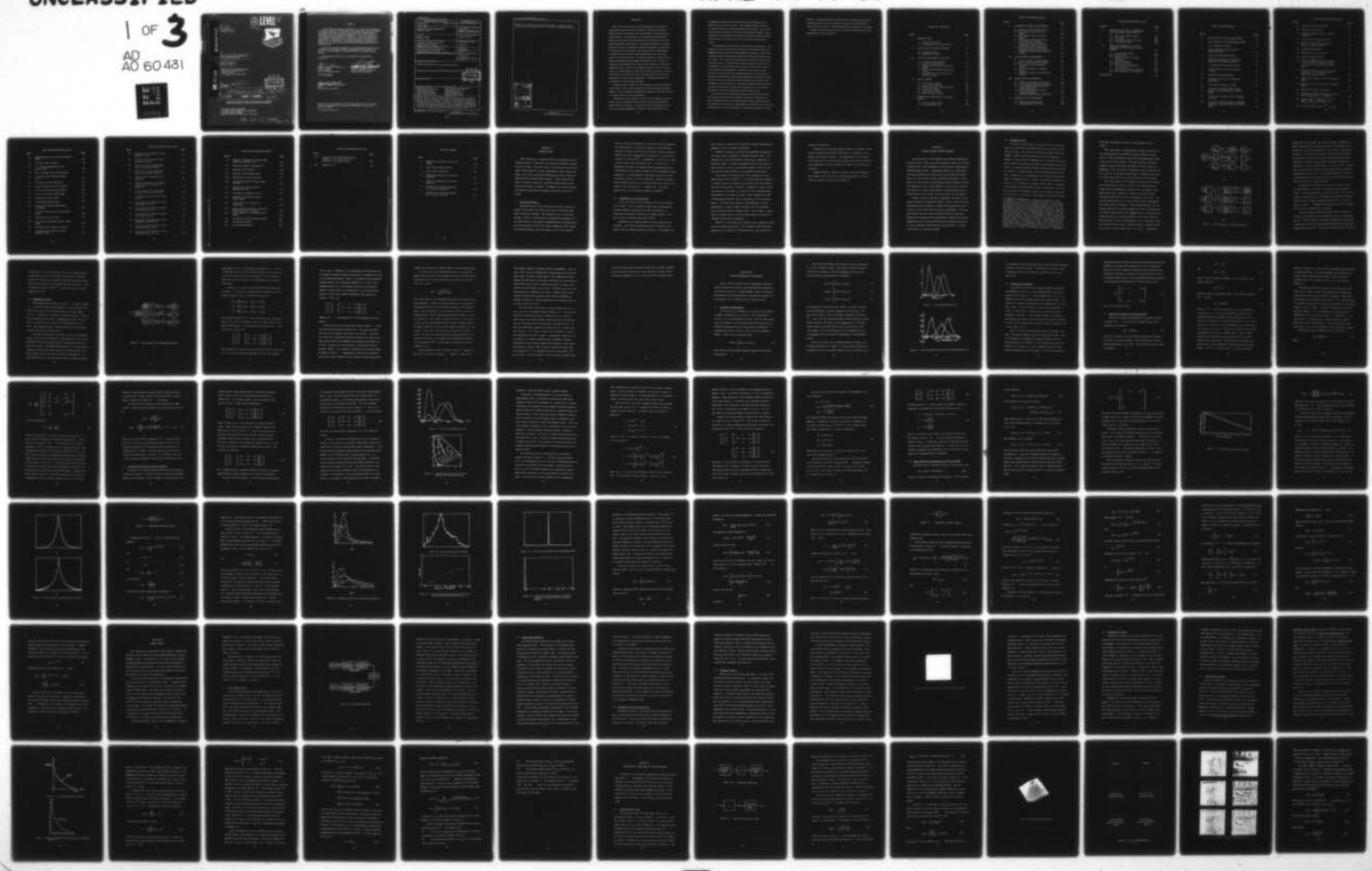
AD-A060 431 UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES IMAGE --ETC F/G 20/6
DIGITAL COLOR IMAGE COMPRESSION IN A PERCEPTUAL SPACE.(U)
JUL 78 H C ANDREWS, C F HALL F33615-77-C-1016

UNCLASSIFIED AFAL-TR-78-80 NL

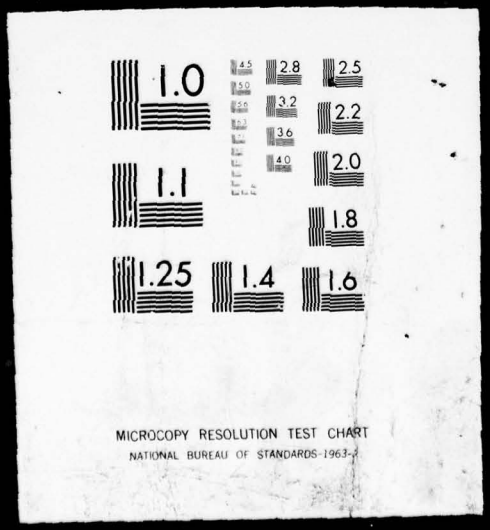
1 OF 3
AD
A060431



SIFIED

1 OF 3

AD
AO 60431



(2) LEVEL III

SC

A053343

AD A060431

DDC FILE COPY

(18) (19) AFAL-TR-78-80



(6) DIGITAL COLOR IMAGE COMPRESSION IN A PERCEPTUAL SPACE

(10) Harry C. Andrews
Charles F. Hall

University of Southern California
Image Processing Institute
University Park
Los Angeles, California 90007

DDC
RECEIVED
OCT 27 1978
B

(11) JUN[redacted]78

(12) 238 p.

(15) F33615-77-C-1016

(16) 2003

(9) Final Report. Jan[redacted]77-Feb[redacted]78

(17) 06

Approved for public release; distribution unlimited.

AIR FORCE AVIONICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

78 10 16 028

391 141

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Information Office (IO) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Peter C Camana

PETER C. CAMANA
Project Engineer
Data Link Technology Group

Charles C. Gauder

CHARLES C. GAUDER, Chief
Avionic Communications Branch
System Avionics Division

FOR THE COMMANDER

Raymond E Siferd

RAYMOND E. SIFERD, Lt Col, USAF
Chief
System Avionics Division

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

AIR FORCE/56780/6 October 1978 — 75

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFAL-TR-78-80	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DIGITAL COLOR IMAGE COMPRESSION IN A PERCEPTUAL SPACE		5. TYPE OF REPORT & PERIOD COVERED Final Report January 1977-February 1978 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Harry C. Andrews Charles F. Hall		8. CONTRACT OR GRANT NUMBER(s) F33615-77-C-1016
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Southern California Image Processing Institute University Park, Los Angeles, Ca. 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2003-06-38
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Avionics Laboratory (AFSC) System Avionics Division (AA) Wright-Patterson Air Force Base, OH 45433		12. REPORT DATE June 1978 13. NUMBER OF PAGES 224
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)

Approved for public release; distribution unlimited.

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

DDC
RECEIVED
OCT 27 1978
B

18. SUPPLEMENTARY NOTES

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

Human Visual System	Discrete Cosine Transform Coding
Video Bandwidth Compression	Human Visual System Coding
Human Perceptual Model	Color Image Coding
Image Evaluation Criterion	Perceptual Space Coding
Fourier Transform Coding	Digital Image Processing

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

Simple mathematical models are developed from the physiological and psychophysical traits of the human visual system. Expressions for the statistical characterization of these models are obtained. When used as a preprocessor, the models are shown to produce images which are statistically compatible with the underlying assumptions necessary to solve the parametric rate-distortion equations. The derived power spectrum equations were used to code black and white and color images with a quality superior to previous results. In

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

78 10 16 026

addition, it is shown that the preprocessor produces a "perceptual space" in which normalized mean square error is an effective image quality measure.

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

PREFACE

This report is the result of research based on the premise that image processing systems can be made more effective if relevant characteristics of the human visual system (HVS) are an integral part of their design. Few would take issue with this general statement; however, prior image processing work has concentrated on the development algorithms and hardware. Few image processing researchers have had the time or inclination to study the physiology or psychophysical characteristics of the HVS. On the other hand, the physiologist is seldom interested in the practical applications of his work with respect to image processing. The difficulty becomes readily apparent when one attempts to find an applicable journal to read. There is no journal (to my knowledge) which spans all of the fields associated with image processing. This is not an atypical situation when one is interested in a multidisciplinary field such as image processing. Because of this problem I have tried to include a layman's guide to the HVS with appropriate references in the Appendices. The interested reader is encouraged to read the Appendices first.

I would also like to point out that the models developed and analyzed in this report are by no means limited to bandwidth compression applications. The experimental applications were limited to this area because of current interests and time limitations.

Unfortunately, time limitations also lead to compromises, particularly in such an exciting field which offers so many research paths. Section VIII is an example. I consider the issue of image quality measures of paramount importance. However, the work that had to be

accomplished leading up to this topic did not leave enough time for experimental work on this subject. The lengthy paradigms required for valid psychovisual results contributed to the problem. As a result, Section VIII presents what I consider to be preliminary results. These comments are not meant to cast doubt on the results reported in Section VIII but rather to encourage the reader to put them in their proper context.

I am indebted to many for their assistance and encouragement. My original interests in the human visual system was kindled by Professor Matthew Kabrisky at the Air Force Institute of Technology. Many of the achromatic model considerations came out of discussions with Professor E. L. Hall. Dr. Werner Frei provided many fruitful discussions on the chromatic model. I would also like to thank Professor Lloyd Welch for "reminding" me that characteristic functions are more than a figment of a mathematicians imagination. Indeed, his help in this area led directly to the power spectrum equations which are of fundamental importance to the bandwidth compression applications discussed in Sections VI and VII. Most importantly, I wish to acknowledge the guidance and assistance of Professor Harry C. Andrews throughout the past two years. I am still amazed that he accepted the challenge of our association and hope that it has been as rewarding for him as it has for me.

The true test of most image processing research is in the viewing. To the extent that this work may appear successful, I am indebted to Mr. Ray Schmidt and the rest of the Image Processing Laboratory staff: Gary Edwards, John Horner, Toyone Mayeda, David Nagai, Clay Olmstead, Patrick Stoliker, and James Tertocha. The artwork was done by Doyle

Howland. Ed Kasanjian and Dave Peck assisted in software development and Marilyn Chan, Amy Yiu and Eileen Jurak provided administrative and secretarial services. The final manuscript was typed by Lucy Cheung.

Finally, I want to thank the U.S. Air Force for providing me with the opportunity to perform this work.

TABLE OF CONTENTS

Section		Page
I	INTRODUCTION	1
	1.1 Research Objectives	1
	1.2 Organization of the Dissertation	2
II	HUMAN VISUAL SYSTEM MODELS	5
	2.1 Biological Model	6
	2.2 Mathematical Model	10
III	CHARACTERIZATION OF IMAGERY	17
	3.1 Continuous Representation	17
	3.2 Discrete Representation	20
	3.3 Spatial Decomposition of Discrete Images	21
	3.4 Spectral Decomposition of Discrete Images	25
	3.5 Some Statistical Characterics of Discrete Images	33
IV	IMAGE CODING	53
	4.1 The Coding Problem	54
	4.2 Pulse Code Modulation	57
	4.3 Differential Pulse Code Modulation	58
	4.4 Transform Coding	59
	4.5 Psychovisual Coders	63
	4.6 Rate Distortion Theory	64
V	STATISTICAL ANALYSIS OF THE HVS MODEL	72
	5.1 The Achromatic Case	72
	5.2 The Chromatic Case	84

TABLE OF CONTENTS (cont'd)

Section		Page
VI	ACHROMATIC CODING EXPERIMENTS	95
	6.1 Block Cosine Transform Coding	95
	6.2 Full Image Cosine Transform Coding	100
	6.3 Full Image Fourier Transform Coding	102
	6.4 Block Cosine Transform Coding in the Perceptual Domain	104
	6.5 Full Image Cosine Transform Coding in the Perceptual Domain	104
	6.6 Full Image Fourier Transform Coding in the Perceptual Domain	104
	6.7 Perceptual Domain Power Spectrum Coding	109
VII	COLOR CODING EXPERIMENTS	116
	7.1 Color Coordinate Transformations	116
	7.2 Block Cosine Transform Coding	123
	7.3 Full Image Cosine Transform Coding	129
	7.4 Full Image Fourier Transform Coding	130
	7.5 Perceptual Domain Power Spectrum Coding	130
VIII	IMAGE QUALITY MEASURES	144
	8.1 Standard Image Quality Measures	144
	8.2 A Perceptual Image Quality Measure	148
	8.3 An Achromatic Subjective Image Quality Experiment	151
	8.4 A Chromatic Subjective Image Qualtiy Experiment	154
	8.5 Computed Image Quality Experiment	156
IX	SUMMARY AND CONCLUSIONS	162
	9.1 Summary and Conclusions	162
	9.2 Recommended Future Work	163

TABLE OF CONTENTS (cont'd)

Appendix		Page
A	SOME PHYSIOLOGICAL PROPERTIES OF THE HUMAN VISUAL SYSTEM	167
	A.1 The Ocular Optical System	169
	A.2 The Retina	172
	A.3 The Lateral Geniculate Bodies	180
	A.4 The Visual Cortex	183
B	SOME PSYCHOPHYSICAL CHARACTERISTICS OF THE HUMAN VISUAL SYSTEM	185
	B.1 A Fundamental Result	186
	B.2 Visual Acuity	189
	B.3 Spatial Frequency Response Functions	193
	B.4 Lateral Inhibition	197
	B.5 Spectral Properties	202
	B.6 Trichromatic and Opponent Color Theories	206
	B.7 Luminosity and Color Constancy	208
	B.8 Some Temporal Considerations	210
REFERENCES		215

LIST OF ILLUSTRATIONS

Figure		Page
1	Block Diagram of the Biological Model	8
2	Block Diagram of the Mathematical Model	11
3	Color Matching Functions	19
4	Color Matching Functions for the 2° Standard Observer	19
5	XYZ Color Matching Functions	28
6	CIE Standard Observer Chromaticity Diagram with MacAdam Ellipses	28
7	Correlation Between Pixels of GIRL	36
8	Autocorrelation and Power Spectrum of GIRL	38
9	Exponential Nonlinear System	39
10	Frequency Curves of the Lognormal Distribution	41
11	First-order Histogram of GIRL	42
12	First-order Histogram Data of GIRL Plotted on Log-probability Coordinate System	42
13	First-order Histogram of the Logarithm of GIRL	43
14	First-order Histogram Data of ℓn [GIRL] Plotted on Linear-probability Coordinate System	43

LIST OF ILLUSTRATIONS (cont'd)

Figure		Page
15	Logarithmic Nonlinear System	47
16	The Image Coding Task	55
17	Mandala Ordered 8 x 8 Block Cosine Transform	61
18	A Typical Rate Distortion Curve	66
19	R(D) for a Memoryless Gaussian Source and MSE Criterion	66
20	Achromatic HVS Model	73
21	Simplified Achromatic Model	73
22	Isotropic Bandpass Filter	76
23	LAX and BRIDGE Image (a) Originals, (b) Processed by Achromatic Model, (c) Processed by Reduced Achromatic Model	78
24	Input Power Spectrum, Filter Function, and Output Power Spectrum of the Reduced Achromatic Model	80
25	Output Power Spectrum of the Achromatic Model	80
26	Rate Distortion Curves for the Achromatic Model	83
27	Simplified Chromatic HVS Model	85
28	Linear-probability Plot of Histogram Data from $GIRL_{\ell}$, $GIRL_{c1}$, and $GIRL_{c2}$	87
29	Adjacent Pixel Correlation Curves for $GIRL_{t_1^*}$, $GIRL_{t_2^*}$, and $GIRL_{t_3^*}$	87
30	Power Spectrum Plots for ℓ, c_1, and c_2	92

LIST OF ILLUSTRATIONS (cont'd)

Figure		Page
31	Rate Distortion Curve for the Chromatic Model	94
32	Low-noise GIRL (original)	98
33	A 1 bit/pixel Bitmap (8 x 8 block size, $\rho = .96$)	98
34	Two 1 bit/pixel Cosine Coded Images	99
35	A 256 x 256 Cosine Domain Bitmap, $\rho = .96$	101
36	A 256 x 256 Cosine Coded Image	101
37	Fourier Domain Frequency Location	105
38	A 256 x 129 Fourier Domain Bitmap	105
39	A 256 x 256 Fourier Coded Image	105
40	Psychovisual Cosine Coder	106
41	A HVS Preprocessed Image	106
42	Psychovisual Cosine Coded Images, 1 bit/pixel	107
43	A 256 x 256 Psychovisual Cosine Coded Image	108
44	Psychovisual Fourier Coder	110
45	A 256 x 256 Psychovisual Fourier Coded Image	110
46	Perceptual Power Spectrum Bitmap	113
47	Perceptual Power Spectrum Coded Images (N = 256)	114

