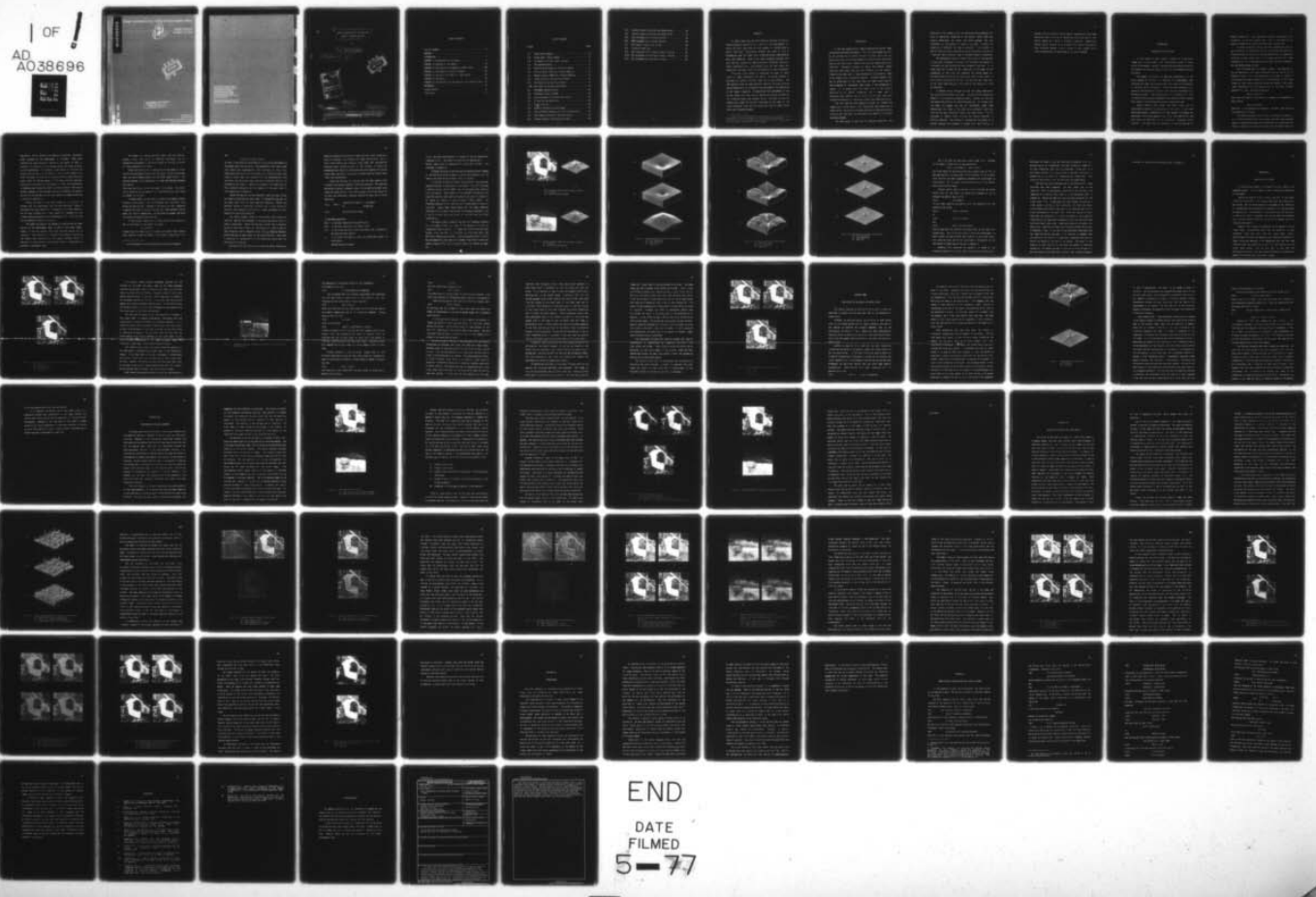
AD-A038 696 UTAH UNIV SALT LAKE CITY DEPT OF COMPUTER SCIENCE F/G 6/4
IMAGE TRANSMISSION AND CODING BASED ON HUMAN VISION.(U)
SEP 75 R J ROM DAHC15-73-C-0363
UNCLASSIFIED UTEC-CSC-75-115 NL



AD A038696

Image transmission and coding based on human vision

RAPHAEL JONA ROM
UNIVERSITY OF UTAH

DDC
APR 29 1976
C

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

SEPTEMBER 1975
UTEC-CSc-75-115
COMPUTER SCIENCE, UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112

DDC FILE COPY

Image transmission and coding based on human vision
RAPHAEL JONA ROM

The views and conclusions contained in this document are those of the author(s) and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency of the U.S. Government.

This document has been approved for public release and sale; its distribution is unlimited.

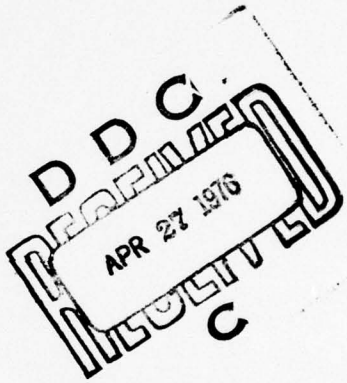
(6)

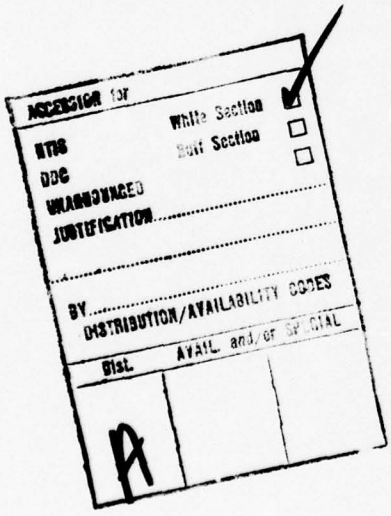
IMAGE TRANSMISSION AND CODING BASED ON HUMAN VISION

by

(10) Raphael Jona/Rom

(9) Technical rept.,





(12) 88 p.

(11) Sep[redacted]1975

(14) UTEC-CSc-75-115

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-73-C-0363, [redacted] F30602-70-C-0300

(15)

404 949

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT *

In recent years more and more attention was paid to digital image processing especially as a result of the developement of highly efficient algorithms and also because of technologically better facilities. Concurrently attempts were made to find a mathematical model for human vision to achieve better understanding about that mechanism. Some of the image processing problems that were (and are) tackled are image enhancement, bandwidth reduction, image transmission etc. Unfortunately very few have taken the mechanism of the human vision into consideration in their processes.

This work is an attempt to incorporate the model of human vision in image transmission and coding. An optimal system is developed to transmit a digital image over a noisy channel. The same system is used for image bandwidth reduction utilizing a simple coding scheme which is not based on the knowledge of the statistics of the image in question. We demonstrate the improvement of the optimal system over other similar systems and provide explanation for situations where other systems failed. The model we use for transmitting images can be also interpreted as the model of the visual mechanism itself and thus shed some light on human vision from a new interesting aspect.

---

*This report reproduces a dissertation of the same title submitted to the Department of Electrical Engineering, University of Utah, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

INTRODUCTION

In the last decade digital image processing and digital image coding have developed considerably. This is a trend caused not only by the availability of better and more efficient facilities but mainly as a result of the developement of very efficient algorithms such as the fast Fourier transform and the high speed convolution algorithm. Successful attempts to deblur and enhance images digitally have drawn more and more attention to the subject. Presently the usual goal of image processing is to produce an image to be looked at by a subjective observer. Although research has, and is, being done to find a good model for human vision very few have attempted to incorporate those models in the processing of images. It is beyond doubt that human vision is not equally sensitive to all spatial frequencies, and it seems that in processing more emphasis has to be put on the "more important" frequencies especially in coding and transmitting images digitally.

The work described here suggests a system for transmitting images over a noisy channel, that incorporates some properties of human vision. We optimize the system according to some optimization criterion and then check its performance and compare it with other available schemes.

The same system is used also for bandwidth reduction. Our

objective in this respect is not to incorporate coding schemes that rely on statistical properties of the source, since those are usually unavailable, but rather use coding schemes that are independent of the ensemble of images to be coded. Although the reduction of bandwidth that may be achieved in this manner is smaller than the reduction achieved by using statistical information it was felt that a practical, easy to implement method is preferred.

The mathematical model for human vision which is incorporated in this work is assumed to be given. Of the models that appear in literature two were selected and used throughout this work. It is shown that the influence of the linear part of the model on the processing is only minor and therefore the choice among all available models (which are of course very close to each other) is not critical from our point of view. The fact that the linear part of the visual model has very little effect on the results will also be discussed.

In chapter one we introduce our model for image transmission and the motivation behind that model. We optimize the system and compare it with previously suggested models. Chapter two analizes the system and points at its important properties. It is shown that the model we suggest may also be interpreted, under some assumptions, as a model of the human vision mechanism itself and thus provide some interesting insight into human vision. This is discussed in chapter three utilizing the results obtained in previous chapters. The problems of implementing the system on a digital computer are discussed in chapter four, and finally, in

chapter five we introduce a coding scheme, independent of the image statistics, that fits into our scheme and allows the use of this optimal system as a means of bandwidth reduction. Comparision between results obtained by this method with results obtained by other available schemes is also included in this chapter which visually demonstrates the advantages of our optimal system.

CHAPTER ONE

INTRODUCTION OF THE SYSTEM

In this chapter we shall suggest a system for transmitting images over a noisy channel that incorporates a model of human vision in the design. We shall introduce the optimization criterion and compare the resultant system with similar ones suggested before by others.

The system in principle is described schematically in the block diagram of fig (1.1). It consists of a preprocessor that processes the image, and a postprocessor that undoes this processing at the other end of the channel. The pre and post processors will depend on some properties of human vision that are represented by a mathematical model. The characteristics of the channel will also influence the pre and post processors thus making the entire system most immune to the disturbances along the transmission path.

Recent research shows [6,10] that human vision can be approximately modelled as consisting of two parts: A nonlinear memoryless system in cascade with a linear system, with memory, as described in the block diagram of fig. (1.2). The function $F(\cdot)$ has been shown by experiments to be a monotonic increasing convex function. this means that no information is lost by passing the

signal through $F(\cdot)$. The logarithmic function, satisfying this constraint, is the most commonly used. This is justified by the physical properties of light as detailed later in this work.

In order to optimize the pre and post processors we have to set a norm for the distortion and try to minimize this quantity. The norm chosen in this work is the mean square error. Although the criterion of mean square error is not always the best for image processing [4] it is mathematically (to date) the most tractable, and was chosen because of that.

Fig (1.3) shows the full system in detail. The quantity I" can be described as the image transmitted from the retina to the observer's brain. I_2" is the perception of the image if looked at directly by the observer whereas I_1" is the perception of the image looked at after being transmitted through the system. The proces itself terminates with the production of I_1' at which a human observer will look. The norm we define is

$$d(x,y) = (I_2"-I_1")^2$$

and since the image is, statistically, a member of an ensemble we shall define

$$M(x,y) = E\{d(x,y)\}$$

Where $E\{\cdot\}$ is the expected value operator. We shall later optimize the system by minimizing M.

The system presented thus far has no constraints imposed on it. In this situation there is no optimal solution since the best system is obtained by making the preprocessor be a pure amplifier that will amplify $I(x,y)$ to such an extent that the channel noise is

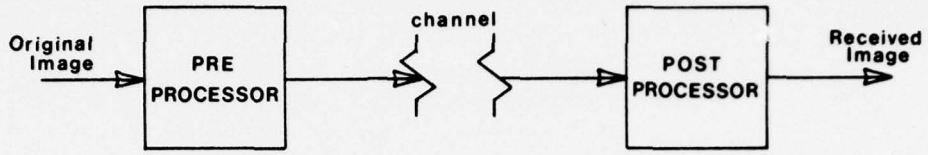


Fig. 1.1 The Principle System

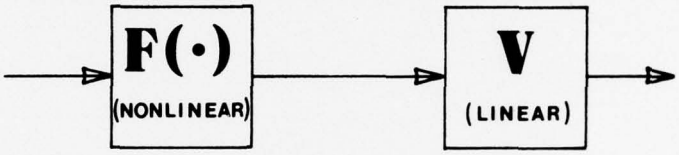


Fig. 1.2 Visual Model - Block Diagram

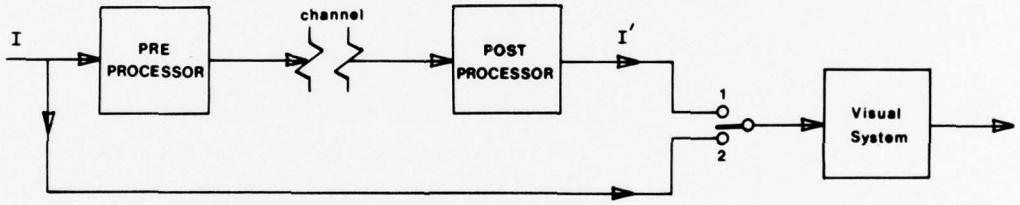


Fig. 1.3 The System with Visual Model Included

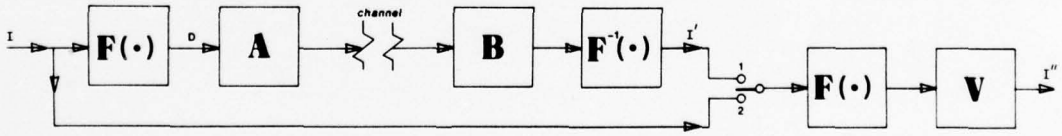


Fig. 1.4 The expended System

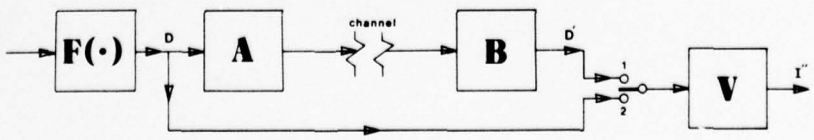


Fig. 1.5 Modified Block Diagram of the entire System

negligible. This of course is not possible in practice. The actual power produced by the preprocessor is limitted. There must therefore be some constraint imposed on the system to make it feasible. We chose to limit the average energy (per image) produced by the preprocessor. This measure is equivalent to limitting the signal-to-noise ratio of the channel. For a given channel, for which the signal to noise ratio is given, and for a (statistically) given noise the average energy in the signal may be computed. Imposing this constraint on the system will force the preprocessor to suppress some frequencies relative to others (some of them may be totally ignored) so that the best fidelity is achieved. An optimal solution to the problem with this last constraint imposed exists and is derived in appendix A.

Taking the model of the visual system as in fig (1.2) we propose that the preprocessor have the same structure, namely a memoryless nonlinear system $F(\cdot)$ in cascade with a linear system A, and the post processor be a linear system B in cascade with the inverse memoryless system of the preprocessor $F^{-1}(\cdot)$. Fig (1.3) can now be expanded to look as in fig (1.4).

The reader may question the necessity of making the nonlinear portion of the preprocessor equal to that of the visual model. Indeed, in the most general case those functions should not be identical. This, however will introduce the problem of obtaining the second order statistics of the signals involved (which is essential to the solution of the problem) which is impossible to evaluate in the general case.

The reader will readily convince himself that the relation between $I''(x,y)$ and $I(x,y)$ as described previously can be schematically expressed in the block diagram of fig (1.5) to which all later computation will refer.

Since the function $F(\cdot)$ is one-to-one the knowledge of $I(x,y)$ implies the knowledge of $D(x,y)$ and vice versa. We can thus assume that the whole process starts with the given signal $D(x,y)$. The relation between $I''(x,y)$ and $D(x,y)$ is linear and analysis becomes much easier.

