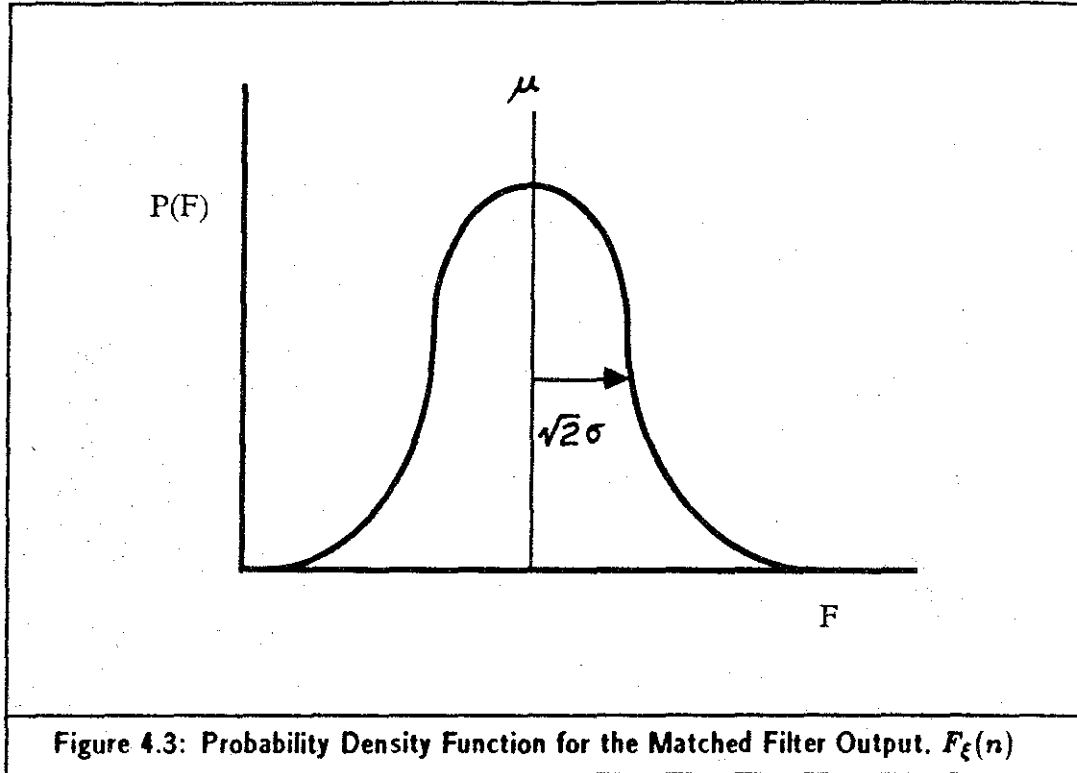
where the sum is taken over all the sensor sizes, $\{S_i\}$. The factor ΔS_i is the difference in sensor size, $S_i - S_{i-1}$ between succeeding sensor arrays and compensates for the fact that only a finite number of sensor sizes are used. The use of Euclidean summation insures that the response function is everywhere positive and weights large responses more than small responses.

4.4.3 Matched Filtering of Response Function

The two-dimensional response function, $R(x, y)$ is a measure of how well the eye responds to the features in the given image. However, another important component of luminance contrast detection appears to be the comparison of the perceived response function with the expected response function. Many authors have suggested that correlation of the perceived signal with the expected signal occurs at some point in the visual system, although there is considerable disagreement over the exact stage of visual processing at which this takes place. In particular, Burgess [Burgess, 1983] has shown that for an ideal observer, the optimal detection strategy is to do matched filtering of the expected and observed signals.

In the current work, a matched filtering approach is used to form the final expected output of the early visual system; however, the fact that the observer knows the location of the signal exactly allows a simplification from a general matched filtering. It is necessary to perform the matching only at the known location of the signal ξ . Hereafter, the term "matched filtering" will refer to this matching performed only at the known location.

Consider an image, I_ξ , which contains only the known signal. If this image is processed by the sensor array, a response function $R_\xi(x, y)$ is produced which is the response



expected by the observer. For an image I_n which produces a response function $R_n(x, y)$, the matched filter output is defined as

$$F_{\xi}(n) = \int_{\text{space}} R_{\xi}(x, y) R_n(x, y) dx dy. \quad [4.7]$$

The fact that the location of the signal is known exactly allows the two response functions to be registered in space. This result tells how well the perceived response function agrees with the expected response function.

The matched filter response $F_{\xi}(n)$ can be considered as a random variable which depends on the exact structure of the image I_n . If in the set of images $\{I_n\}$ the observer's task is similar from image to image (the same target embedded in similar backgrounds in the different images), $F_{\xi}(n)$ will have some probability density function which gives the likelihood that a given value of $F_{\xi}(n)$ will be observed. The exact form of this density function is unknown; it will be assumed that it is Gaussian with some unknown mean and standard deviation (Figure 4.3):

$$p(F_{\xi}(n)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(F_{\xi}(n) - \mu)^2/2\sigma^2]. \quad [4.8]$$

4.4.4 Definition of $P(D_i|T_i, I_n)$

For the current experimental paradigm, two cases may be distinguished for a given signal ξ : either the signal is present or it is not. For $\{I_n\}$, it is assumed that the images are similar, but that in some subset of the images, $\{I_p\}$, the signal is present and in the remainder of the images, $\{I_a\}$, the signal is absent. Let the number of images in $\{I_p\}$ be N_p and the number of images in $\{I_a\}$ be N_a . For those images in which the signal is present, the density function is

$$p(F_\xi^p(n)) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp[-(F_\xi(n) - \mu_p)^2/2\sigma_p^2], \quad [4.9]$$

and for those where the signal is absent

$$p(F_\xi^a(n)) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp[-(F_\xi(n) - \mu_a)^2/2\sigma_a^2]. \quad [4.10]$$

A schematic plot of these two density functions is shown in Figure 4.4. The mean and variance of these density functions are unknown, but can be estimated by standard techniques.

For $p(F_\xi^p(n))$,

$$\hat{\mu}_p = \frac{1}{N_p} \sum_{n=1}^{N_p} F_\xi^p(n) \quad [4.11]$$

$$\hat{\sigma}_p^2 = \frac{1}{N_p - 1} \sum_{n=1}^{N_p} (F_\xi^p(n) - \mu_p)^2. \quad [4.12]$$

Similar expressions hold for $p(F_\xi^a(n))$.

In order to make a decision whether the signal is in fact present in a particular image I_n , the observer must pick some decision threshold such that if $F_\xi(n)$ is greater than the threshold, he says "yes" and otherwise says "no". The performance of the observer will be optimized if the decision threshold is taken as the value of $F_\xi(n)$ when

$$P(a) p(F_\xi^a(n)) = P(p) p(F_\xi^p(n)).$$

Let this value be denoted by θ_ξ ; if it is assumed that $P(a) = P(p)$, θ_ξ can straightforwardly be calculated by solving analytically

$$\frac{1}{\sqrt{2\pi\sigma_p^2}} \exp[-(\theta_\xi - \mu_p)^2/2\sigma_p^2] - \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp[-(\theta_\xi - \mu_a)^2/2\sigma_a^2] = 0. \quad [4.13]$$

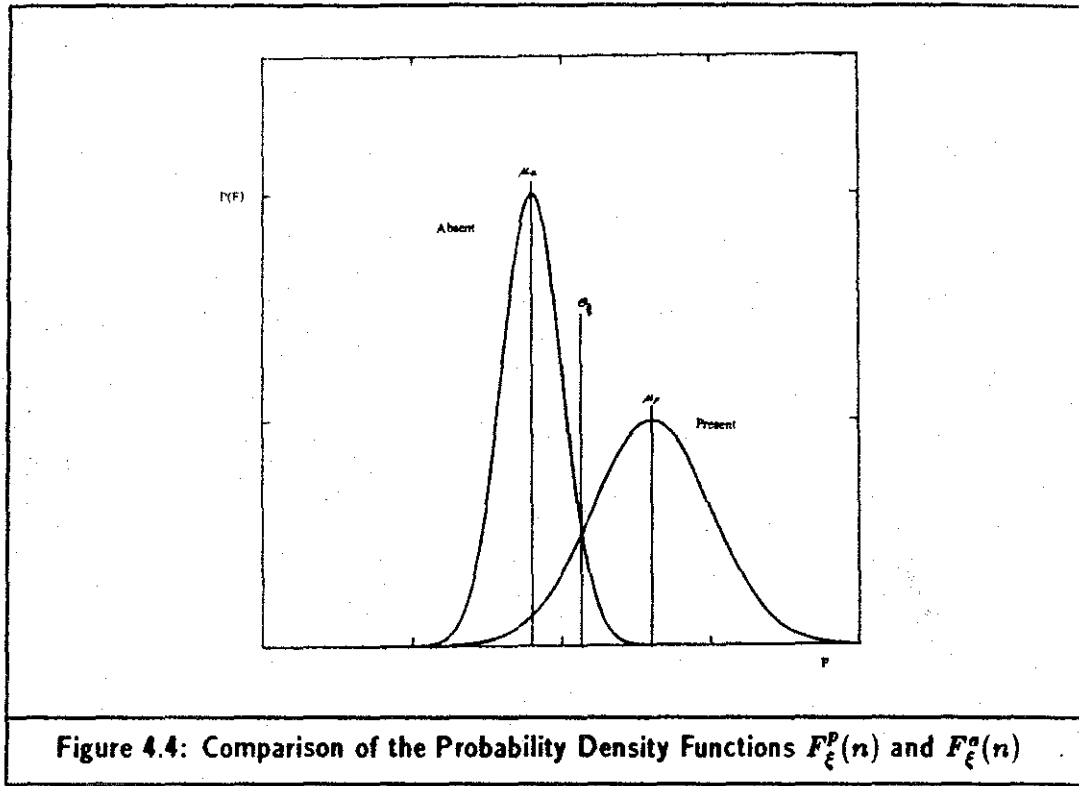


Figure 4.4: Comparison of the Probability Density Functions $F_{\xi}^p(n)$ and $F_{\xi}^a(n)$

This yields a quadratic equation in θ_{ξ} ; after some algebra, the results are seen to be

$$\theta_{\xi} = \frac{1}{\sigma_p^2 - \sigma_a^2} \left[\sigma_a \sigma_p \left[2(\sigma_p^2 - \sigma_a^2) \log[\sigma_p/\sigma_a] + (\mu_p - \mu_a)^2 \right]^{1/2} + (\mu_a \sigma_p^2 - \mu_p \sigma_a^2) \right] \quad [4.14]$$

$$= \frac{-1}{\sigma_p^2 - \sigma_a^2} \left[\sigma_a \sigma_p \left[2(\sigma_p^2 - \sigma_a^2) \log[\sigma_p/\sigma_a] + (\mu_p - \mu_a)^2 \right]^{1/2} - (\mu_a \sigma_p^2 - \mu_p \sigma_a^2) \right] \quad [4.15]$$

The choice of which root should be used as the threshold can be made on the basis of where the roots lie relative to the means of the density functions. If a root lies between the means, it is chosen. Otherwise, the choice of root depends on the relative heights of the two distributions. If $\sigma_a < \sigma_p$, the larger root is chosen, otherwise the smaller root.

The conditional probability $P(D_i|T_i, I_n)$ can now be constructed. There are two possible true situations, so let $\{T_i\} = \{T_p, T_a\}$ and $\{D_i\} = \{D_p, D_a\}$. Define

$$P(D_p|T_p, I_n) = P(F_{\xi}(n) > \theta_{\xi}) \quad [4.16]$$

$$P(D_a|T_a, I_n) = P(F_{\xi}(n) < \theta_{\xi}) \quad [4.17]$$

and with Q now equal to 2,

$$IQM = \frac{1}{2N} \left[\sum_{\{I_p\}} P(F_{\xi}(n) > \theta_{\xi}) + \sum_{\{I_a\}} P(F_{\xi}(n) < \theta_{\xi}) \right] \quad [4.18]$$

4.4.5 Calculation of the IQM

The value of the IQM can now be calculated for a set of images $\{I_n\}$; define $N_>$ as the number of images in the set $\{I_p\}$ such that $F_\xi^p(n) > \theta_\xi$ and $N_<$ as the number of images in $\{I_a\}$ such that $F_\xi^a(n) < \theta_\xi$. Then

$$P(F_\xi^p(n) > \theta_\xi) = N_>/N_p, \quad [4.19]$$

$$P(F_\xi^a(n) < \theta_\xi) = N_</N_a; \quad [4.20]$$

substitution into Eq. 4.18 yields

$$\begin{aligned} IQM &= \frac{1}{2N} \left[N_p \left(\frac{N_>}{N_p} \right) + N_a \left(\frac{N_<}{N_a} \right) \right] \\ &= \frac{1}{2N} (N_> + N_<) \end{aligned} \quad [4.21]$$

It is also useful to determine the variance that should be expected in the calculation of the IQM. Each of $N_>$ and $N_<$ can be considered as a sample of a binomially distributed random variable; hence, the variance of each term is

$$\begin{aligned} \sigma_{N_>}^2 &= N_> P(F_\xi^p(n) > \theta_\xi) (1 - P(F_\xi^p(n) > \theta_\xi)) \\ &= N_> \frac{N_>}{N_p} \left(1 - \frac{N_>}{N_p} \right), \end{aligned} \quad [4.22]$$

$$\begin{aligned} \sigma_{N_<}^2 &= N_< P(F_\xi^a(n) < \theta_\xi) (1 - P(F_\xi^a(n) < \theta_\xi)) \\ &= N_< \frac{N_<}{N_a} \left(1 - \frac{N_<}{N_a} \right), \end{aligned} \quad [4.23]$$

and the variance of the IQM is then

$$\sigma_{IQM}^2 = \frac{1}{4N^2} (\sigma_{N_>}^2 + \sigma_{N_<}^2) \quad [4.24]$$

4.4.6 Implementation

The preceding model was implemented using the C programming language. Most of the model was straightforward to implement; here a few details of the program which calculates the response function of the sensor array are described.