LIST OF ILLUSTRATIONS (cont'd)

Figure		Page
48	Perceptual Power Spectrum Coded Images (N = 512)	115
49	The 256 x 256 Color GIRL Image (Luminance Planes)	118
50	The 256 x 256 Color GIRL Image (First Chrominance Planes)	119
51	The 256 x 256 Color GIRL Image (Second Chrominance Planes)	120
52	Cosine Coded 1 bit/pixel 16 x 16 Blocksize	126
53	Monochrome Display of YIQ Space (Cosine Coded 1 bit/pixel 8 x 8 Blocksize)	128
54	Cosine Coded and Fourier Coded (1 bit/pixel 256 x 256 Blocksize)	126
55	Perceptual Power Spectrum Coded (N = 256)	136
56	Perceptual Power Spectrum Coded ANN Image (N = 512)	136
57	Perceptual Power Spectrum Coded LAKE Image (N = 512)	139
58	Perceptual Power Spectrum Coded F16 Image (N = 512)	139
59	Perceptual Power Spectrum Coded BUILDING Image (N = 512)	141
60	Perceptual Power Spectrum Coded BABOON Image (N = 512)	141
61	Perceptual Power Spectrum Coded PEPPERS Image (N = 512)	143

LIST OF ILLUSTRATIONS (cont'd)

Figure		Page
62	Difference Images From ANN Coding Results (see Figure 7.8)	143
A.1	Horizontal Section of Human Eye	168
A.2	Rayleigh Limit Diagram	168
A.3	Structure of the Human Retina	173
A.4	Distribution of Rods and Cones	173
A.5	Absorption Spectra for Three Types of Cones	177
A.6	Photopic and Scotopic Spectral Sensitivity Curves	177
A.7	Nonlinearity Comparison Curves	181
B.1	Modulation Transfer Function of Ocular Media	188
B.2	Visual Acuity as a Function of Luminosity	188
B.3	Spatial Response of HVS as a Function of Mean Luminance in Trolands	196
B.4	Dark Adaptation and Equivalent Background	196
B.5	Illustration of Lateral Inhibition Effects	199
B.6	The Mach Band Effect	200
B.7	Color Naming Functions	205

LIST OF ILLUSTRATIONS (cont'd)

Figure		Page
B.8	Psychophysically Determined Spectral Response of the Photoreceptors (includes ocular media effects)	205
B.9	Benham's Disc	212

LIST OF TABLES

Table		Page
1	Statistical Parameters From GIRL Image	89
2	Color Plane Energy Packing	121
3	Color Plane Correlation	122
4	Subjective Rankings for Achromatic Image Set	155
5	Subjective Rankings for Chromatic Image Set	157
6	Computed and Subjective Rankings for Achromatic Image Set	158
7	Computed and Subjective Rankings for Chromatic Image Set	160

SECTION I

INTRODUCTION

This dissertation is concerned with the processing of discrete, sampled imagery, in particular, the coding and bandwidth compression of such data. The major thesis of this work is that the human visual system (HVS) has certain characteristics which, when quantified, can be used to formulate mathematical models suitable for analyzing and processing digital imagery. These models should lead to a fidelity criterion for visual data which matches human subjective evaluation of images. In addition, more efficient coding and bandwidth compression techniques should evolve from such models.

1.1. Research Objectives

The primary goal of this research in the above context is to quantify these models and verify their utility in coding and bandwidth compression systems. The emphasis here is on the word quantify. Several researchers have recognized the importance of the characteristics of the HVS in implementing and evaluating image processing systems [1] thru [15]. This recognition is often limited to an acknowledgement of the importance of one or two specific

facets of the HVS accompanied by a heuristic argument supporting the implementation of a particular image processing technique. This type of approach (which could be called a top down approach) assumes one knows a priori which characteristics of the HVS are relevant to the task at hand. Unfortunately, the HVS is a complex nonlinear system with interrelated traits. As shown in [1], a simplifying assumption with regard to the nonlinearity alters the characteristics of the system and fails to reveal important contrast properties. To take advantage of the entire system it is more reasonable to study or model the HVS with a bottom up approach. After analyzing the effects of the entire system, the model may be reduced to one appropriate for a specific task. This latter approach will be used during the present investigation.

1.2. Organization of the Dissertation

In the next section we will develop a model for the human visual system. First a biological model based on physiological and psychophysical properties of the HVS is presented. A mathematical homologue -- which can be readily analyzed -- will then be used to quantify the biological model.

In Section III the characterization of visual images will be presented. The statistical properties will be developed in consonance with the model generated in Section II. The spectral (or

color) content of images will be discussed in detail with particular emphasis on color coordinate conversions.

Section IV contains a brief survey of bandwidth compression and image coding, including so called psychovisual coders. The emphasis will be on rate distortion theory and its application to transform coding. The basic assumptions which are necessary to find a solution to the set of parametric equations, which are the heart of rate distortion theory, are presented and discussed.

The results of Sections II thru IV are combined with some experimental results in Section V. It is shown that the mathematical models derived in Section II are consistent with the measured statistical characteristics of images. Furthermore, when a statistical analysis of the model is carried out, with a standard image representation as an input, the output of the complete HVS model is statistically compatible with the assumptions of rate distortion theory. This latter point cannot be overemphasized. Several assumptions are made to obtain solutions to the rate distortion theory equations which are seldom met for "raw" images. Images which are preprocessed by the HVS model satisfy all of these assumptions except that of stationarity.

Section V is followed by two sections which contain the results of several coding experiments. The achromatic (black and white) experiments are reported in Section VI and the color results are

contained in Section VII.

In Section VII a new image quality measure is presented. This measure evolved from the HVS model and is a "subjective" mean square error fidelity criterion. The applicability of this criterion to rate distortion theory and image evaluation will be discussed and some experimental results indicating the utility of this measure are presented.

Finally, Section IX contains a review of the major findings of this research and a discussion of possible applications. Several areas for continued research are pointed out.

SECTION II

HUMAN VISUAL SYSTEM MODELS

In the past two or three decades visual system modeling has come into vogue. There are several reasons for this, not the least of which is the recent availability of large amounts of physiological and psychophysical data. Technological advances, in both laboratory instrumentation and communication of the spoken and written word, are prime factors in this "information explosion" no doubt. Indeed, the vast amount of information presently available has taxed the imaginations of the "model builders" in some cases. However, the literature is replete with models of the HVS and just like the proverbial bus, wait a while, the one you want will come along.

Another reason for this age of modeling is the advent of computerized image processing and analysis. Prior to this time vision modeling was done primarily to explain and understand the interworkings of the system with little practical application. The biological models which are being conjectured today are quite often quantified and transformed into mathematical models which become integrated parts of complex software and/or hardware systems (in our case, bandwidth compression and coding systems). We will now formalize our biological model.

2.1. Biological Model

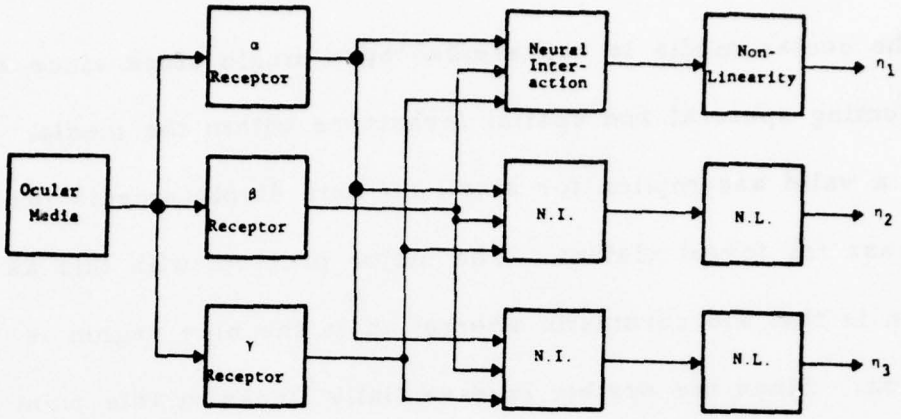
We begin by indicating all of the areas which our model will not cover including any assumptions which will be made in developing the model. The model is for processing single frame color imagery, therefore temporal aspects will not be considered. In addition, we will assume the images will be viewed with an illuminant of 5500°K at intensity levels which assure photopic (cones only) vision. The viewing distance to image size ratio will be such that we subtend a 2° field and hence, we are considering foveal vision only. Furthermore, no consideration will be given to stereoscopic, depth, or disparity effects. In short, our model will assume monocular, color, single-frame, photopic, foveal vision. In addition, we will assume the ocular media and the retinal mosaic to be spatially isotropic and homogeneous (which is a reasonable assumption for the fovea [16 , pp. 47-50]).* The biological model which follows

* Perhaps a comment on the isotropic assumption is in order. As pointed out in Section B.3, the sensitivity of the visual system to contrast gratings varies with angular orientation of the gratings. The response to vertical and horizontal gratings is the same, but sensitivity decreases for rotations less than 90 degrees. The minimum sensitivity occurs at 45 degrees rotation and at this point the response of the system to a 30 cycles/degree grating is 3dB below that at zero degrees rotation. The decrease in sensitivity is less for spatial frequencies below 30 cycles/degree. Thus, the describing function variation with rotation is minimal. One may question this conclusion since we obviously do not "see" as well upside down as we do upside right. The difference is that "seeing" involves cognition and the higher level mechanisms which are the precursors of cognition are not rotationally invariant. Since we are modeling only the preprocessor functions, the isotropic assumption is reasonable.

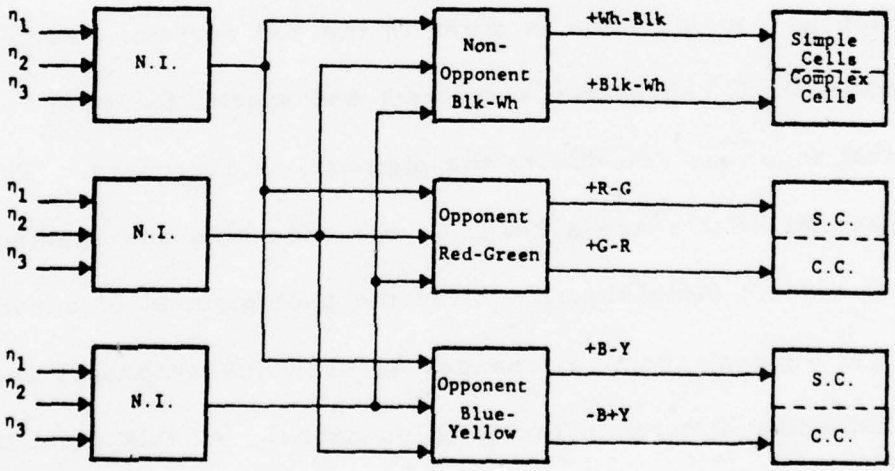
from these assumptions is shown in block diagram form in Figure 1.

The ocular media is represented by a single block since we are assuming spectral and spatial invariance within the media. This is a valid assumption for small off axis displacements (which is the case for foveal vision). The major problem with this assumption is that the chromatic aberration in the blue region is significant. Since the system is essentially linear at this point we have chosen to include the resultant loss of resolution in this spectral region with that due to blue cone spacing in the next stage.

The ocular media block is followed by three blocks representing the three types of cones. Since we are modeling photopic foveal vision no consideration is given to the rod system. Each photoreceptor block represents a spectral and spatial function. The spectral functions are due to the pigments of the cones. The low-pass spatial effects are a result of the cone size and spacing (the retinal mosaic dimensions). After the photopigment of a cone absorbs light several chemical changes occur which eventually lead to electrical spike activity in the ganglion axons. At this point the neuronal signals are a nonlinear function of the visual stimulus. The actual site at which the nonlinearity occurs in the human retina is not known; however, there is evidence that it is after the receptors and prior to the ganglion cells [17, p. 251]. Jameson has



(a)



(b)

Figure 1. Block Diagram of the Biological Model

argued that if the receptors are linear and linear summations occur before the nonlinearity, then the trichromatic and opponent color theories of vision are compatible [18, pp. 391-397]. Indeed, physiological recordings from the retinas of several species indicate the horizontal cells may be the site of spectral summation which produces the luminance signal and the chromaticity signals are generated in the outer plexiform layer [19, pp. 199-200]. In addition, recordings from the inner nuclear layer indicate a nonlinear transformation has occured. The biological model of the retina is completed by the neural interaction (NI) blocks which represent the rich interconnectivity within the retina.

The ganglion cell axons form the optic nerve which carries the output signals of the retina to the lateral geniculate bodies (LGB). The processing which occurs at this point is still a matter of debate (see Section A.3). Neurological recordings in primates have revealed a response organization at this level. The LGB blocks in Figure 1 represent this organization with four opponent cell and two non-opponent cell structures.

From the lateral geniculates the three pairs of outputs go directly to the visual cortex, in particular, area 17 of the striate cortex. This last block in the diagram represents the simple and complex cells which have been investigated primarily by Hubel and Wiesel (see Section A.4). The cells are located in area 17 and 18

of the cortex. The characteristics of these cells suggest higher cortical processes are involved and at this point the transition between the "preprocessor elements" and the functional processing which includes cognition and perception becomes prominent. Given the biological model of Figure 1 we will now develop a concise mathematical model.

2.2. Mathematical Model

The mathematical homologue of Figure 1 is shown in Figure 2. The ocular media is represented by an ideal low-pass filter which is invariant over the spectral range of the input signal, $f(r, \theta, \lambda)$. Furthermore, the system is assumed isotropic, therefore the line spread function (LSF) is rotationally invariant. The LSF for a 3mm pupil has been shown to be approximately $\exp(-.7r)$ [20]. This formulation also compares favorably with the data of Campbell and Gubisch (see Section B.1).

The spectral sensitivities of the three cones can be quantified by the curves shown in Figure B.8. Note that these curves include the effects of the ocular media which is consistent with the structure of our model. The spatial characteristics of the red and green channels at this point have been shown to be effectively that of the ocular media, hence, they require no further modification. The blue channel however, has been shown to have a contrast sensitivity

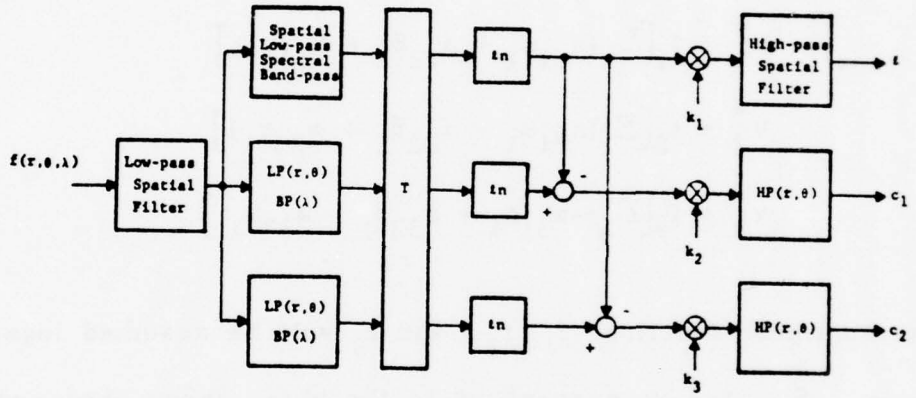


Figure 2. Block Diagram of the Mathematical Model

which peaks at only 2 cycles/degree (see Section B.7). The increased high frequency loss of the blue channel is due to the scarcity of blue cones and can be represented by an ideal low-pass filter with a cutoff frequency of 2 cycles/degree and a slope of -6dB/octave.