Note also that $D'(x,y)$ is not the signal to be viewed. The signal to be looked at by the observer is $I'(x,y)=F^{-1}[D'(x,y)]$ which does not appear in fig (1.5).

It seems proper, at this point, to mention two papers closely related to this subject. One is by Stockham [11], the other is by Mannos and Sakrison [6]. Stockham in his early work [12], suggested a system like that of fig (1.4) but set the linear systems A and B equal to V and V^{-1} respectively. No optimization attempt was done but rather the argument was as follows.

If the original image $I(x,y)$ is viewed by the visual system we get, at the output of that system, the signal

$$I_1'' = F(I) \circledcirc V \quad \dagger$$

whereas when the image $I(x,y)$ is viewed by the same visual system after passing through the channel (with channel noise $N(x,y)$) we

---

† In this document the $\circledcirc$ sign denotes the convolution operator.

get

$$I_2''(x,y) = I_1''(x,y) + N(x,y)$$

so that if the noise is white there will be a white disturbance to the signal sent from the retina. The parameters of the visual model were based upon psychophysical data available at that time. Experiments carried out did not produce nice looking pictures when using a visual model that complied with the psychophysical data available, and the actual model used was a corrected version. No explanation was given, in terms of the system, for those results. More elaborate discussion for the reasons will be given later in this chapter and in chapter 3.

Mannos and Sakrison carried Stockham's ideas one step further by trying to optimize the visual model in a subjective way and, at the same time, allow for some bandwidth reduction. Mannos and Sakrison defined a distortion criterion and a rate distortion function, and produced a collection of images processed by their system with controlled distortion.

The coding scheme, based on the defined rate distortion function, requires the knowledge of the source statistics which is usually unavailable. Mannos and Sakrison noticed that a Gaussian source is the worst to code i.e. any source will yield a smaller rate distortion than a Gaussian would, thus by assuming Gaussian source they set an upper bound for the rate distortion. Note that in practice for implementation of this method one would need the statistics of the source.

By controlling the distortion, as described above, Mannos and

Sakrison produced a collection of images that were later viewed by a group of observers, who refereed the images subjectively. One of their final results was finding a visual model that consistently produced better looking pictures in their method. It should be emphasized that neither of those works optimize a system for a given visual model (as done in this work) but rather used the visual model itself as a preprocessor.

Using the calculus of variation and the mean square error criterion the optimum systems A and B are derived. The detailed derivation is given in appendix A and, in this section we shall only summarize those results. The linear systems A and B are specified by their two dimensional frequency response and obey the following equations:

$$A^2(f) = \begin{cases} (S_n/S_d)^{1/2}\,[\mu V(f) - (S_n/S_d)^{1/2}] & \text{if } \mu V > (S_n/S_d)^{1/2} \\ 0 & \text{ootherwise} \end{cases} \tag{1.1}$$

$$B(f) = A^*(f) / [|A(f)|^2 + S_n/S_d] \tag{1.2}$$

In the above equations:

$S_d(f)$ is the power spectrum of $D(x,y) = F[I(x,y)]$

$S_n(f)$ is the power spectrum of the channel noise

$V(f)$ is the linear portion of the visual model, and is assumed to be known [6,10 and others].

μ is a scalar evaluated to yield the prescribed signal to noise ratio.

* denotes complex conjugate.

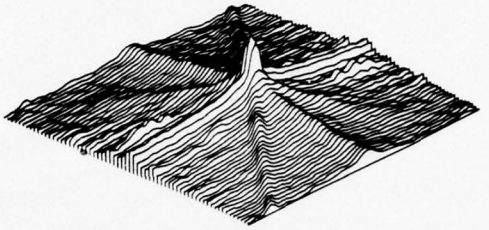
In all the above and hereafter (f) stands for the two dimensional frequency (f_x, f_y). The signal to noise ratio is denoted by P. Special cases such as frequencies for which $S_d(f) = 0$ etc. are discussed in appendix A.

Although the noise is arbitrary and no constraints are imposed on its statistics we shall assume in the following section, for the sake of the qualitative discussion, that it is white.

In fig(1.6) we have an original (called "MILL") and the power spectrum estimate it generates (on a db scale). Fig (1.7) displays a second original (called "BARN") and its estimated power spectrum (on a db scale). Some remarks on the problems involved and the algorithm used for power spectrum estimation are given in chapter 4. We tested our method on three different visual models. The frequency response of the linear portion of these models is given in fig (1.8). Using these three models and the power spectrum estimates of fig (1.6b) we calculated the frequency response of the A and B filters which are given in fig (1.9) and fig (1.10) respectively.

The shape of $S_d(f)$ resembles very much the frequency response of a low pass filters i.e. most of its energy is in the low frequencies (both f_x and f_y). In images that contain more man-made objects (like the MILL in fig (1.6)) we find that $S_d(f)$ has additional energy along the axes. Since by assumpton $S_n(f)$ is white S_n/S_d has essentially the shape of a highpass filter which is also the general shape of V(f), up to the point where V(f) starts to taper off.

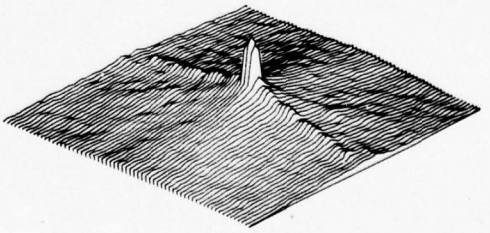
(a)



(b)

First original (MILL) and its power spectrum
(a) Original
(b) Power Spectrum estimate

(a)



(b)

Fig. 1.7 Second original (BARN) and its power spectrum
(a) Original
(b) Power Spectrum estimate

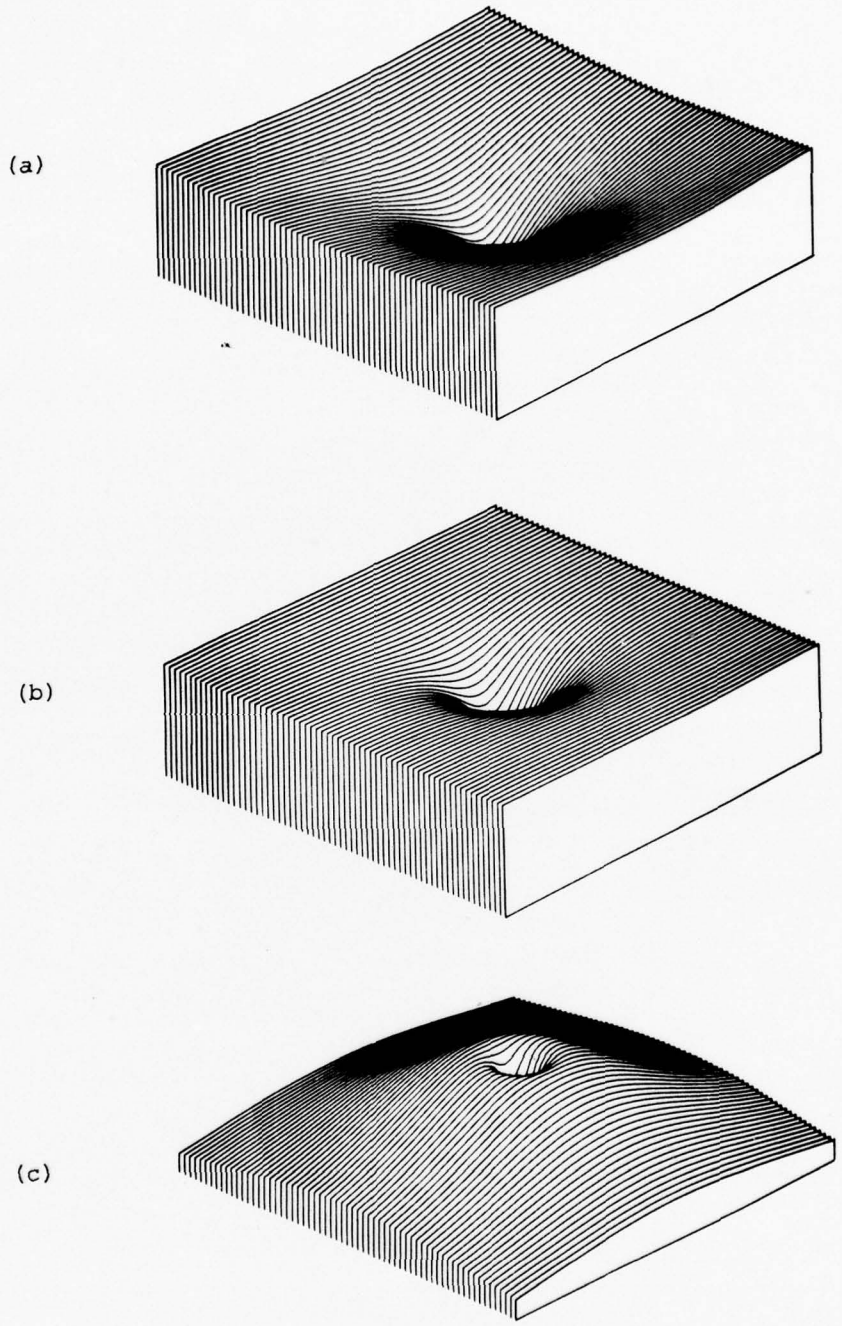
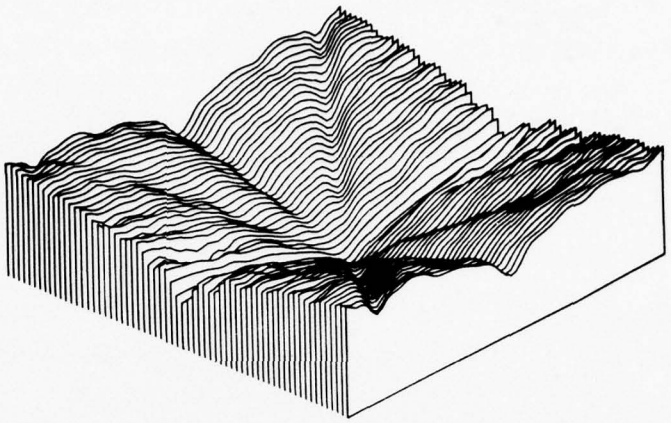
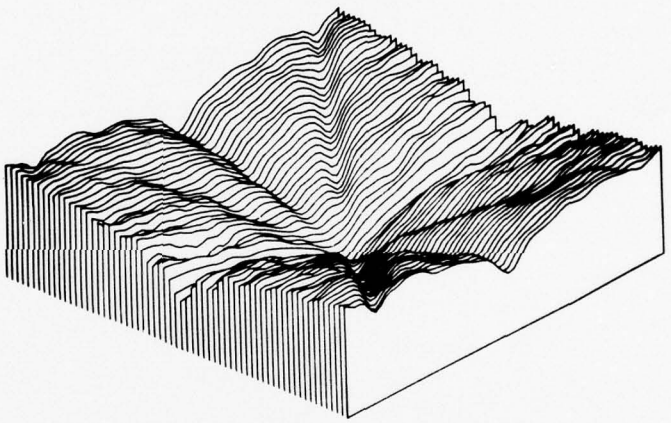


Fig. 1.8 Linear portions of various Visual Models
(a) Slow saturating
(b) Fast saturating
(c) Taper-off

(a)



(b)



(c)

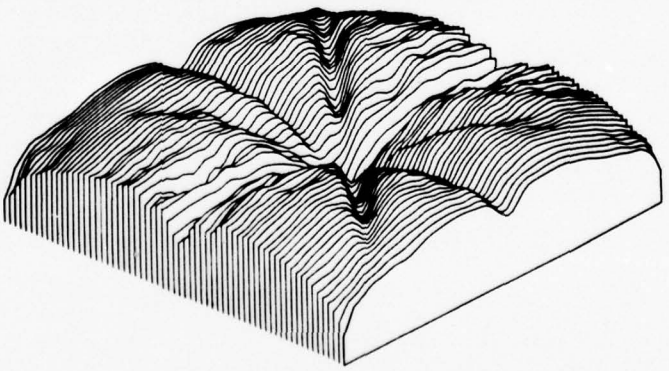
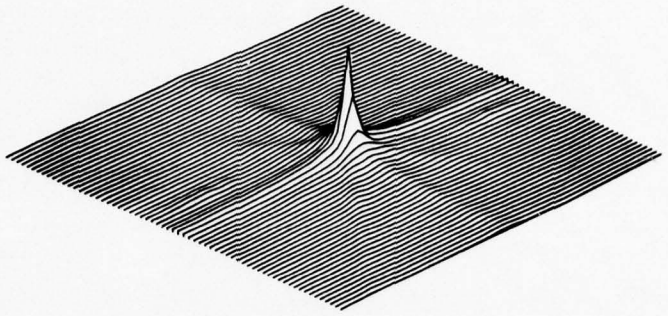
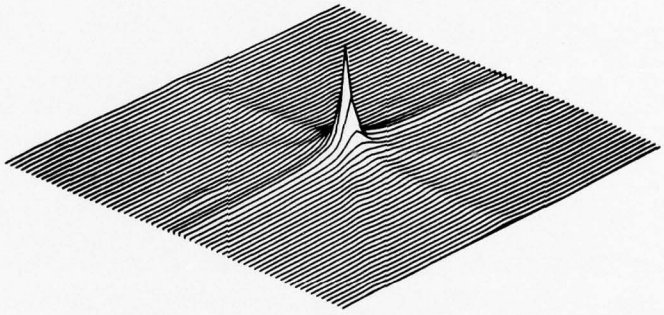


Fig. 1.9 A-filters for various Visual models
(a) Slow saturating
(b) Fast saturating
(c) Taper-off

(a)



(b)



(c)

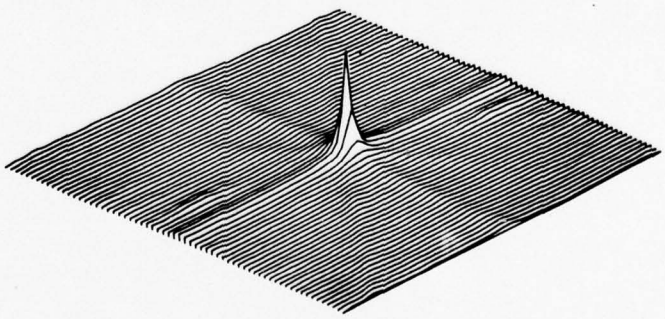


Fig. 1.10 B-filters for various Visual Models
(a) slow saturating
(b) Fast saturating
(c) Taper-off

Now if we take the case where μ gets larger (i.e. increase in the signal to noise ratio) we may approximate

$$A^2(f) = \mu V(f)(S_n/S_d)^{1/2} \; ; \quad B(f) = A^{-1}(f)$$

and since $(S_n/S_d)^{1/2}$ as mentioned before has a shape close to V(f) we may say that A(f) is proportional to V(f) and B(f) is its inverse. This justifies the model suggested by Stockham for high signal to noise ratios. For low signal to noise ratios this will, of course, yield worse looking images.

Another special case to discuss is the following; We showed before that V(f) and $(S_n/S_d)^{1/2}$ are essentially highpass so let us consider the special case in which

(1.3) $$V(f) = (S_n/S_d)^{1/2}$$

In this case, substituting equation (1.3) into equation (1.1) and equation (1.2) we get

(1.4) $$A^2(f) = (\mu-1)V^2(f)$$

or

(1.5) $$A(f) = (\mu-1)^{1/2}V(f)$$

and

(1.6) $$B(f) = [(\mu-1)^{1/2}/\mu]V^{-1}(f)$$

Thus we see that A(f) and B(f) are proportional to V(f) and $V^{-1}(f)$ respectively. This is the only case in which the preprocessor is equal (except for some constant amplification) to the visual model. This case has some qualitative significance. Discussion on the implication of these results is given in chapter 3.