The program takes as input a two-dimensional image illumination distribution and produces as output a two-dimensional representation of the response function of the sensor array. In the implementation, the response function is calculated for each of a set of individual sensor sizes; the illumination distribution is laid down on an array of sensors of the appropriate size and for each sensor the response is calculated and the result spread

equally over the center region of the sensor. The resulting response function is a two-dimensional array of floating-point numbers with a resolution of one pixel; thus, the output is an image of the same size as the input. Each floating-point image is then summed as in Eq. 4.6, producing a two-dimensional image which is the total response function of the self-similar sensor array. For each individual sensor, the dimensions of the central region were chosen to equal to $S/3$ on each side. The size of the sensors, S , was allowed to vary from 1 to 9 pixels, with the size of the sensor in each succeeding plane a factor of $\sqrt{3}$ larger than the preceding plane; thus, there were five distinct sensor sizes used.

Since the properties of the target lesion are known, it is possible to then calculate the lesion. The resulting expected response function was then multiplied times the actual response function and summed as in Eq. 4.7 to obtain the matched filter output. This process is repeated for each image in a test set and the resulting distributions (Eqs. 4.9 and 4.10) calculated.

4.4.7 Algorithm for Calculating the IQM

In this section, a brief summary is given of the algorithm for calculating the IQM for a set of images $\{I_n\}$ and a given signal, ξ .

- 1) The response function is calculated for the known signal, ξ , before its insertion into the images.
- 2) For each image in $\{I_n\}$, the two-dimensional response function is calculated through the use of the sensor array.
- 3) The response function for the processed image is multiplied by the normalized response function that was obtained in step 1; the two response functions are placed into proper registration using the known location of the target.
- 4) Steps 2 and 3 are repeated for all images in the test set $\{I_n\}$.
- 5) The resulting values for the matched filter output are used to calculate the statistics of the probability density functions given in Eqs. 4.9 and 4.10.
- 6) The threshold for optimum performance, θ_ξ , is determined from Eq. 4.14.
- 7) The IQM and its variance are calculated from Eqs. 4.21 and 4.24.

4.4.8 Extensions and Testing of the IQM

The use of matched filtering to compare the expected signal with the perceived signal is useful and agrees with processes which are thought to occur in the visual system. However, as developed so far, the IQM requires that the target signal be completely specified in advance; this precludes the use of realistic targets in which the target is not precisely specified, although certain partial descriptions may be known.

To avoid this problem, one may attempt to determine tentative image objects as they might be determined by the visual system. A promising approach to determining these objects which was investigated as part of this work is the "stack"-based description which follows from ideas given by Koenderink, 1984 and was developed and implemented by Pizer *et al.*, 1985. In this work, the image is considered as a nested set of dark and light blobs at different levels of scale. Such objects, or more precisely their boundaries, could be used to determine the expected response to tentatively seen regions. Using a *priori* knowledge of the feature to be detected, the technique is likely to yield only a small number of possible target boundaries.

Unfortunately, because the targets to be detected are small both in spatial extent and intensity range, the stack method did not produce useful boundaries in the image. This probably resulted from pixel artifacts; the low contrast Gaussian lesions used in this study covered only a few pixels. While the stack method holds promise as a tool for tentative object definition, its current state of development was insufficient for reliable use with the targets used in this research.

Finally, the performance of the IQM must be validated; this can be done for a restricted range of tasks using observer studies, but generalization to other tasks will remain somewhat dubious. In Chapter 5, an attempt is made to validate the IQM for one particular task through the use of observer experiments.

Chapter 5

Experimental Results

*If your experiment needs statistics,
you ought to have done a better experiment.*

—Lord Ernest Rutherford [1871–1937]

5.1 Purpose and Goals

The experiments described below have two purposes: to compare the effectiveness, in the detection of well-localized objects, of a particular adaptive contrast enhancement method, Adaptive Histogram Equalization, with that of a common global contrast enhancement method, interactive linear min-max windowing, and to evaluate the performance of the measure described in Chapter 4 as an objective test of image quality. Careful experimentation is the only way to examine these problems, since the ultimate arbiter of image quality is the human observer.

5.1.1 Comparison of Contrast Enhancement Modalities

In Chapter 3, Adaptive Histogram Equalization (AHE) was described and its properties examined. The experience so far gathered with this method shows it to have considerable promise as a contrast enhancement technique. In several branches of medical imaging, notably computed tomography, the contrast enhancement technique in common use at this time is global linear min-max windowing, as described in Chapter 2. AHE offers several advantages over this technique, *e.g.*, it allows the simultaneous viewing of different image structures, it can be applied without manual intervention, and its effects are repeatable. However, if AHE does not provide local diagnostic information roughly equal to that of windowing, its usefulness in a clinical setting is severely limited. While it is effectively impossible to show that AHE is diagnostically equivalent to windowing in all situations of clinical interest, it is possible to compare the two methods for a single well-defined task similar to those encountered in clinical practice.

In the current experiment, the task is the detection of simulated lesions in clinical images by trained observers. A number of chest CT scans were used to construct a set of test images; these test images included both the original images and the same images with artificial lesions inserted into the lung and mediastinum areas. All the images in the test set were then processed with AHE. Both the AHE'd images and the unprocessed images were presented to radiologists; they were allowed to perform center-width windowing on the unprocessed images. With these two types of images, they were asked to rank their confidence that a lesion was present at a designated place in each image. Their rankings were evaluated using ROC analysis.

5.1.2 Evaluation of the IQM

The image quality measure described in the last chapter is intended to allow the objective comparison of different contrast enhancement methods. However, before the measure can be used, it must be ascertained whether its results correspond to reality, that is, does the IQM rank the methods in the same order as human observers. In the current work, the images used in the experiment described above were evaluated using the IQM; its predictions were compared to the performance of the observers for the particular task at hand. In the remainder of this chapter, the experimental procedures used are described in detail and the results obtained evaluated.

5.2 Experimental Methodology: Contrast Enhancement Comparison

In this section, the experimental methodology used for comparing the two contrast enhancement methods is described. There are two stages in designing an acceptable experimental methodology. First, the overall goals of the design must be clearly formulated. Next, a large number of individual design decisions must be made to implement the goals. For each such decision, there is the possibility of conflict among the design goals; thus a number of compromises must be made. In the following sections, as each individual design decision is discussed its relationship to the overall design goals will be stated.

5.2.1 Design Goals

At the highest level, the goal of the current experiment is to allow the comparison of two contrast enhancement methods in as realistic a setting as possible. To accomplish this task, three design aims can be stated.

Clinical Realism. Since the overall goal is to evaluate the relative usefulness of each contrast enhancement method, the experiment must be designed to be as realistic as possible. If the experiment is not closely related to the daily tasks of the radiologist, the results will be unconvincing.

Analyzability. In order to derive statistically meaningful results, the experimental data must be analyzable, preferably by a convincing, well-understood method. Note that this goal will often be in conflict with the first goal of clinical realism, since an absolutely realistic experiment will be almost impossible to analyze, due to the inability of currently available processing technique to deal with situations where the true presence or shape of the lesion is unknown.

Specificity of Task. In doing the experiment, it is important that the task to be done by the observer be as precisely defined as possible. If this is not true, there is the possibility that the resulting data will reflect a mixture of different goals for each observer or the results from different experiments. Again, this goal may be in conflict with the first goal of clinical realism.

A fourth design aim may be implicitly added to the three above: the experiment must be practicable in a reasonable length of time and without undue demands upon the observers. If, as P. B. Medawar [Medawar, 1982] says, science is the art of the soluble, then experimental design is surely the art of the practical. Again and again, design decisions must be made to fit the experiment into the available resources. For example, while the aim of statistical accuracy demands as much data as possible, the practical limitations involved in using human observers puts a strict ceiling on the amount of obtainable data.

Within these design aims, a number of design decisions must be made. Of these, the two most important are the exact task of the observer and the experimental paradigm which will allow the data to be analyzed.

5.2.2 Design Decisions

The Task. In measuring the ability of observers to detect contrast, there are a number of different tasks which can be performed. The simplest task is detection, where both the location and the size and shape of the possible lesion are known to the observer. Such signal known exactly (SKE) experiments measure the fundamental limit of contrast detection for a given target and background, since there are no complicating tasks to be performed. The analysis of such experiments is relatively straightforward, since it is easy to distinguish between correct and incorrect responses. There is only one variable—whether the known signal is present or not. Additional levels of complexity can be introduced by having the location of the signal be unknown to the observer (localization experiment) or forcing

the observer to characterize the detected signal in some way (recognition experiment). The methods of analysis for these tasks are not as advanced as that for the detection experiment. In real clinical cases, a mixture of all three tasks are present.

In the current work, a pure detection task has been chosen. This selection has two principal advantages. First, the detection task paradigm is well-understood and hence reliable estimations can be obtained from the data. Second, the task of the observer is simple; this is important in that it allows all the observers to have a clear understanding of the decision criteria and allows the observer to keep a relatively consistent standard of judgement.

Here is a clear example of conflict between design aims. Clinical realism demands as realistic a task as possible, but for tasks which involve localization and recognition, the analysis is much complicated, as will be seen in the next section. However, while detection is a simplification of the general diagnostic task, there is evidence to suggest a correlation between observer performance for a detection task and that for localization and recognition tasks.

In order that the observer can know the exact size and shape of the target, simulated lesions were used for the experiment. While the use of real clinical pathology would be more satisfactory from the standpoint of realism, the difficulty is that both for the analysis of the data and the evaluation of the IQM, it is necessary to know the exact characteristics of the lesions. This information is not easily obtainable for true clinical lesions.

The lesions used in this experiment are of considerably lower contrast than would ordinarily be encountered in clinical practice. This is unrealistic but necessary in order that the observer's performance not be too good; it was found during testing that lesions of ordinary clinical appearance are almost always detectable both with windowing and with AHE. The use of very subtle lesions tends to bias the experiment somewhat in favor of windowing, since it is possible for some of the lesions that only a very finely adjusted window can show the lesion; this factor must be taken into account when comparing the diagnostic power of AHE and windowing.

Experimental Paradigm. The observer was asked to perform a rating experiment. An image was presented which may or may not have contained an artificial lesion of a size and shape known to the observer. Crosshairs were used to show the observer the exact location of the possible lesion and a representation of the lesion's appearance was shown inset into the upper left corner of the image. Thus the experiment was of the SKE form; the observer

was asked to give a rating of his confidence that the lesion was actually present on a scale of 0 (lesion certainly not present) to 4 (lesion probably present). This rating scale data can be used to measure the diagnostic efficiency of a contrast enhancement method by using the technique of receiver operating characteristic (ROC) curves.

A basic introduction to the ROC analysis method is given by Metz, 1978. The best detailed description of ROC analysis is that of Green and Swets, 1974, while the application of ROC analysis to the evaluation of diagnostic systems can be found in Swets and Pickett, 1982. The ROC method gives a curve for each diagnostic method (AHE or windowing); these curves can be evaluated at any desired false positive fraction (degree of conservatism) to determine the superior method at that conservatism level. This technique is widely accepted for measuring the diagnostic performance of a method, in this case for contrast enhancement to allow the discrimination of contrast in a medical image.

The ROC technique is very powerful, but must be used with some care. It is necessary that a sufficient number of experimental trials be used in order to obtain adequate statistical confidence in the results. Oftentimes, this means that the data from trials with several different observers or images must be pooled. This can only be done if there is evidence that the individual statistics of the experiments to be pooled are similar. Thus, the careful selection of case studies and artificial lesions and the exact specification of the task and ranking criteria are essential.

The ROC analysis technique is well-understood for the detection task; its theoretical basis for more complex tasks is not as well-known. For localization of a lesion, the Localized ROC (LROC) paradigm must be used. Here again, the correctness of the observers response must be known, but there is difficulty in deciding whether a given response is correct or incorrect. The observer is required to specify the location of the lesion; however, if the specification is not at the exact site of the lesion, it is difficult to decide how far off it must be before it is judged as an incorrect response. Thus, there is a tradeoff here between choosing a realistic task and having the ability to analyze the data.

5.2.3 Selection of Case Sample

Given the decisions about the task and analysis paradigm, a case sample was constructed for presentation to the observer. The images used were CT scans of the of the chest at the level of the major vessels above the heart. The principal areas of diagnostic interest are the lungs and mediastinum; often the pathology of a disease will involve both of these areas and adequate contrast enhancement is desired in both fields.

Selection of Base Images. The base images were chosen from CT scans of five separate patients who were classified as having no pathology in the areas of interest. The images were obtained on a Technicare 2060 CT scanner located at the North Carolina Memorial Hospital in Chapel Hill. Dr. Edward Staab of the Department of Radiology did the preliminary selection of these images. The images were taken from the scanner in digital form; the intensities in each image were calibrated in Hounsfield units, with the CT intensities approximately in the range -1000 to $+1200$, thus implying almost 12 bits of information; their spatial resolution was 512×512 pixels. From about 100 slices in five patients, 32 slices were chosen. Adjacent slices were avoided to maximize anatomical differences between slices and reduce the possibility of memorization of normal variation by the observer. Of the 32 slices, 8 were chosen for use as a training set, while the remaining 24 were used as the base images for the actual study.

Lesion Site Selection. In each image, four sites, two sites in the lungs and two in the mediastinum, were chosen for the insertion of artificial lesions. The criteria for site selection were the presence of appropriate natural anatomy and the prevalence of real lesions at that site in clinical practice. Similar, but not identical, sites were chosen in each base image. This selection scheme insures that the observer's task was approximately the same at each lesion site across the base set of images, a circumstance which enhances the probability that the rating scale data can be pooled across observers and images for each site. Selection of the sites was done with the collaboration of a board-certified radiologist.

The selection of multiple sites within the same image in two different image fields allows the comparison of a single AHE image with multiple different windowed images; this is important in that one of the apparent diagnostic advantages of AHE is that it allows the replacement of multiple windowed images with a single view which shows diagnostic information in many fields.

Generation of Artificial Lesions: Shape. The artificial lesions inserted into the chosen sites are Gaussian lesions with an appropriate size and intensity. The characteristics of these lesions are given in Table 5.1. A Gaussian profile was chosen as approximately that which would be generated by a spherical tumor; the widths of the lesions were chosen as appropriate to appear in the given field (lung or mediastinum).