The first set of neural interconnections and the nonlinearity of Figure 1 are due to the linear spectral summations as proposed by Jameson [18, p. 392] and are of the form

$$\begin{aligned} V_1^* &= f_1\left[\Sigma_\lambda(a_{11}\alpha_\lambda + a_{12}\beta_\lambda + a_{13}\gamma_\lambda)\right] \\ V_2^* &= f_2\left[\Sigma_\lambda(a_{21}\alpha_\lambda - a_{22}\beta_\lambda + a_{23}\gamma_\lambda)\right] \\ V_3^* &= f_3\left[\Sigma_\lambda(-a_{31}\alpha_\lambda + a_{32}\beta_\lambda + a_{33}\gamma_\lambda)\right] \end{aligned} \tag{1}$$

The nonlinear functions f_1, f_2, and f_3 will be assumed logarithmic. The α_λ, β_λ, and γ_λ correspond to the blue, green, and red cone spectral sensitivities. The linear portion of equation (1) may be written in matrix form

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & -a_{22} & a_{23} \\ -a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \alpha_\lambda \\ \beta_\lambda \\ \gamma_\lambda \end{bmatrix} \tag{2}$$

This formulation is similar to the first stage of Frei's color model [21, p. 116] and it satisfies Grassmann's laws of color mixture

[22, p. 233]. In addition, V_1 is interpreted as luminance like, V_2 is redness for positive values and greenness for negative ones, and V_3 is yellowness-blueness. Since V_1 is luminance this formulation satisfies Abney's law of luminance addition [23, p. 370]. The weighting factors, a_{ij}, are dependent upon the set of functions chosen to represent the cone distributions. If the Konig distributions are used for the receptor sensitivities, then equation (2) becomes [18, p. 395]

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 0.0 & .15 & .85 \\ .37 & -2.23 & 1.66 \\ -.71 & .06 & .34 \end{bmatrix} \begin{bmatrix} \alpha_\lambda \\ \beta_\lambda \\ \gamma_\lambda \end{bmatrix} \tag{3}$$

Equation (3) is represented by T in the mathematical model of Figure 2.

The last block of the retina model shown in Figure 1 represents spectral and spatial characteristics. The spectral portion accounts for the opponent color traits of the system. Cornsweet has shown that a logarithmic difference operation can produce chromatic signals which are compatible with human hue perception [17, p. 248]. In particular, hue perception is relatively invariant to intensity changes. This operation is performed by the linear adders as shown in Figure 2. Multiplicative constants have been introduced at this point to adjust the color balance so that an incremental

change in the chrominance signals results in an equivalent hue shift.

The last block in the mathematical model is the high-pass filters which provide the low-frequency roll-off of the HVS contrast sensitivity curves. These filters have been shown to be of the form [1, p. 166]

$$H(\omega) = \frac{10^{-4} + \omega^2}{4 \times 10^{-3} + .8\omega^2} \tag{4}$$

The actual location of the differencing points and the filters has not been established. Indeed, as noted in Section B.7, the presence of the high-pass filters in the chrominance channels is still being debated. The configuration for the luminance channel is well established however, and the last filtering operation probably occurs at the retina level. The signal ℓ is fed to the LGB. In the case of c_1 and c_2 (if the filtering takes place), there is evidence that the filtering is under the control of more central mechanisms (cortical control). These filters may actually be located in the striate cortex. The inputs to chrominance filters may be derived in the LGB's, since there is some indication the differencing networks are located there [24]. In any case, the sequence as shown is probably correct.

The mathematical model as shown in Figure 2 appears to fall short of the model in Figure 1. Figure 2 shows only

three output variables, luminance and two chrominances. This is not a defect since the complements of these signals can be derived quite easily. Of more concern might be the nonexistence of the simple cell and complex cell behavior exhibited in the cortical area. These effects have not been included since they are, again, considered to be under higher order control. Therefore, they do not fit the preprocessor definition of our model. Indeed, there is much evidence indicating the responses at this level are modified by heredity, environment, cultural background, and conscious effort on the part of the viewer [25].

We would like to add however, that the eventual use of ℓ, c_1, and c_2 will be in the spatial frequency domain, i.e., we will work with the two-dimensional Fourier transforms of ℓ, c_1, and c_2. Some authors have argued that the cortical areas of the visual system are performing such a transformation (see Section A.4). In fact, the simple and complex cell behavior can be explained using such a theory. As a result, several "Fourier Models" of vision have appeared in the past 10 years. Unfortunately, matters are not so simple as to validate completely such a simplistic viewpoint. Although the Fourier Models explain many nonintuitive visual phenomena and are consistent with a wealth of psychovisual data, they are considered to be "an outlandish notion" by some authors [26, pp. 210-214]. For other reasons, which will become apparent later,

we will use the Fourier transform domain; and, because it appears to be the domain of the brain in many respects, we shall refer to the Fourier transforms of ℓ, c_1, and c_2 as the "perceptual space."

SECTION III

CHARACTERIZATION OF IMAGERY

In this section we will present a mathematical characterization of images which will be used throughout this dissertation. The basic ground work in image sampling, spatial and spectral decompositions and transformations, and statistical analysis will be developed. We will begin with the continuous image.

3.1. Continuous Representation

Let $\mathcal{I}(x, y, t, \lambda)$ be the intensity of an image source defined at spatial coordinates (x, y), at time t, and of wavelength λ. $\mathcal{I}(x, y, t, \lambda)$ is a real and positive function. For the "still-image" case, the intensity is time invariant and we may write $\mathcal{I}(x, y, \lambda)$. The spectral dependence of the image may be eliminated by integrating the product of $\mathcal{I}(x, y, \lambda)$ and a luminous efficiency function. Thus, for the achromatic case

$$\mathcal{I}(x, y) = \int_0^\infty \mathcal{I}(x, y, \lambda)\, V_t(\lambda)\, d\lambda \tag{5}$$

where $V_t(\lambda)$ is the achromatic spectral response of the human visual system.

The color representation of an image is usually accomplished by a set of tristimulus values. The luminous efficiency function in this case is defined over three overlapping spectral regions. The three image representations are defined by

$$\mathcal{R}(x,y) = \int_0^\infty \mathcal{I}(x,y,\lambda)\, V_{\mathcal{R}}(\lambda)\, d\lambda \tag{6}$$

$$\mathcal{G}(x,y) = \int_0^\infty \mathcal{I}(x,y,\lambda)\, V_{\mathcal{G}}(\lambda)\, d\lambda \tag{7}$$

$$\mathcal{B}(x,y) = \int_0^\infty \mathcal{I}(x,y,\lambda)\, V_{\mathcal{B}}(\lambda)\, d\lambda \tag{8}$$

In this particular tristimulus space, which is commonly referred to as the RGB-space, the peak responses of $V_{\mathcal{R}}(\lambda)$, $V_{\mathcal{G}}(\lambda)$, and $V_{\mathcal{B}}(\lambda)$ fall at 600nm, 530nm and 440nm, respectively (see Figure 3). Thus, the red luminousity function peaks between pure green (530nm) and pure red (650nm), in the yellow region. The green function peaks at mid-green and the blue function peaks in the violet region. The label RGB-space can be misleading if one is not cognizant of the true spectral characteristics of the defining curves.

There are several color coordinate systems currently in use in image processing [27, Chapter 3]. Each of these systems can be defined by a set of luminosity functions, as in the RGB case, or

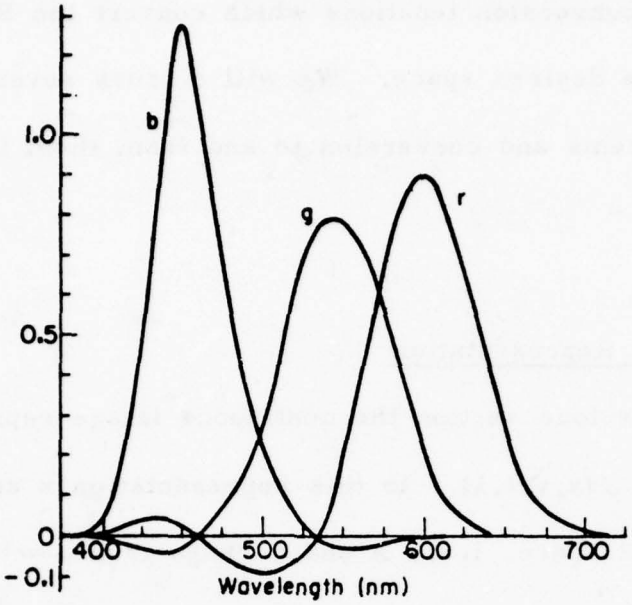


Figure 3. Color Matching Functions

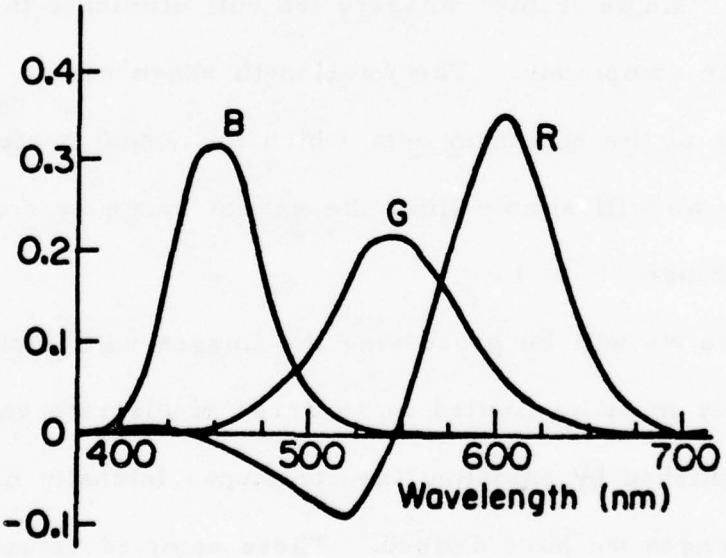


Figure 4. Color Matching Functions for the 2^{o} Standard Observer

by coordinate conversion functions which convert the RGB luminosity functions to the desired space. We will discuss several of these coordinate systems and conversion to and from them in more detail later.

3.2. Discrete Representation

In the previous section the continuous image representation was defined as $\mathcal{I}(x, y, t, \lambda)$. In this representation x and y are defined over all space, i.e., x and y range from $-\infty$ to $+\infty$. In addition, time and wavelength also have this infinite range. The first step to be taken in discretizing our representation is to limit these bounds. Since the primary concern in this dissertation is "still" or "single frame" imagery we will eliminate the time dependence completely. The wavelength range can be reduced to that range of the spectrum over which the visual system responds. For now, we will simply limit the spatial range by confining x and y to the range -L to L.

Since we will be processing the images with a digital computer, they must be limited to an array of discrete values. This is accomplished by sampling the continuous intensity over the limited ranges we have defined. These sampled values are then quantized with a number of levels compatible with the accuracy desired and digital word size available. For the imagery used in

the experimental work of this research 256 level quantization was used. In addition, all images were sampled over a 512 x 512 linear grid. Where 256 x 256 size images are specified in this dissertation, said images were obtained by averaging a 512 x 512 picture with a 2 x 2 picture element (pixel) square. We will represent discretized imagery by a two-dimensional matrix denoted by a bracketed letter, hence

$$[f] = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N} \\ f_{2,1} & & & \\ \vdots & & & \\ f_{N,1} & f_{N,2} & \cdots & f_{N,N} \end{bmatrix} \tag{9}$$

is an N x N discretized image.

3.3. Spatial Decomposition of Discrete Images

Assume the discrete representation of an image as defined by equation (9). We may write a separable linear transformation on the image as

$$[F] = [u]^t [f][v] \tag{10}$$

where [F] is called the unitary transform domain of the image, [u] and [v] are unitary operators, and the superscript t denotes matrix transposition [28, p. 30]. If [u] and [v] are unitary, then by definition

$$[u]^{-1} = [u^*]^t$$

and

$$[v]^{-1} = [v^*]^t \tag{11}$$

where * denotes complex conjugation. For the case of a real unitary matrix [u],

$$[u]^{-1} = [u]^t \tag{12}$$

and [u] is called an orthogonal matrix. The inverse of equation (10) becomes

$$[f] = [u][F][v]^t \tag{13}$$

Equation (10) is commonly referred to as an orthogonal decomposition of [f]. Since the decomposition is over the two-dimensional spatial representation of the image in this case, it may also be referred to as a spatial transformation. Such transformations are useful for image representation to the extent that they "average" the energy or information contained in the original representation into a more "compact" space. Hence, certain elements of the transformed space may be set to zero with a minimal loss of information. This attribute of orthogonal transformations is useful in bandwidth compression and coding applications. There are an infinite number of possible orthogonal

systems; however, only a few have been formally defined and used in image processing. The most commonly used transforms are: Fourier, Haar, Hadamard/Walsh, Slant, Cosine, Sine, and Karhunen-Loeve [27, Chapter 10], [28, pp. 33-38], and [29, Chapter 6].

The optimum statistical transform for minimizing the mean square error criterion between the original and a reconstructed image (formed with a reduced number of transform coefficients) is the Karhunen-Loeve transform (KLT) [29, p. 123]. This transform is composed of eigenvectors of the correlation matrix of the original image, or class of images. There are two problems associated with this transform. The first problem is the large number of computations which must be performed to: (1) determine a correlation matrix, (2) diagonalize it to obtain eigenvectors, and (3) perform the actual transformation. The second problem is that mean square error is not necessarily a valid criterion for imagery.

The discrete Fourier transform (DFT) is defined as

$$[F] = [A][f][A] \tag{14}$$

where

$$[A] = \frac{1}{\sqrt{N}} \begin{bmatrix} W^0 & W^0 & & \cdots & W^0 \\ W^0 & W^1 & W^2 & \cdots & W^{N-1} \\ W^0 & W^2 & W^4 & \cdots & W^{2(N-1)} \\ W^0 & W^3 & & & \\ \vdots & \vdots & & & \\ W^0 & W^{N-1} & & \cdots & W^{(N-1)^2} \end{bmatrix} \tag{15}$$

For this representation

$$W \equiv \exp\left[-\frac{2\pi i}{N}\right] \tag{16}$$

where N is the number of samples in each direction and $i = \sqrt{-1}$. Note that [A] is symmetric and therefore equation (14) follows directly from equation (10). Several features make the DFT appealing. Firstly, the transform can be implemented with a fast computational algorithm, the fast Fourier transform (FFT). The FFT requires a number of computations proportional to $2N^2 \log_2 N$ rather than $2N^3$ as for the Karhunen-Loeve transform (assuming an N x N image) [30, p. 49]. A second favorable trait of the DFT is that under the proper statistical assumptions, as N grows, the DFT approaches the optimum decomposition [ibid]. Another somewhat mundane reason for representing images in this form is the compatibility with linear systems analysis and the direct analogy

between the time-frequency and space-spatial frequency domains. A problem which is often cited for not using the DFT is that the kernel -- defined by equation (16) -- is complex.

The discrete cosine transform (DCT) obviates the complex problem. This transform is defined on the reals only and is given by

$$G_x(0) = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} x(n)$$

$$G_x(k) = \frac{2}{N} \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)k\pi}{2N}\right], \quad k = 1, 2, \ldots, (N-1) \qquad (17)$$

where $G_x(k)$ is the kth DCT coefficient [31]. Ahmed has shown the DCT is closer to the optimum (KLT) than the FFT for the statistical assumptions of a first order Markov system with correlation equal to .9 [ibid]. Jain has shown this to be true for correlations greater than .5; however, for correlations less than .5 other sinusoidal transforms perform better than the DCT [32].

3.4. Spectral Decomposition of Discrete Images

In Section 3.1 we briefly touched on the spectral decomposition of continuous images. As pointed out then, several color-coordinate systems may be defined. These "rotations" of the color axes can

produce various energy packing and/or decorrelation properties which make one system more appropriate than another for a specific task. For discrete images, conversion between linear systems involves a single matrix multiplication

$$\begin{bmatrix} CO_1 \\ CO_2 \\ CO_3 \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \begin{bmatrix} CI_1 \\ CI_2 \\ CI_3 \end{bmatrix} \tag{18}$$

where CI_i and CO_i are color input and color output tristimulus values. Many conversion matrices are defined in terms of the RGB functions shown in Figure 4. One such conversion which has found wide applicability is the National Television Systems Committee (NTSC) receiver primary color coordinate system. The three coordinates of this system are referred to as Y, I, and Q; hence, the system is sometimes called the YIQ system. The conversion is defined by

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} .299 & .587 & .114 \\ .596 & -.273 & -.322 \\ .212 & -.522 & .315 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{19}$$

The Y signal represents luminance and the I and Q are chrominance signals which are linear functions of R-Y and B-Y respectively.

As can be seen from Figure 4, the Commission Internationale

de l'Eclairage (CIE) standard observer curves have some negative areas. Thus, some tristimulus values are negative which is a nonrealizable situation. To eliminate this problem, the XYZ primary system was developed by the CIE. The color matching functions of this system are shown in Figure 5. This set of curves can be produced from those of Figure 4 by the conversion

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} .607 & .174 & .201 \\ .299 & .587 & .114 \\ 0.0 & .066 & 1.117 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{20}$$

Note that Y in this system is equivalent to Y in the NTSC YIQ system.