Equation (1.1) specifies the system A in terms of its frequency response (for a given signal to noise ratio affecting μ).

Obviously for large μ, the right hand side of equation (1.1) will be positive for all frequencies. For small values of μ there will be some frequencies that would be suppressed. Since A(f) is a continuous function of μ there exists a smallest (critical) μ, denoted by μ_c for which all frequencies are transmitted. This notion of the critical μ will be used later for comparisions.

The last item to be discussed here is the function $F(\cdot)$. Many functions have been suggested. The most common one is the logarithm. There is some physical justification for the use of the logarithm or a function very close to it (as explained later in chapter 3). *Mannos and Sakrison using the same structure for the* visual model (i.e. a memoryless nonlinear system followed by a linear system) noticed, in their experiment, that there was almost no difference which function was used in the nonlinear part with a slight preference of the cubic root. Since the results of their experiment are subjective it is hard to comment or give physical interpretation to those results especially that the functions used were very close to each other. Experiments done by this author with square root, logarithm, and some other functions yielded very slight differences if any. In the rest of the experiments described here the logarithm function was used. It is important to remember that the choice of the function F is completely independent and does not effect the design of the rest of the system. The answer to the question of what function $F(\cdot)$ optimizes the system is therefore subjective. No attempt was made in this work to find the best $F(\cdot)$ and the choice of the logarithmic function, that is used throughout

this work, is justified by the reasons given in chapter 3.

CHAPTER TWO

ANALYSIS OF THE SYSTEM

In the previous chapter we introduced the main ideas of the suggested system. In this chapter we shall analyze the components of the system.

Looking at equation (1.2) we notice that B(f) is the Wiener filter for a given A(f). This implies some known facts about B(f) as follows. For low noise levels B(f) converges to the inverse of A. Frequencies that are zeroed out by system A will also be zeroed by system B etc. The dependence of the system B upon the eye model is implicit in that A is dependent upon the eye model and B depends on A. We shall devote, therefore, the rest of the chapter mainly to analyzing the system A.

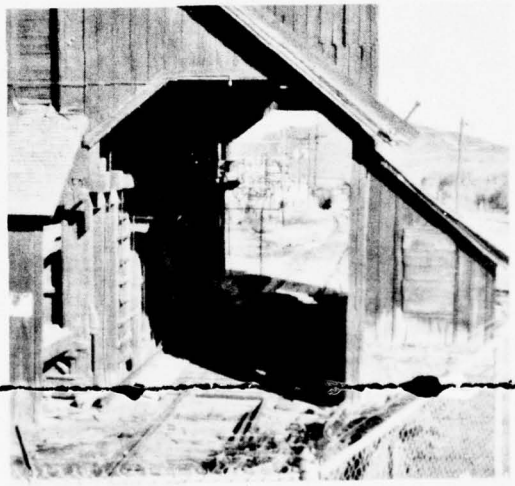
Equation (1.1) shows the dependence of the system A on the visual model. We mentioned before that the system A, in most instances, looks very much like the visual model itself. Qualitatively this should be expected since frequencies important to human vision are expected to be emphasized over the the less important ones thus shaping the filter A in the direction of the visual model chosen. Also one must bear in mind that the optimization criterion, the least squares estimation, influences the shape of A no matter what visual model is used.

The question that arises immediatelly is the one of choosing an eye-model. There are a few models to consider, most of them suggested previously by various researchers. Stockham proposed two different models [10,11,12] Mannos suggested another [6] and many others appear in literature. Althogh the visual system has been shown not to be isotropic all the models suggested are circularly symmetric for reason of ease of implementation. The models used in this research, although they do not have to be symmetric, were chosen to be such. In any case all the models suggested are some sort of high or bandpass filters. This stems from psychophysical experiments and are justified physically by the desire to seperate illumination from reflectance (which implies emphasis of high frequencies over low frequencies) and by the imperfectness of the lens system (which causes tapering off at very high frequencies). We shall elaborate more on this subject in chapter 3.

For a given signal to noise ratio and a given visual model, the system, namely the filters A and B, are uniquely defined (under the chosen optimization criterion) and are given in equations (1.1) and (1.2).

It is possible, for a given signal to noise ratio, to vary the visual model and produce a collection of images to be viewed by subjective observers and determine the "best" visual model for this system. Preliminary experiments showed that the difference in choosing between the results for the various models tested are very slight, and a more careful experiment has to be conducted in order to come up with a conclusive result relating to the visual model.

(a) (b)



(c)

Fig. 2.1 Processed versions of MILL using various Visual models and $P=P_c=14.5$
(a) Slow saturating
(b) Fast saturating
(c) Taper-off

In fig (2.1) three different processed versions (of the original of fig (1.6a)) are shown. Each of the three processed versions was produced using a different visual model with the same signal to noise ratio of P_c=14.5. Isometric plots of the visual models used are shown in fig (1.8). Most conspicuous in comparing the processed versions of fig (2.1) is the fact that the pictures appear only slightly different from one other. In fig (2.2) we have the second original (of fig (1.7a)) processed by the system using the visual model of fig (1.8c) and $P=P_c$=14.5.

The fact that the system is not very sensitive to changes in the visual model is somewhat encouraging. Definitely the true visual model is slightly different for different observers. It is therefore desired that the system be insensitive to the visual model so that its performance will be close to optimal for as wide a class of observers as possible. On the other hand the fact that the entire system is so slightly dependent on the visual model (in the final subjective appearance of the images produced) makes the finding of a "best model" a much harder problem to solve.

The insensitivity to the visual model raises the question of the importance of incorporating a visual model in the transmission scheme. In our case there is a slight improvement in performance when using a visual model over the case where a flat visual model was used. However, this author believes that the visual model must be included in any kind of image processing and that its influence may be enhanced when a different fidelity criterion is used.

Some remarks are appropriate here to point out the nature of

Fig. 2.2 Processed versions of BARN using the Taper-off Visual model and $P=P_c=14.5$

the dependence of the system A and B on their parameters.

From Appendix A we find

(2.1) $$\mu = (P+1)\cdot\int S_n df / \int |V(f)| (\eta/g_d)^{1/2} df$$

Let us assume that two different experiments are conducted with the same signal to noise ratio P (that yields $\mu > \mu_c$), but changing only the visual model in such a way that

(2.2) $$V_2(f) = K\cdot V_1(f)$$

Where $V_1(f)$ and $V_2(f)$ are the visual models for the first and second experiments respectively and K is a positive constant. Using equation (2.1) we find

(2.3) $$\mu_2 = \mu_1/K$$

which in turn yields

(2.4) $$\mu_2 V_2(f) = [\mu_1/K]\,[KV_1(f)] = \mu_1 V_1(f)$$

Looking at equation (1.1) we notice that A(f) depends only on the product $\mu V(f)$ and considering the result of equation (2.4) we conclude that for a given signal to noise ratio the system is invariant under scaling of the the visual model. Obviously since B(f) depends only on A(f) there will be no change in B(f) in this case.

Another property is the following. Assume that in two different experiments we use the same visual model but increase the power of the noise (not changing its power spectral shape) in such a way that

(2.5) $$S_n'(f) = K^2 S_n(f)$$

and retaining in both experiments the same signal to noise ratio. Equation (2.1) yields

(2.6) $$\mu' = K\mu$$

and after some simple algebra we find

(2.7) $$A'(f) = K \cdot A(f)$$

This means that for a given signal to noise ratio an increase in the noise level amounts to corresponding amplification in the system A.

Substituting equation (2.5) and (2.7) into (1.2) reveals that

(2.8) $$B'(f) = B(f)/K$$

We conclude that the increase in noise level does not affect the shape (or performance) of the entire system except for a constant amplification.

In the experiments conducted in this research the ensemble of images was scaled in such a way that D(x,y) [in fig (1.5)] ranges between 0 and 511. The noise was chosen to range between -51 and +51 (i.e. peak value of the noise is 10% of the signal). We have just shown that the result, for which $\mu > \mu_c$, will not change with a change in the normalization factor.

We can provide at this point some explanation to the results encountered by Stockham when using the taper-off visual model. Stockham used, as mentioned before V(f) and $V^{-1}(f)$ as the A and B filters. Thus the preprocessing in his experiment yielded results close to those obtained by our system. The postprocessing however, is quite different since there is no specific noise handling mechanism in Stockham's scheme and it therefore produced completely different results. The problem was that the tapered portion of the visual model amplified parts of the noise which made the entire image bad looking. It turns out that the system B is not less

important then the system A and in fact may be more important in some cases since it processes not only the signal but also the noise. It is important at least to have the systems A and B be a matching pair in the Wienner sense.. Looking at B(f) for the various systems (figs (1.10)) reveals that the B filter looks more like the inverse of the visual model of fig (1.8b) than that of fig (1.8c). What happenned in these experiments is that the inverse of the visual model of fig (1.8c) enhanced high noise frequencies that irritated the human observer. Those frequencies should have been suppressed since in this domain the power of the noise may exceed that of the image to such an extent that the image is dominated by noise. A better system that should have been suggested is one that uses the eye model as the system A and its matching Wienner filter as system B. At any rate this explanation is now of minor importance since it is proved in this work that for optimum results the system A should not be made equal to the visual model.

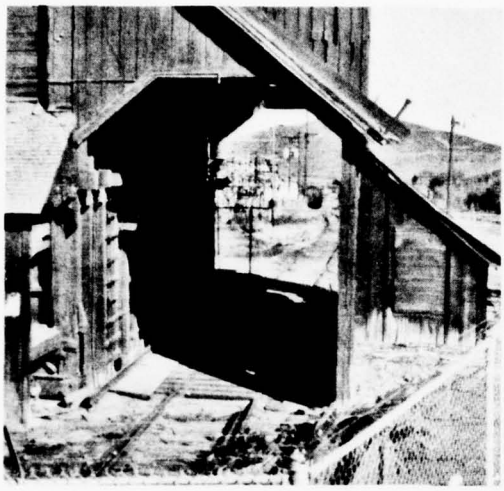
In general, when using the minimum least squares criterion for optimization, one should remember, as noticed by Costas [1] and others that the importance lies in making the pre and postfilters a matched pair, and that the prefilter have the desired general shape. The fine structure of the filter will usually not improve the performance of the entire system significantly.

To demonstrate the effect of the signal to noise ratio on the results the following experiment was conducted. The image of fig (1.6) was processed (see fig (2.3)) using $\mu=\mu_c$. Using μ_c yields some signal to noise ratio that we term P_c. (For the "MILL" and the

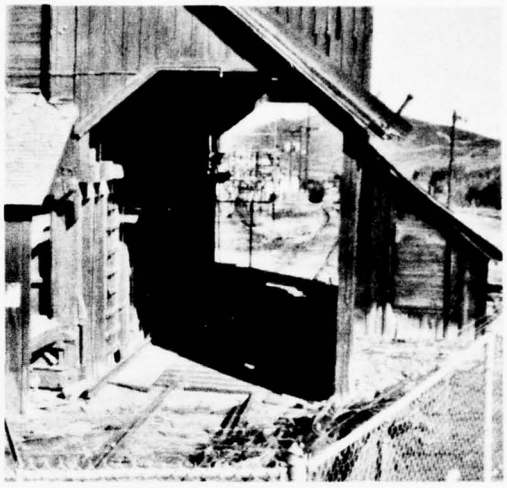
"taper-off" visual model P_c was calculated to be 14.5). The same image was also processed using $P=2/3P_c$ and $P=1.5P_c$. The A filter for $P=1.5P_c$ is essentially a scaled version of the A filter for P_c. Not so in the case of $2/3P_c$ where the A (and therefore also the B) filter which resulted looked essentially as for the P_c case except that some high frequencies (both f_x and f_y) where zeroed out. The result of the processing with these signal to noise ratios is given in fig (2.3). Although fig. (2.3b) is undoubtedly better then fig. (2.3a) the latter is strikingly good in the preservation of details. In fig. (2.3c) we used a ratio of P=1 which is much below the critical power. As can be seen the picture is much noisier, dynamic range has decreased but in spite of this low signal to noise ratio most of the detail, even the fine detail, is well preserved and demonstrates the capability of the system to transmit images under extremely noisy circumstances.

The improvement one gets when using our system over regular transmission is demonstrated by comparing fig. (2.3) with figs. (5.9a) and (5.10a). These latter pictures are statistically equivalent to what would happen if the original image had been transmitted through the same noisy channel without the processing done by the pre and postprocessors.

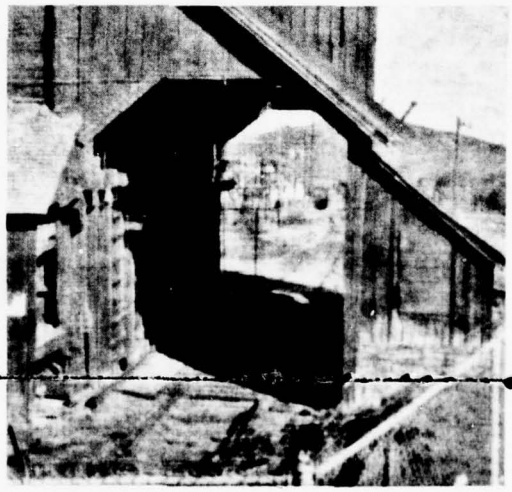
The choice of what ratio to use depends upon the subjective distortion one would tolerate. If detail is important one might reduce the signal to noise ratio and if "cleanliness" of the resultant picture is important a high ratio is necessary.



(a)



(b)



(c)

Fig 2.3 MILL processed with different S/N ratios
(a) $2/3P_c$
(b) $1.5P_c$
(c) $P = 1$

CHAPTER THREE

IMPLICATION OF THE RESULTS ON HUMAN VISION

The results obtained by calculating the A and B filters as described in chapter one may shed some light on the mechanism of human vision

Although, as mentioned before, the criterion of mean square error is not always appropriate for image processing, and some of the results we obtained are strongly dependent upon to the optimization criterion, it still seems that some consideration should be given to explain the results on a more qualitative basis, and to try and suggest some new points of view about the nature of human vision.

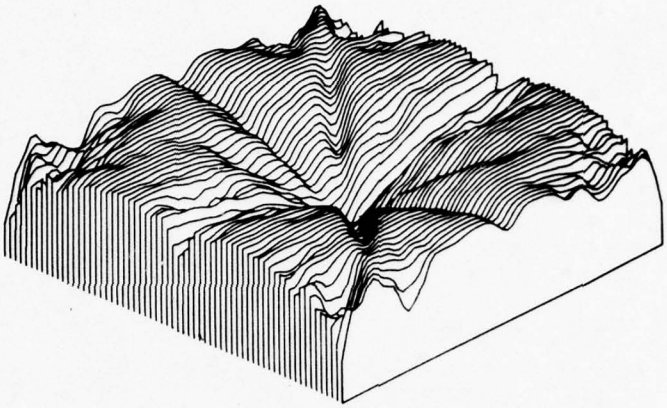
Looking at equation (1.1) there are two major components influencing the design of the system A, namely the visual model V(f) and the quantity S_n/S_d . If the human vision were equally sensitive to spatial frequencies or alternately if we were not to consider the visual system at all in the process of optimizing the pre and post processors we would get results that are worth some special consideration. Quantitatively this means replacing V(f) in equation (1.1) with

(3.1) $$V(f) = 1 \quad \text{for all frequencies.}$$

An isometric plot of A(f) resulting from such substitution is shown in fig (3.1a). Comparing the result with the one derived for a true visual model (fig (1.9)) reveals that the shape of A(f) has not changed much. This implies that the shape of A(f) is controlled mainly by the shape of the quantity S_n/S_d . This suggests that the system is more sensitive to the information itself (implicitly represented by S_d) and to the channel (characterized by S_n) than to the mechanism of vision. In this case, where V(f) is absent from the system, the A filter that results looks very much like some models of the visual system that are suggested in the literature. This fact may provide us with some explanation on the behaviour of human vision.