The intensity profile for a given lesion is given by

$$I(x, y) = A(\Omega) \exp\{-(x^2 + y^2)/2\sigma^2\}. \quad [5.1]$$

Lungs		
Lesion Number	Width (σ)	Height factor
0	1.41	0.85
1	2.0	1.15
2	2.0	0.85

Mediastinum		
Lesion Number	Width (σ)	Height factor
0	2.0	0.85
1	2.82	1.15
2	2.82	0.85

Table 5.1: Parameters for the Artificial Lesions.
Widths are given in pixels; heights in Housfield units.

The intensity of the exponential varies between 0 and 1; thus the function $A(\Omega)$ determines the height (intensity) of the lesion; it depends on the local neighborhood Ω of the lesion site. The criteria used to specify $A(\Omega)$ are given in the next section. The variance, σ^2 , determines the width of the lesion. When $r^2 = x^2 + y^2 = 2\sigma^2$, the lesion intensity has fallen to e^{-1} of its maximum intensity. There are two variances given in the table for each field (lung or mediastinum); the lesions corresponding to these two variances will be referred to as the narrow and wide lesions. For a given field, the ratio of the standard deviations of the wide lesion, σ_w , and the narrow lesion, σ_n is

$$\sigma_w/\sigma_n = \sqrt{2}. \quad [5.2]$$

This ratio was chosen for two reasons. First, it corresponds visually to a noticeable difference between the wide and narrow lesions; second, it causes the two lesions to differ in signal to noise ratio by a factor of 2 if they are of equal intensity. Burgess *et al.*, 1982 have shown that this ratio is more than adequate for a significant difference in detectability of Gaussian spots embedded in white noise. An interesting question which is raised by their work is whether this result holds for images with more complex backgrounds; this issue will not be treated in detail here.

Generation of Artificial Lesions: Intensity. The intensity of a lesion will be defined by the maximum intensity (in Hounsfield units) of the lesion before its insertion into a base image. The selection of the correct lesion intensities is crucial to the success of the experiment. If the lesion intensities are too small, then insufficient information will be conveyed to the observer, and the observer's decisions will proceed by guesswork. If the lesion intensities are too large, the observer will make a correct choice on every trial and no discrimination of the contrast enhancement methods will be possible. Unfortunately, choosing the correct intensity for a lesion is difficult; it is not sufficient to pick a fixed lesion intensity applicable to all images and sites. The ability of the eye to detect contrast is strongly dependent on the nearby background of the image about the lesion site. Both the mean intensity of the near background and the presence of structure in the image change the minimum detectable contrast of the lesion. In addition, there is evidence [Van der Wildt and Waarts, 1983] that the far background of the image exerts considerable influence on the detectability of the lesion.

Thus, the intensity of a given lesion is dependent on a number of factors; preliminary tests performed on the images used in this experiment indicated that the appropriate lesion intensities must be chosen on an image-by-image basis. There are two possible methods here: either the appropriate intensities can be determined manually or a simple measure of the image complexity can be formulated to select an intensity based upon the local background structure of the image site. In the current experiment, the second choice was made. As an approximate measure of image complexity, the average value of the Laplacian in the neighborhood of the lesion site was determined and appropriate lesion intensities were generated on the basis of this measure.

To determine the appropriate parameters for the complexity measure, a small number of the images (4) were examined by hand; for each of these a method of limits experiment was performed to determine the minimum detectable lesion intensity at each site. Since the detectability of a given lesion will be affected by the processing (windowing or AHE) done on the image at observing time, it was necessary to perform the method of limits determination after processing of the image. In this instance, an appropriate windowing of the image field independent of the lesion presence was chosen and the detectability of the lesion then determined.

To measure the structural complexity of the site, the Laplacian of the image was calculated in a square of side 2σ centered on the lesion position, where σ is the standard deviation of the lesion. At each pixel within the square, the Laplacian value was determined from

$$\nabla^2 I(x, y) = I(x+1, y) + I(x-1, y) + I(x, y+1) + I(x, y-1) - 4I(x, y); \quad [5.3]$$

these values were then averaged over the entire square and the result used as the average value of the Laplacian:

$$|\overline{\nabla^2 I}| = \frac{1}{4\sigma^2} \sum_{x, y=-\sigma}^{\sigma} |\nabla^2 I(x, y)|. \quad [5.4]$$

A plot was made of the lesion intensities determined by the method of limits against the average Laplacian. A straight line fit of this data was made and the resulting line equation was used to predict an appropriate lesion size for the remaining lesion sites in all the images. The resulting values were checked during the execution of the training runs (to be described in the next section) to verify their appropriateness. Separate equations were used for the lung and mediastinum field; however, no significant difference between the wide and narrow lesions was discernible, and the same equation was used for both. The plots used and the resulting coefficients of the linear equations are shown in Figure 5.1 and Figure 5.2. These linear equations constitute the definition of $A(\Omega)$ as given in Eq. 5.1, where the neighborhood Ω is the region over which the average Laplacian is calculated.

Two different possible lesion intensities, referred to as the bright and dim lesions, were chosen at each site. These lesions were respectively 85% and 115% of the intensity predicted by the average Laplacian measure. Thus, there are four possible lesions which can be inserted at a given lesion site, corresponding to two choices for the width and two choices for the intensity. In this experiment, only three of these possible lesions were used in the interest of reducing the time required to perform the experiment. It can be seen in Table 5.1 that the bright, wide lesion has been omitted. The intensity of the lesion actually inserted into the image must be adjusted for the windowing which was used in the method of limits experiment previously described. Thus, the lesion was scaled by an appropriate amount before insertion into the image.

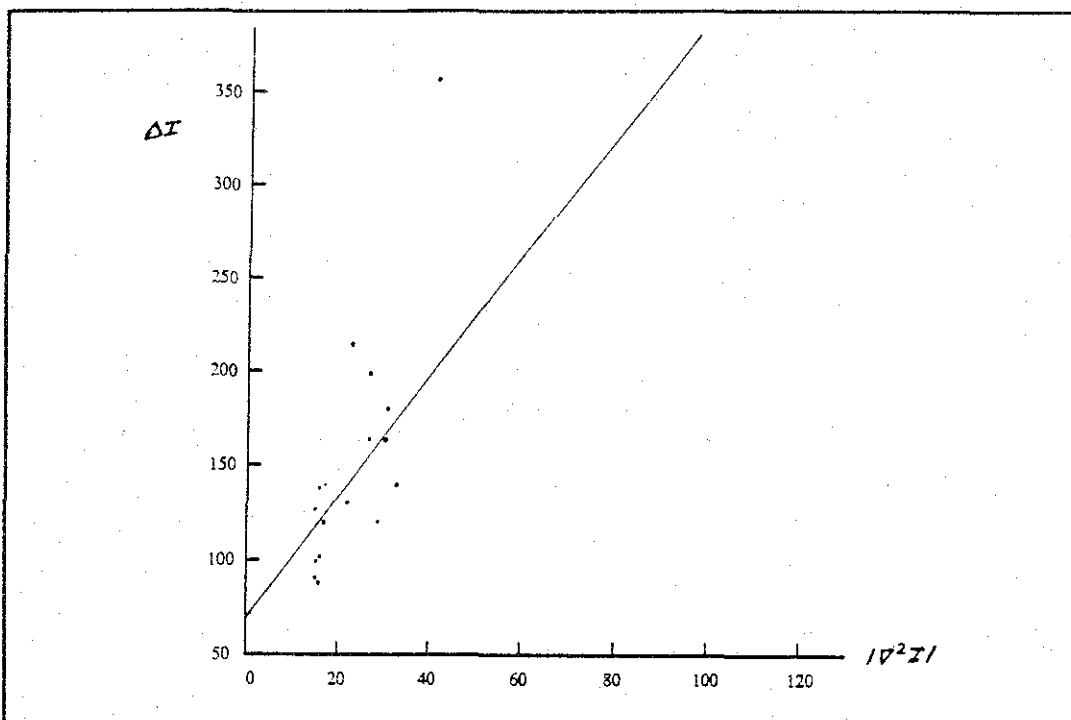


Figure 5.1: Plot of the the minimum detectable lesion intensity against average Laplacian for the lung field.

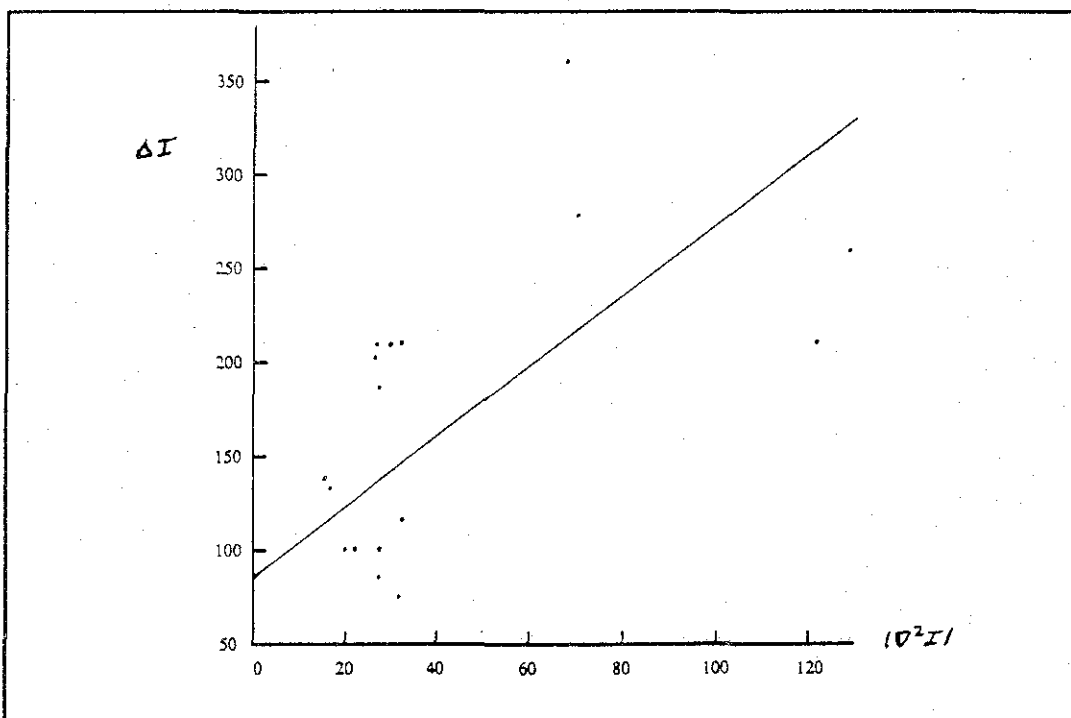


Figure 5.2: Plot of the the minimum detectable lesion intensity against average Laplacian for the mediastinum field.

Lesion Parameter	Values	Comments
Base Image	24	
Lesion Field	2	Lungs or Mediastinum
Lesion Type	3	
Lesion Site	2	
Processing	2	AHE or windowing
Lesion Presence	2	Present or absent
Redundancy Factor	10%	Number of repeated images
Total	1280	All combinations

Table 5.2: Parameters for the Trial Images

5.2.4 Preparation of the Trial Images

After the selection of the base images, the lesion sites, and the lesion characteristics, the set of trial images was generated. Table 5.2 shows the possible characteristics of a given trial image. Each image is described by a 6-tuple of parameters that gives the base image, the lesion site, the type of lesion, *etc.* Thus, there are 1152 possible trial images; all these were included in the trial set. In addition, approximately 10% (128) of the possible images were chosen at random and included twice, giving a total of 1280 images in the complete set. Hence, some images were seen twice by each observer, giving a check on the consistency of a given observer in rating the images. The 6-tuples for the 1280 images were generated and randomly permuted, then divided into 20 runs of 64 trials each. A given observer was asked to rank one run (64 images) in a single session. The random permutation of the 6-tuples guarantees no discernible pattern in the order of presentation. The 6-tuples were then used to generate the trial set. Each trial image was identified by an absolute sequence number; identifying information was included in the file containing the processed image to preclude misidentification.

A number of other processing steps were required for the generation of the final images; it is worthwhile to detail the exact preparation of the images in order to make clear the reasons for these manipulations. The preparation was done using a suite of programs; all calculations were done in the original range of the CT data and then scaled to the range 0-255 for display on the Comtal frame buffer (described below).

For images which were to be windowed, the inserted lesion was prepared by prescaling it to correspond to the local image structure and the windowing of the image used in the method of limits experiment as previously described. The prescaled lesion was then added to the image at the correct site. The same lesion was inset into the upper left corner of the image to serve as a reference for the observer.

Unfortunately, the Comtal frame buffer used in this experiment is able to display and window its images with only 8 bits of intensity, compared to a typical CT console device which allows the observer to window the full 12 bits of information before it is scaled into 8 bits for display. This allows the user of the CT console to have finer control over the displayed image than is possible with the Comtal frame buffer. In order to preserve as much of the original information as possible in the Comtal image, a prewindowing operation was performed on the original data before it was transformed into the Comtal display range. Here, a broad window was chosen to include the data of interest (the lungs or mediastinum) and applied with the full 12 bits of precision; the resulting image was then scaled into the Comtal display range; the effect of this was to reduce the amount of data compression involved in the transformation from 4 bits (12 bits to 8 bits) to 2 or 3 bits (10 or 11 bits to 8 bits). The result for the observer was finer control over the exact windowing of the image.

For AHE'd images, the inserted lesion was also prescaled to be identical to the lesion used for the windowed image. The image was then processed with a well-tested implementation of the AHE algorithm developed by the author using a square grid of 8×8 sample points; the calculations were done using the full range of the original CT data. After the AHE was performed, a reference lesion was inserted into the upper left corner of the image. This reference was not subjected to the AHE process, but was rather prescaled so that its visual appearance was similar to the reference inserted into the analogous windowed image. Thus, the reference lesion was not an exact copy of the lesion that was in the AHE image; because of the nonlinear nature of the AHE process, extraction of the appearance of the processed lesion was not possible.

In addition to the trial set described above, the 8 test images were used to generate 4 runs of 64 images each for use in training the observers. One of the 4 runs was used exclusively to familiarize the observers with the equipment and experimental procedure; the other three formed an exhaustive set of three test images over all six parameters. These test runs were used to evaluate the choice of lesion parameters (shape and intensity) to insure that they were reasonable. It was found that the scaling based on the average Laplacian produced acceptable results.