In order to evaluate the effectiveness of a color coordinate system, one may devise a color-order system which specifies all object colors within the limited domain under consideration. There are three general categories these systems may be grouped under: additive color, subtractive color, and perceptual color. For obvious reasons, we are concerned with the latter. One system of this group has gained wide popularity among researchers, the Munsell Color System [22, p. 476]. The Munsell Book of Color contains color patches arranged in equal visual spacings of hue, luminance, and saturation. This arrangement yields color solids with loci of constant hue and saturation on a surface of constant

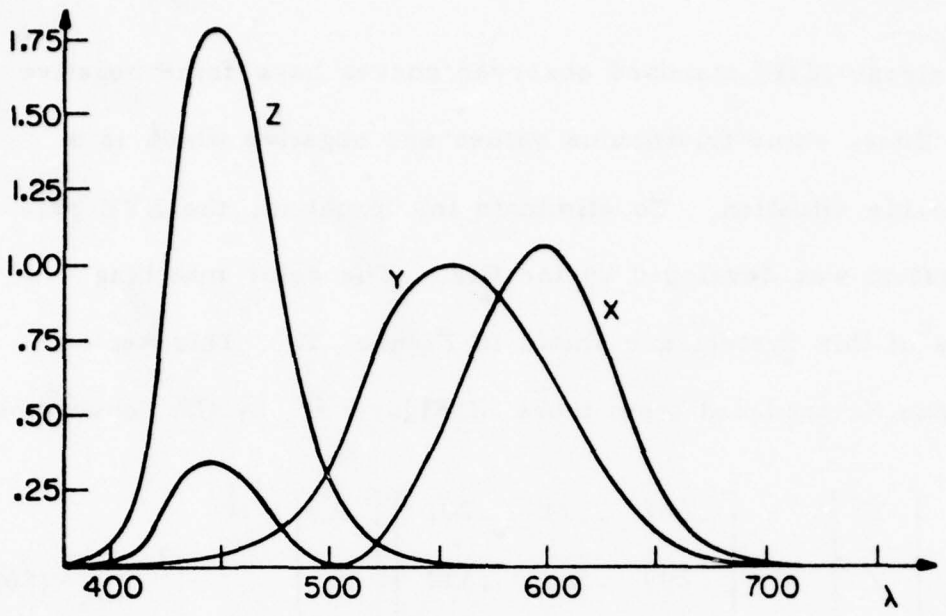


Figure 5. XYZ Color Matching Functions

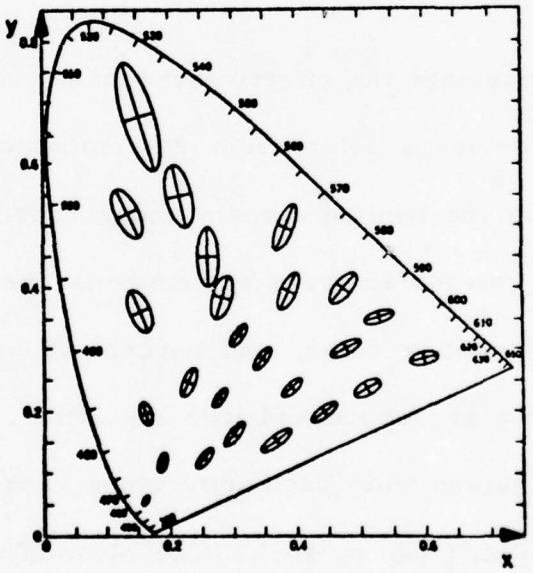


Figure 6. CIE Standard Observer Chromaticity Diagram with MacAdam Ellipses

luminance. These loci form a polar coordinate system.

Given that a Euclidean property of color perception is approximately valid, a chromaticity scale based on this polar coordinate system can be converted to a uniform scale. Analytic expressions that transform the CIE standard observer tristimulus values to three new variables which define a "distorted" space can then be defined. In this space the chromatic difference between any two samples in an equi-luminance plane corresponds to the same distance separation of their representation points. Thus, the vector distance between two colors corresponds to their perceived difference. With the Munsell system as a basis, a number of attempts at acceptable -- but simple -- analytic transformations have been made [22, p. 454]. The more recently developed cube-root coordinate system [33] has received much attention because of its simplicity and good approximation to the spacing provided by the Munsell system.

The RGB system can be represented in a chromaticity diagram as shown in Figure 6. The outer horseshoe shaped curve, the chromaticity curve, is the locus of wavelength points for the gamut of saturated hues in the system. Overlayed on this curve is a set of MacAdam ellipses which represent the regions within which chrominance can be varied without perceptible color shifts. The actual size of these ellipses has been exaggerated.

The important point is their size varies over the range of spectra shown. The blue region, for example, is much more sensitive to shifts than the green region. An ideal perceptual color-coordinate system would map these areas into circles of equal radii.

The cube-root or Lab color coordinate system was developed with this idea in mind [33]. In addition, the system is based on simple conversion formulas. In terms of RGB, the system is defined as

$$\begin{aligned} L &= 25.29G^{1/3} - 18.38 \\ a &= 106.0(R^{1/3} - G^{1/3}) \\ b &= 42.34(G^{1/3} - B^{1/3}) \end{aligned} \tag{21}$$

where $R = 1.02X$, $G = Y$, and $B = .847Z$ [33]. The set of equations can be rewritten as

$$\begin{aligned} L &= 25\left(100\,\frac{Y}{Y_0}\right)^{1/3} - 16 \\ a &= 107.72\left[\left(100\,\frac{X}{X_0}\right)^{1/3} - \left(100\,\frac{Y}{Y_0}\right)^{1/3}\right] \\ b &= 43.08\left[\left(100\,\frac{Y}{Y_0}\right)^{1/3} - \left(100\,\frac{Z}{Z_0}\right)^{1/3}\right] \end{aligned} \tag{22}$$

where X_0, Y_0, and Z_0 are the tristimulus values for the reference white. Several factors should be noticed. First of all, this

system provides a set of coordinates in close agreement with the Munsell system. The three coordinates L, a, and b correspond to lightness, redness-greenness, and yellowness-blueness (just as our color model requires). In addition, the formulation contains a nonlinearity, and in particular, one which has been proposed as the "correct" nonlinearity for the HVS [34, p. 15]. Thus, the Lab space has strong physiological and psychophysical support.

Another color system which is based on the visual system is the retinal cone color system [27]. This system is based on functions for normal, deuteranopic and protanopic vision which were developed by Judd [35]. The conversion is defined as

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ -.460 & 1.359 & .101 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$= \begin{bmatrix} .299 & .587 & .114 \\ .127 & .724 & .175 \\ 0.0 & .066 & 1.117 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad (23)$$

Note that T_1 in this system is equivalent to Y in the XYZ and YIQ systems and is luminance. T_2 and T_3 can be seen to be chrominance signals which are greenish and blueish respectively. Frei has used this coordinate system in the development of a HVS color model [36].

The Frei color system can basically be represented by the set of equations

$$
\begin{aligned}
g_1 &= 21.5\ \ell n\ Y \\
g_2 &= 41\ \ell n\left[\frac{-.46X + 1.359Y + .101Z}{Y}\right] \\
g_3 &= 6.27\ \ell n\left[\frac{Z}{Y}\right]
\end{aligned}
\tag{24}
$$

The similarity between this system and the Lab system is readily apparent. Furthermore, this set of equations is consistent with the color model developed in Section II. Frei's complete model also contains spatial filters in the last stage giving

$$
\begin{aligned}
g_1^* &= g_1 \circledast h_1(x,y) \\
g_2^* &= g_2 \circledast h_2(x,y) \\
g_3^* &= g_3 \circledast h_3(x,y)
\end{aligned}
\tag{25}
$$

where $\circledast$ denotes convolution. The filter functions used by Frei were of the bandpass type [3].

One last system which is also based on a model of the HVS will be discussed, that due to Faugeras [37]. Faugeras developed a matrix based on the uniform color scale conversion of Stiles [37, p. 103]. This matrix defined a cone absorption stage in his color model and it is given by

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} .363 & .610 & .026 \\ .125 & .814 & .062 \\ .001 & .060 & .939 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{26}$$

The three signals A, c_1, and c_2 which correspond to luminance, red-green, and yellow-blue, respectively are then given by

$$\begin{aligned} A &= 13.83 \ln F_1 + 8.34 \ln F_2 + .429 \ln F_3 \\ c_1 &= 64 \ln\left[\frac{F_1}{F_2}\right] \\ c_2 &= 10 \ln\left[\frac{F_1}{F_3}\right] \end{aligned} \tag{27}$$

This set of equations can be seen to be similar to those of the Frei model, equation (24). The most important difference is between A and g_1. Recall that g_1 derives its luminosity character from the linear equation which defines Y. In the case of A, the constants multiplying each logarithmic function provide the correct mixture for an approximation to luminance.

3.5. Some Statistical Characteristics of Discrete Images

The mean value of a discrete image is a matrix of the form

$$[\mu] = E\{[F]\} = [E\{F(x,y)\}] \tag{28}$$

where $E\{\cdot\}$ denotes the expected value operator. The correlation

can be defined as

$$R(x_1, y_1 ; x_2, y_2) = E\{F(x_1, y_1)\, F^*(x_2, y_2)\} \qquad (29)$$

and similarly the convariance becomes

$$C(x_1, y_1 ; x_2, y_2) = E\{[F(x_1, y_1) - E\{F(x_1, y_1)\}]$$
$$[F^*(x_2, y_2) - E\{F^*(x_2, y_2)\}]\} \qquad (30)$$

If the image array, F, comes from a wide sense stationary process, the correlation function is a function of $k = x_1 - x_2$ and $\ell = y_1 - y_2$, thus

$$R(x_1, y_1 ; x_2, y_2) = R(x_1 - x_2, y_1 - y_2) = R(k, \ell) \qquad (31)$$

and similarly for the covariance,

$$C(x_1, y_1 ; x_2, y_2) = C(x_1 - x_2, y_1 - y_2) = C(k, \ell) \qquad (32)$$

The two matrices will be of block Toepliz form under these conditions [27]. When the correlation between the elements of the array is separable in the x and y direction then the correlation matrix can be expressed as a direct product of row and column matrices. If we consider the special case of a Markov process with the adjacent pixel correlation equal to ρ we get the covariance matrix

$$[C_R] = \sigma_R^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & & & & \\ \rho^2 & & & & \\ \vdots & & & & \\ \rho^{N-1} & & & & 1 \end{bmatrix} \qquad (33)$$

where the subscript R denotes row statistics and σ_R^2 is the variance of pixels along a row. Again, for the x and y separable case, the covariance can be expressed as a direct product of the row and column matrices, C_R and C_c.

The Markov process assumption is valid for many types of images. The computed correlation of adjacent pixels in the Kodak GIRL picture are plotted in Figure 7. The slope of this curve is the ρ for a first order Markov process and for this data $\rho = .96$. Habibi and Wintz have reported ρ's in the range .78 to .92 for four data sets [38]. Note that the data points are very close to the straight line and since the ordinate in Figure 7 is logrithmic this indicates the data is Markov.

Once the stationary covariance function has been determined the discrete power spectral density may be computed. The power spectral density in this case is the two-dimensional DFT of the covariance function, thus

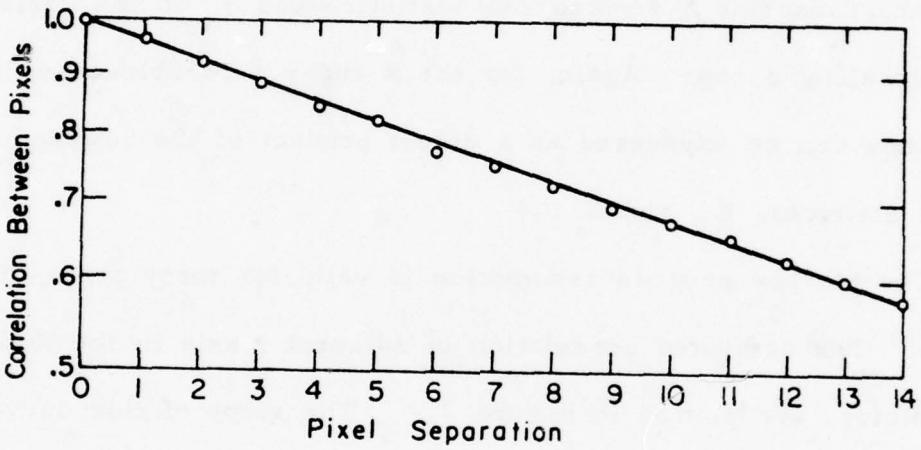


Figure 7. Correlation Between Pixels of GIRL

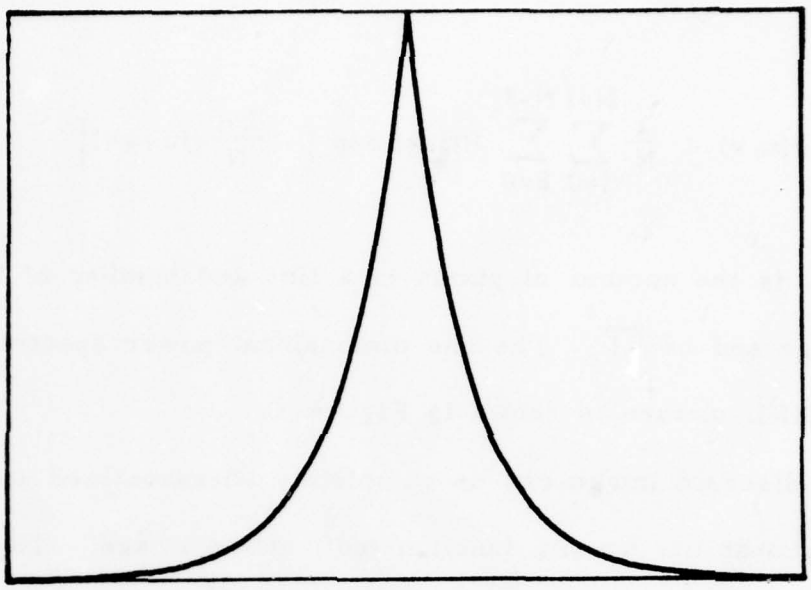
$$S(u,v) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} R(j,k) \exp\left\{ - \frac{2\pi i}{N} (ju + kv) \right\} \qquad (34)$$

where N is the number of pixels in a line and number of lines in the image and $i = \sqrt{-1}$. The one-dimensional power spectral density of the GIRL picture is shown in Figure 8.

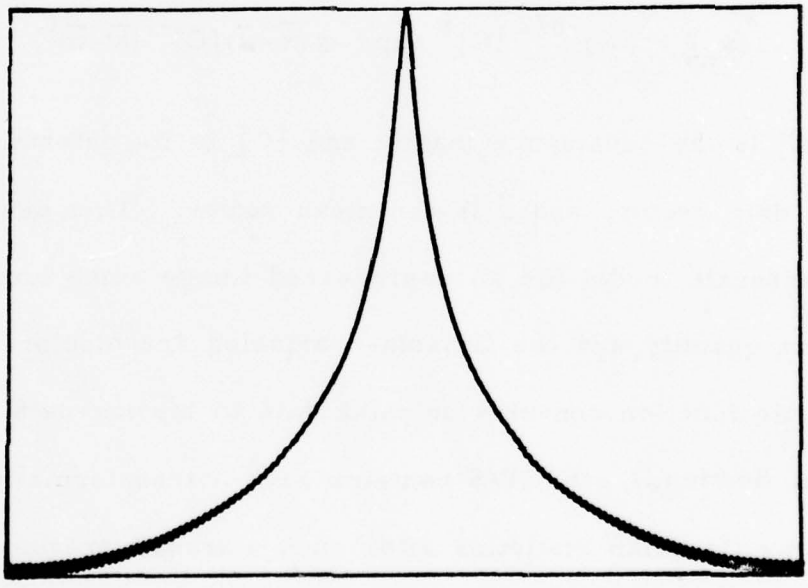
A discrete image can be completely characterized statistically by the probability density function (pdf) of the image. The most common pdf is the joint Gaussian which can be defined by [39, p. 255]

$$p(x_1, \ldots, x_n) = (2\pi)^{-n/2} |C|^{\frac{1}{2}} \exp\{ -\tfrac{1}{2} (\bar{x} - \bar{\mu})[C]^{-1} (\bar{x} - \bar{\mu})^t \} \qquad (35)$$

where $[C]$ is the covariance matrix and $|C|$ is its determinant, $\bar{x}$ is the data vector, and $\bar{\mu}$ is the mean vector. This density is not an adequate model for an unprocessed image since luminance is a positive quantity and the Gaussian variables are bipolar. The logarithmic function converts unipolar data to bipolar data and, as shown in Section II, the HVS contains such a transformation. If we assume Gaussian statistics after such a transformation what would be the pdf of the input? This question can be answered quite simply by considering Figure 9 and applying a fundamental theorem discussed in Section 5.2 of Papoulis [39, pp. 126-127].