Many assumptions have been made about the nature of information transmission from the eyeball to the brain. The one that seems very logical is that the optic nerve, connecting the eyeball to the brain, acts like a noisy channel (i.e. that information is processed in the eyeball and transmitted to the brain through the noisy optic nerve). If we assume this kind of visual system it is easy to model such a system. In fact the system we suggested in chapter one becomes a good model for the entire visual system. Specifically the preprocessor corresponds to the eyeball in which initial processing is performed. The processed signal is then transmitted through the noisy channel namely the noise optic nerve. The brain, at the other end of the channel is the postprocessor but since there is no "visual system" at the other end as in the system described in chapter one and fig (1.3), the substitution suggested

(a)



(b)

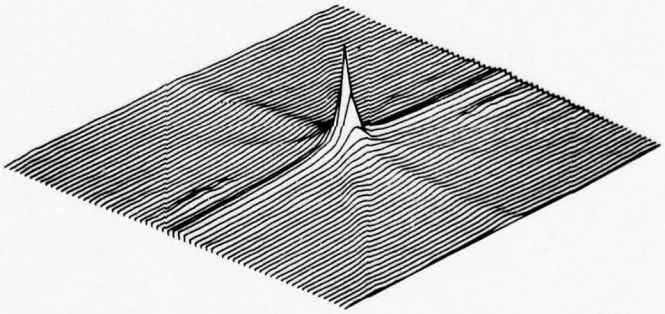


Fig 3.1 Filters for the case V(f)=1
(a) A-filter
(b) B-filter

in (3.1) is appropriate. The result is the system A shown in fig (3.1). We noticed before the resemblence between the A filter and the visual model, which strengthens the assumption that part of the eyeball's processing of images is directed towards a safer transmission through the optic nerve much like the task of the preprocessor in our system. The fact that A(f) is not circularly symmetric strengthen the observation that the human visual mechanism is not quite isotropic.

Early experiments in image enhancement showed that a highpass filter usually reveals or enhances details that could not be clearly seen in the original image. Apart for the reduction in dynamic range due to highpass filtering of the log illumination the following arguments were made. An image is composed of two components, illuminaton and reflectance that are multiplied to form the image. It is the reflectance that we are interested in since it describes the objects we are looking at. The illumination has a strong dependence on light sources, which are slowly varying across the scene, and has therefore most of its energy in low frequencies. The reflectance, on the other hand, has considerable amounts of energy in high frequencies since it "represents" textures, sharp edges of objects, and fine detail. Unfortunately illumination and reflectance cannot be absolutely separated because illumination has some energy in high frequencies and reflectance has some important low frequency components. This confirms that the separation between illumination and reflectance cannot be made by a pure highpass filter but only optimally approximated by a filter that would be

close to the one shown in fig (1.8b).

Giving this argument some mathematical ground we may write

(3.2) $$I(x,y) = i(x,y)\cdot r(x,y)$$

where $I(x,y)$ is the image, $i(x,y)$ and $r(x,y)$ are the illumination and reflectance components respectively. Using the logarithm as the nonlinear function $F(\cdot)$ we get

(3.3) $$D(x,y) = \mathrm{Log}[I(x,y)] = \mathrm{Log}[i(x,y)] + \mathrm{Log}[r(x,y)]$$

or

$$D(x,y) = i'(x,y) + r'(x,y)$$

Looking at (3.3) and remembering that $i'(x,y)$, the illumination component, is the one we want to get rid of, we realize that some sort of highpass filtering on $D(x,y)$ is necessary. (This is in short the principle of homomorphic image enhancement as originally stated by Oppenheim et al [7]). This analysis provides us with the physical justification for the use of the logarithm as the nonlinear portion of the visual system.

The previous arguments pose some interesting questions. We note that the visual system is a kind of highpass filter. We have shown that highpass filtering achieves best transmission of images through the optic nerve and we have also shown that a highpass (homomorphic) filter achieves good separation between illumination and reflectance that is desired for good quality vision. This suggests that the visual mechanism achieves two goals: it separates reflectance from illumination and encodes the image to be transmitted in a safer way through the optic nerve. The question is whether or not these are really two separate issues or two aspects

of the same phenomenon.

At the time of this writing there is no unique decisive answer to this question. It is the author's belief that the true situation is some kind of a compromise between the two arguments for the reasons outlined in the next paragraph.

If the eye system were based only on safe transmission consideration we should have ended up with an eye model that obeys

$$V(f) = (S_n/S_d)^{1/2}$$

since in this case the signal, after passing through filter A would have the same statistics and therefore optimally immune to the noise of the optic nerve. An experiment that was conducted using that kind of visual model produced very bad looking pictures. The reason that the results were bad looking is that such a visual model suppresses low frequencies too much (thus getting rid of some needed reflectance component) and does not work right for high frequencies either, because of the high amplification it introduces in this range. This suggests that the mechanism of the vision is not based upon transmission consideration alone.

The fact that the visual system is not based solely on the separation of reflectance and illumination can be proved by the fact that for such a separation we need a highpass filter (like that of fig (1.8b)) whereas psychophysiological data shows that the visual system is not really highpass but rather a sort of bandpass i.e. tapers off at the very high frequencies. (this tapering off is not due to the finite aperture of the eyeball lens, that should have occured at a much higher frequency but because of some imperfectness

in the lens system and retinal imaging surface).

It is therefore the auther's belief that human vision is a compromise between safe transmission of the image through the (noisy) optic nerve and the desired separation of illumination and reflectance. Because of the importance of the issue it seems worthwile for future researchers to focus more attention in trying to solve this question in order to acquire better understanding of human vision and its mathematical modelling.

CHAPTER FOUR

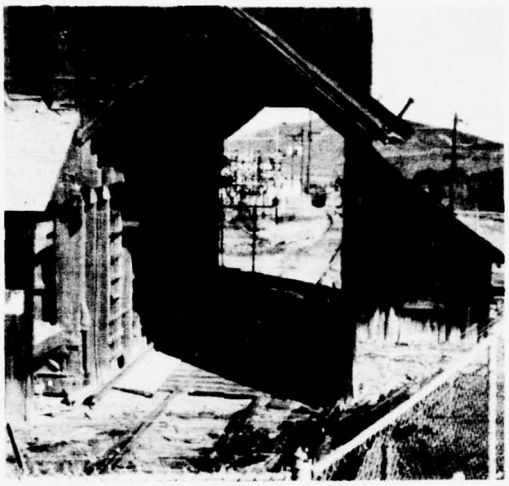
THE SYSTEM ON A DIGITAL COMPUTER

The system described and analized in the previous chapters and the equations derived in appendix A are all true for continuous functions. Because of the isomorphism established between the continuous and the sampled functions [9] those results are true also for the sampled case. Note that this isomorphism relates the continuous-space function to the discrete-space function (i.e. functions of a continuous or discrete domain) and does not relate to the question of whether the range of the function is continuous or discrete. We shall assume that the image in the computer is sampled fine enough and will serve as our original, and that the equations apply to this original image. In this chapter we shall outline the process of implementing the system on a digital computer. The chapter will not deal with problems that arise because of the facilities (and their hardware limitations) but rather with some theoretical problems.

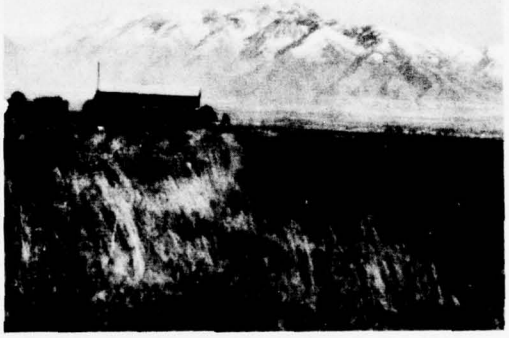
The first problem is the one of estimating the power spectrum of the image ensemble. It is obvious that the true power spectrum is not available to us and we have to be satisfied with some close estimate. The algorithm used, based on averaging periodograms, was

suggested (for one dimension) by Welch[13]. The picture is broken up into (possibly overlapping) sections. Each section is windowed (we used a two dimensional Hanning window) and then the magnitude square of the discrete Fourier transform of that section is calculated. This quantity is then averaged over all sections. The final result is then divided by the energy in the window. An example of the power spectrum estimate (on a db scale) of the logarithm of an image is given in fig (1.6) and fig (1.7).

The accuracy of such an estimate is discussed in Welch [13]. Using the above algorithm we may take one of three approaches. The first uses the estimate taken from one image as the estimated power spectrum of the entire ensemble. The second approach averages estimates over a collection of images. This author's experience shows that four or five carefully selected images suffice for such an estimate. The last approach is to divide the ensemble into sub-ensembles each of which has a power spectral estimate, and then using one of those estimates for the given image. (The sub-ensembles should depend heavily on the amount of man-made objects in the images since this is the main contributer to the differences in the power spectra). The third approach seems to be the best one. However, it is difficult to automate because of some unsolved problems in image classification. In general the results are not very heavily dependent upon the kind of algorithm used. The image of fig (1.6) was processed by filters designed using the power spectrum estimate of the image of fig (1.7) and the results are given in fig (4.1). [compare with the image of fig (2.1c)].



(a)



(b)

Fig 4.1 Cross filtering (using P_c)

(a) MILL processed with filters of BARN
(b) BARN processed with filters of MILL

Another important problem is that of windowing. By windowing we refer to the process of modifying the impulse response of a system in such a way that its frequency response will change only slightly. The need for windowing arises mainly because we are dealing with an infinitely long impulse response that has to be truncated for the implementation on a (finite memory) digital computer. Our windowing process was chosen to be the multiplication of the impulse response of the system with a bell shaped function whose circumferencial points are zero. Undoubtedly no windowing at all yields bad results but on the other hand when a(x,y) is windowed it is no more the optimal filter. One has to learn, when using digital computers, to compromise optimality of the solution for the use of this powerful device. A few experiments were made to find the best way of windowing out of the following five possibilities:

1. Window filter A only.
2. Window filter B only.
3. Window filter A and B (after the optimal filters have been calculated.
4. Window filter A. Calculate filter B as the Wiener filter of the windowed A.
5. Window as in (4) but apply window to filter B as well

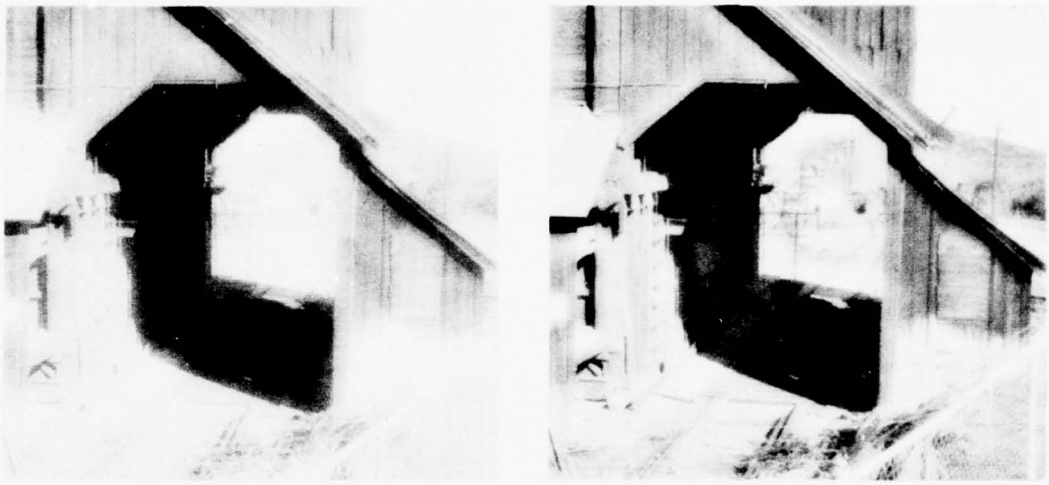
From all these possibilities the one found most satisfactory is the fifth method suggested above. Whenever windowing was needed this method was used. Some examples of various windowing procedures

related to the methods 1,3 and 4 above are shown in fig (4.2). The window itself is always a two dimensional Hanning window.

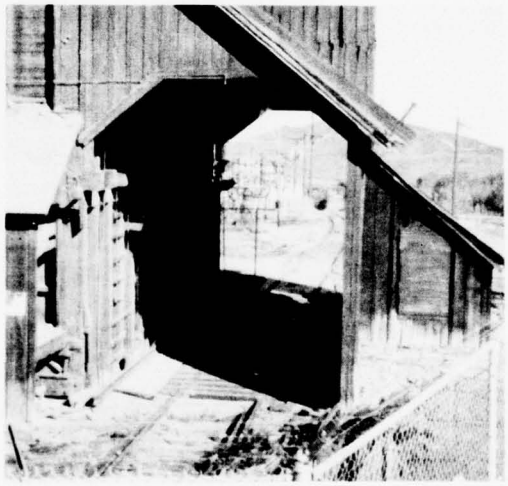
The only way to avoid windowing (and its side effects) is by using periodic convolution, as oppsosed to aperiodic convolution which requires windowing. The periodic convolution although freeing us from window worries introduces some other side effects. More time is needed to estimate large-record power spectra (Although this is a one time problem) and the result converges more slowly. In general, the results were not of much difference when both methods had been compared. A slight preference was noticed for the periodic convolution. Compare fig (4.3) which used circular convolution with fig (2.1c) and fig (2.2c) where aperiodic convolution was performed with the same set of filters.

Another theoretical issue is the sensitivity of B(f) [or b(x,y)] to the number of bits used to describe it. In general all the computatuion was done in standard floating point arithmetic but experiments were performed to see what kind of accuracy is needed for b(x,y). The issue is of some importance for implementing the system on a short word computer or if attempts are made to use fixed rather then floating point arithmetic. The experiments show that for 512 by 512 picture elements and filters that are 64 by 64 the use of five bits per sample of the filter B suffices, four bits give good results but three bits introduce extra conspicuous distortion.

We would like to pause at this point and make some accounting for the actual number of bits to be transmitted. Two additional pieces of information have to be added to the actual data

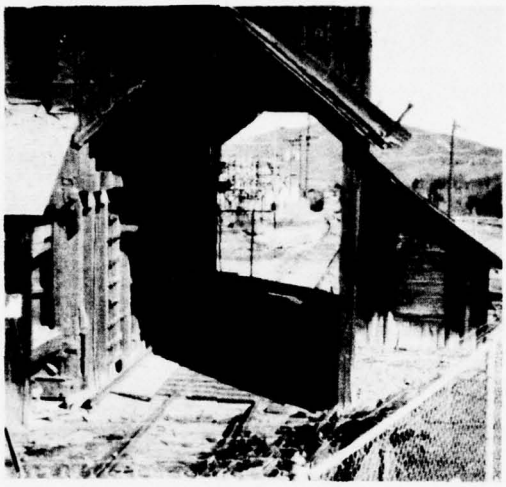


(a) (b)



(c)

Fig. 4.2 Effect of windowing on process
(a) Only A-filter windowed
(b) A and B-filters windowed separately
(c) B-filter as a Wiener filter of the windowed A-filter



(a)

(b)

Fig 4.3 MILL and BARN processed with circular convolution

transmitted. The first one is the average of the image, which is usually not zero (in the log domain). This is done because some coding schemes work better on a zero average signal and thus the original average has to be separately transmitted. Experiment show that this average is in the range in which 9 bits will usually suffice. The second and more important issue is the B filter. If we decide to transmit the B filter along with the image itself the number of extra bits needed is computed in the following. We mentioned before that a B filter with 64X64 elements quantized to have 5 bits per sample yields good results. Thus the total number of bits to be added to the transmission is 64X64X5. If the image is a 512X512 this means an addition of less than 0.08 bits per picture element transmitted. Although this quantity is quite small one would like to avoid this transmission. One way of doing it is creating a "library" of B filters. We have demonstrated before in fig (4.1) that the A and B filters need not be matched to the image in question. Thus one can decide to use A and B filter pairs out of a "standard" small library so that the only information to be added to the actual bits of the image is the index into the library which is a negligible addition to the total.