5.2.5 Selection of Observers

Three observers were used in this study, all board-certified radiologists and experienced in the reading of chest CT scans. The observers were guaranteed that no results of the experiment would be associated with a particular observer. The experiment was designed to occupy no more than 25 hours of reading time, including the necessary training time. The observers were trained in the use of the experimental programs and allowed to do test runs to familiarize themselves with the appearance of AHE images.

5.2.6 Experimental Equipment and Layout

In this section, the experimental apparatus used for the studies is described; a more complete description can be found in Appendix 1.

Experimental Apparatus. The trial images were displayed on a Tektronix 690SR RGB monitor using a Comtal 10/24 frame buffer. A VAX11/730 running the Berkeley 4.2BSD version of the UNIX* operating system served as the host computer to control the presentation of the images and collect the data. The Tektronix monitor has very stable performance; our experience shows that its intensity display characteristics are stable on a time scale of weeks. The monitor was calibrated and converged prior to the beginning of the experiment. The images were displayed in monochrome mode (with the red, green, and blue inputs identical); the display scale was linearized using the procedures previously described. The luminance range of the monitor was 6×10^{-4} foot-lamberts for a driving intensity of 0 display units and 26.2 foot-lamberts for a driving intensity of 255.

The Comtal 10/24 is able to display images of 512×512 pixels and contains sufficient image memory for 12 512×512 images; it includes various function memories for table lookup and pipelined processors which were used to implement the linearization of the displayed image and the windowing specified by the observers. The VAX computer was equipped with a data tablet to allow the observers to supply input to the program while they were observing.

Experimental Software. A program was written to control the experiment and collect the rating scale data. The external interface to this program, *zdos*, is described in Appendix 1. Every attempt was made to insure that the program was robust and sufficiently automated to allow the observers to conduct trial runs on their own time scale. After the observers were trained, they were able to operate the program unattended.

* UNIX is a trademark of AT&T Bell Laboratories

Experimental Layout. The observers were seated before the Tektronix monitor with the data tablet on a table in front of them. A light box was placed on top of the monitor facing away from the observers to provide an ambient light intensity similar to that used for the linearization of the monitor. A video display terminal was provided for interaction with the host operating system; however, once the trial run was underway the VDT was not used, all interaction being performed using the data tablet. The observers were allowed to position themselves comfortably and no attempt was made to constrain their movements.

Physical Environment. The physical environment of the observers was controlled insofar as this was possible in the experimental area. The room lights were extinguished and all extraneous sources of light shielded except for the light box used for ambient illumination. Measurement of the ambient light using a photometer with cosine diffuser directed at the center of the monitor screen showed an illumination of 3.3 lux. The observers were seated at an average distance from the screen of 1 m, with the displayed image subtending an angle of 15°. The observers were required to acclimate themselves to the environment for a period of one to two minutes before beginning the experiment; in the actual procedure, the period of acclimation was somewhat longer.

5.2.7 Observer Procedure

The experiment was divided into 20 sessions of length approximately one hour each. At each session, the observer was asked to rate 64 images. In addition, each observer performed 4 additional one hour runs for training purposes. After familiarization with the interaction devices and the use of the software, the observers did one training run of 64 images in which they received feedback on the correctness of their answers. This allowed them to calibrate their perceptions against the appearance of the displayed images. Following this, they performed three training runs without feedback to provide them with further experience and to allow the calibration of the experimental procedure against their performance. Both the actual scores achieved on the training runs and the comments of the observers concerning the physical layout of the experiment, the convenience of the software interface, and the clinical realism of the experiment were recorded. The results of these training runs allowed minor adjustments in the experimental procedure and the techniques for preparation of the trial runs to be made before the beginning of actual data collection.

Before the beginning of their training, the observers were required to read a document explaining the purpose and methods of the experiment and detailing the criteria on which their responses were to be based. This document is included as Appendix 1 of this

dissertation; it contains considerable detail about the actual conduct of the experiment which will not be repeated here. The observers were allowed to consult Appendix 1 during their observer sessions and were urged to refresh themselves periodically on the criteria used in making a rating decision.

The observers were shown each image in the trial set and asked to provide a rating of their confidence that the simulated lesion was present. Table 5.3 shows the five categories and the criteria that were to be used to choose each rating. It should be noted that the rating scale is not symmetric; there is no response in the category "Lesion definitely present", nor is there any equivocal category. The observer must make a decision as to which belief is stronger, the presence or absence of the lesion. The observers were urged to use all five rating categories, scaling a 0 response to the occasions when they were most certain the lesion was not present and a 4 response to their greatest certainty that the lesion was present.

During a typical AHE trial, the observer was shown the AHE'd image with a representation of the inserted lesion in the upper left corner (as described earlier); crosshairs were superimposed on the image at the presumed location of the lesion. The observer was allowed to remove the crosshairs at will by pressing a button on the data tablet puck. A 15 second time limit was imposed on the observer for the observation period, although the observer was allowed to terminate the observing period earlier if he wished. When the observation period was finished, the observer was asked to give his ranking.

For a windowed image, the procedure was somewhat different. The observing period was divided into two parts. During the first part, the observer was shown the image with no indication of the location or appearance of the inserted lesion. The observer was told which field of the image (lungs or mediastinum) in which the target would appear; he was then asked to choose a normal clinical windowing of that field using the data table to control the center and width of the window. The observer was given at most 25 seconds to accomplish this procedure, although generally less time was used. During the second part of the observing period, the crosshairs were placed on the image and the representation of the inserted lesion was made visible in the upper left corner of the image. The trial then proceeded as in the AHE case.

Rating	Confidence
0	Definitely not present
1	Probably not present
2	Possibly not present
3	Possibly present
4	Probably present

Table 5.3: Rating Criteria for the ROC Response Categories

Analysis of Results. The result of the experiment was a set of 1280 ratings for each observer; these can be considered as the outcome of 24 simultaneous experiments (2 fields \times 2 sites per field \times 3 lesions per site \times 2 enhancement modalities). Each such experiment was analyzed separately using programs written by Metz and his collaborators [Metz *et al.*, 1983; Metz *et al.*, 1984], CORROC and ROCFIT. ROCFIT uses the maximum likelihood estimation technique of Dorfman and Alf, 1969 to estimate the ROC curve due to a given set of data. CORROC allows the analysis of data in the case where the images are correlated. In the present experiment, the images are correlated, since each image is shown at least twice, once as a windowed image and once as an AHE'd image. For the 10% of the images were shown more than once, each repeated trial was matched with its corresponding image of the opposite contrast enhancement methodology and included in the analysis separately. For experiments which produced similar statistical results, such as trials which used the same field and site but different lesions, the results were pooled and reanalyzed. The programs also produced goodness of fit estimations allowing an analysis of statistical confidence.

The principal statistic used in comparing two similar experiments using different modalities was the area under the ROC curve; this can be shown analytically to be equal to the percentage of correct responses one would obtain from a two alternative forced choice experiment and gives an overall measure of the performance of a given methodology. This approach precludes the comparison of the two enhancement methodologies at different levels of confidence; however, the integrated area is more stable against statistical fluctuations in the data. With the relatively small sample size used in this experiment (approximately 50 data points per experiment), this is an important criteria.

5.3 Results: Comparison of Contrast Enhancement Modalities

The principal results for each observer are shown in Tables 5.4-5. The headings of each column describe the parameters of a given experiment (lesion type, site and processing modality), the integrated area under the resulting ROC curve, and the standard deviation of the area. The standard deviations shown have been corrected for the correlation of the data by the CORROC program. There is no data pooling in these results. The sixth column shows the number of corrected standard deviations, n_{σ} , by which the areas of the two ROC curves differ. The seventh column is the result of a two-tailed t test; it represents the confidence with which the null hypothesis, in this case that the two ROC curves have the same area, can be accepted. The "Results" column indicates which, if either, method was found preferable. In the current analysis, the results were evaluated at two levels; if the entry has no asterisk, the areas under the ROC curve differ by more than 1.5 corrected standard deviations but less than 2.5 corrected standard deviations. If the asterisk is present, the areas differ by more than 2.5 corrected standard deviations. The data is arranged so that experiments with the same situation (field, lesion type, lesion site) are adjacent with only the contrast enhancement modality different to allow ready comparison. The values of 1.5 and 2.5 standard deviations correspond to two-tailed p values of 0.1336 and 0.0124 respectively.

For three cases, the CORROC program failed to converge. In these cases, the data was analyzed as if it were uncorrelated using the ROCFIT program; these results are shown with an "I" in the Results column. Experience showed that the values for σ derived using this method did not differ substantially from those derived with CORROC; thus it can be asserted that none of these three results shows any significant difference between the two contrast enhancement modalities.

Interpretation of Results. The results of the full experiment show that in most cases there is no significant difference in the two modalities. For a difference of 1.5 standard deviations in the lung field, windowing was superior in two experiments, AHE in two, and no significant difference was seen in 14 experiments. For 2.5 standard deviations in the lung field, two experiments showed windowing preferable; the rest showed no significant difference. For 1.5 standard deviations in the mediastinum, windowing was preferred in six experiments, AHE in none, and no difference was seen in 11 cases. For 2.5 standard deviations in the mediastinum, three cases using windowing showed a significant difference and the remaining 13 showed no difference. Windowing did better in more cases in the mediastinum than in the lungs; this may be attributable to the fact that the overall data

range in the mediastinum is usually less and the width of the window used is limited by the noise in the image, whereas in the lung the limiting factor is the range of the data present rather than the noise. In the case of Observer 1, conversations subsequent to the experiment showed that he had used a different strategy while performing the mediastinum windowing, using a window that was somewhat narrower than would be acceptable in clinical practice. This resulted from his recognition that the lesions were likely to lie in fairly well determined areas of the mediastinum; hence he adjusted his window to compensate for this. As can be seen from the data, he was more likely to find windowing significantly better than AHE in the mediastinum than Observers 2 and 3. This result is further borne out in the pooled data described below. Nevertheless, Observer 1's overall performance on the windowed mediastinum images was not significantly better than that of Observers 2 and 3. Little difference is seen between observers compared to the intra-observer variation.

Results Pooled Across Lesion Sites. Given the full results, it appears to be valid to pool the data for the same lesion when shown in the same field (lungs or mediastinum) but at different sites in the field. The results from this pooling are shown in Tables 5.7-5.9. These results bear out the results from the full data in that in the lung field for 1.5 standard deviations no methodology seems preferable (one experiment favored windowing, one AHE, and seven showed no significant difference) while in the mediastinum for 1.5 standard deviations five experiments showed windowing to be preferable while four showed no significant difference. For 2.5 standard deviations, only one experiment in the mediastinum showed windowing to be superior. Again, Observer 1 was found to prefer windowing over AHE in the mediastinum, though the overall performance of Observer 1 (as reflected by the area under the ROC curve) was not significantly better than the other observers.

One difference which is easier to see in the pooled data is the difference in detectability of the different lesions. Lesions 0 and 2 seem to be about equally detectable regardless of image field, while lesion 1 appears to be more obvious than the other two. Lesion 1, as seen in Table 5.1, has the greatest intensity. The difference in width of lesions 0 and 2 seems to make no significant difference.

It appears that there is little or no significant difference in the two contrast enhancement modalities in the current experiment.

Time Series Analysis. One question of interest is whether the performance of the observers improved over time while conducting the experiment. It might be expected that as their familiarity with the AHE methodology increased, their performance might improve. In order to investigate this possibility, the data pooled across the sites was arranged in the order in which the observer performed the runs; the rearranged data was divided into quarters and re-analysed. The ROCFIT program was used since the division of the data precluded correlated analysis; since the purpose of these tables is to show trends in the data rather than absolute performance, this is not a drawback. The tables shown give the parameters of the experiment, the area under the ROC curve, an estimate of σ for each curve, the χ^2 fit of the ROC curve to the data, and σ_{tot} , the standard deviation of the difference in the curves when analyzed as independent. The results are shown in Tables 5.10-12; there is no apparent increase in performance for the later runs. That is, the 4 hours of training seems to have been adequate for the observers to learn fully to judge AHE'd images.

Observer 1							
Lungs							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7352	0.0694	0.6380	0.5235	
0	0	A	0.6808	0.0764	0.6380	0.5235	
1	0	W	0.7877	0.0628	-0.1635	0.8701	
1	0	A	0.7980	0.0650	-0.1635	0.8701	
2	0	W	0.6478	0.0747	-0.5953	0.5516	
2	0	A	0.7038	0.0741	-0.5953	0.5516	
0	1	W	0.7152	0.0680	-1.5024	0.1330	
0	1	A	0.8303	0.0633	-1.5024	0.1330	A
1	1	W	0.8866	0.0472			I
1	1	A	0.8836	0.0512			
2	1	W	0.7478	0.0682	-0.2562	0.7978	
2	1	A	0.7694	0.0672	-0.2562	0.7978	

Mediastinum							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.8567	0.0513	2.5621	0.0104	W*
0	0	A	0.6619	0.0761	2.5621	0.0104	
1	0	W	0.8205	0.0652			I
1	0	A	0.7499	0.0735			
2	0	W	0.6616	0.0748	0.5819	0.5606	
2	0	A	0.6031	0.0893	0.5819	0.5606	
0	1	W	0.7293	0.0705	0.6991	0.4845	
0	1	A	0.6720	0.0748	0.6991	0.4845	
1	1	W	0.7526	0.0693	2.5052	0.0122	W*
1	1	A	0.5795	0.0797	2.5052	0.0122	
2	1	W	0.6703	0.0769	1.0135	0.3108	
2	1	A	0.5993	0.0791	1.0135	0.3108	

Table 5.4: ROC Experiment Data for Observer 1.

The columns indicate the lesion type, site, and processing; the area of under the ROC curve and its standard deviation; the number of corrected standard deviations by which the areas differ; the two-tailed p value; and the result. The lesion types are as given in Table 5.1. No data pooling was performed.