(a)



(b)

Figure 8. Autocorrelation and Power Spectrum of GIRL



Figure 9. Exponential Nonlinear System

Referring to Figure 9, let x be a Gaussian process; therefore,

$$f_x(\cdot) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{36}$$

Now

$$y = g(x) = e^x \tag{37}$$

or

$$x = \ell n\ y \tag{38}$$

and

$$g'(x) = \frac{dy}{dx} = e^x \tag{39}$$

From Papoulis

$$f_y(\cdot) = \frac{f_x(x_1)}{|g'(x_1)|} \tag{40}$$

which becomes (after appropriate substitutions)

$$f_y(\cdot) = \frac{1}{\sigma y_1\sqrt{2\pi}} \exp\{-(\ell n\ y_1 - \mu)^2/2\sigma^2\} \tag{41}$$

where $y_1 \geq 0$. This pdf is known as the lognormal distribution and it has several interesting properties [40]. Plots of this function for several values of μ and σ^2 are shown in Figure 10.

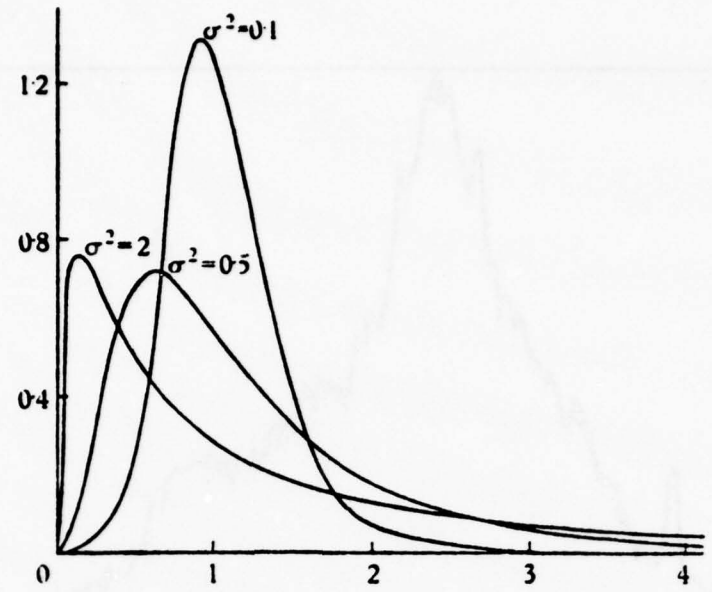
A plot of the first order histogram values obtained from the GIRL picture is shown in Figure 11. The similarity between Figures 10 and 11 are readily apparent. If we plot the histogram data on log-probability paper the curve of Figure 12 is obtained. Straight lines on this type of plot indicate lognormal data and the parameters μ and σ can be estimated from the curve by

$$\mu = \ln \xi_{50\%} \tag{42}$$

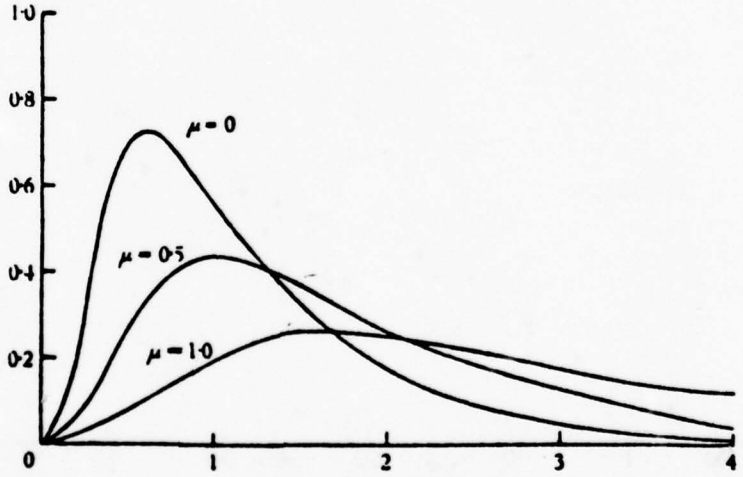
and

$$\sigma = \ln\left\{\frac{1}{2}\left(\frac{\xi_{50\%}}{\xi_{16\%}} + \frac{\xi_{84\%}}{\xi_{50\%}}\right)\right\} \tag{43}$$

where $\xi_{x\%}$ indicates the value at x% [40, p. 32]. The data points are essentially a straight line over the 1% to 99% range which indicates the image is strongly lognormal. If the GIRL image is processed by the logarithmic point nonlinearity and the histogram computed, the curves of Figures 13 and 14 can be obtained. Figure 13 has the characteristic "bell" shape of the Gaussian pdf. Since the abscissa represents equivalent normal deviates and the ordinate is linear, straight lines on Figure 14 indicate Gaussian like behavior. The slopes of the lines are equal to the



(a)



(b)

Figure 10. Frequency Curves of the Lognormal Distribution

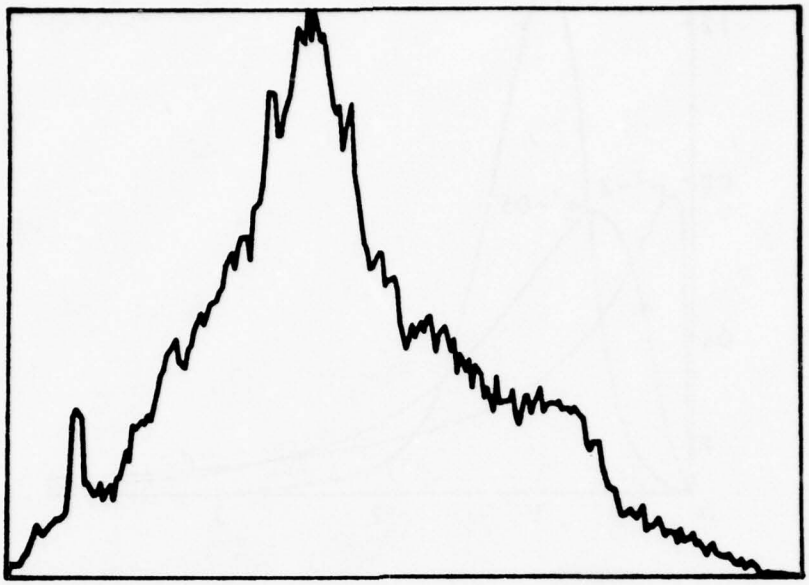


Figure 11. First-order Histogram of GIRL

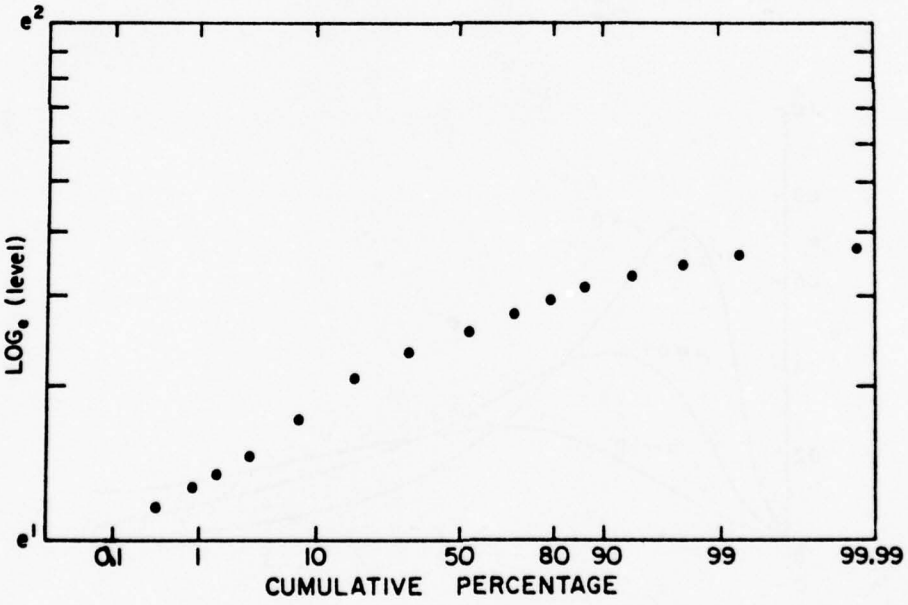


Figure 12. First-order Histogram Data of GIRL Plotted on Log-probability Coordinate System

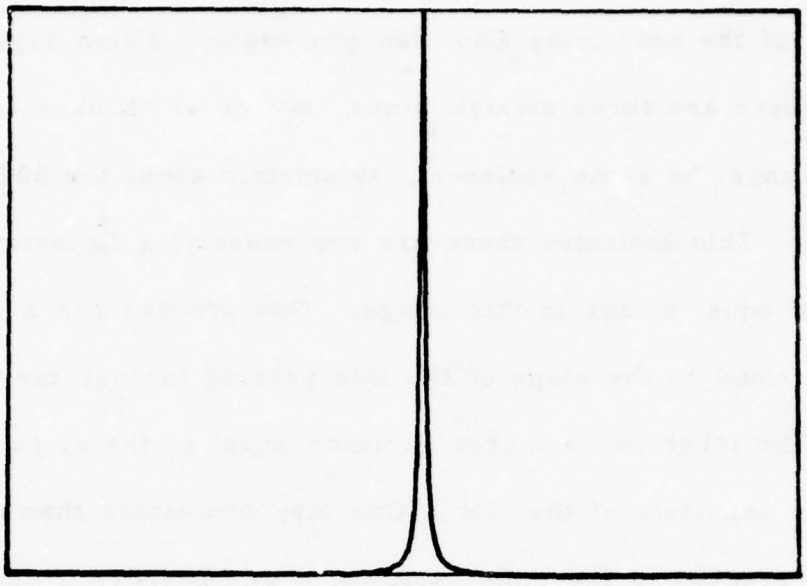


Figure 13. First-order Histogram of the Logarithm of GIRL

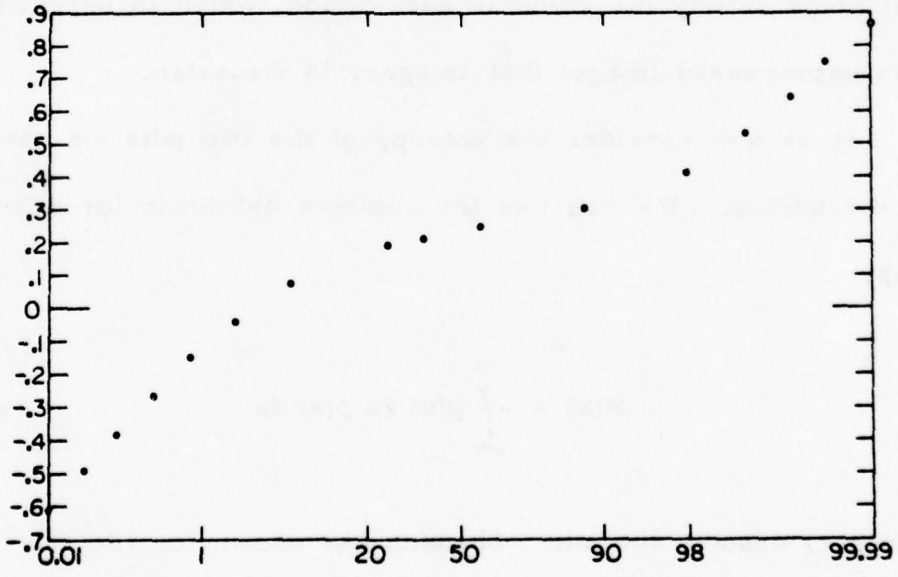


Figure 14. First-order Histogram Data of ℓn [GIRL] Plotted on Linear-probability Coordinate System

variance of the underlying Gaussian processes. From Figure 14 we see there are three straight lines, two of which have equal slopes (hence the same variance), symmetric about the 50% point or mean. This indicates there are two underlying Gaussian processes of equal means in this image. One process has a low variance equal to the slope of the line passing through the 50% point. The other has a higher variance equal to the slope of the two outer segments of the plot. One may conjecture that the low variance process is from the basic form or "gestalt" of the image; whereas, the high variance process is a result of the edge information and/or noise. From this discussion we see that the HVS model helps satisfy the common assumption (which is unrealistic for an unprocessed image) that imagery is Gaussian.

Let us now consider the entropy of the two pdfs we have been discussing. We will use the common definition for differential entropy

$$H(x) = -\int_{-\infty}^{\infty} p(x) \ln p(x)\, dx \qquad (44)$$

where $p(\cdot)$ denotes the pdf. Shannon has shown for the Gaussian case we get [41]

$$H(x) = \sigma\sqrt{2\pi e} \qquad (45)$$

where e is the base of natural logarithms. Consider the lognormal distribution,

$$p(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\{-(\ell n\, x - \mu)^2/2\sigma^2\} \tag{46}$$

The logarithm of this distribution is

$$\ell n\, p(x) = -\ell n\,(\sigma x\sqrt{2\pi}) - \frac{(\ell n\, x - \mu)^2}{2\sigma^2} \tag{47}$$

and the entropy becomes

$$H(x) = \int_0^\infty p(x)\left[\ell n(\sigma x\sqrt{2\pi}) + \frac{(\ell n\, x - \mu)^2}{2\sigma^2}\right]dx \tag{48}$$

where the lower limit of integration has been changed to 0 since x ranges from 0 to ∞ for the lognormal pdf. Equation (48) can be rewritten as

$$H(x) = \int_0^\infty p(x)\ \ell n\,(\sigma\sqrt{2\pi})\,dx + \int_0^\infty p(x)\ \ell n\,(x)\,dx + \int_0^\infty p(x)\,\frac{(\ell n\, x - \mu)^2}{2\sigma^2}\,dx \tag{49}$$

But for any valid pdf

$$\int_{-\infty}^{\infty} p(x)\,dx = 1 \tag{50}$$

Therefore,

$$H(x) = \ell n\,(\sigma\sqrt{2\pi}) + \int_0^\infty p(x)\,\ell n\,(x)\,dx$$

$$+ \frac{1}{2\sigma^2}\int_0^\infty p(x)\,(\ell n\,x - \mu)^2\,dx \qquad (51)$$

Now let $y = \ell n\,x$, which implies $x = e^y$ and hence, $dx = e^y dy$. Also, when $x = 0$, $y = -\infty$ and when $x = \infty$, $y = \infty$. Substituting into equation (46) gives

$$p(\cdot) = \frac{1}{\sigma\sqrt{2\pi}\,e^y}\,\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \qquad (52)$$

Making this substitution in equation (51) gives

$$H(\cdot) = \ell n\,(\sigma\sqrt{2\pi}) + \int_{-\infty}^\infty \frac{y}{\sigma\sqrt{2\pi}}\,\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}dy$$

$$+ \frac{1}{2\sigma^2}\int_{-\infty}^\infty \frac{(y-\mu)^2}{\sigma\sqrt{2\pi}}\,\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}dy \qquad (53)$$

The first integral is the mean and the second the variance of a Gaussian pdf; hence,

$$H(\cdot) = \ell n\,(\sigma\sqrt{2\pi}) + \mu + \frac{1}{2\sigma^2}\,\sigma^2$$

$$= \ell n\,(\sigma\sqrt{2\pi e}) + \mu \qquad (54)$$

Thus, for a nonzero μ we have an entropy change after passing

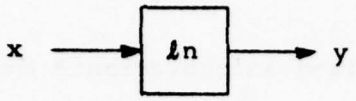


Figure 15. Logarithmic Nonlinear System

through the nonlinearity which is equal to μ, the mean of the output Gaussian pdf.