The last item we would like to comment on is the noise. Usually when talking about a channel, there is no control over the noise. In experimental work one must choose the noise. We simulated the channel by uniformly distributed zero mean random numbers. Later on when this scheme is used for image coding we shall introduce some differennt kinds of noise more commonly known

as dither.

CHAPTER FIVE

THE USE OF THE MODEL FOR IMAGE CODING

Up to now we considered the system for transmitting images in a general sense. One other very important use of such a system is bandwidth reduction or, how to encode digital images with as few bits per picture element as possible and such that the closest possible version of the original may be retrieved . In theory, for a given ensemble of signals we can define a distortion measure d and a rate distortion function R(d) based on this distortion. It can be shown [2] that for any distortion d there exists a coding scheme that will code the image with R(d) bits per sample (or unit area, in case of images). Unfortunately the creation of this most efficient code requires the knowledge of the joint probability density function of the ensemble, which is beyond our reach. Some assumptions have been made about this function but none was close enough to give a good approximation. Beyond the knowledge of the source statistics one needs to define a rate-distortion function. The author is unaware of any rate distortion function for sources other than Gaussian. The direct consequence is that although in theory we can get as low as R bits per sample (for a given distortion) in practice this is unachievable. The problem is thus

to find a reasonably efficient coding scheme that would be practical.

One of the first schemes suggested in the image processing context is the one introduced by Roberts[8]. The algorithm goes as follows. To the image that has to be encoded we add pseudo random noise of zero mean and peak value K. The result is then quantized (in quantization intervals of 2K) and truncated to yield the desired number of levels. At the other end the pseudo random noise is subtracted to yield the final retrieved image. This subtraction introduces a synchronization problem between the sending and receiving stations which, if possible, should be avoided.

The system can be viewed as a "black-box" that produces the output signal from the input signal by adding noise, quantizing, truncating and subtracting the same noise again. This black box can be also modelled as a simple addition of noise (not the one actually added) to the truncated input signal. Roberts investigated the statistics of the output signal and states that if truncation is ignored (or possibly avoided) the equivalent noise (i.e. the difference between the output and input signals), although different from the one actually added, has the same distribution. As to the second order statistics it has to be determined how much is the equivalent noise correlated with the original and what is its "color".

Lippel and Kurland [5] carried Roberts' ideas one step further. They questioned the whole idea of using pseudo random noise and suggested instead using a stylized pattern which is termed

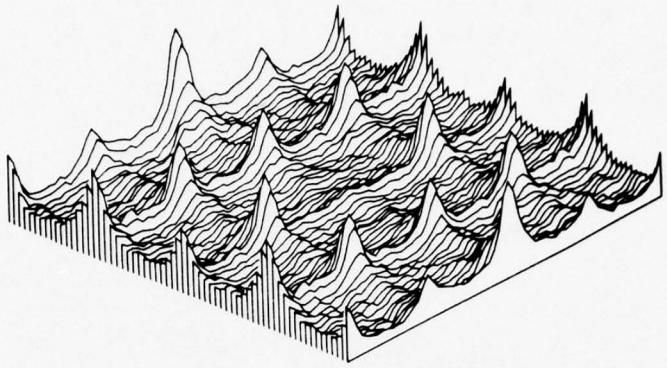
"dither". An attempt was made to find optimal patterns according to some criteria and a few of the results are given in [5], one of which is used throughout this work. The idea of the dither is of importance because of a partial solution it provides to the synchronization problem. The pattern Lippel and Kurland suggested consists of a (relatively) small kernel that is repeated over and over again (the kernel used in this work is a 4X4). This means that the synchronization is limitted to the knowledge of the kernel only and not to a big record like in the pseudo random noise method. Lippel and Kurland also noticed that using dither without the subtraction at the other end yield satisfactory results. In the work done here the use of dither proved to be superior to the use of pseudo random noise especially in the case where no subtraction was performed. Lippel and Kurland, interested only in the final appearance of the image, did not investigate the statistical properties of the resultant image which are obviously different from those obtained by Roberts' method.

A few experiments were conducted to find (empirically and theoretically) the color of the equivalent noise and its correlation with the input signal, especially when deterministic dither patterns are used. When random white noise was used the equivalent noise was white even for colored inputs (for example fig. (1.6b)). The use of dither pattern is more troublesome in that it is hard to talk about its statistics, especially second order. Even if one assumes that averaging periodograms produces some information about the spectral properties one is faced with a more severe problem - the fact that

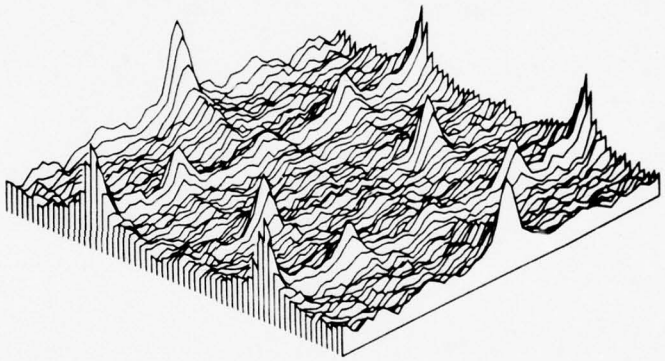
deterministic patterns are not space invariant and thus making the entire idea of power spectral density invalid. However, the fact that the power spectrum of the dither pattern itself is never calculated, but rather the power spectrum of the equivalent noise, leaves some hope for good results. It is expected though that the equivalent noise be correlated with the input signal even for quite fine quantization. Indeed, this fact was experimentally confirmed and the power spectrum of the equivalent noise is a mixture of the power spectrum of the input signal and the Fourier components of the dither pattern.

We took the MILL, added a dither pattern, quantized to 1 bit/pel and then subtracted the dither pattern and the original to leave us with the equivalent noise. The power spectrum estimate of the noise is shown in fig (5.1a) and clearly demonstrates that this is a blend of the image and the dither. The power spectrum of the image shows up in the shape of the peaks, and the symmetric distribution of the peaks is due to the symmetry that the dither has in the frequency domain. When quantizing finer the shape of this power spectrum becomes more flat (i.e. smaller max/min ratio) but retains the shape of symmetrical peaks as shown in fig (5.1b) where the same experiment was repeated with quantization of 3 bits/pel. If the noise that is actually added is white and does not have a stylized Fourier transform like dither patterns, the equivalent noise is white as shown in fig (5.1c). In either case, if the original signal is white the peaks disappear and the equivalent noise is white. When using our system the signal entering the

(a)



(b)



(c)

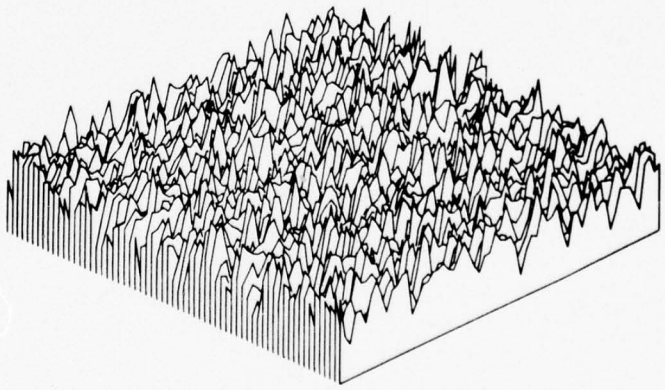


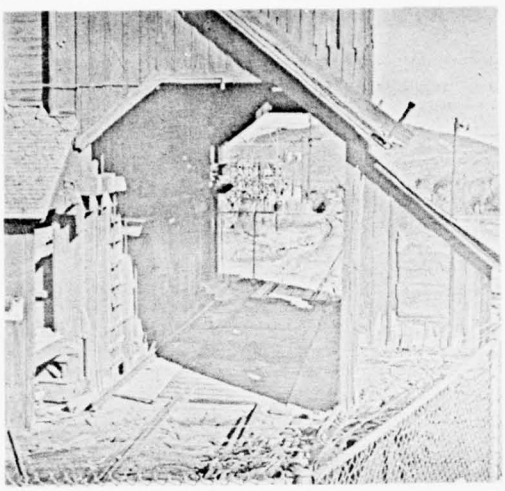
Fig. 5.1 Power spectrum estimate of Equivalent Noise using:
(a) Dither and 1 bit/pel quantization
(b) Dither and 3 bit/pel quantization
(c) Random noise and 1 bit/pel quantization

quantizer is preprocessed and as described before the A filter whitens the signal thus making the equivalent noise white, which is what we assumed when designing the preprocessor.

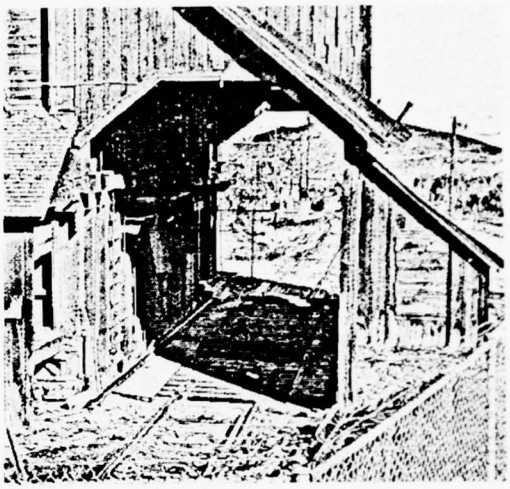
The amount of correlation between the signal and the the equivalent noise varies again depending upon what noise is actually added. The amount of corelation turns out to be the smallest when the input signal is white which is what happens when applying our preprocessor before quantization.

The two originals of fig (1.6a) and fig (1.7a) (with 9 bits/pel) where each processed using 1 bit/pel and applying pseudo random noise or dither with and without subtraction at the retrieval time. The originals used were carefully selected. The first picture is a very "busy" one with a lot of detail. The piece of sky at the top right is of great importance because all the side effects of the process will be clearly seen there whereas they are hidden (although existing, of course) in the other busy sections of the picture. The power spectrum of the image (an estimate of which is given on a db scale in fig (1.6b)) has a lot of energy in off-axes frequencies because of the huge amount of diagonal structure in the original image. The second picture differs quite a bit from the first in that there are almost no man made objects in the scene. This manifests itself in the fact that most of the energy is concentrated along the axes of the power spectrum (given on a db scale in fig (1.7b)).

To demonstrate visually the operation of the system some "internal" signals of the process (designed for $P=P_c$) are shown in



(a)

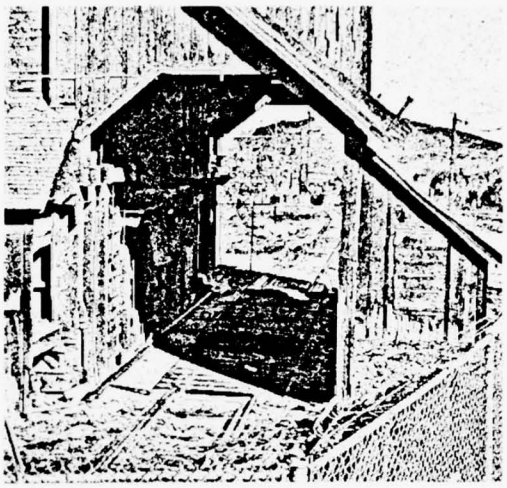


(b)

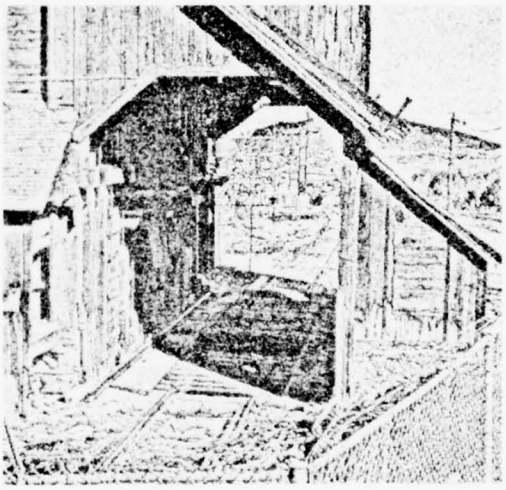
(c)

Fig. 5.2 Internal signals (using $P=P_c$ and dither)

(a) After preprocessing
(b) After quantizing to 1 bit/pel
(c) After subtraction of dither



(a)



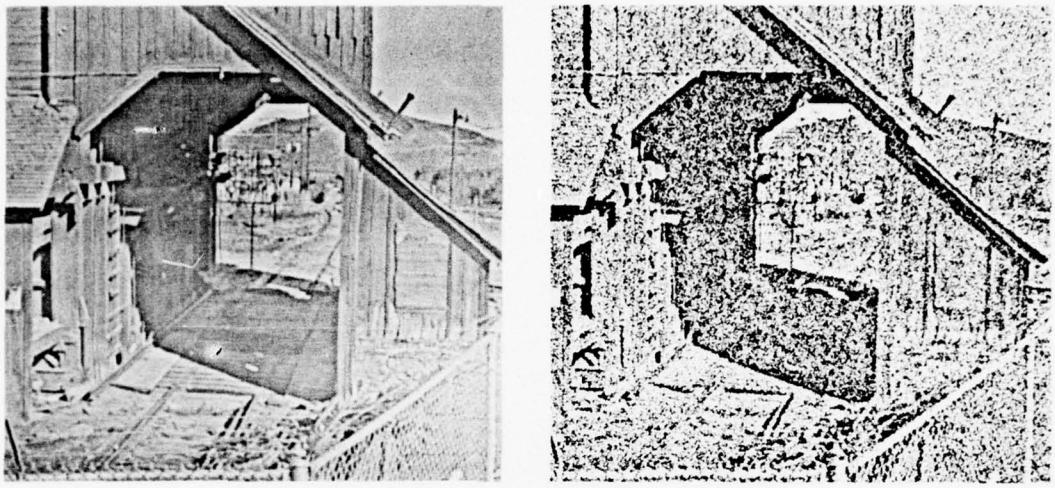
(b)

Fig 5.3 Internal signals (using $P=P_C$ and random noise)

(a) After quantizing to 1 bit/pel
(b) After subtracion of random noise

fig. (5.2). Fig. (5.2a) shows the signal after preprocessing where the contours have been enhanced and the low frequencies almost removed. Fig (5.2b) shows the signal after being quantized to 1 bit/pel (dither was added before quantization took place), and fig. (5.2c) shows the signal prior to postprocessing i.e. after dither was subtracted. The same internal signals when random white noise was used instead of dither are shown in fig. (5.3). The experiment was repeated for a signal to noise ratio of P=1. The signal past the preprocessor, past the quantizer, and with the random noise (which was added before quantizing) subtracted, are shown in fig. (5.4).