Observer 2							
Lungs							
Type	Site	Proc	Area	σ	n_{σ}	pval	Result
0	0	W	0.6865	0.0736	-0.4104	0.6815	
0	0	A	0.7237	0.0720	-0.4104	0.6815	
1	0	W	0.8679	0.0465	0.1426	0.8866	
1	0	A	0.8609	0.0496	0.1426	0.8866	
2	0	W	0.8506	0.0524	2.6395	0.0083	W*
2	0	A	0.6378	0.0728	2.6395	0.0083	
0	1	W	0.7973	0.0573	-0.7875	0.4310	
0	1	A	0.8440	0.0516	-0.7875	0.4310	
1	1	W	0.9371	0.0318	1.3777	0.1683	
1	1	A	0.8779	0.0456	1.3777	0.1683	
2	1	W	0.7558	0.0708			I
2	1	A	0.8268	0.0624			

Mediastinum							
Type	Site	Proc	Area	σ	χ^2	σ_{tot}	Result
0	0	W	0.9061	0.0503	1.9669	0.0492	W
0	0	A	0.7840	0.0609	1.9669	0.0492	
1	0	W	0.8929	0.0430	2.5087	0.0121	W*
1	0	A	0.7323	0.0693	2.5087	0.0121	
2	0	W	0.6924	0.0713	-0.1845	0.8536	
2	0	A	0.7067	0.0709	-0.1845	0.8536	
0	1	W	0.7587	0.0650	0.4053	0.6852	
0	1	A	0.7249	0.0704	0.4053	0.6852	
1	1	W	0.7698	0.0632	0.3883	0.6978	
1	1	A	0.7321	0.0704	0.3883	0.6978	
2	1	W	0.6363	0.0754	-0.6317	0.5276	
2	1	A	0.6881	0.0744	-0.6317	0.5276	

Table 5.5: ROC Experiment Data for Observer 2.
No data pooling.

Observer 3							
Lungs							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7827	0.0640	2.5205	0.0117	W*
0	0	A	0.5963	0.0789	2.5205	0.0117	
1	0	W	0.6363	0.0960	-0.7300	0.4654	
1	0	A	0.7088	0.0683	-0.7300	0.4654	
2	0	W	0.5517	0.0837	-1.0492	0.2941	
2	0	A	0.6380	0.0737	-1.0492	0.2941	
0	1	W	0.7760	0.0897	-0.3544	0.7230	
0	1	A	0.8074	0.0577	-0.3544	0.7230	
1	1	W	0.6269	0.0955	-1.7126	0.0868	A
1	1	A	0.7915	0.0597	-1.7126	0.0868	
2	1	W	0.8184	0.0782	0.3866	0.6990	
2	1	A	0.7851	0.0619	0.3866	0.6990	

Mediastinum							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7544	0.0635	0.9862	0.3241	
0	0	A	0.6807	0.0705	0.9862	0.3241	
1	0	W	0.8094	0.0618	1.2161	0.2239	
1	0	A	0.7258	0.0685	1.2161	0.2239	
2	0	W	0.6595	0.0733	-0.1898	0.8495	
2	0	A	0.6768	0.0812	-0.1898	0.8495	
0	1	W	0.8078	0.0611	2.2891	0.0221	W
0	1	A	0.6362	0.0766	2.2891	0.0221	
1	1	W	0.6350	0.0771	1.6268	0.1038	W
1	1	A	0.5249	0.0830	1.6268	0.1038	
2	1	W	0.6073	0.0767	-0.3827	0.7020	
2	1	A	0.6405	0.0771	-0.3827	0.7020	

Table 5.6: ROC Experiment Data for Observer 3.
No data pooling.

Observer 1							
Lungs							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.7230	0.0486	-0.5697	0.5689	
0	*	A	0.7555	0.0482	-0.5697	0.5689	
1	*	W	0.8356	0.0385	-0.3939	0.6936	
1	*	A	0.8530	0.0366	-0.3939	0.6936	
2	*	W	0.6891	0.0509	-0.5465	0.5847	
2	*	A	0.7238	0.0500	-0.5465	0.5847	

Mediastinum							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.7807	0.0451	2.2882	0.0221	W
0	*	A	0.6544	0.0534	2.2882	0.0221	
1	*	W	0.7830	0.0470	3.0435	0.0023	W*
1	*	A	0.6260	0.0540	3.0435	0.0023	
2	*	W	0.6564	0.0525	1.2007	0.2299	
2	*	A	0.5903	0.0576	1.2007	0.2299	

Table 5.7: ROC Experiment Data for Observer 1.
Data pooled across observing sites.

Observer 2							
Lungs							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.7442	0.0460	-0.7951	0.4265	
0	*	A	0.7866	0.0435	-0.7951	0.4265	
1	*	W	0.9047	0.0276	1.2076	0.2272	
1	*	A	0.8654	0.0341	1.2076	0.2272	
2	*	W	0.7912	0.0437	0.9378	0.3483	
2	*	A	0.7404	0.0469	0.9378	0.3483	

Mediastinum							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.8192	0.0402	1.2445	0.2133	
0	*	A	0.7576	0.0463	1.2445	0.2133	
1	*	W	0.8301	0.0389	1.7445	0.0811	W
1	*	A	0.7306	0.0495	1.7445	0.0811	
2	*	W	0.6565	0.0521	-0.6566	0.5114	
2	*	A	0.6940	0.0514	-0.6566	0.5114	

Table 5.8: ROC Experiment Data for Observer 2.
Data pooled across observing sites.

Observer 3							
Lungs							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.7945	0.0462	1.7873	0.0739	W
0	*	A	0.7070	0.0495	1.7873	0.0739	
1	*	W	0.6426	0.0694	-1.5216	0.1281	
1	*	A	0.7504	0.0455	-1.5216	0.1281	A
2	*	W	0.6816	0.0594	-0.4427	0.6580	
2	*	A	0.7077	0.0490	-0.4427	0.6580	

Mediastinum							
Type	Site	Proc	Area	σ	n_σ	pval	Result
0	*	W	0.7706	0.0444	2.2180	0.0266	W
0	*	A	0.6515	0.0522	2.2180	0.0266	
1	*	W	0.7187	0.0503	1.8186	0.0690	W
1	*	A	0.6275	0.0547	1.8186	0.0690	
2	*	W	0.6308	0.0531	-0.2332	0.8156	
2	*	A	0.6454	0.0557	-0.2332	0.8156	

Table 5.9: ROC Experiment Data for Observer 3.
Data pooled across observing sites.

Time Data Pooled by Quarters							
Observer 1							
Lungs							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.8236	0.0481	0.1002	0.0775
1	*	*	A	0.7652	0.0608	1.3414	0.0775
2	*	*	W	0.6915	0.0599	0.6713	0.0861
2	*	*	A	0.6651	0.0618	6.7438	0.0861
3	*	*	W	0.7745	0.0531	3.5241	0.0767
3	*	*	A	0.8021	0.0554	1.7291	0.0767
4	*	*	W	0.7518	0.0593	0.1288	0.0917
4	*	*	A	0.5711	0.0700	0.0959	0.0917
Mediastinum							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.6686	0.0643	0.0908	0.0910
1	*	*	A	0.6783	0.0644	1.3628	0.0910
2	*	*	W	0.7959	0.0545	9.9357	0.0839
2	*	*	A	0.7364	0.0638	2.3212	0.0839
3	*	*	W	0.8192	0.0544	0.8691	0.0698
3	*	*	A	0.8432	0.0438	0.5667	0.0698
4	*	*	W	0.7954	0.0563	0.3607	0.0898
4	*	*	A	0.5832	0.0700	9.9191	0.0898

Table 5.10: Observer 1 ROC Data Pooled by Quarters.
Data also pooled by site and lesion type.

Time Data Pooled by Quarters							
Observer 2							
Lungs							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.7487	0.0559	0.8118	0.0813
1	*	*	A	0.7347	0.0591	0.1534	0.0813
2	*	*	W	0.7999	0.0514	2.9206	0.0815
2	*	*	A	0.6917	0.0632	5.4734	0.0815
3	*	*	W	0.8803	0.0403	1.6327	0.0645
3	*	*	A	0.7779	0.0503	0.6259	0.0645
4	*	*	W	0.7939	0.0563	2.3956	0.0821
4	*	*	A	0.7174	0.0597	0.4459	0.0821
Mediastinum							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.8386	0.0464	2.4341	0.0571
1	*	*	A	0.9102	0.0332	7.1943	0.0571
2	*	*	W	0.7449	0.0552	3.4127	0.0776
2	*	*	A	0.7776	0.0545	0.3016	0.0776
3	*	*	W	0.8478	0.0445	0.8789	0.0749
3	*	*	A	0.7457	0.0602	0.9413	0.0749
4	*	*	W	0.7812	0.0538	2.9538	0.0752
4	*	*	A	0.7716	0.0526	0.1451	0.0752

Table 5.11: Observer 2 ROC Data Pooled by Quarters.
Data also pooled by site and lesion type.

Time Data Pooled by Quarters							
Observer 3							
Lungs							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.6965	0.0771	7.0711	0.0971
1	*	*	A	0.7388	0.0590	2.0208	0.0971
2	*	*	W	0.7017	0.0628	3.6967	0.0871
2	*	*	A	0.6645	0.0604	0.2155	0.0871
3	*	*	W	0.8510	0.0541	1.3543	0.0806
3	*	*	A	0.6380	0.0597	4.4030	0.0806
4	*	*	W	0.7434	0.0576	2.1472	0.0866
4	*	*	A	0.7121	0.0646	0.8262	0.0866
Mediastinum							
Quarter	Type	Site	Proc	Area	σ	χ^2	σ_{tot}
1	*	*	W	0.6895	0.0727	2.7665	0.0932
1	*	*	A	0.7272	0.0583	0.8874	0.0932
2	*	*	W	0.6634	0.0654	8.2336	0.0886
2	*	*	A	0.7052	0.0597	0.9830	0.0886
3	*	*	W	0.7330	0.0713	2.4015	0.0926
3	*	*	A	0.7778	0.0591	0.4882	0.0926
4	*	*	W	0.7632	0.0539	0.0120	0.0869
4	*	*	A	0.5850	0.0682	0.8133	0.0869

Table 5.12: Observer 3 ROC Data Pooled by Quarters.
Data also pooled by site and lesion type.

5.4 Experimental Methodology: IQM Validation

The second experiment performed in this work evaluated the predictions of the image quality measure against the results from the first experiment. The set of 1280 test images used for the observer experiment described above were used as input to the image quality measure; for those images which were processed by windowing, the window used was that for the determination of lesion parameters described in Section 5.2.4. For each set of data which described a single experiment (identical image field, lesion type, lesion site, and processing method), the statistics of the probability density functions were determined and an overall IQM and its standard deviation were determined as given in Eqs. 4.21 and 4.24.

5.5 Results: IQM Validation

Preliminary evaluation of the IQM using obvious lesions embedded in real medical images showed that the measure was able to correctly identify the edge features of interest in the image. A full validation attempt was then performed using the images described in Section 5.2.4. The results of these validation runs are shown in Table 5.13. The columnar headings show the parameters of the images used, the means and standard deviations of the two distributions defined in Chapter 4, and the IQM value where it was determinable. As can be seen in the table, in 18 out of the 24 cases, no IQM could be calculated due to the results of calculations of μ_p and μ_a . For these 18 cases, $\mu_p < \mu_a$, contrary to expectation. It would appear that the current experiment is not sufficiently sensitive to differentiate between images which have an inserted lesion and those which do not. In the remaining cases, there are two situations where a calculation of the IQM was possible both for AHE and windowing; in both instances, the IQM differs very little for the two modalities. In order to obtain better statistical accuracy, the data were also pooled across observing sites; the results of these calculations are shown in Table 5.14. No significant improvement is seen by increasing the size of the statistical sample.

5.5.1 Possible Factors Influencing the IQM Calculation

In this section, some of the reasons are examined why the IQM calculation may have failed to yield an acceptable result.

Statistical Accuracy. The matched filtering calculation was carried out for a set of 1280 images; these images were then divided into 24 sets of approximately 50 images per set and used to calculate the numbers in Table 5.13. Of the 50 images in each set, approximately half contained an artificial lesion and half did not. Thus, the size of the

statistical sample is very small; this is reflected in the fact that the standard deviations of the probability distributions are large compared to the difference between their means, in most cases between six and ten times as large. Since there is such a large uncertainty in the determination of the means, it is not surprising that in many cases (here a majority), the mean of the distribution with the lesion present is smaller than that when the lesion is absent. However, increasing the size of statistical sample in Table 5.14 does not result in any significant improvement in the results.

Image Characteristics. The variation in the output of the matched filter depends not only on the presence or absence of the lesion, but also on the structural background of the image into which the lesion may be inserted. It can be seen in Tables 5.13 and 5.14 that the instances where an IQM could be calculated all lie in the mediastinum where the structured background is less complicated than in the lungs; the standard deviations of the results in the mediastinum also tend to be smaller than the equivalent results in the lungs. Since the inherent structural variation in the lungs is large, it is probable that the standard deviation will remain substantial even with an increase in the size of the statistical sample. In addition, the lesions which were used are very small, both in spatial extent and in intensity range. Since only a limited number of intensity levels (256) were available in the image, the lesions necessarily occupied only a very few intensity levels. Both the spatial and intensity resolution of the lesion were degraded by quantization effects. Thus, the majority of the response which is calculated by the IQM is due to the background structure rather than the presence of a lesion. The lesion represents only a minor perturbation of the overall response function. Again, this causes the difference in the means of the two distributions to be small.

In spite of these considerations, a larger sample size would result in an increased confidence in the determination of the mean and the ability to determine whether the failure of the IQM to discriminate between images in which the lesion is absent or present is due to a poor statistical estimate of the parameters or to fundamental difficulties in the current model.

5.5.2 Conclusions

The current experiment was unable to provide an adequate verification of the validity of the IQM due to the considerations outlined above; while the statistical sample is small, it would appear that the current model is insufficiently sensitive to the presence of contrast in the image. A logical next step in evaluating the current model is to repeat the calculations using a larger statistical sample.