Next we will consider the autocorrelation and power spectrum for the system of Figure 15. The autocorrelation is (assuming y is a Gaussian process)

$$R_x(\tau) = E\{x(t)\, x(t+\tau)\} = \int_{-\infty}^{\infty} \cdots \int \frac{e^{y_1} e^{y_2} e^{-\frac{1}{2}(\bar{y}-\bar{\mu})^t [C]^{-1} (\bar{y}-\bar{\mu})}}{(2\pi)^{N/2} |C|^{\frac{1}{2}}} \, d\bar{y} \tag{55}$$

where N is the dimension of the system and is equal to two for the following discussion. Also,

$$\bar{y}^t = (y_1 \; y_2) \tag{56}$$

and

$$[C] = \begin{bmatrix} \sigma^2 & \sigma^2 \rho_y(\tau) \\ \sigma^2 \rho_y(\tau) & \sigma^2 \end{bmatrix} \tag{57}$$

where $\rho_y(\tau)$ is the normalized autocovariance defined as

$$\rho_y(\tau) = E\{[y(t) - \mu][y(t+\tau) - \mu]\} \qquad (58)$$

Letting $\lambda_1 = \lambda_2 = -i = -\sqrt{-1}$, equation (55) can be rewritten as

$$R_x(\tau) = E\{e^{i\lambda_1 y_1} e^{i\lambda_2 y_2}\}$$

$$= \int_{-\infty}^{\infty}\int \frac{e^{i\lambda_1 y_1 + i\lambda_2 y_2}}{\sqrt{2\pi}\,|C|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{y}-\bar{\mu})^t [C]^{-1} (\bar{y}-\bar{\mu})} dy_1\, dy_2 \qquad (59)$$

The above equation is in the form of a characteristic equation. The characteristic function for a two-dimensional Gaussian of nonzero mean is [39, p. 255]

$$\Phi(\lambda) = e^{-\frac{1}{2}\bar{\lambda}^t [C] \bar{\lambda}}\, e^{j\bar{\lambda}^t \bar{\mu}} \qquad (60)$$

where $\bar{\lambda}^t = (-\sqrt{-1}, -\sqrt{-1})$. Therefore, equation (59) reduces to

$$R_x(\tau) = \Phi\begin{pmatrix} -i \\ -i \end{pmatrix} = e^{\sigma^2[1 + \rho_y(\tau)] + \mu_y(\tau) + \mu_y(t+\tau)} \qquad (61)$$

where we have added the subscript y to the means for clarity. Equation (61) gives the autocorrelation of x in terms of the statistics of y.

In general, the autocorrelation can be expressed in terms of the covariance of a process as

$$R_x(\tau) = \mu_x^2 + C_x(\tau) = \mu_x^2 + \sigma_x^2 \rho_x(\tau) \tag{62}$$

From equation (61) we have

$$R_x(\tau) = \mu_x^2 + \sigma_x^2 \rho_x(\tau) = e^{2\mu_y + \sigma_y^2 [1 + \rho_y(\tau)]} \tag{63}$$

Now

$$E\{x\} = \mu_x = E\{e^y\} = e^{\mu_y + \frac{1}{2}\sigma_y^2} \tag{64}$$

(the later equality follows since y is Gaussian) which implies

$$\mu_x^2 = e^{2\mu_y + \sigma_y^2} \tag{65}$$

Substituting this form into equation (63) gives

$$\mu_x^2 + \sigma_x^2 \rho_x(\tau) = \mu_x^2 e^{\sigma_y^2 \rho_y(\tau)} \tag{66}$$

or

$$1 + \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau) = e^{\sigma_y^2 \rho_y(\tau)} \tag{67}$$

Expanding the right side of this equation gives

$$1 + \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau) = 1 + \sigma_y^2 \rho_y(\tau) + \sum_{k=2}^{\infty} \sigma_y^{2k} \frac{\rho_y^k(\tau)}{k!} \tag{68}$$

The sum in equation (68) represents the error if we use only

the first two terms of the expansion. The normalized covariance of any process will have an upper bound of 1. The value for σ_y^2 is typically .5. Thus, the worst case expansion is on $e^{.5}$ and the error introduced by using the first two terms is less than 10%. This is a very conservative error bound, particularly since it assumes the data are completely correlated. Neglecting the error term gives

$$1 + \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau) = 1 + \sigma_y^2 \rho_y(\tau) \qquad (69)$$

From equation (67) we could have approximated the logarithm

$$\ln\left[1 + \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau)\right] = \sigma_y^2 \rho_y(\tau) \qquad (70)$$

Typical image data will give a $\frac{\sigma_x^2}{\mu_x^2}$ ratio of .16. For the worst case $\rho_x(\tau) = 1$ we get $\ln(1 + .16) = .14842$ which is within 8% of .16. Thus, within experimental error, we get the previous result

$$\ln\left[1 + \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau)\right] = \frac{\sigma_x^2}{\mu_x^2} \rho_x(\tau) = \sigma_y^2 \rho_y(\tau) \qquad (71)$$

Now $\rho_x(0) = \rho_y(0) = 1$ for any valid covariance function; therefore,

$$\frac{\sigma_x^2}{\mu_x^2} \approx \sigma_y^2 \qquad (72)$$

Substituting into equation (71) gives

$$\sigma_y^2 \rho_x(\tau) = \sigma_y^2 \rho_y(\tau) \tag{73}$$

which implies that $\rho_x(\tau) = \rho_y(\tau)$. Thus, the output autocorrelation becomes

$$R_y(\tau) = \mu_y^2 + \sigma_y^2 \rho_x(\tau) \tag{74}$$

By definition, the power spectrum of the y process is

$$S_y(\omega) = \int_{-\infty}^{\infty} R_y(\tau)\, e^{-j\omega\tau}\, d\tau \tag{75}$$

therefore

$$S_y(\omega) = \int_{-\infty}^{\infty} [\mu_y^2 + \sigma_y^2 \rho_x(\tau)]\, e^{-j\omega\tau}\, d\tau$$

$$= 2\pi \mu_y^2\, \delta(\omega) + \sigma_y^2 \int_{-\infty}^{\infty} \rho_x(\tau)\, e^{-j\omega\tau}\, d\tau \tag{76}$$

This relationship is of great importance in rate distortion applications. Given an input autocorrelation, we can compute the output power spectrum which can be used in the equations [42, p. 117]

$$D_\theta = \frac{1}{2\pi} \int_{-\infty}^{\infty} \min [\theta, S(\omega)]\, d\omega \tag{77}$$

and

$$R(D_\theta) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \max \left[0, \log\left(\frac{S(\omega)}{\theta}\right)\right] d\omega \tag{78}$$

Thus, the rate distortion curve of a process which has been passed through a logarithmic nonlinearity can be specified. A detailed discussion of rate distortion theory and the implications of the result just obtained is contained in the following section. To complete the present analysis let us return to the Markov assumption,

$$\rho_x(\tau) = e^{-\alpha|\tau|} \tag{79}$$

Substituting this form into equation (76) gives

$$S_y(\omega) = \sigma_y^2 \int_{-\infty}^{\infty} e^{-\alpha|\tau|} e^{-j\omega\tau} d\tau + 2\pi\mu_y^2 \delta(\omega)$$

$$= \frac{2\alpha\sigma_y^2}{\alpha^2+\omega^2} + 2\pi\mu_y^2 \delta(\omega) \tag{80}$$

We have shown that if an image source is lognormal and Markov, then after passing through a logarithmic nonlinearity it will be Gaussian Markov with a power spectrum defined by equation (80). Furthermore, the entropy of the original source will be changed by μ, the mean of the resultant Gaussian process. The importance of these results will be explored in Section V.

SECTION IV

IMAGE CODING

The rapid growth in high speed, large storage, computational facilities in recent years has made sophisticated digital image processing a reality. The degree of success which can be achieved was demonstrated world wide when pictures were transmitted to earth from the moon and Mars. Two of the major problems that occur in projects such as the Apollo moon missions are effective data reduction and noise free transmission.

The first problem arises due to the bandwidth constraints that exist on any practical communications channel. A standard NTSC television frame contains 525 scan lines of 525 pixels, or approximately 2^{18} data points. The human visual system can resolve from 16 to 256 intensity levels depending on subject matter, type of quantization, and viewing conditions. For the worst case 8×2^{18} or 2^{21} bits would be required to define a single monochrome image. For flicker free television, we need approximately 30 frames per second which gives a bit rate of 2^{26} bits/sec. If we consider color, another factor of three is required; hence, $\sim 2^{27}$ bits/sec or 10^{8} bits/sec would be the final required rate.

The second problem, that of noise susceptibility, is equally

important. If one can transform the image in such a way as to make it less sensitive to noise in the channel, then the signal to noise ratio is increased. This leads to a lower power requirement and a simpler channel coder-decoder design, which results in a lower cost system.

The transmission of images is not the only application for image coding. Obviously, with such a large number of bits per image, storage (particularly high speed rapid access storage) becomes a problem. For example, a single frame of the color image discussed earlier would require approximately 2^{22} 32-bit words of storage (packed as four 8-bit bytes/word), or 4×10^6 words of core on a PDP-10 computer.

4.1. The Coding Problem

In the preceding paragraphs the applicability of image coding was discussed in general terms. We will now present the basic coding problem in more definitive terms. An image coding task may be illustrated as shown in Figure 16. The scanner may be one of many types depending on the source of the original image (the "real world") and it will not be considered in detail. The important point is that in most situations the scanner performs an analog to digital conversion. Thus, X is an estimate (sampled and quantized version) of the original object. The source encoder

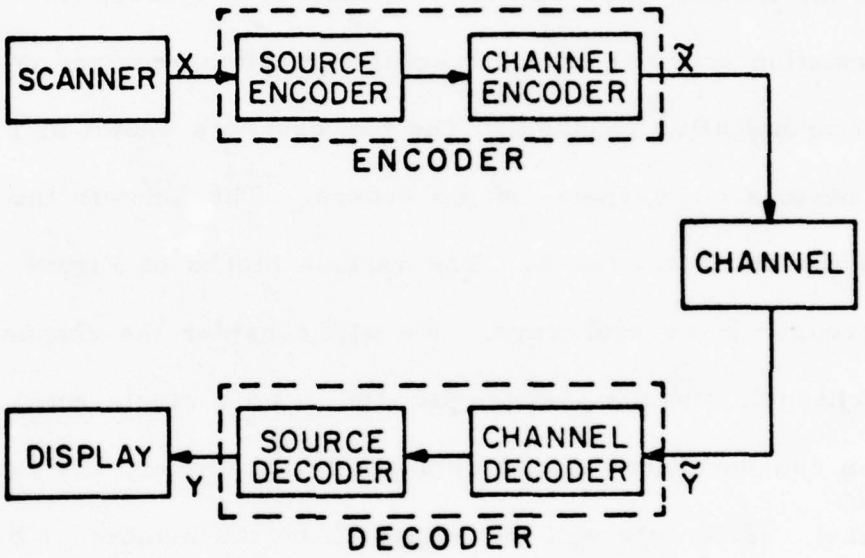


Figure 16. The Image Coding Task

transforms X into a sequence of binary digits. The goal is to make this transformation optimal in terms of data reduction and image fidelity. The channel encoder codes the output of the source in such a way as to insure the binary sequences can be reliably reproduced after passing through the channel. The channel can take on many forms. For example, it may be a storage device (an information channel) or a transmitter-medium-receiver combination (a communication channel). The two decoders shown in Figure 16 are obvious counterparts of the coders. The input to the display, Y is the reconstructed X. The various blocks of Figure 16 can be grouped in several ways. We will consider the channel coder, the channel, and the channel decoder to be a single entity. This group can be characterized by a single parameter, the rate of the channel. This rate will be defined to be the number of bits per picture element (bits/pixel) which can be passed through a given channel. We wish to find the source coding scheme which minimizes the number of bits/pixel required to represent an image and thereby reduce the required channel rate to a minimum. Given this rate, the picture size, and pertinent time factors, the required channel capacity can be computed. The image coding problem therefore centers around the design of an "optimal" source encoder-decoder.

4.2. Pulse Code Modulation

An encoding scheme which has been widely used is pulse code modulation (PCM). This technique, in its simplest form, involves the sampling of an analog signal at a uniform rate and encoding these samples in a binary coder. An adequate number of quantization levels is required to maintain a good signal-to-noise ratio. For most images this requires a minimum of 64 levels; therefore, 6-7 bits/pixel is the normal rate of such a system. If too few levels are used, the images will contain false contours. This type of noise is more annoying to a viewer than additive random noise of the same rms value. Roberts has used this trait in a pseudo-random noise modulation technique which lowers the rate to 4 bits/pixel [43]. Since the noise that remains in a picture processed by the Roberts method is random it can be reduced by averaging. Sawchuk has found that a modified Roberts method which uses averaging and edge detection will produce 3.1 bit/pixel images "almost" as good as the original [44] and [45] Another approach has been to use non-uniform quantization [46] and [47]. These techniques minimize the quantization error by taking advantage of the statistical character of the image. For the Max quantizer, optimum decision and reconstruction tables are computed by using the probability distribution, p(f). Alternatively, a non-linear transformation based on p(f) can be performed and the result

linearly quantized. This latter procedure is called companding. The companding and Max quantizer methods can reduce the rate by as much as 1 bit/pixel.

Conventional PCM makes no assumption about the relationship between adjacent pixels in an image. By default, the pixels are taken to be uncorrelated and as Habibi and Robinson have pointed out, pictures satisfying this assumption occur in places such as television screens after station sign off and are of little interest [48]. Schreiber has shown the conditional entropy of a PCM signal (for the case of uniform amplitude distribution and picture correlation so high that pixel to pixel variations are primarily due to random Gaussian noise equal to one quantization level) is 1.12 bit/pixel [13]. This value represents a lower bound to the required channel capacity for PCM regardless of the statistical relationships employed. If the imagery being coded is multiframe, the rate can be reduced by as much as a factor of five because of the interframe redundancies [48].

4.3. Differential Pulse Code Modulation

As pointed out earlier, PCM makes the assumption that the data is uncorrelated and the same number of bits is assigned to every data point. Since picture data is obviously correlated this procedure is inefficient. One way to obtain less correlation

between the points to be coded is to use a linear predictor to generate a difference signal and quantize this difference signal with a Max quantizer based on the appropriate probability density function. This type of coder is referred to as a differential pulse code modulator (DPCM). Several different types of DPCM systems have been used with the basic differences lying in the predictor design [48, pp. 25-28]. The rates achieved with DPCM are about one half those obtainable with PCM [27].

4.4. Transform Coding

Another way to decorrelate image data is to perform a two-dimensional spatial transformation. As discussed in Section 3.3, the optimum transform would be the KLT; however, the large number of required computations make it a poor choice for coding. Several of the fast transform algorithms have been used [29, Chapter 7]. Transform coders perform two significant operations which make them more efficient than most other types of coders. The first operation is that of performing the linear transformation which maps the statistically dependent pixels into a set of "more independent" (decorrelated) pixels. The second operation is to code each transformed pixel independently, assigning the number of bits according to the variance of that coefficient and/or the location of the coefficient in the transform domain. The first criterion gives

more bits to those pixels with the highest variance or information. The second criterion (particularly for the Fourier domain) assigns more bits to those areas in which the HVS sensitivity is highest.

A major disadvantage of transform coding techniques is the entire image must be available before processing begins. Thus, large amounts of buffering are required for a "real-time" transform coding system. One solution to this problem is to process the image in blocks. For example, rather than compute the 256×256 DCT of a 256×256 image one may compute 1024 8×8 cosine transforms by performing a 32×32 partition of the original image. Only eight lines of the image are required for processing to begin and, in addition, the two covariance matrices which need to be diagonalized to determine bit assignments are only 8×8. In other words, a single 8×8 bitmap is sufficient for coding the entire 256×256 image. To visualize how the 8×8 bitmap is used, the partioned cosine transform domain may be reordered as shown in Figure 17. The 32×32 subpicture shown in the upper left was formed with the 1024 "DC" terms of the 8×8 block transforms, the next subpicture is from the (0, 1) harmonics, etc. In this manner the 8×8 block transform produces an 8×8 array of subpictures. The reordered transform is called a Mandala transform and Kajiya [49] has suggested that the transition to higher harmonic subimages rotates the "feature" space into a "texture" space.

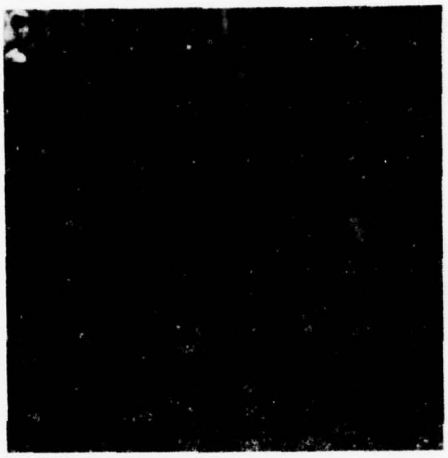


Figure 17. Mandala Ordered 8 x 8 Block Cosine Transform

In Figure 17, the subpictures have been scaled individually for viewing purposes. There is more than six orders of magnitude difference between the coefficients of the upper left and lower right subpictures. When coding this image every term in each subimage is coded with the same number of bits; therefore, only an 8x8 bitmap is required. Note how the increasing harmonics (left to right and top to bottom) represent more and more "edge" information and the highest harmonic is almost random noise or, if you like, texture.