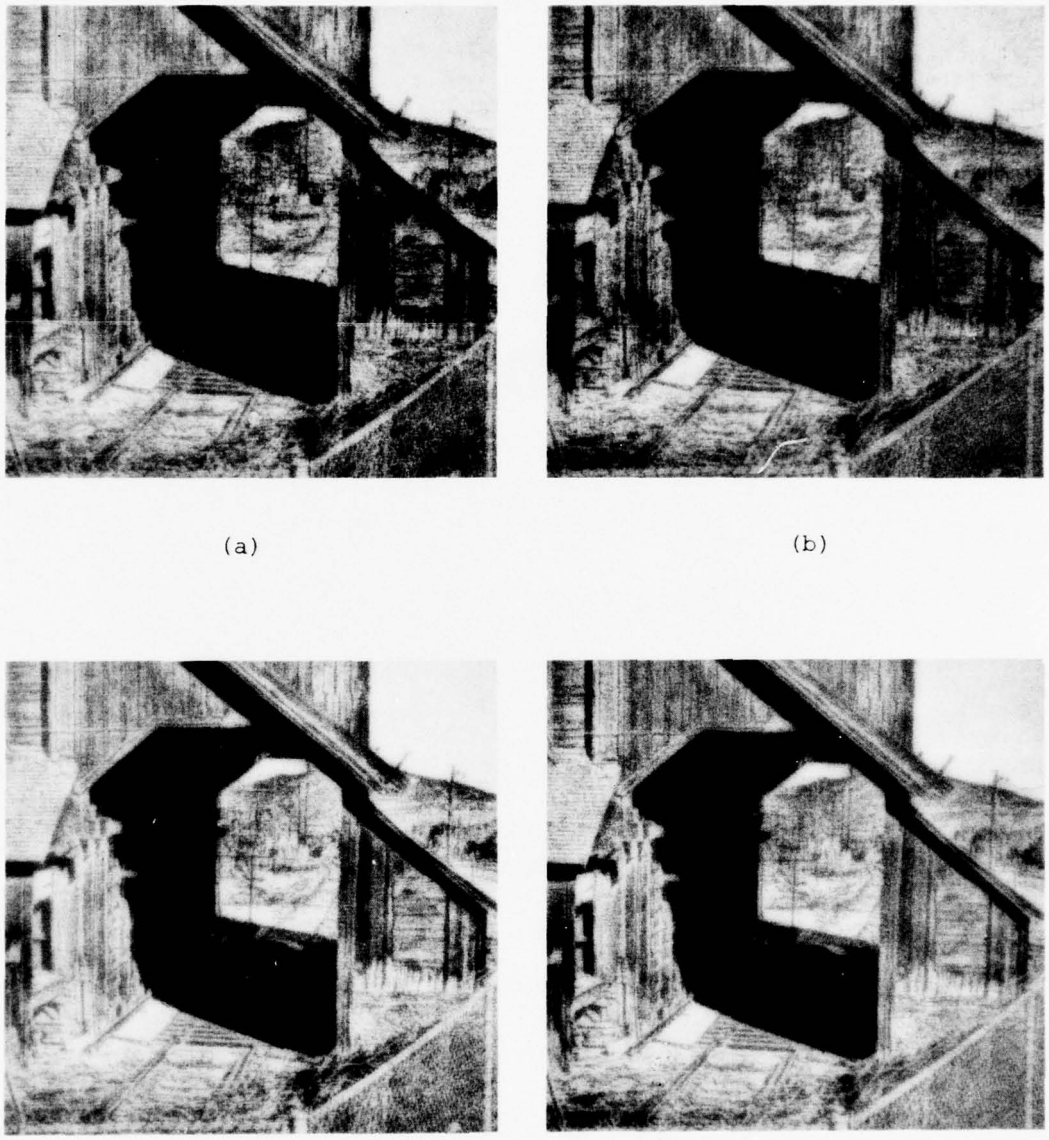
In figures (5.5) and (5.6) we have four processed versions of each of the two originals using the scheme just presented. In all versions we preprocessed (using $P=P_c$) the original, added noise (random or a dither pattern), and quantized to 1 bit/pel. In figs. (5.5a), (5.5c), (5.6a), and (5.6c) we also subtracted the noise that was previously added, from the output of the quantizer. All signals were then postprocessed by the B filter. The version processed with dither and with subtraction seems to be the best looking of all. In all images details are very well preserved. Enhancement done by the author on the shadowed region shows that details in that area, although hardly visible in the original, are all present in the processed versions. Note that the version processed with pseudo random noise looks a little "dirty" especially in the upper right section of the picture. In that domain, in the version processed with dither, the careful observer will find a



(a) (b)

(c)

Fig. 5.4 Internal signals (using P=1 and random noise)
(a) After preprocessing
(b) After quantizing to 1 bit/pel
(c) After subtraction of random noise



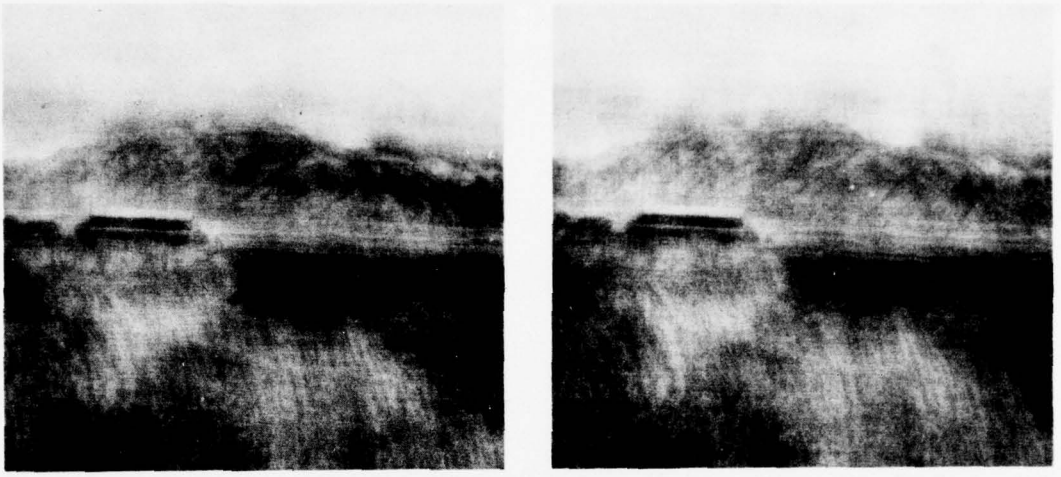
(a) (b)

(c) (d)

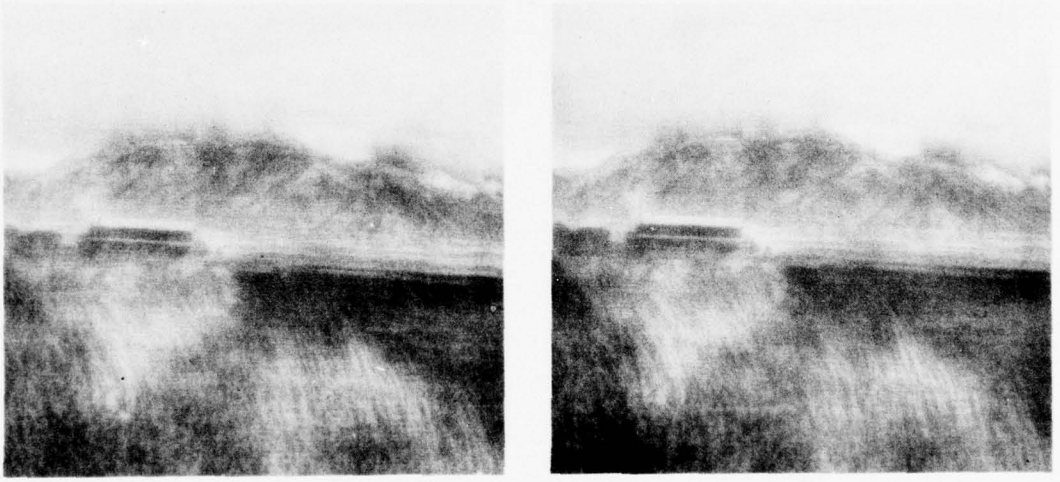
Fig. 5.5 MILL processed with 1 bit/pel and $P=P_c=14.5$.

Preprocessed, noise added, quantized, and postprocessed using:

(a) Pseudo random noise with subtraction
(b) Pseudo random noise without subtraction
(c) Dither pattern with subtraction
(d) Dither pattern without subtraction



(a) (b)



(c) (d)

Fig. 5.6 BARN processed with 1 bit/pel and $P=P_c=14.5$

Preprocessed, noise added, quantized, and postprocessed using:

(a) Pseudo random noise with subtraction
(b) Pseudo random noise without subtraction
(c) Dither pattern with subtraction
(d) Dither pattern without subtraction

dither pattern somewhat embedded in the background. The small difference between the results using dither with and without subtraction suggests of course the use of the method without the subtraction of the dither.

The reason that subtraction in the case of dither improves the final image only slightly is the fact that the dither pattern was chosen such that its energy is entirely concentrated in the very high frequencies which have the lowest visibility. In those frequencies the B filter is attenuating the signal considerably thus reducing the influence of the subtracted dither which as indicated before have low visibility to start with. In the case of pseudo random noise the energy is distributed among all frequencies and the low-frequency component will get through the B filter and show up in the final image.

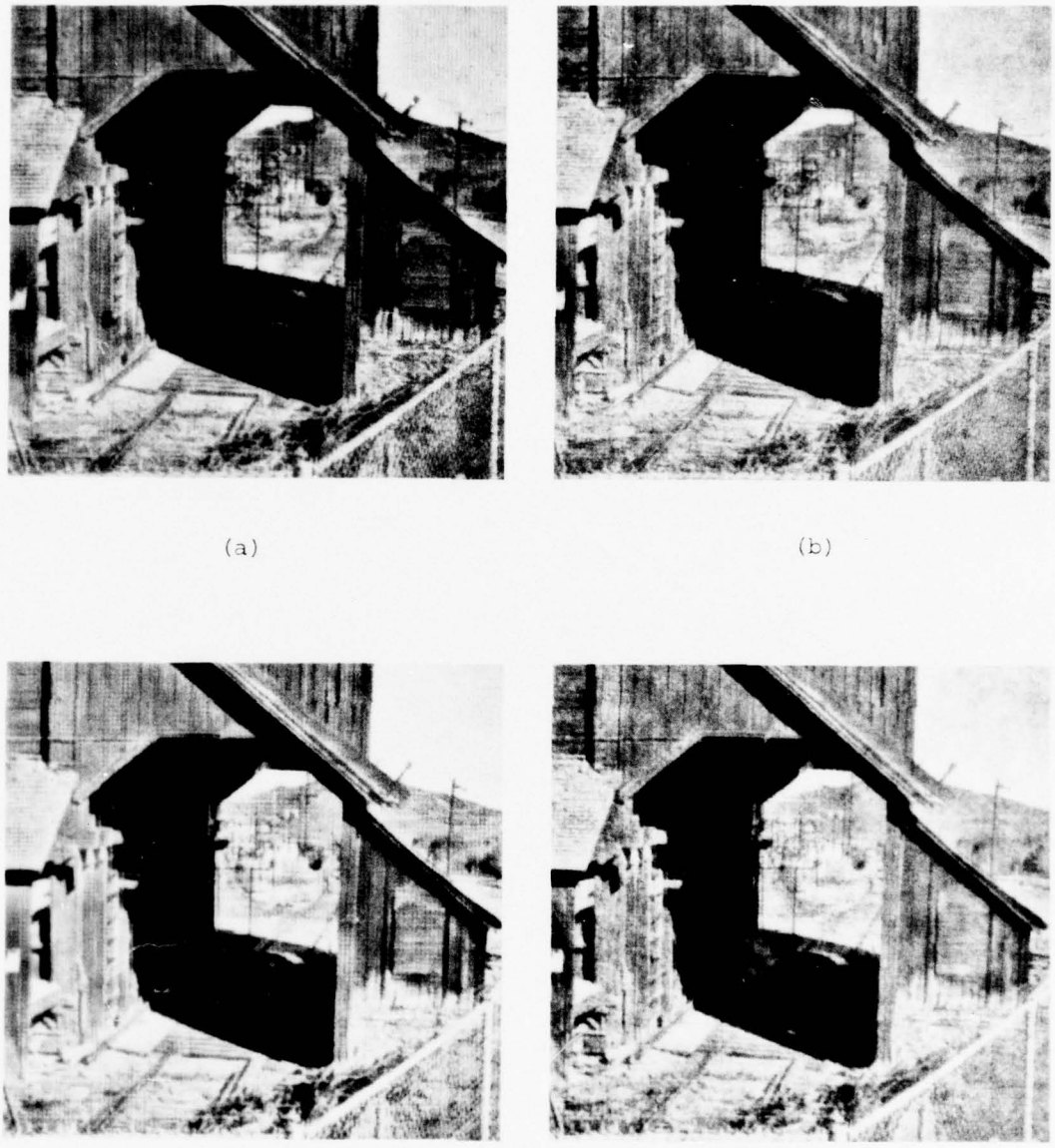
In the original Roberts' scheme the subtraction of the noise plays an important role, whereas in our scheme it seems to have a much smaller effect. This big difference lies in the fact that we have the postprocessor at our disposal to reduce the effect of the subtracted noise. Following the path this noise goes through, we find that it is first lowpassed by the B filter to leave mainly low frequencies. Next, when viewing the picture with a human eye, we pass the noise through a filter which suppresses low frequencies and thus reducing the effect of the components left by the postprocessor.

The dither pattern seen on these images is not the one subtracted but is a pattern created by the process from the dither

added to the signal before being quantized. Attempts to "tailor" the B filter to reduce the visibility of this pattern failed, mainly because the equivalent noise in this rough quantization was too correlated with the signal. In finer quantization this pattern has much less effect.

The result using our coding scheme with much lower S/N ratios are demonstrated in fig. (5.7). Here we have processed the MILL with 1 bit/pel using a signal to noise ratio of 2 in figs. (5.7a) and (5.7b) with dither and random noise respectively, and a ratio of 1 in fig. (5.7c) and (5.7d) with dither and random noise respectively. The observer will notice immediatelly the presence of a dither pattern (or noise) all over the scene even in busy regions. The detail, though, is preserved far better than in the straight Roberts method.

The reduction of the S/N ratio implies in our case the reduction of the entropy, which when applying a source coding scheme plays an important role in the number of bits/pel needed. When using the quantization method the way we did, this ratio is much less important since the quantizer cannot go below 1 bit/pel. It should, however, be chosen to match the "Roberts Box" with the rest of the system as explained later in this chapter. Reduction of the S/N ratio can be achieved by either increasing the noise level or decreasing the total signal power. The reduction in signal power is not done by merely attenuating the signal but also by changes in the shape of the filter (to retain the optimality of the design) which may possibly inhibit some of the frequencies from being transmitted



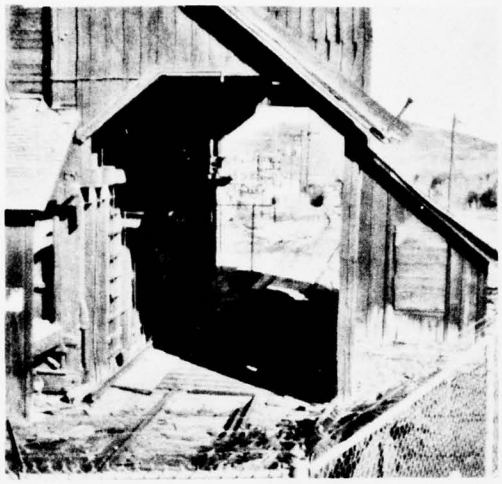
(a) (b)

(c) (d)

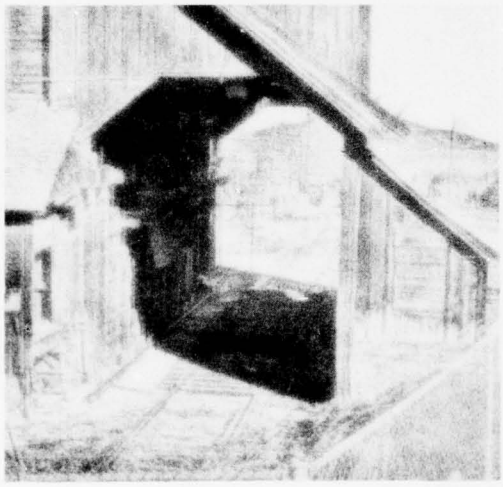
Fig. 5.7 MILL coded at 1 bit/pel and low S/N ratios with subtraction
(a) P=2 and dither
(b) P=2 and random noise
(c) P=1 and dither
(d) P=1 and random noise

and thus causing the resultant image to be more blurred. As results indicate, when using the quantizer method (rather than source coding) it is not advisable to reduce the S/N ratio to such low levels but rather choose some intermediate value.