IQM Validation

Lungs							
Type	Site	Proc	μ_a	σ_a	μ_p	σ_p	IQM
0	*	W	0.4377	0.3214	0.3939	0.2719	NA
0	*	A	1.1513	1.0681	0.9829	0.8331	NA
1	*	W	0.4300	0.3325	0.3859	0.2414	NA
1	*	A	1.1360	1.0589	0.9239	0.8069	NA
2	*	W	0.3379	0.2355	0.3021	0.1909	NA
2	*	A	0.7936	0.6590	0.6379	0.5479	NA

Mediastinum							
Type	Site	Proc	μ_a	σ_a	μ_p	σ_p	IQM
0	*	W	0.9466	0.3418	0.9605	0.3356	0.2476
0	*	A	0.7888	0.2474	0.7788	0.2336	NA
1	*	W	0.9358	0.3466	0.9716	0.3482	0.2644
1	*	A	0.7811	0.2461	0.7635	0.2379	NA
2	*	W	2.1907	0.7371	2.1783	0.7000	NA
2	*	A	1.9576	0.6030	1.9925	0.6191	0.2477

Table 5.14: Results of IQM Validation Experiment
Pooled by site.

IQM Validation							
Lungs							
Type	Site	Proc	μ_a	σ_a	μ_p	σ_p	IQM
0	0	W	0.5103	0.3723	0.4438	0.3154	NA
0	0	A	0.9691	1.0138	0.8698	0.9704	NA
0	1	W	0.3728	0.2580	0.3510	0.2254	NA
0	1	A	1.3475	1.1096	1.0920	0.6754	NA
1	0	W	0.5105	0.3730	0.4439	0.2879	NA
1	0	A	0.9312	0.9926	0.7790	0.8991	NA
1	1	W	0.3495	0.2694	0.3299	0.1740	NA
1	1	A	1.3645	1.1025	1.0688	0.6893	NA
2	0	W	0.4023	0.2605	0.3642	0.2187	NA
2	0	A	0.6377	0.6080	0.6349	0.6305	NA
2	1	W	0.2823	0.1998	0.2400	0.1362	NA
2	1	A	0.9820	0.6813	0.6414	0.4463	NA
Mediastinum							
Type	Site	Proc	μ_a	σ_a	μ_p	σ_p	IQM
0	0	W	0.9019	0.2652	0.8836	0.2716	NA
0	0	A	0.7922	0.2351	0.7449	0.2071	NA
0	1	W	0.9930	0.4068	1.0318	0.3766	0.2830
0	1	A	0.7848	0.2658	0.8114	0.2562	0.2885
1	0	W	0.8655	0.2811	0.9190	0.2959	0.2778
1	0	A	0.7413	0.2189	0.7394	0.2209	NA
1	1	W	1.0145	0.3990	1.0264	0.3940	0.2500
1	1	A	0.8208	0.2689	0.7859	0.2546	NA
2	0	W	2.2097	0.6714	2.1387	0.5922	NA
2	0	A	1.9851	0.5234	1.9562	0.5392	NA
2	1	W	2.1708	0.8134	2.2180	0.8031	0.2788
2	1	A	1.9300	0.6825	2.0274	0.6960	0.2778

Table 5.13: Results of IQM Validation Experiment
No data pooling.

Chapter 6

Summary and Directions for Further Work

*Se no è vero
Ma è ben trovato.*

—Giordano Bruno [1548-1600]

6.1 Summary of the Current Work

The work presented in this dissertation can be classified into two areas: an evaluation of the effectiveness of a particular adaptive contrast enhancement method, AHE, and the development of an objective measure of image quality. In this section, the results of these efforts are briefly reviewed.

Effectiveness of AHE. Adaptive Histogram Equalization (AHE) was investigated as a promising method for performing contrast enhancement on digital images. Empirical studies showed that AHE is effective in all image areas and produces few artifacts when used in a reasonable manner. Observer studies were then carried out to compare the effectiveness of AHE with interactive global min-max windowing in presenting low contrast lesions. Three trained observers were shown computed tomography images into some of which artificial lesions had been inserted. The observers were allowed to perform arbitrary interactive windowing on the images; they were also shown the same images after processing with AHE. In all cases, they were asked to rate their confidence that a lesion was present in a given image. Their answers were then analyzed using the Receiver Operating Characteristic (ROC) method to allow the comparison of their performance using the two methodologies. A number of experiments were carried out simultaneously, testing the contrast enhancement methods for two fields (the lungs and mediastinum) of the images using a variety of artificial lesions. There were 1280 observations for each observer in the test set, grouped into 24 experiments. It was found that there was essentially no difference in observer performance using the two contrast enhancement methodologies, though the observers performed slightly better using interactive windowing in the mediastinum field.

Thus, since AHE is fast, repeatable, requires no manual intervention, and is able to depict contrast in many areas of the image simultaneously, it offers an excellent alternative to interactive windowing as a diagnostic technique.

Development of an Objective Image Quality Measure. An Image Quality Measure (IQM) was developed to allow the objective measurement of image quality for different contrast enhancement methods without the use of observer studies. The IQM was based upon models of visual processing. The desired characteristics of these models were that they should be sensitive to edge features in the image and that the response of the model be scale-invariant over a wide range of scales. The Koenderink-van Doorn model for retinal sensors was used to predict the response of the retina to varying illumination patterns; this model considers the retina as overlying mosaics of self-similar sensor arrays, each sensor being sensitive to edges in the illumination field having a scale similar to its self. The sensor mosaic yields a two-dimensional response function indicating the sensitivity of the mosaic to features present in the retinal plane. This response function was then compared to the response function of the lesion which the observer would expect to see using a matched filtering operation. The IQM was defined as the probability that the observer would correctly determine the true state of the image, that is whether the lesion was in fact present or not.

The IQM was evaluated using the images developed for the observer studies previously described; each image was used as input for the IQM program with parameters for the inserted lesion supplied as in Chapter 5. The results were inconclusive in that the IQM was unable to reliably distinguish between images in which the lesion was present and images in which it was absent; it was often found that the value of the IQM averaged over a set of images was less for images which contained a lesion than for those which did not. It was concluded that this was the result of a number of factors: 1) the statistical accuracy of the evaluation was limited due to the small number of images used for the verification of the IQM, 2) the intrinsic variation of the background structures in the images caused the IQM to have a wide variation, and 3) the lesions used for the evaluation were very near the limits of detectability and hence pixel artifacts may have caused the IQM to be insufficiently sensitive to the presence of the lesion contrast in the image. No conclusions could be drawn as to the appropriateness of the current approach to the definition of an Image Quality Measure; further studies must be done to increase statistical accuracy.

6.2 Future Directions

In this section, a brief outline is given for pursuing the current work in the future.

Further Development of AHE. The current results are sufficiently promising that a full clinical trial of AHE is appropriate. Since AHE has shown itself to be able to perform equivalently to windowing even in the case of lesions which are much more subtle than those commonly encountered in clinical practice, AHE should be offered as an alternative methodology for clinicians to use in their daily work, in which at most institutions diagnosis is made from film in which a technician-chosen or preset window has been used. A careful study should be made of the physicians' performance in these clinical circumstances. The results of the experiments previously described do not indicate that the physician will require substantial training in order to utilize the new methodology; however, in a few of the many cases which have been examined by clinicians in our laboratories, the effect of AHE was to produce structures which if interpreted as if they were in a windowed image appear to indicate a pathological condition when in fact none is present. No further observer studies with artificial lesions are indicated at this time; since it is possible to conceive an infinite series of such studies, it is probable that daily clinical use will be of more utility in determining the clinical viability of AHE.

Another direction which should be pursued is the investigation of modified AHE methods. While AHE has not proven to produce inferior diagnostic results, aesthetic objections have been raised to the appearance of images produced by AHE; in particular, the obvious appearance of noise in AHE'd images is found to be annoying by some physicians. A version of AHE which limits the amount of contrast enhancement so that the noise is not as prominent in homogeneous regions (*Noise-limited Adaptive Histogram Equalization*) is currently under investigation by Pizer and his collaborators. If this alternative methodology proves attractive, observer studies may be an appropriate next step. A comparison of NLAHE with AHE would serve to illustrate whether diagnostic ability has been changed by limiting the amount of contrast enhancement.

Further Development of the IQM. The IQM was shown in preliminary studies to possess many desirable characteristics; it was able to correctly determine the edge features of interest in the image. However, it was not as sensitive as the human in detecting the subtle lesions used in the observer experiments. An immediate first step in pursuing the question of whether the current visually based approach to defining the IQM is efficacious will be to prepare a much larger sample set to minimize the statistical effects of the

small sample size in the current experiment. It can then be determined if the intrinsic variation in image structure is the limiting factor in the ability of the current IQM to discriminate between contrast enhancement methods. In either case, attention can then be paid to improving the model to incorporate more of our knowledge of visual processing, in particular as regards image structure. Clearly, there are many mechanisms of the visual system that have been ignored in the current model; it would appear that any satisfactory measure of image quality must be based upon the way humans process visual information.

Via oviclipitum dura est.

—Adlai Stevenson [1900–1965]

*One thing I have learned in a long life:
that all our science, measured against reality,
is primitive and childlike— and yet
it is the most precious thing we have.*

—Albert Einstein [1879–1955]

References

- Abramson, N. 1963. *Information Theory and Coding*, McGraw-Hill, New York.
- Andrews, H. C. 1982. *Interactive Real-time Adaptive Nonlinear Convolution*, Comtal Corporation, Pasadena, California.
- Armstrong, J. D., J. A. Sorensen, J. A. Nelso, I. Tocino, P. D. Lester, J. O. Janes, L. T. Niklason, and W. Stanish. 1981. *Clinical Evaluation of Unsharp Masking and Slit Scanning Techniques in Chest Radiography*, University of Utah School of Medicine.
- Barrett, E. B. 1976. "Image Enhancement by Retinal Operators," *Proceedings of the Conference on Image Science Mathematics, November 10-12, 1976, Monterey, California*, C. Wilde and E. Barrett (eds.), 19-23 .
- Burgess, A. E., R. F. Wagner, and R. J. Jennings. 1982. "Human Signal Detection Performance for Noisy Medical Images," *Proceedings of the International Workshop on Physics and Engineering in Medical Imaging, Asilomar, March 10-18 1982*, IEEE Computer Society, 99-105.
- Burgess, A. E. 1983. *Observer Performance Testing for Medical Imaging*, American Association of Physicists in Medicine. Notes for A.A.P.M. Refresher Course.
- Castleman, K. R. 1979. *Digital Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Cormack, J. and B. F. Hutton. 1980. "Minimisation of Data Transfer Losses in the Display of Digitised Scintigraphy Images," *Physics in Medicine and Biology*, **25**, 271-282.
- Cormack, J. and B. F. Hutton. 1981. "Quantitation and Optimization of Digitized Scintigraphic Display Characteristics Using Information Theory," *Medical Image Processing: Proceedings of the VIIth International Meeting on Information Processing in Medical Imaging, 22-26 June, 1981, Stanford, California, Stanford University*, 240-263.
- Dorfman, D. D. and E. Alf. 1969. "Maximum-Likelihood Estimation of Parameters of Signal-Detection Theory and Determination of Confidence Intervals - Rating Method Data," *Journal of Mathematical Psychology*, **6**, 487-496.
- Driscoll, E. C. and C. Walker. 1981. "Evolution of image processing algorithms from software to hardware," *Display Technology II*, **271**, 43-50, Society of Photo-Optical Instrumentation Engineers.
- Driscoll, E. C. and C. Walker. 1983. "Image Enhancement Using A General Purpose Image Processor," *Electronic Imaging*, **2(5)**, 52-56.

- Fahnestock, J. D. and R. A. Schowengerdt. 1983. "Spatially variant contrast enhancement using local range modification," *Optical Engineering*, **22**(3), 378-381.
- Fu-Tse, K. 1895. "The Analects," *Sacred Books of the East*, J. Legge (ed.), Oxford University Press, Oxford. J. Legge Translation.
- Garibotto, G. 1984. "Adaptive and Interactive X-ray Enhancement Techniques: Comparison of Performances," *Proceedings of the Digital Signal Processing Conference*, Florence, Italy.
- Green, D. M. and J. A. Swets. 1974. *Signal Detection Theory and Psychophysics*, Robert E. Krieger Publishing Company, New York.
- Hall, E. L. 1979. *Computer Image Processing and Recognition*, Academic Press, New York.
- Harris, J. L. 1977. "Constant variance enhancement: a digital processing technique," *Applied Optics*, **16**(5), 1268-1271.
- Hummel, R. A. 1975. "Histogram Modification Techniques," *Computer Graphics and Image Processing*, **4**, 209-224.
- Hummel, R. A. 1977. "Image Enhancement by Histogram Transformation," *Computer Graphics and Image Processing*, **6**, 184-195.
- Ketcham, D. J., R. W. Lowe, and J. W. Weber. 1976. "Real-time Image Enhancement Techniques," *Seminar on Image Processing*, Pacific Grove, California, Hughes Aircraft Company, 1-6.
- King, M. A., T. R. Miller, D. A. Jacobs, P. W. Doherty, K. S. Sampthkumaran, and K. L. Johnson. 1983. *Nonstationary Image Processing in the Frequency Domain*, Department of Nuclear Medicine, The University of Massachusetts, Worcester, MA.
- Koenderink, J. J. and A. J. van Doorn. 1978. "Visual Detection of Spatial Contrast; Influence of Location in the Visual Field, Target Extent and Illuminance Level," *Biological Cybernetics*, **30**, 157-167.
- Koenderink, J. J. 1984. "The Structure of Images," *Biological Cybernetics*, **50**, 363-370, Springer-Verlag.
- Lee, J. 1980. "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2**(2), 165-168.
- McDonnell, M. J. 1981. "Box-Filtering Techniques," *Computer Graphics and Image Processing*, **17**, 65-70.
- Medawar, P. B. 1982. *Pluto's Republic*, Oxford University Press, London.
- Metz, C. E. 1978. "Basic Principle of ROC Analysis," *Seminars in Nuclear Medicine*, **8**(4), 283-297.
- Metz, C. E., P. Wang, and H. B. Kronman. 1983. *Statistical Issues in ROC Analysis of Medical Imaging Procedures*, Preprint.