Since an 8x8 block transform coding technique uses an 8x8 covariance matrix, this method does not take full advantage of the redundancies of the image. The performance of block coders improves with increasing block size; however, correlation between adjacent pixels is small for shifts greater than 20 [38]. This reduces the error due to block size to an insignificant amount for n>16 [11, p. 815]. For a 16x16 block size and at 1.5 bits/pixel the Slant, Haar, Hadamard and Fourier transforms have been shown to give results similar to the KLT [50]. Achromatic pictures have been coded at 1 bit/pixel with a root mean square error of .8% [51]. Since transform coding techniques usually involve some type of spatial filtering, they are a type of adaptive or psychovisual coder.

4.5. Psychovisual Coders

Psychovisual coders attempt to take advantage of the limitations of the HVS and code only that data which can be perceived or is meaningful. Since the visual system is the means by which most imagery is ultimately used, compared, and/or judged, psychovisual coding should prove effective. A common point which is used to support the importance of this technique is that the human observer can only absorb about 50 bits/sec [27]. When compared to 10^8 bits/sec (color television) the reduction is six orders of magnitude! But the human observer is usually in a cognitive mode, absorbing the bits of interest. When one views a scene, the entire scene -- in complete detail -- is not perceived at once. If we know exactly where the viewer will look and what mode he is in, the "image" coding problem would be substantially reduced. However, the cost of coding this peripheral information would place the rate well above the 50 bits/sec bound. Thus, the bound is interesting but far from obtainable.

Nonetheless, psychovisual coding is important from two aspects. The first, as previously mentioned, is "why transmit or store that which is not used anyway?" The second, and perhaps more important aspect, involves errors in and the fidelity of the coded images. If we implement a coder in a "perceptual space" which minimizes the visual effects of error, i.e., maintains a

maximum image fidelity, by how much can we reduce the rate and still obtain usuable reconstructions? A concomitant benefit of such an implementation is the definition of a valid error criterion. At present, most coding results are judged subjectively or with an image space mean square error (MSE) criterion (which is known not to be valid). However, if a MSE criterion is used in a perceptual space -- hence a perceptual MSE (PMSE) -- its utility should be increased significantly. Once such a fidelity criterion is precisely defined, development of optimal coders with specified rates and distortion levels becomes possible. Such an approach can be couched quite nicely in terms of rate distortion theory.

4.6. Rate Distortion Theory

Berger has pointed out that there are two basic problems to be coped with when designing a coding system; (1) what information should be transmitted? and (2) how should it be transmitted [42, p.2]? Early work in information theory concentrated on the second problem. In 1959 Shannon addressed the first problem [52]. He defined the rate distortion function of an information source with respect to a fidelity criterion and established the fundamental theorems basic to rate distortion theory. Stated simply, the basis of this theory is the rate distortion function of a source with known probability distribution determines the

minimum channel capacity required to transmit the source output as a function of the desired minimum average distortion [53]. The distortion function, or fidelity criterion, is a measure of agreement between the source and system output specified by the user. The theory is covered in detail in Berger [42] and Gallager [54, Chapter 9]. A fundamental result is if D is the desired average distortion and R(D) is the rate distortion function, then a system can be designed that achieves the distortion D if and only if the capacity of the channel between the source and user is greater than R(D). Thus, R(D) is the effective rate at which the source produces information subject to a distortion D. For $D = 0$, $R(0) \leq H(\cdot)$, where $H(\cdot)$ is the entropy of the source. As D increases R(D) decreases monotonically and -- more importantly -- in a convex manner, usually becoming zero at some finite value of distortion, D_{max}. A typical R(D) versus D curve is shown in Figure 18.

There are two key points in applying rate distortion theory. First, the probability distribution of the source is required. Secondly, the rate distortion function must be defined. Finding the probability distribution of a class of images is not a simple task, particularly for the sources with memory (the more interesting ones as noted earlier). Once the distribution is determined and a distortion criterion selected, the problem of deriving R(D) usually

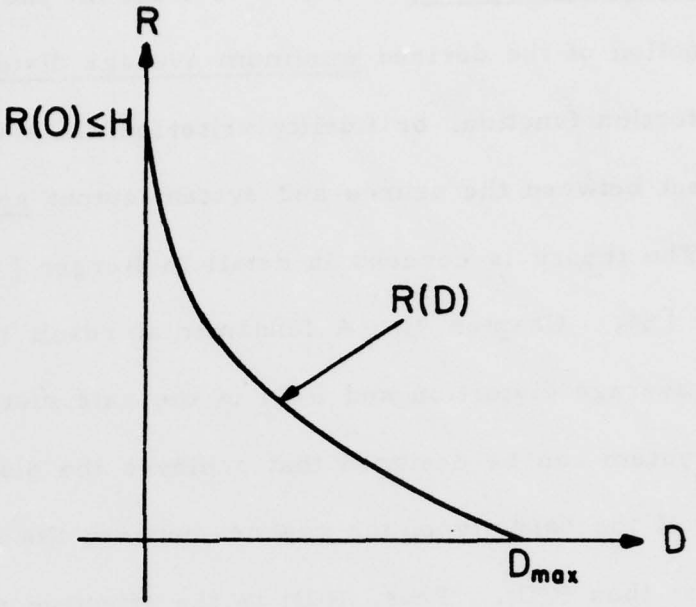


Figure 18. A Typical Rate Distortion Curve

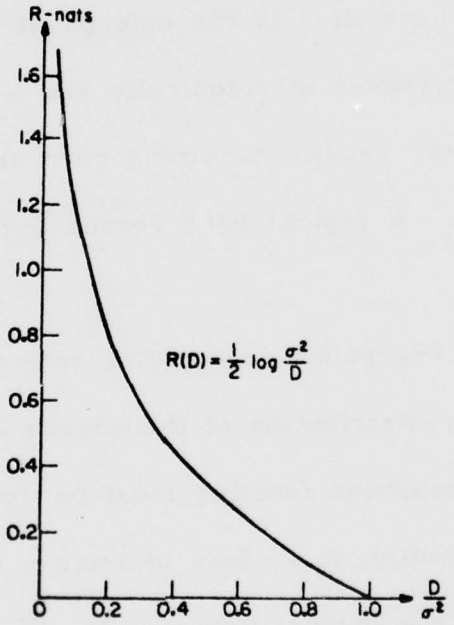


Figure 19. R(D) for a Memoryless Gaussian Source and MSE Criterion

proves to be unsolvable. One combination which is tractable is an independent Gaussian source with a MSE distortion measure. The independent Gaussian assumption is certainly not valid for image sources; however, this particular combination is an upper bound on achievable performance for any source with common second moments [53, p. 802].

To obtain this simple result we must first define the source. Let $X = \{x_i, i = 1, 2, \ldots, N\}$ be the set of independent source samples which are Gaussian with zero mean and variance σ^2. The output of the source decoder (see Figure 16) will be represented by $Y = \{y_i, i = 1, 2, \ldots, N\}$. The distortion measure is defined as

$$d(X, Y) = \sum_{i=1}^{N} (y_i - x_i)^2 \tag{81}$$

so that the average MSE becomes

$$D = \frac{1}{N} \sum_{i=1}^{N} E\{(y_i - x_i)^2\} \tag{82}$$

where $E\{\cdot\}$ denotes the expected value operator. The rate distortion function corresponding to these conditions has been shown to be [42, p. 99]

$$R(D) = \begin{cases} \frac{1}{2} \ln \frac{\sigma^2}{D} & , \quad 0 \le D \le \sigma^2 \\ 0 & , \quad D > \sigma^2 \end{cases} \tag{83}$$

This particular function is illustrated in Figure 19. The rate predicted by equation (83) is achieved theoretically by encoding in such a way as to produce an output error which is Gaussian with variance D and is independent from sample to sample. In practice, the rate is approached within 1/4 bit per pixel by optimum quantization (via a Max quantizer [47] and noiseless coding [53], p. 803]. Davisson has given the following rough intuitive justification of the rate distortion function in terms of quantizing [53, p. 803]. The noise standard deviation, as a function of the signal amplitude, is inversely proportional to the number of quantization levels. Therefore, the number of levels should be proportional to $\sigma/\sqrt{D}$ and the number of information bits should be the logarithm of this quantity [55]. If the distortion is greater than the variance of the signal, the transmission rate should be zero since nothing need be transmitted. This is the relationship established in equation (83).

In the preceding discussion we defined a source X and an encoded output Y and obtained a set of parametric equations which define the rate distortion relationship. Let us now consider the input to be a raster scanned image, u(x, y). Further, assume that

this image is passed through a linear system defined by the transfer function $A(f_x, f_y)$, thus

$$v(x, y) = u(x, y) \circledast a(x, y) \tag{84}$$

where $a(x, y)$ is impulse response corresponding to $A(f_x, f_y)$. The encoded output, denoted as $\tilde{u}(x, y)$ will yield a similar result, therefore the MSE distortion becomes

$$d(v, \tilde{v}) = \iint [v(x, y) - \tilde{v}(x, y)]^2 \, dx \, dy \tag{85}$$

$$= \iint \{[u(x, y) \circledast a(x, y)] - [\tilde{u}(x, y) \circledast a(x, y)]\}^2 \, dx \, dy$$

$$= \iint \{[u(x, y) - \tilde{u}(x, y)] \circledast a(x, y)\}^2 \, dx \, dy$$

$$= \iint [\Delta u(x, y) \circledast a(x, y)]^2 \, dx \, dy \tag{86}$$

where $\Delta u(x, y)$ denotes the difference picture formed by subtracting the coded image from the source image. Now that the distortion measure has been defined we need only specify the probability distribution of the source to be able to calculate the rate distortion function. We will take U_{xy} to be a two-dimensional random field representing the random source (a collection of random variables parameterized by two independent variables). Let the estimate of the mean be

$$m = E\{U_{xy}\} \tag{87}$$

and the correlation function be

$$R_u(\tau_x, \tau_y) = E\left\{U_{x+\tau_x, y+\tau_y} U_{xy}\right\} \tag{88}$$

We will assume the joint distribution of U_{xy} to be Gaussian. Again, even though this may not be the correct distribution, this is a worst case assumption [41]. Sakrison and Algazi have shown that for a raster scan large compared to the correlation distance of the image, the rate distortion function is given parametrically by [56].

$$R(\theta) = \frac{1}{2} \iint_{S_v(f_x, f_y) > \theta} \log_2 \left[\frac{S_v(f_x, f_y)}{\theta}\right] df_x\, df_y \tag{89}$$

$$d(\theta) = \int_{-\infty}^{\infty} \min\left[S_v(f_x, f_y), \theta\right] df_x\, df_y \tag{90}$$

in which $S_v(f_x, f_y)$ is the power spectral density of $v(x, y)$ and is defined as the Fourier transform of $R_v(\tau_x, \tau_y)$.

Briefly reviewing, the following assumptions were made in obtaining equation (89) and equation (90):

(1) The class of images can be represented by a uniform, homogeneous, and stationary random field U_{xy}.

(2) The probability distribution of U_{xy} is a two-dimensional joint Gaussian distribution.

(3) The autocorrelation, $R_u(x,y)$, and the corresponding power spectral density, $S_u(f_x, f_y)$, of U_{xy} are known.

(4) The system transfer function $A(f_x, f_y)$ and hence, the power spectral density, $S_v(f_x, f_y)$ are known.

Given these assumptions, we may compute a rate distortion curve similar to that in Figure 19 by varying distortion, θ, in equations (89) and (90). This curve will represent a theoretical bound by which the performance of any coder implemented within the system can be judged.

SECTION V

STATISTICAL ANALYSIS OF THE HVS MODEL

In Section II we developed a mathematical model for the HVS (see Figure 2). Subsequently, some statistical properties of images were discussed in Section III and the basics of rate distortion theory were presented in the previous section. In this section we will bring these ideas together and develop a set of rate distortion curves which are valid for a perceptual domain defined by our HVS model. We will begin with an achromatic model.

5.1. The Achromatic Case

If we assume a black and white image, then the two chrominance signals c_1 and c_2 in Figure 2 become zero. Thus, the luminance signal, ℓ, is the only output of our model and the model reduces to that shown in Figure 20. This model has been discussed extensively and analyzed by Hall and Hall [1]. A fundamental result of the analysis was that the high frequency roll-off of the overall describing function for this system is a function of contrast. In particular, as the contrast of the input increases, the system sensitivity to high spatial frequencies decreases. This

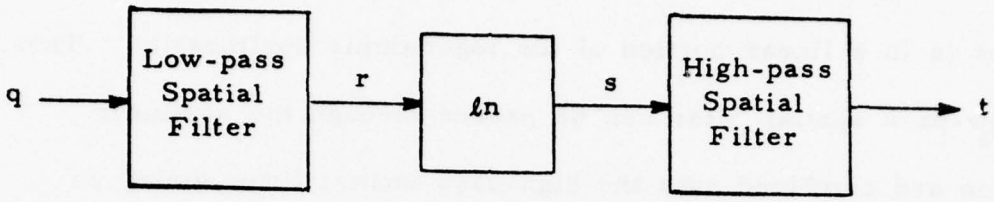


Figure 20. Achromatic HVS Model

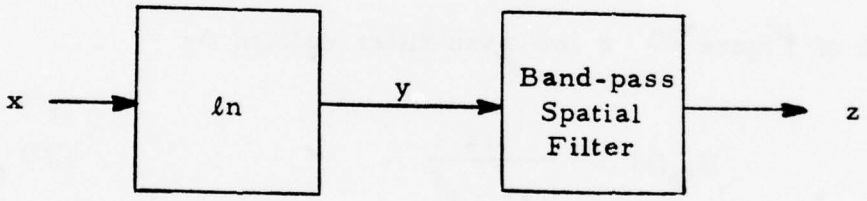


Figure 21. Simplified Achromatic Model

particular characteristic is not present in the model of Figure 21.

The simplified model in Figure 21 is obtained from the model of Figure 20 by assuming that the intensity range of input images is in a linear portion of the logarithmic nonlinearity. Thus, the low-pass spatial filter can be passed through the nonlinear function and combined with the high-pass spatial filter giving an overall bandpass function. This particular type of argument is used in justifying the contrast sensitivity functions which are obtained from sine-wave grating experiments. Indeed, the bandpass filter of Figure 21 would be of the form shown in Figure B.3.

We have previously compared the results of processing black and white images through these two achromatic models [57]. For the model of Figure 20 a low-pass filter defined by

$$H_{lp}(\omega) = \frac{.14}{.49 + \omega^2} \qquad (91)$$

was used. This function corresponds to a 3mm pupil and it is -3dB at 6.6 cycles/degree. The high-pass filter was defined by

$$H_{hp}(\omega) = \frac{10^{-4} + \omega^2}{4 \times 10^{-3} + .8\omega^2} \qquad (92)$$

The model shown in Figure 21 was implemented with a filter function developed by Mannos and Sakrison [7] and it was defined as

$$H_{bp}(\omega) = 2.6[0.0192 + 0.018\omega] \exp[-(0.018\omega)^{1.1}] \qquad (93)$$

This particular function peaks at 8 cycles/degree and an isotropic version is shown in Figure 22. Two 512x512 images (one an aerial photograph of Los Angeles International Airport [LAX] and the other a country bridge scene) were processed with the two achromatic models. The results are shown in Figure 23. From the pictures in Figure 23 it can be seen that for practical purposes the two models produce equivalent results. The only difference is in the peak frequency response which gives slightly more blur in the full achromatic model case. Thus, it appears that the bandpass model is valid for "real-world" achromatic images.