A very important issue in Roberts' method is the truncation. Roberts assumed that the entire range is devided into the desired number of levels, and the truncation occurs when the addition of noise causes overflow of this range. In our case the noise is given in advance and thus the quantization levels are predetermined. In the optimum case the preprocessor should take precaution to compensate for (or avoid) truncation. However this is a nonlinear operation and the A filter is not optimized for this truncation. Moreover, the fact that A is designed for an average image may cause more than average truncation for some members of the ensemble. We can demonstrate the effect of truncation by the following experiments. We let the quantizer have infinitely many levels (i.e. no truncation) but otherwise leave the process unchanged. The result of this experiment is demonstrated in fig (5.8a) and should be (ensemble wise) equivalent to fig. (2.1c) since the Roberts process without truncation is equivalent to the addition of noise. Alternately one can simulate the Roberts process by inserting a "clipper" before random noise is added. This is shown in fig. (5.8b) which should be (ensemble wise) equivalent to fig. (5.5a). Thus the differences between fig. (2.1c) and (5.5a) or between figs. (5.7d) and (2.3c) are all due to truncation. The flare that is seen near some of the contours is due to severe



(a)



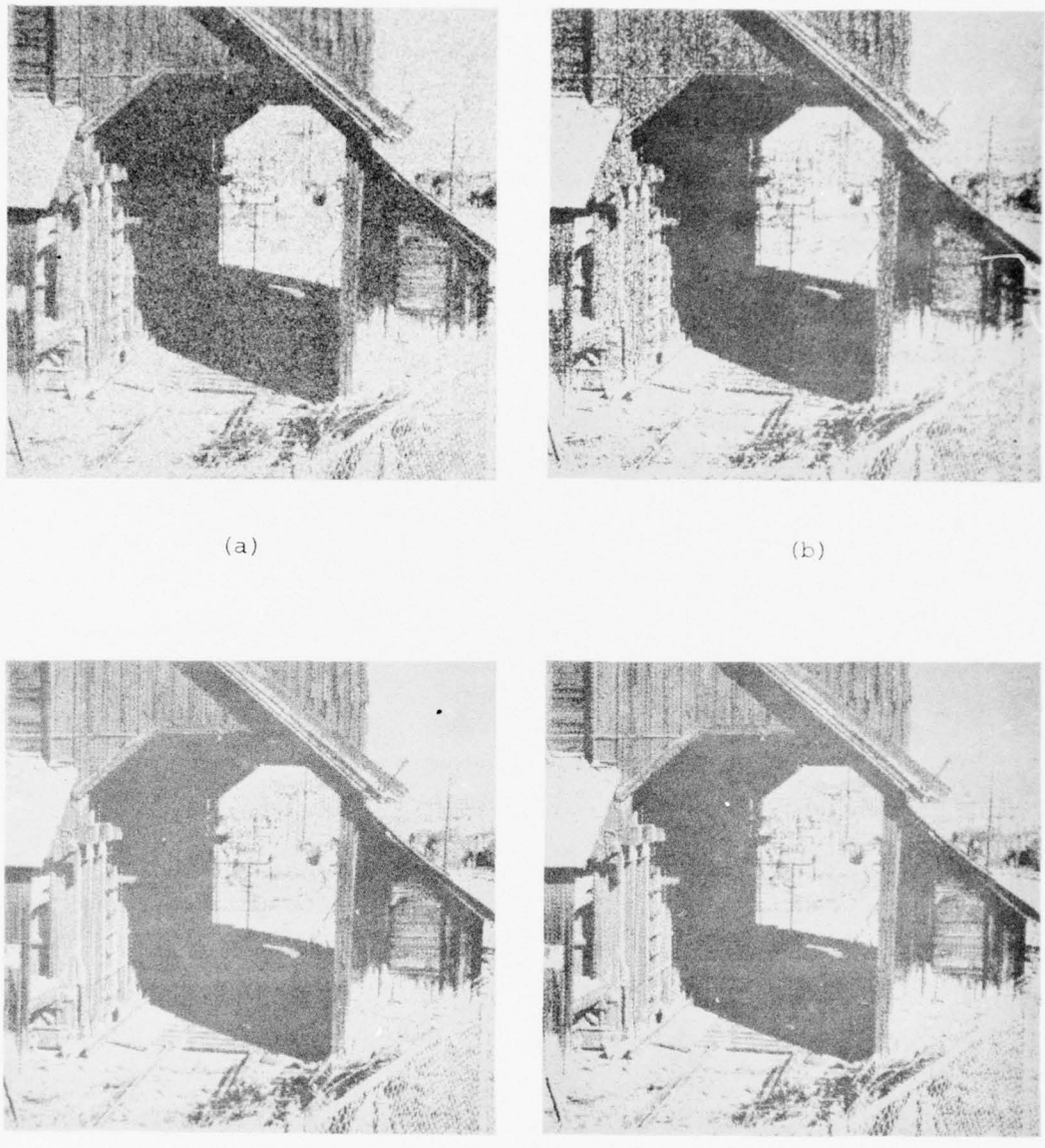
(b)

Fig. 5.8 Truncation experiment
(a) Using random noise, quantizing without truncation, and subtraction of noise
(b) Using random noise with clipping (without quantizing)

truncation in those regions.

As mentioned before there are two signal to noise measures to be considered and in some way to be matched. The first quantity P (or alternately μ) mentioned in chapter one, effects the design of the filters. The second is a signal to noise ratio implied by the Roberts box and the truncation that takes place inside it. Those two quantities need not necessarily be the same but one ought to choose P (or μ) to yield a signal that will suffer least truncation. It is up to the user to decide whether fig. (5.7) (with very little truncation) or fig. (5.5) (with more truncation but better dither suppression) is more to his liking.

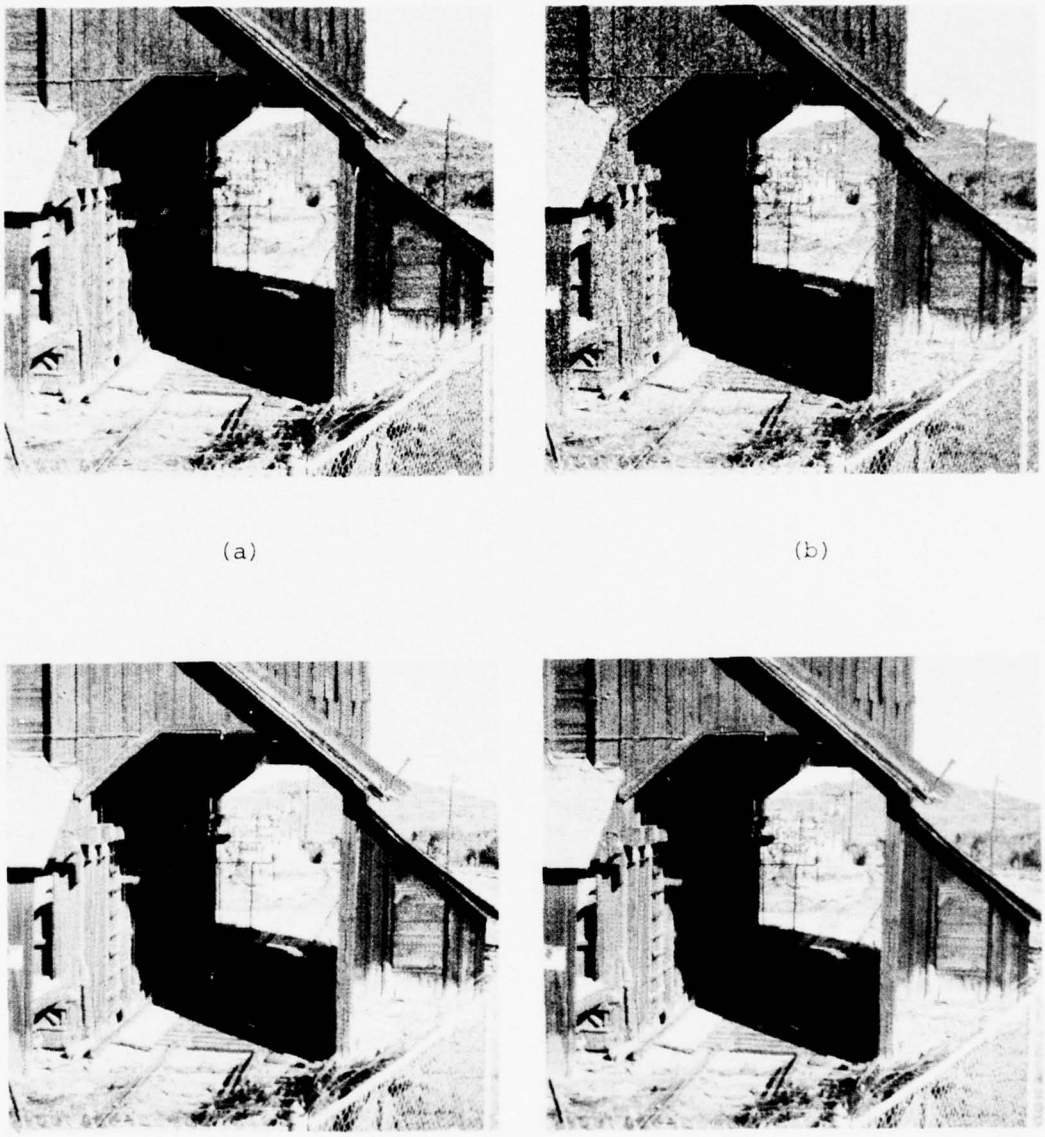
Some comparision between the results using Roberts' method and the system we suggested will be outlined here. In figs. (5.9) and (5.10) we have processed the MILL by Roberts' method using 1 and 2 bits/pel, applying pseudo random noise and a dither pattern and in each case we demonstrate the results with and without subtraction, i.e. the way figs. (5.5) and (5.6) were produced except without pre and post filtering. All the versions of fig. (5.9) look quite dull and a lot of information is gone. The pictures of fig. (5.10) look much better and the user whose objective is to produce sharp looking images will probably prefer this kind of image. However, some information is lost due to this kind of processing. Any observer will note immediatelly the presence of the dither pattern all over the image. Its existence is seen not only in the upper right corner but rather all over the scene. Some information appearing on the original is lost. The careful observer will note that the



(a) (b)

(c) (d)

Fig. 5.9 MILL processed with 1 bit/pel without filtering and
(a) Pseudo random noise with subtraction
(b) Pseudo random noise without subtraction
(c) Dither pattern with subtraction
(d) Dither pattern without subtraction



(a) (b)

(c) (d)

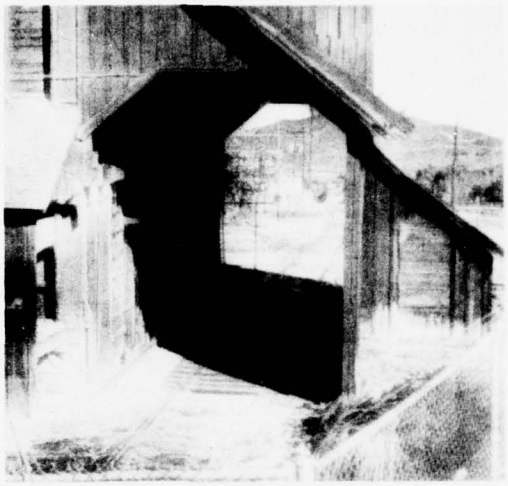
Fig. 5.10 MILL processed with 2 bit/pel without filtering and
(a) Pseudo random noise with subtraction
(b) Pseudo random noise without subtraction
(c) Dither pattern with subtraction
(d) Dither pattern without subtraction

electrical wires that go accross the sky in the upper right corner have disappeared and also some detail in the background, seen through the building, is gone.

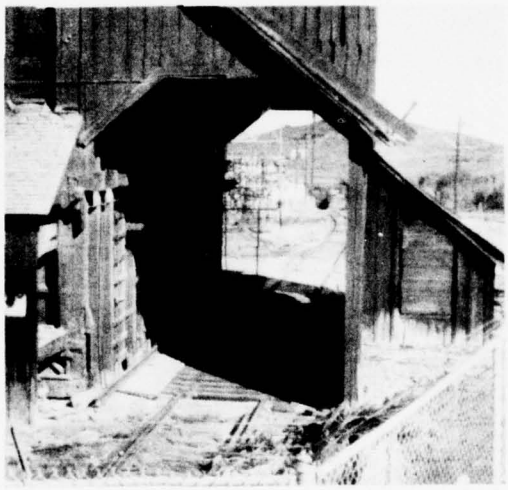
The images processed with our system (we refer, for example, to fig. (5.5c)) tend to be more smeared and hazy. The total appearance of the image is not so sharp (remember, though, that fig. (5.5c) is processed with 1 bit/pel) but information is preserved far better. Note for example that the dither pattern is much less conspicuous. This stems from the fact the the B filter acts as a blurring system on the dither and considerably attenuates the frequencies in which the dither has most of its energy. The pattern that is still seen on the picture results from the dither added before the quantization and not the one that was subtracted later. The electrical wires mentioned before are clearly seen in this image.

We can summarize the comparision by stating that the Roberts method creates (for 2 and more bit/pel, but not for 1 bit/pel) sharper looking images but on the other hand tend to omit some of the fine detail and makes the pseudo random noise or dither pattern more prominent. The system we suggest produces pictures that look more hazy but which preserve more detail. The dither (or any other noise pattern) in our method is much less conspicuous than in the regular Roberts method.

To demonstrate the power of this method and the improvement one gets when more than 1 bit/pel is used we have processed the "MILL" using dither and cuantizing to 2 and 3 bits/pel. The results



(a)



(b)

Fig. 5.11 MILL processed with more than 1 bit/pel ($P=P_c$) using dither pattern with subtraction
(a) 2 bit/pel
(b) 3 bit/pel

are given in fig (5.11). Compare figs. (5.9) and (5.10) (that use Roberts' method with no prefiltering) with fig. (5.11) to verify the improvement one gets when using our method over the regular Roberts method or the dither method.

Note how close these versions are to the original and how with 3 bits per picture element most of the flare, typical to this processing, is almost gone since less truncation took place.

CHAPTER SIX

CONCLUSIONS

This work attempted to incorporate some properties of human vision into the processing of images specifically for image transmission and bandwidth reduction.

The incorporation of a model of human vision seemed to be important since the goal of any image processing is to produce an image to be viewed through this mechanism. The system we suggested includes a preprocessor the output of which is transmitted through a noisy channel whose statistics is assumed to be known and a postprocessor that undoes the preprocessing taking into account the disturbances that might have occurred in the transmission process. Our goal in bandwidth reduction was to employ a coding scheme that is independent of the source so that the source statistics is not required since it is usually not available.

By employing the least squares criterion for optimization we derived the optimal pre and post processors and investigated the sensitivity of the resultant system to the visual model used. As it turns out there is very little dependence of the system on the visual model and a much havier dependence on the statistics of the channel and the ensemble of images.