- Metz, C. E., P. Wang, and H. B. Kronman. 1984. "A New Approach for Testing the Significance of Differences Between ROC Curves Measured from Correlated Data," *VIIIth Conference on Information Processing in Medical Imaging: Brussels, Belgium, 29 August - 2 September, 1983*, F. Deoninck (ed.), Martinus Nijhoff, The Hague.
- Narendra, P. M. and R. C. Fitch. November, 1981. "Real-Time Adaptive Contrast Enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(6), 655-661.
- Peli, T. and J. S. Lim. 1982. "Adaptive Filtering for Image Enhancement," *Optical Engineering*, 21(1), 108-112.
- Pirsig, R. M. 1974. *Zen and the Art of Motorcycle Maintenance*, William Morrow & Company, New York.
- Pizer, S. M. 1981. "Intensity Mappings for the Display of Medical Images," *Functional Mapping of Organ Systems: 11th Annual Symposium on the Sharing of Computer Programs and Technology in Nuclear Medicine, 5-8 February, 1981*, New Orleans, Louisiana, P. D. Esser (ed.), Society of Nuclear Medicine, 205-218.
- Pizer, S. M. 1981. "Intensity Mappings to Linearize Display Devices," *Computer Graphics and Image Processing*, 17, 262-268.
- Pizer, S. M. 1981. "An Automatic Intensity Mapping for the Display of CT Scans and Other Images," *Medical Image Processing: Proceedings of the VIIth International Meeting on Information Processing in Medical Imaging, 22-26 June, 1981*, Stanford, California, Stanford University, 276-309.
- Pizer, S. M., J. B. Zimmerman, and E. V. Staab. 1984. "Adaptive Grey Level Assignment in CT Scan Display," *Journal of Computer Assisted Tomography*, 8(2), 300-305.
- Pizer, S. M., J. J. Koenderink, L. M. Lifshitz, L. Helmink, and A. D. J. Kaasjager. 1985. "An Image Description for Object Definition Based on Extremal Regions in the Stack," *Proceedings of Information Processing in Medical Imaging IX, June 10-14, 1985*, Washington, D.C..
- Plato 1928. "Phaedrus," *The Works of Plato*, I. Edman (ed.), The Modern Library, New York, 263-332. Jowett Translation.
- Rutherford, E. 1967. *The Mathematical Approach to Biology and Medicine*, John Wiley and Sons, New York.
- Schwartz, A. A. and J. M. Soha. 1977. "Variable threshold zonal filtering," *Applied Optics*, 16(7), 1779-1781.
- Shakespeare, W. 1971. "Hamlet," *The Complete Works of Shakespeare*, I. Ribner and G. L. Kitteridge (eds.), Xerox, Lexington, Massachusetts, 1042-1103.
- Sorensen, J. A., L. T. Niklason, and J. A. Nelso. 1981. "Photographic Unsharp Masking in Chest Radiography," *Investigative Radiology*, 16(4), 281-288.

- Sorensen, J. A., J. D. Armstrong, L. T. Niklason, and J. A. Nelso. 1981. "Letter to the Editor," *Investigative Radiology*, **16**(6), 529-530.
- Stevenson, R. L. 1922. "El Dorado," *Virginibus Puerisque*, **2**, Charles Scribner's Sons, New York, *Works of R. L. Stevenson*.
- Swets, J. A. and R. M. Pickett. 1982. *Evaluation of Diagnostic Systems*, Academic Press, New York.
- Todd-Pokropek, A. E. 1976. "Image processing in nuclear medicine: an examination of the quest for a measure of clinical quality," *7th L.H. Gray Conference: Medical Images*, 278-292.
- Van der Wildt, G. J. and R. G. Waarts. 1983. "Contrast Detection and Its Dependence on the Presence of Edges and Lines in the Stimulus Field," *Vision Research*, **23**(8), 821-830, Pergamon Press, Ltd..
- Wallis, R. 1976. "An approach to the space variant restoration and enhancement of images," *Proc. Symp. on Current Mathematical Problems in Image Science*, Monterey, California, Naval Postgraduate School.
- Zuidema, P. 1984. "Development and Present Status of the Quantum Concept in Visual Psychophysics," *Limits in Perception*, A. J. van Doorn, W. A. van de Grind, and J. J. Koenderink (eds.), VNU Science Press, Utrecht, The Netherlands, 5-48.

Appendix 1

AHE-Windowing Observer Study (ZDOS)

This document is designed to acquaint you with the purpose and conduct of the observer study for comparison of linear min-max windowing (herein, *windowing*) and Adaptive Histogram Equalization (AHE). Please read carefully. In order for the study to derive valid results, it's necessary that all the observers start with the same aims in mind. If everyone is looking for something different when observing the images, it's likely they'll all find it.

This document is in four parts. The first section describes in brief what you have to do and what you'll see when you sit down in front of the monitor. A fuller description of the study appears in the second section, which also discusses the purpose of the observer study and what we hope to find out. The third section tells exactly what you should be looking for when conducting the trials and how to formulate your judgements; this section constitutes the actual definition of the experiment. The fourth section details the way you actually go about manipulating the computer, the terminal, and the data tablet in the course of the experiment and includes a checklist of what to do if (when) things go wrong.

1.1 The Experiment in Brief

During each experimental session, the observer will be presented with a series of clinical CT scans of the chest in the region of the mediastinum. To about half of these images, a simulated lesion has been added, either in the mediastinum or the lungs. The other images have had no lesion added and are, we hope, normal. In either case, crosshairs are superimposed on the image, and the observer is asked to express his confidence that a simulated lesion is present at the location of the crosshairs. A copy of the simulated lesion is present in one corner of the screen, so the observer knows exactly what he is supposed to be looking for. The observer's response takes the form of a number in the range 0 to 4, with 0 meaning that the lesion is certainly not present and 4 meaning that there is a good probability that the lesion is present. The criteria for deciding on a rating for a particular image is further discussed in Section 3.

The images which are presented are in one of two forms: either the image has been processed with the AHE algorithm or it is the raw CT data. In the former case, the observer need only look at the image and make a rating of his confidence of the lesion presence. In the later case, the observer is asked to do windowing on the image before observing. The observer is told which field (lungs or mediastinum) is to be windowed.

Our objective is to see which contrast enhancement technique produces the better performance by the observer in reading the images over a large number of trials. Because the abilities of observers differ and because we can't count on any one image (or small group of images) to be "typical", it is necessary to use a large number of readings to get good statistical accuracy.

1.2 Description of the Study

1.2.1 What the study is designed to show

This observer study is designed to compare two methods of contrast enhancement, windowing and AHE, in situations which are not enormously different from those encountered in clinical practice. The principal result that we hope to derive is a measure of the effectiveness of AHE as compared to windowing in a semi-realistic situation. AHE has been shown to be quite effective in informal studies; the current experiment is designed to formally evaluate that effectiveness.

If the results of this experiment show that for the particular task of finding subtle lesions in the lungs and mediastinum AHE performs as well or better than windowing, then this will be incentive to further develop AHE in order to build a clinically useful tool. Unfortunately, this study can't imply that AHE would have performed equally well for other imaging tasks or modalities. A good performance here only says that AHE performed in a particular way for a particular task. Nevertheless, a good performance on both the lungs and mediastinum would be quite interesting, particularly since for these images multiple windows are typically required to view all the regions of diagnostic interest.

This study will also provide data for other investigations I'm carrying out in the area of automatically evaluating image quality. The image quality measure that I'm developing as another part of my dissertation will be compared to the results of the observer study in order to calibrate its performance against that of real observers.

1.2.2 Theoretical basis of the study

The evaluation of the observer data obtained in this study will be done using the Receiver-Operator Characteristic (ROC) technique. This method is a powerful tool for discriminating between the effectiveness of two diagnostic methods. The analysis of the data is done by comparing the observer's confidence rating against the known "truth" of the situation, *i.e.*, whether the simulated lesion is in fact present or not on a given trial. This requirement that the truth be known in advance is one reason why simulated lesions are used rather than cases with actual clinical pathology.

The ROC technique yields a measure of goodness for each of the two modalities and a statistic describing the uncertainty in the results. This uncertainty is the reason a large number of trials must be used. For more trials, the uncertainty is smaller.

1.2.3 Limitations

Notice that while every attempt has been made to simulate the clinical situation as closely as possible, it is inevitable that there are areas of the study that are not very realistic. The lesions are simulated and of far too regular a shape to be realistic. The locations and appearances of the lesions are not what would be expected in practice. The clinical task (that of simply detecting a known lesion in a known location) is unrealistic. These approximations to reality are unavoidable if the data is to be analyzable. In many cases there is evidence that the task to be performed is well correlated with the more difficult situations encountered in the clinic; hence the results will be related more closely to reality than might first be apparent. In any case, an increase in realism would be desirable, but if any sense is to be made of the data, tradeoffs such as those used in this experiment are necessary. The observer has to be careful not to let the lack of clinical realism change the task which he is to perform; the observer's task is very specific (as will be detailed below) and should not be affected by the lack of realism which may occasionally be apparent.

1.2.4 Frustration

Because we are operating at the limit of what the visual system can detect, the simulated lesions are of necessity not very prominent. If the lesions were obvious, then no statistical differentiation of the two methods would be possible; every call by the observer would be correct. We have found in previous studies that the correct lesion intensity to give good statistical results is so small that at first glance the lesion is essentially invisible. This can lead to frustration on the part of the observers, who feel they are operating

by complete guesswork; they can't see anything! Nevertheless, the visual system is even better at it's work than we think it is; we have found that even when the observer is convinced that his answers are completely random, a considerably higher percentage of correct calls are made than could be accounted for by chance. Thus, the observer should relax and make the best call he can without being overly concerned by his indecisiveness. You are meant to get a proportion of these wrong. The lesions are really there half of the time, though they may appear not to be.

1.2.5 Terminology

For the purpose of this study, we'll define the following terms:

study or experiment: The totality of this work. Either of these words refers to the whole experiment.

run: A single session of observing. The experiment consists of 20 runs, each lasting one hour.

trial: A single image seen during a run. A run consists of 64 trials; the trial is the basic unit of data. For each trial, the observer gives a *call*, which is the rating of the observer's confidence that the lesion is present. Altogether then, there are 1280 trials in the study.

1.2.6 What you'll see

This section describes the observer experiment in considerable detail. First, we'll examine more closely how the images to be presented were prepared; second, we'll describe what you'll see when you sit down at the observing workstation.

Facts about the set of trial images. For each trial, an image is presented on the video monitor. The observer is to evaluate the likelihood that the simulated lesion is present. We start with some facts about how the trials are put together.

The case sample (trials) that you'll see started with 24 base images, selected to have a good likelihood of being normal (no apparent pathology). Each image has two *fields*, the lungs and mediastinum. In each field of each base image, two *sites* were chosen where artificial lesions could be inserted. These sites were selected as places where real lesions could possibly appear. Various simulated lesions were constructed which could be inserted at the sites. Each trial then is an image constructed from a particular base image, with a particular site in one of the fields chosen for the possible insertion of a simulated lesion.

One of the lesions is chosen and then either inserted or not. The image is then either processed with AHE or left alone. Thus, associated with each trial is a set of parameters which describe which base image was used, which site was chosen, etc.

All the possible combinations of these parameters were generated, describing all the images it is possible to create. The possible images were then created in a random order. Approximately ten percent of the resulting set of images were chosen at random and included in the test set twice, thus giving some redundancy. This allows a check of the consistency of an observer with himself, since he sees some of the images in the test set twice. It is important to note that the order in which the images are presented to the observer is completely random; you may see a dozen non-AHE images followed by a dozen AHE images as a result of random chance. The random ordering helps reduce the probability of introducing bias into the results.

Interacting with the equipment. The observing workstation consists of a Comtal frame buffer attached to a VAX11/730 computer along with various peripheral devices that allow you to interact with the program running the observer studies. This program is named `zdos`, and it has been designed to be as robust as possible; that is, it's not supposed to fail in a disastrous way while you're doing the experiment. To start the a run of the observer study, you invoke `zdos` (in much the same way that one invokes `daemons` and `evil spirits`, by a mixture of arcane incantations and foul language). Once `zdos` is running, you'll be confronted by a video monitor on which the images will appear, an ordinary CRT computer terminal, and a data tablet with a movable puck. Most of the things you have to do during the run are controlled by the data tablet and puck; you push buttons on the puck to start the observation of each trial, and the windowing of an image is controlled by the position of the puck on the tablet. The terminal screen reveals various information about the progress of the run, including information about the current window being used and menus describing the functions of the various puck buttons. The terminal keyboard is used only to start the run. Most of this will be self-explanatory after a little practice, but details are given in the last section of this document, the Observer's Cookbook. One important point: essentially nothing happens until you give a positive go-ahead signal by pushing a button on the puck or a key on the keyboard. The experiment won't run off and leave you. After every trial, you'll be given the opportunity to rest for a moment and collect your wits.

How it goes per trial--an AHE trial. zdos manages the various devices and controls the display of images on the screen. When started, zdos performs various initialization routines and then starts presenting trials to the observer. At each trial, a new image is displayed on the monitor screen. This image will either have been processed with AHE or will appear as the raw CT data. Each trial can then be divided into two periods: a *windowing* period and an *observing* period. If the image has been AHE'd, the windowing period of the trial is omitted; only the observing period is performed. When you're ready to start the observing period, you push button 1 on the puck. A green cursor appears on the screen at the location of the potential lesion. By pushing button 1 again, you can make the cursor disappear. On the screen, you see the image, the cursor, and in the upper left corner of the screen, a picture of the lesion that may have been inserted into the image. Thus, you not only know where to look but what to look for. This type of experiment is called a signal known exactly (SKE) experiment.

At this point, two buttons on the puck are active. Button 1 toggles the green cursor on and off so you can get a good look at where the lesion is supposed to be. Button 4 signals that you're ready to give your score. During the observing period, you're trying to determine your confidence that a lesion is present at the designated site. There's a time limit on how long you get to observe (15 seconds). If you don't press button 4 before the 15 seconds elapses, zdos will black the screen anyway and ask for your score. Button 4 is if you finish making your decision early and don't want to wait for the timer to expire. When you're ready, you should go ahead and push button 4 rather than waiting until the timer expires.

When the screen goes black (after the timer expires or the ready button is pushed), you will be prompted to enter your score. The data tablet is used to indicate your call; on the data tablet is an overlay which is divided into six regions. Five of the regions are numbered with the values 0 through 4; the sixth region is labeled "OK". To enter your score, you move the data tablet puck into the appropriate area on the data tablet surface and press any button. The score you have selected will appear on the Tektronix monitor screen. You can change your mind by moving the puck to another selection and pressing a button. Before zdos will accept the score, you have to confirm that it's your final selection by moving the puck into the "OK" area and pressing a button. Thus, the sequence you use for entering your score is 1) Place the puck into the appropriate score area and press a button; 2) Move the puck to the "OK" area and press a button to confirm the score. There is no time limit on how long you can take to enter your score. The score you select

will be shown on both the Tektronix display monitor and the HP terminal. After entering the score, you're ready to start the next trial.

At this point, you're given the option to terminate the session early. Button 3 quits the run. You can proceed with the next trial by pressing button 1. If you terminate early, zdos will automatically pick back up at the place you quit the next time you come in for a run.

How it goes—a windowing trial. For trials which have not been processed by AHE, there is a windowing period before the observing period. During this time, you must select a window before you are allowed to look for the lesion. When the image comes up on the screen, it will be immediately apparent if the image needs to be windowed. You will be told on the terminal screen which field of the image you should window. When you're ready to window, hit button 1 on the puck. The windowing is of the center-width variety; the position of the puck on the tablet surface controls the center and width of the window. As you move the puck in the top-bottom direction on the surface of the tablet, the width of the window is changed; moving the puck toward the bottom narrows the window, to the top widens the window. Moving the puck from left to right controls the center of the window. Moving the puck to the left moves the center of the window toward the lower intensities, to the right toward the upper intensities. On the HP terminal screen, the current values of the window parameters are shown at all times.

You'll notice that the green cursor is not present on the screen. You won't know where the prospective lesion will be beyond a general area; thus the window you choose should be optimal for examining all of the field (*e.g.*, the mediastinum) simultaneously. The lesion picture which was in the upper left corner has also been covered up, either with a white or green patch. Thus, you don't know the appearance of the lesion at this time either.

Again, all the action is controlled by the data tablet and puck. Now button 1 is inactive, while button 2 "freezes" the windowing. If you press button 2, the puck no longer controls the window, which freezes where it was when the button was pushed. Thus, you can pause during windowing to examine a specific window more closely or freeze your final choice. A timer is again running; you have 30 seconds to choose the window. If you have finished choosing the window before the time elapses, push button 4 to proceed. At this point the observation sequence starts just as for the AHE trial. When you push button 1 to start observing, the green cursor will appear and the lesion image will appear in the upper left corner. Everything then proceeds as before.

Learning curve. This may all seem rather mystifying, but it will rapidly become clear in practice. At all stages, zdos prompts for the correct input on the terminal screen. As a last resort, see Section 4, the Observer's Cookbook. In the next section, we discuss what you should look for at each trial and explicate the instructions to observers. The next section is rather formal, since it is important that the ground rules be clearly spelled out.

1.3 Experimental Procedure

This section contains the formal specification for this observer study. It is important that the observer continually keep in mind the criteria for performing the experiment outlined in this section. If the observers are using different criteria, the results of the experiment may be considerably degraded. The observer has control over the following areas of the experiment: the physical environment in which the experiment takes place, the choice of a linear min-max window for those images which are to be windowed, and the rating assigned to the observer's confidence that the lesion is or is not present in the images.

1.3.1 Physical Environment

It is essential that the physical environment be as constant as possible across observing sessions. The following areas are especially important.

Lighting conditions. A standard configuration of room lights is to be used in the experiment. A light box has been installed behind the Tektronix display monitor. Only this light source should be on during the experiment. All other room lights and lamps should be extinguished.

Light adaptation. Since the experiment is carried out in subdued lighting, it is important that the observer be acclimatized to the lighting environment before beginning the experiment. The observer should allow at least one minute to elapse from the time the room lights are extinguished before beginning the observer experiment.

1.3.2 Windowing

For each trial which is to be windowed, the observer will be directed which field (lungs or mediastinum) of the image to window. Since the observer does not know exactly where the simulated lesion may be located, a general window appropriate for the given area should be selected. The general areas which may be specified are "left lung", "right lung", or "mediastinum". In this case, "left" and "right" refer the anatomical position of the organ; thus the right lung is on the left side of the screen and the left lung on the

right. This is in accordance with the usual CT practice. The simulated lesion will lie immediately in or at the periphery of the organ in question. Again, it is important not to select a window which will emphasize a particular area of the organ at the expense of other areas, even if the observer thinks he sees a lesion. Within this constraint, the chosen window should conform as closely as possible to good clinical practice.

1.3.3 Choice of Rating Values

At each trial, the observer will be asked to evaluate his or her confidence that a lesion is present at the designated site. To aid in this determination, removable crosshairs will be present at the center of the location of the possible lesion. A representation of the lesion in isolation is also given in the upper left corner of the monitor screen. For windowed images, both the lesion representation and the image itself will be affected by the chosen window. For AHE images, the lesion representation has been prescaled to approximately conform to the intensity of the lesion which may be present in the AHE image. Both the cursor and the lesion representation should be used in determining the confidence rating.

In examining the image for the lesion, the following criterion should be kept in mind: the clinical realism of the situation is not at issue here. The observer's task is to evaluate the likelihood that the lesion is present, regardless of the probability that a true clinical lesion of the given type would be present at the given location.

After each image has been windowed (if necessary) and observed, the observer will be asked to express his or her confidence that the lesion was present. This rating takes the form of a number in the range 0 to 4 inclusive; thus there are five possible rating categories. In assigning a numerical rating to his or her confidence, the observer should use the following criteria:

<i>Rating</i>	<i>Confidence</i>
0	Definitely not present
1	Probably not present
2	Possibly not present
3	Possibly present
4	Probably present

Note that there is no equivocal category. Thus in each case the observer must decide which confidence is stronger, the presence or non-presence of the lesion. The

criteria enumerated above should be applied at every decision as consistently as possible. Notice also that the rating scale is not symmetric; there is no category entitled "Certainly present". It is not expected that such a category would be much used; if the observer's opinion is that the rating should be "Lesion certainly present", he should choose rating value 4. The observer is expected to use every rating category; answers should not be restricted to a subset of the possible ratings. That is, the observer should choose his answers such that the trials he is most confident about should be rated as 4; those he is least confident about as 0.

1.3.4 Conduct of the Experiment

The results of this observer study will be published in the open literature as a portion of a doctoral dissertation. With the permission of the observer, acknowledgement will be made of the observer's contribution to the dissertation. However, the performance of the individual observer in these trials will be held in strictest confidence. Published references will not identify any observer with any result by name.

1.4 The Observer's Cookbook

This last section contains the nitty-gritty of what the observer actually has to do to conduct a run of the study. There are three sections. In the first, we describe the basic stuff about logging in to the computer, running *zdos*, etc. The second section gives some information about things the observer should never have to worry about—just in case. Finally, the third section is a checklist of things that should be done at every run.

1.4.1 Getting Started

In order to do a run, the observer has to arrange the physical environment, log in to the UNCPACS VAX, start *zdos*, and observe the trial images. Let's consider each of these in turn.

The physical environment. Since a controlled environment is necessary to provide for the reproducibility of results from run to run, it's necessary to set up a few things before starting a run. First, you should place the prepared sign on the outside of the door to the VAX room that notifies people that an observing run is in progress. This (hopefully) will prevent you from being disturbed while you're working. Next, there is a light box positioned behind the Tektronix display monitor. This box should be turned on. All the other room lights should be extinguished. Finally, make sure the Tektronix monitor controls are at their preset condition. There are two knobs on the front of the monitor "Black Level" and "Contrast". Both knobs should be in the fully

counter-clockwise position (there are detents in the control knobs so you'll know when you're there).

Both the VAX and Comtal should be running. Listen for the fans in the Comtal to make sure it's powered up; if it's not, press the red button on the top right corner of the cabinet and listen for the fans. In any case, you should reset the Comtal to it's initial condition. This is done by typing on the Comtal keyboard (large white boxy thing) the sequence `~r` (that's "tilde r"). To type tilde, you have to hold down the shift key. A green star should appear at the lower left corner of the Tektronix monitor; this indicates all is probably well.

Logging in. UNCPACS is a VAX11/730 running the UNIX (trademark of Bell Laboratories) Operating System. You should have to interact very little with the operating system in order to run `zdos`. A special login to the system has been installed for the use of observers. The Hewlett-Packard terminal adjacent to the Tektronix monitor should be used as your control terminal. To log in, make sure that the HP terminal is powered on (a blinking white line on the screen is the cursor; you should be able to see this), and hit *RETURN* on the keyboard. The terminal should respond with the login banner:

UNC Department of Radiology (uncpacs)

login:

Respond to the login prompt by typing "observer". This is the login id for the observer study. No password is required.

A number of strange messages will appear on the terminal; after a minute or two, you should get the system prompt:

p%

At this point, you're logged in and ready to start `zdos`. In response to the prompt, type

p% zdos

`zdos` will then start and you can proceed with the current run.

Running zdos. When zdos starts, it will clear the screen, put up various status items, and ask for your observer id. This is so it can figure out what run it should do next. Type in your last name in response (all small letters). For example, I would type in "zimmerman" as my observer id. If you make a mistake, zdos will reject your login; don't worry, you get three chances to enter it correctly. zdos will then calculate the necessary parameters for this run and ask you whether these are acceptable:

Run parameters are

Run number = xxx

Trial number = yyy

Are these ok?

Check the run and trial numbers against the log sheet you should have filled out last time. If these are correct, type "y", otherwise "n". If you type "n", zdos will prompt for a new run and trial number. This permits you to back up and do a run over again if for some reason the last session got botched (say by hardware problems). After this business, zdos will go away for a while and then print a message that it is loading images:

Loading image n out of 8

The images load at the rate of about 5 seconds each, so there will be about a 40 second delay. Sit tight. When the images are all loaded, zdos will start the observing sequence as described in Section 2. After every 8 trials, zdos will pause again to reload the frame buffer. Again, this will take about 40 seconds.

After you've done 64 trials (the current trial number is shown on the terminal screen in the upper left), the run is over and zdos will terminate. You'll be back in the operating system and will receive the p% prompt. Type **bye** and you'll be logged out of the machine. Before leaving, turn off the light box and log the current session on the log sheet in the clipboard.

1.4.2 Things You Shouldn't Need to Know

Any number of things can go wrong. Every effort has been made to protect the programs and data against disaster, but both minor and major errors could occur. What do you do in case of disaster?

Emergency halt. First, you should be aware that you can do an emergency halt of *zdos* at any time by typing an interrupt character on the terminal keyboard. This character is generated either by holding down the *BREAK* key momentarily or typing *CTRL-C*. *CTRL-C* is generated by holding down the key labeled *CTRL* on the left side of the keyboard and then simultaneously typing *C*. *zdos* will ask whether you want to continue or not. If not, type "n" and *zdos* will halt. Send me mail (see below) about the problem and I'll try to correct it.

Sending mail. If something goes wrong during a session, send me electronic mail and I'll try to fix it. To send mail, type in response to the *p%* prompt

p% mail jbz

You can then type in your message. End the message by typing a "." (period) alone at the beginning of the line. For example:

p% mail jbz

Subject: Botched observer session

**The Comtal caught fire during this morning's
observer session.**

p%

Comtal problems. One potential area of trouble is the Comtal has a propensity to go away and not come back. If *zdos* is just sitting there doing nothing for a long time, and *BREAK* does not cause the termination message to appear, reset the Comtal using a *r*. Nasty messages will start coming out on the printer, but you can now type *BREAK* and terminate the program. This should rarely be necessary, but it's very difficult for the program to recover. You can then start *zdos* over again, entering the trial number where the program previously bombed out.

Tape loading. In order for *zdos* to conduct the observer run, the trial images must be located on the disk. Since the disk storage is very limited, an elaborate dance is necessary with the system manager (me) loading the necessary images from tape at the correct time only hours before the observing run. It may happen that you will occasionally come in and start *zdos* and it will type a message like this:

Load tape 3 on the tape drive. Type "y" when ready.

This means the necessary images are not on disk and must be loaded from tape. Since this can take 15 or 20 minutes, it's a disaster to a tight schedule. A program exists to check whether the necessary images are on disk. If you log in and type **p% preload**

the **preload** program will start. It will look much like **zdos**, but its function is only to check for the presence of the necessary images.

Other disasters. Unusual things can happen. If something does, send me mail.

1.4.3 Checklist

Here's the checklist of things to do when you start an observer run.

1. Post the **Experiment in Progress** sign on the door.
2. Turn on the light box.
3. Extinguish the room lights.
4. Check the controls on the Tektronix monitor.
5. Reset the Comtal using `r`.
6. Log in to the VAX using the "observer" login.
7. Start `zdos`.
8. Enter your observer id (last name in all small letters).
9. Verify the run information presented by `zdos` with the log sheet.
10. Perform the trials.
11. Log out the terminal by typing `bye`.
12. Turn off the light box.
13. Record the run information on the log sheet.
14. Take down the **Experiment in Progress** sign.

1.4.4 Summary of commands

This section briefly describes the commands available at various stages in zdos.

Startup

CTRL-C: Interrupt zdos no matter what it's doing.

terminal keyboard: Used to respond to zdos prompts.

Windowing

CTRL-C: Interrupt zdos no matter what it's doing.

Button 1: Not used.

Button 2: Freeze windowing.

Button 3: Not used.

Button 4: Ready button. Proceed to observing period.

Between windowing and observing

CTRL-C: Interrupt zdos no matter what it's doing.

Button 1: Proceed to observing

Button 2: Not used.

Button 3: Not used.

Button 4: Not used.

Observing

CTRL-C: Interrupt zdos no matter what it's doing.

Button 1: Toggle cursor on and off.

Button 2: Not used.

Button 3: Not used.

Button 4: Ready button. Proceed to score entering period.

Entering score

CTRL-C: Interrupt zdos no matter what it's doing.

Buttons: Any button enters the score.

Between trials

CTRL-C: Interrupt zdos no matter what it's doing.

Button 1: Proceed to start next trial

Button 2: Not used.

Button 3: Quit zdos.

Button 4: Not used.