An important point to consider is the optimization criterion itself. The minimum least squares criterion is not always adequate for image processing. Some of the results obtained, depend on the criterion chosen. Unfortunately there are not very many criteria whose characteristics have been thoroughly investigated and whose mathematical properties are in some way tractable. One of the most important consequences of using this criterion is the fact that the entire system is not very sensitive to the fine structure of the filters. As long as the filters are a matching pair and the prefilter follows the general outlines desired, the performance of the system will be satisfactory. The fine structure of the prefilter will usually not improve the performance of the system significantly. This point has been noticed before by others and is emphasized in this work by comparing the processed versions of the same original with two different sets of filters.

The attempt to employ a coding scheme has been shown to be worthwhile. We have used Roberts' method for quantization and the result, going down to 1 bit per picture element are better than anticipated. Comparision of images coded by Roberts' method and images coded by our scheme show quite an improvement in the numbers of bits/pel needed.

Comparision of the system suggested here with the one suggested by Mannos and Sakrison calls for more detail. Mannos and Sakrison's work uses a source coding scheme and therefore may yield results with an average of less than 1 bit per picture element which is the lower bound of our scheme. Mannos and Sakrison's work sets

an upper bound on the number of bits per sample needed to code that source, but unfortunately the code requires the knowledge of the source statistics which is unavailable. Our scheme, though requiring more bits for the code (per sample) than the bound set by Mannos and Sakrison, is very easy to implement and produces satisfactory results.

In view of the previous discussion it is suggested to merge the two schemes. Some of the equations derived in the two works bear a lot of resemblence which emphasizes their closeness. As of this writing there was no measurement made to compare the results of Mannos and Sakrison with those obtained in this work on a quantitative basis. It is suggested to apply Mannos and Sakrison's work to the optimal system derived here. This may relate the signal to noise ratio to a distortion measure and will yield better understanding on a quantitative basis, of how good this coding scheme does compared to the theoretical bound.

For the purpose of coding, it is not obvious that the random noise or dither pattern used produce best results. A different pattern may improve the performance. Moreover, a special investigation of the quantization levels is required. We chose the quantization level to be 10% of the maximum value of the original image in question. The choice is arbitrary and more experiment and research are needed to establish firm rules.

As to the comments on the visual system, very few have tried to analyze the visual system from a physical point of view. Most of the explanations, as should be, are results of psychophysical

experiments. In this author's opinion those psychophysical results should be analyzed also by physical consideration. The remarks made in this work are by no means complete but should be viewed as suggestions for future researchers in this area. The physical explanation of natural phenomena, of which human vision is one instance, is of importance not only for understanding of the mechanism as such but to assist the design of artificial mechanisms that interact with nature.

APPENDIX A

DERIVATION OF THE EQUATIONS FOR A AND B FILTERS

In this appendix we shall derive and detail the restrictions on the systems A and B. The derivation refers to the block diagram of fig (1.5). †

We shall denote by $a(x,y)$, $b(x,y)$, and $v(x,y)$ the impulse response of the systems A, B, and V respectively †† and by $n(x,y)$ the channel (random) noise. From fig (1.5) we have

(A1) $$D_2' = D(x,y);\ D_1'(x,y) = [a \circledcirc D + n] \circledcirc b$$

(A2) $$D_2' - D_1' = [a \circledcirc b - \delta(x,y)] \circledcirc D + b \circledcirc n$$

where $\delta(x,y)$ is a two dimensional impulse function. Also we have

(A3) $$I''(x,y) = D'(x,y) \circledcirc v(x,y)$$

Our aim is to find the linear systems A and B such that the quantity M defined in (A4) is minimized.

(A4) $$M = E\{[I_2'' - I_1'']^2\} = E\{[(D_2' - D_1') \circledcirc v]^2\}$$

where $E\{\cdot\}$ is the expected value operator and the random processes

---

† Results similar to those obtained here may be partially found in Costas [1].

†† Throughout this appendix, to alleviate the tedious typing procedure, we shall frequently delete the arguments of the functions. In cases where ambiguity may arise we adopt the notation of using lower case letters for space domain functions and upper case letters for the corresponding frequency domain function. The two dimensional frequency (f_x, f_y) will be denoted by (f).

are $D(x,y)$ and $n(x,y)$ which are assumed to be statistically independent. Therefore, using (A2)

(A5) $$M = E\{(a\circledcirc b-\delta)\circledcirc v\circledcirc D + n\circledcirc b\circledcirc v]^2\}$$

$$=E\{[(a\circledcirc b-\delta)\circledcirc v\circledcirc D]^2\} + E\{[n\circledcirc b\circledcirc v]^2\}$$

Using Parseval's theorem and switching to the frequency domain we get

(A6) $$M = \int|A(f)B(f)-1|^2|V|^2S_d df + \int|B|^2|V|^2S_n df \ \dagger$$

where $S_d(f)$ and $S_n(f)$ are the power spectra of the signal $D(x,y)$ and $n(x,y)$ respectively. As explained in chapter one we also want the energy of the signal being transmitted to be constrained thus requiring:

(A7) $$\int|A|^2S_d df = K$$

or by using noise units we define

$$K=p\cdot\int S_n df$$

Where p is a positive integer.

Let us define the quantity

(A8) $$H(f) = [|AB-1|^2S_d+|B|^2S_n]\,|V|^2+\lambda|A|^2S_d$$

in which λ is a constant called Lagrange multiplier. Using the calculus of variations for minimising M in (A6) with the constraint specified by (A7) yields a regular min-max problem on the function H in terms of A and B [3]. Since A and B are in general complex functions we have to solve the following set of four equations:

---

† All the integrals in this appendix have the limits $-\infty$ to $+\infty$ unless otherwise stated.

(A9) $$\partial H/\partial \mathrm{Re}\{A\}=0;\ \partial H/\partial \mathrm{Im}\{A\}=0;$$

$$\partial H/\partial \mathrm{Re}\{B\}=0;\ \partial H/\partial \mathrm{Im}\{B\}=0;$$

Since each of the quantities in the left hand sides of equations (A9) is real (note that H itself is real) we can equivalently solve the following set of two complex equations

(A10) $$\partial H/\partial \mathrm{Re}\{A\} + j\partial H/\partial \mathrm{Im}\{A\}=0$$

$$\partial H/\partial \mathrm{Re}\{B\} + j\partial H/\partial \mathrm{Im}\{B\}=0$$

Performing the derivation of equation (A10) we get:

(A11) $$S_d|V|^2(AB-1)B^*+\lambda S_d A=0$$

(A12) $$S_d|V|^2(AB-1)A^*+|V|^2S_n B=0$$

We shall introduce the following notation in the rest of this appendix:

$$\mu=1/(\lambda)^{1/2};\ S=(S_n/S_d)^{1/2};\ \alpha=1/(\mu|V|)$$

equations (A11) and (A12) can now be rewritten as

(A13) $$(AB-1)B^* = -\alpha^2 A$$

(A14) $$(AB-1)A^* = -S^2 B$$

Deviding (A13) by (A14) we get

$$B^*/A^* = (\alpha^2/S^2)(A/B)$$

or

(A15) $$|B|^2/|A|^2 = \alpha^2/S^2$$

and multiplying (A13) by the complex conjugate of (A14) yield

$$(AB-1)B^*\cdot(AB-1)^*A = (-\alpha^2 A)(-S^2 B^*)$$

(A16) $$|AB-1|^2 = \alpha^2 S^2$$

Eliminating A(f) from (A13) and B(f) from (A14) yield

(A17) $$A = B^*/(|B|^2+\alpha^2)$$

(A18) $$B = A^*/(|A|^2+S^2)$$

Equation (A18) is quite important. It states that B(f) is the Wienner filter for a given A(f).

Multiplying (A17) by the conjugate of (A18) gives

$$AB^* = [B^*/(|B|^2+\alpha^2)]\,[A/(|A|^2+S^2)]$$

$$(|B|^2+\alpha^2)(|A|^2+S^2) = 1 \qquad \text{(A19)}$$

and substituting (A5) in (A19) yields the final expression:

$$|A|^2 = S/\alpha - S^2 = S/\alpha(1-\alpha S) \qquad \text{(A20a)}$$

The left handside of the latter equation is obviously (real and) positive whereas the right handside is not necessarily so. Equation (A20a) should be interpreted (see [1], [3]) as

$$|A|^2 = \begin{cases} S/\alpha(1-\alpha S) & \alpha S < 1 \\ 0 & \text{Otherwise} \end{cases} \qquad \text{(A20b)}$$

Equation (A15) states that whenever |A| is zero so is |B|. At those frequencies the phases of A and B are unimportant. We shall now focus our attention to those frequencies for which $|A| \neq 0$ in trying to find their phases.

Multiplying (A17) by (A18) yields

$$AB(|B|^2+\alpha^2)(|A|^2+S^2) = A^*B^*$$

and using the result of (A19) we get

$$AB = (AB)^*$$

which means that the quantity A·B is real thus

$$\angle A = -\angle B \qquad \text{(A21)}$$

This is the only constraint imposed on the phases of A(f) and B(f) and therefore we may set the phases of both systems to zero at all frequencies. Note that since $S_s(f)$ and $S_n(f)$ are even functions the choice of even V(f) makes A(f) even and thus a(x,y) will also be

real and even.

We now have to evaluate μ (or equivalently λ). We have only to satisfy equation (A7) hence

(A22) $$K = \int |A|^2 S_d df = \int S/\alpha(1-\alpha S) S_d df$$

$$= \mu \int |V| (S_n S_d)^{1/2} df - \int S_n df$$

This integral is evaluated at the region for which $\mu > (S_n/S_d)^{1/2}/|V|$. This yields:

(A23) $$\mu = (p+1) \int S_n df / \int |V| (S_n S_d)^{1/2} df$$

There exists a smallest μ such that $\mu > (S_n/S_d)^{1/2}/|V|$ for all frequencies. This is called The critical μ (denoted by μ_c). For $\mu > \mu_c$ all integrals are evaluated from $-\infty$ to $+\infty$ and all the results may be obtaind in close form. For $\mu > \mu_c$ we have

$$|A|^2 = (S_n/S_d)^{1/2} (\mu V - (S_n/S_d)^{1/2}]$$

$$B = A^*/[\mu V (S_n/S_d)^{1/2}]$$

$$M = [\int |V| (S_n S_d)^{1/2} df] / \mu$$

The only items that remain to be taken care of are the special cases: $S_n=0$, $S_d=0$ and $V=0$. By looking at equations (A11) and (A12) we note that if for some frequency $V(f)=0$ equation (A12) holds and equation (A11) yields $A(f)=0$, and therefore $B(f)$ may be set to zero for that frequency. This is only a hypothetical case since the human visual system is not "blind" to any frequency.

If $S_d(f)=0$ for some frequency equation (A11) holds and equation (A12) yields $B(f)=0$ and $A(f)$ is arbitrary at that frequency

and may thus be set to zero to save power. This case means that in the entire ensemble there is not one single member that has a nonzero component in this frequency. In the ensemble of regular images, with which we are dealing, this does not happen.

If $S_n=0$ for some frequency we have from equation (A14) $B(F)=A^{-1}(f)$ which when substituted into (A13) gives $A(f)=0$ and $B(f)$ is undefined (also A has no inverse). This situation has to be interpreted in the following way. If $S_n(f)=0$ it means that there will never be a noise component in this frequency and the information conveyed by this frequency can be accurately restored. But since we want to use the least power possible for sending this information we may set $A(f)=\epsilon$ where ϵ is arbitrarily small and then make $B(f)=1/\epsilon$ in this frequency, so that the information is truely restored and a very small amount of power used. Practically we do not expect nature to be that courtous so this situation, although possible, is very rare.

REFERENCES

1. Costas J.P. "Coding With Linear Systems", Proceedings of the IEEE Vol 48 No. 9,September 1952 pp. 1101-1103.

2. Berger T. "Rate Distortion Theory", Prentice Hall, New-Jersy, 1971.

3. Hildebrand F.B. "Methods of Applied Mathematics", Prentice Hall, Englewood Cliffs, N.J., 1952.

4. Huang T.S. et al, "Image Processing", Proceedings of the IEEE Vol 59, No. 11, November 1971.

5. Lippel B. and Kurland M., "The Effect Of Dither On Luminance Quantization Of Pictures", IEEE Trans. Comm. Tech. Vol COM-19, No. 6, December 1971, pp. 879-888.

6. Mannos J.L. and Sakrison D.J. "The Effect Of A Visual Fidelity Criterion On The Encoding Of Images", IEEE Transc. on Information Theory, Vol IT-20, No. 4, July 1974, pp. 525-536.

7. Oppenheim A.V., Schafer R.W., and Stockham T.G.Jr., "Nonlinear Filtering of Multiplied and Convolved Signals", Proceedings of the IEEE Vol 56, August 1968, pp. 1264-1291.

8. Roberts L.G. "Picture Coding Using Pseudo Random Noise", IRE Transc. on Information Theory, Vol IT-8, No. 2, February 1962.

9. Stieglitz K. "The Equivalence Of Digital And Analog Signal Processing", Inform. Contr., Vol 8, 1965, pp. 455-467.

10. Stockham T.G. Jr. "Image Processing In The Context Of Visual Model", Proceeding of the IEEE Vol 60, No. 7, July 1972, pp. 828-842.

11. Stockham T.G. Jr. "Intra-Frame Encoding For Monochrome Images By Means Of A Psychophysical Model Based On Nonlinear Filtering of Multiplied Signals", Proceedings of the Symposium on Picture Bandwidth Compression, M.I.T. Cambridge, Mass. April 1972, pp.415-442.

12. Stockham T.G.Jr. "Natural Image Information Compression with a Quantitative Error Model", Proceedings of the second University of Illinois conference on computer graphics, 1969, p. 67.

13. Welch P.D. "The Use Of Fast Fourier Transform For The Estimation Of Power Spectra: A Method Based On Time Averaging Over A Short, Modified Periodograms", IEEE Trans. on Audio Electronics Vol AU-15 June 1967, pp. 70-73.

ACKNOWLEDGEMENT

My deepest gratitude to Dr. T.G. Stockham for suggesting the subject and for his guidance and advice throughout this research. The creative and stimulating atmosphere he creates and the personal help he provided were essential in making this work possible.

I would like also to thank Dr. C. Rushforth for the guidance criticism and help that brought about this work. Thanks also to Dr. R. Plummer and Dr. L. Timothy who helped in evaluating this work. Special thanks are due to M. Milochik for his expert photographic work.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER UTEC-CSc-75-115	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMAGE TRANSMISSION AND CODING BASED ON HUMAN VISION		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Raphael Jona Rom		8. CONTRACT OR GRANT NUMBER(s) DAHC15-73-C-0363
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Utah Salt Lake City, Utah 84112		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order #2477
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209		12. REPORT DATE September 1975 13. NUMBER OF PAGES 86
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In recent years more and more attention was paid to digital image processing especially as a result of the development of highly efficient algorithms and also because of technologically better facilities. Concurrently attempts were made to find a mathematical model for human vision to acheive better understanding about that mechanism. Some of the image processing problems that were (and are) tackled are image enhancement, bandwidth reduction, image transmission, etc. Unfortunately very few have taken the mechanism of the human vision into consideration in their processes.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

This work is an attempt to incorporate the model of human vision in image transmission and coding. An optimal system is developed to transmit a digital image over a noisy channel. The same system is used for image bandwidth reduction utilizing a simple coding scheme which is not based on the knowledge of the statistics of the image in question. We demonstrate the improvement of the optimal system over other similar systems and provide explanation for situations where other systems failed. The model we use for transmitting images can be also interpreted as the model of the visual mechanism itself and thus shed some light on human vision from a new interesting aspect.

are demonstrated