In Section 3.5 we found that an input process with first-order Markov statistics produced a power spectrum (out of a logarithmic nonlinearity) given by equation (80). The output power spectrum from the reduced achromatic model is simply

$$S_z(\omega) = S_y(\omega)|H_{bp}(\omega)|^2 \qquad (94)$$

where

$$S_y(\omega) = \frac{2\alpha\sigma_y^2}{\alpha^2 + \omega^2} + 2\pi\mu_y^2\delta(\omega) \qquad (95)$$

and $H_{bp}(\omega)$ is given by equation (93) Habibi and Wintz have

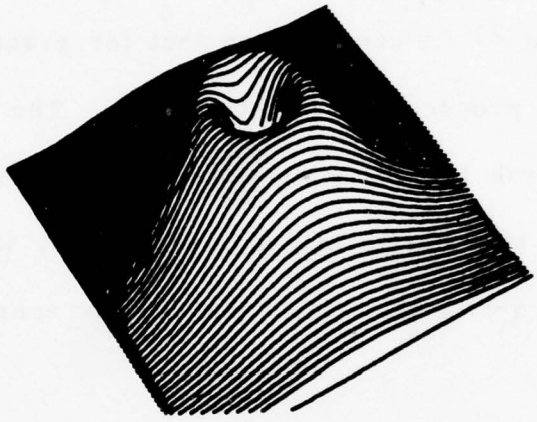


Figure 22. Isotropic Bandpass Filter

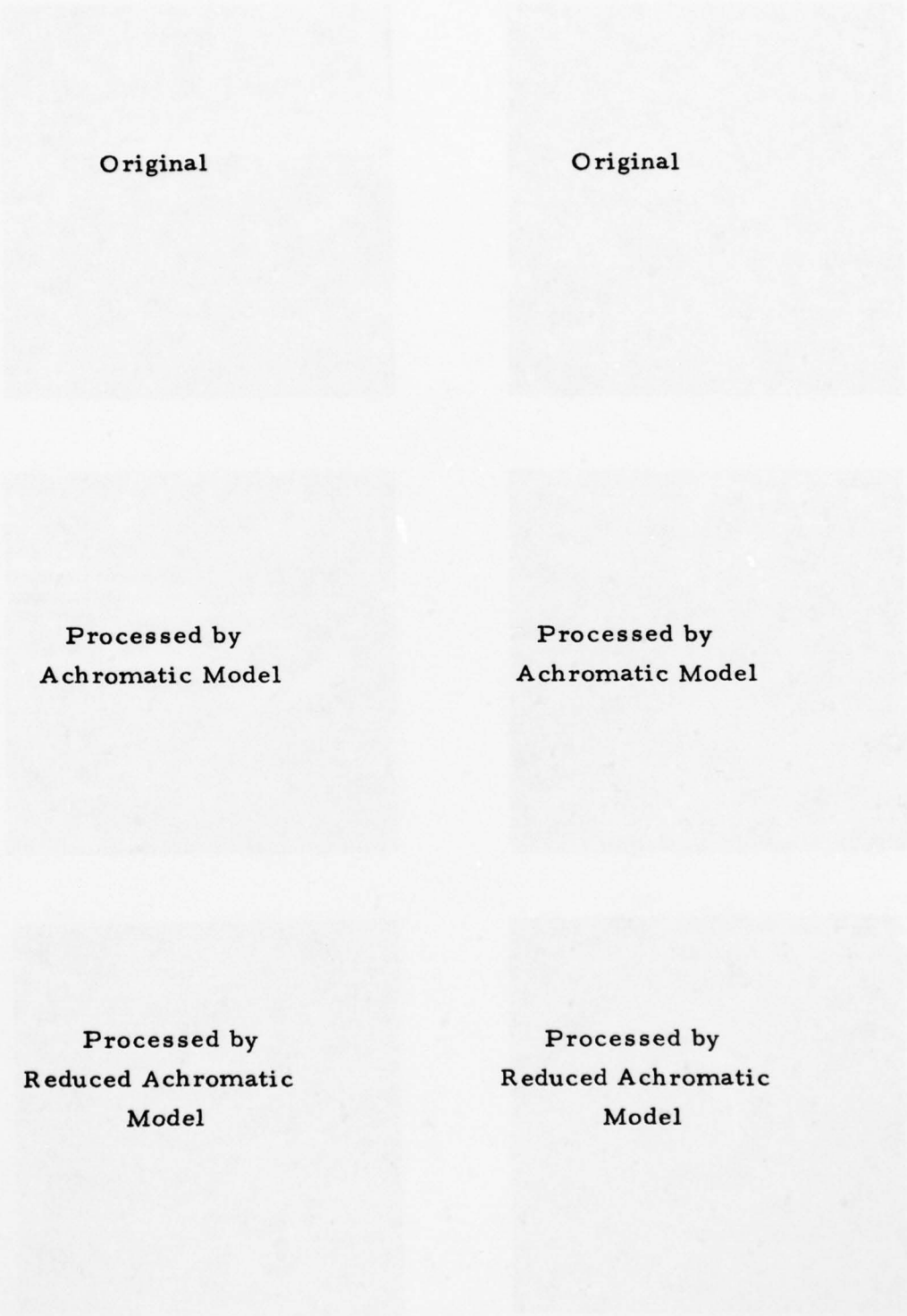


Figure 23. LAX and BRIDGE Images

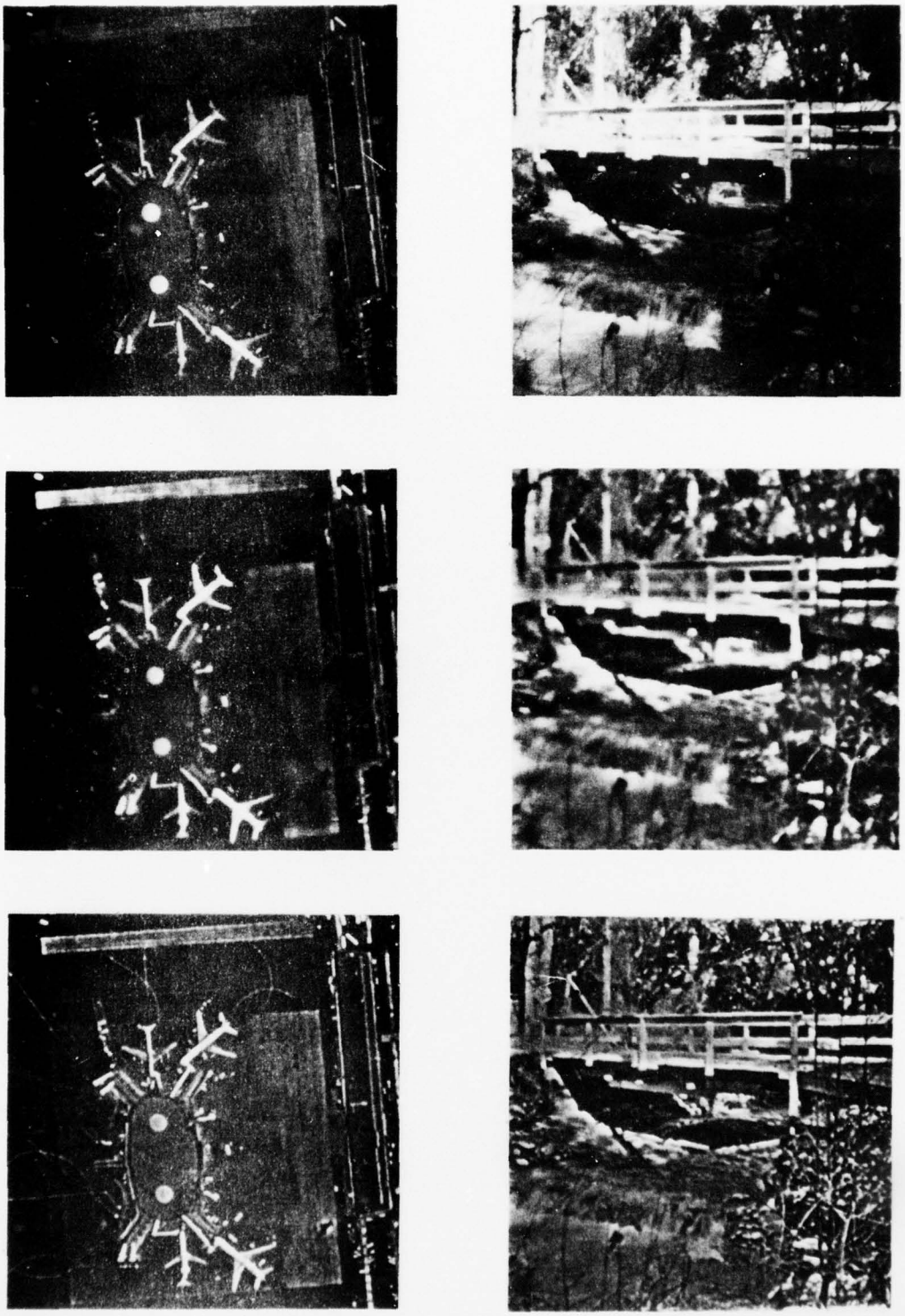


Figure 23. LAX and BRIDGE Images

shown the Markov assumption to be valid for raw images and a typical value for α is .1 [38]. Typical values for σ_y^2 and μ_y^2 are .5 and 16.8 respectively. These parameters give the set of curves for $S_y(\omega)$, $H_{bp}(\omega)$, and $S_z(\omega)$ shown in Figure 24.

The power spectrum defined by equation (94) is valid for the model shown in Figure 21. Although the experimental system comparison shown in Figure 23 indicates the effects of the models in Figures 20 and 21 are similar, a question of interest is how does $S_z(\omega)$ compare to $S_t(\omega)$? From Figure 20, given an input process q with autocorrelation $R_q(\tau)$ and power spectrum $S_q(\omega)$, the power spectrum $S_r(\omega)$ is defined by

$$S_r(\omega) = S_q(\omega) \left| H_{lp}(\omega) \right|^2 \qquad (96)$$

where $H_{lp}(\omega)$ is given by equation (91). By definition, $R_r(\tau)$ is the inverse Fourier transform of $S_r(\omega)$, hence

$$R_r(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_q(\omega) \left| H_{lp}(\omega) \right|^2 e^{j\omega\tau} \, d\omega \qquad (97)$$

We also known that in general

$$R_r(\tau) = \mu_r^2 + \sigma_r^2 \rho_r(\tau) \qquad (98)$$

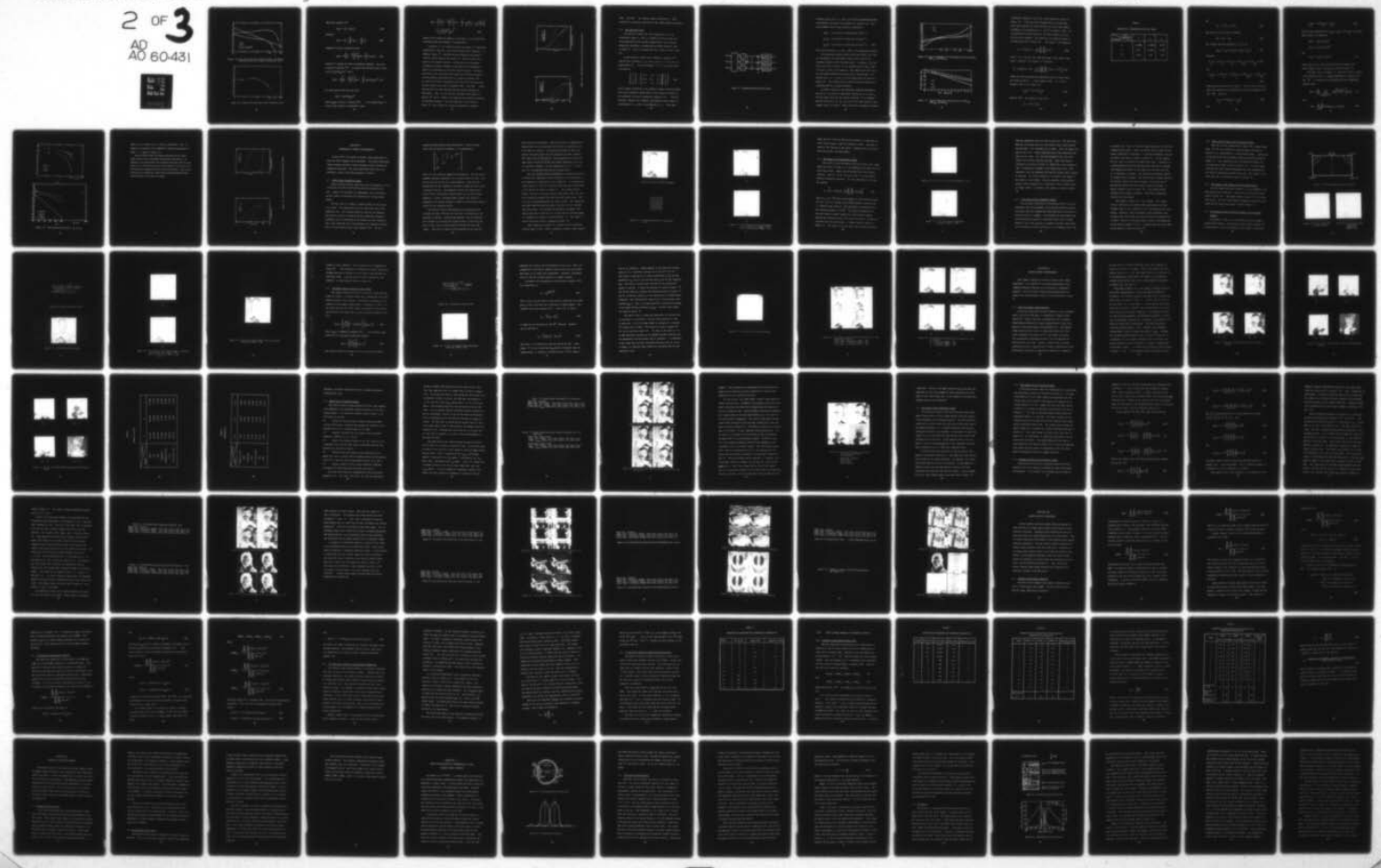
which implies

$$\rho_r(\tau) = \frac{R_r(\tau) - \mu_r^2}{\sigma_r^2} \qquad (99)$$

AD-A060 431 UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES IMAGE --ETC F/G 20/6
DIGITAL COLOR IMAGE COMPRESSION IN A PERCEPTUAL SPACE.(U)
JUL 78 H C ANDREWS, C F HALL F33615-77-C-1016

UNCLASSIFIED AFAL-TR-78-80 NL

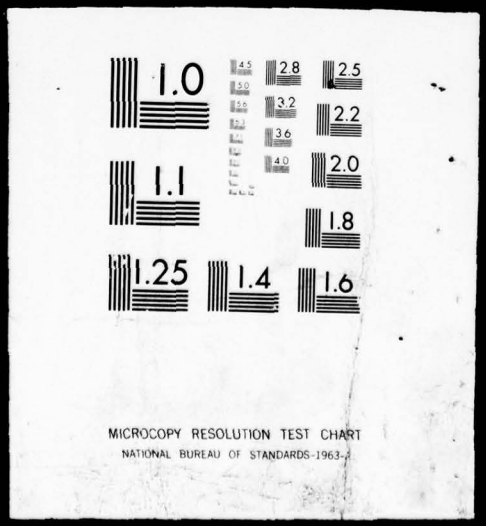
2 OF 3
AD
A060431



SIFIED

2 OF 3

AD
AO 60431



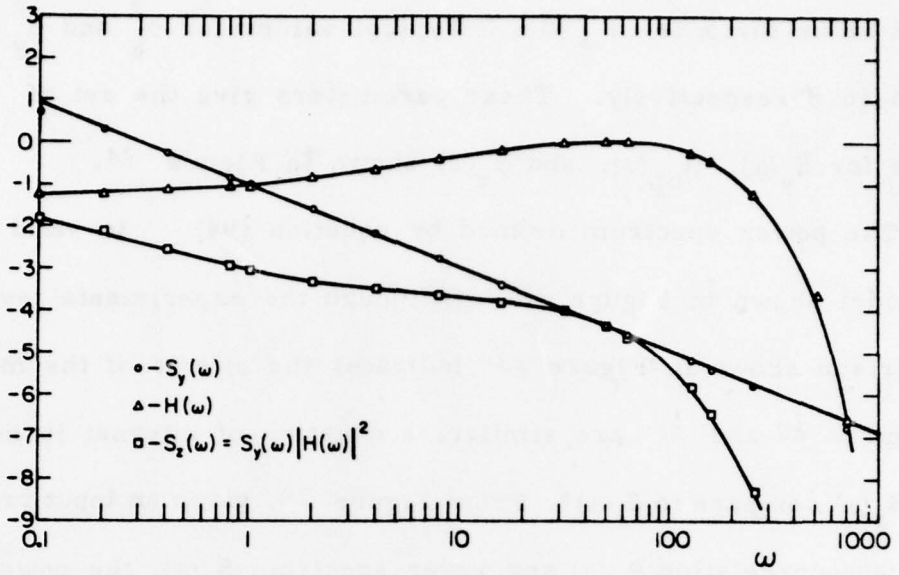


Figure 24. Input Power Spectrum, Filter Function, and Output Power Spectrum of the Reduced Achromatic Model

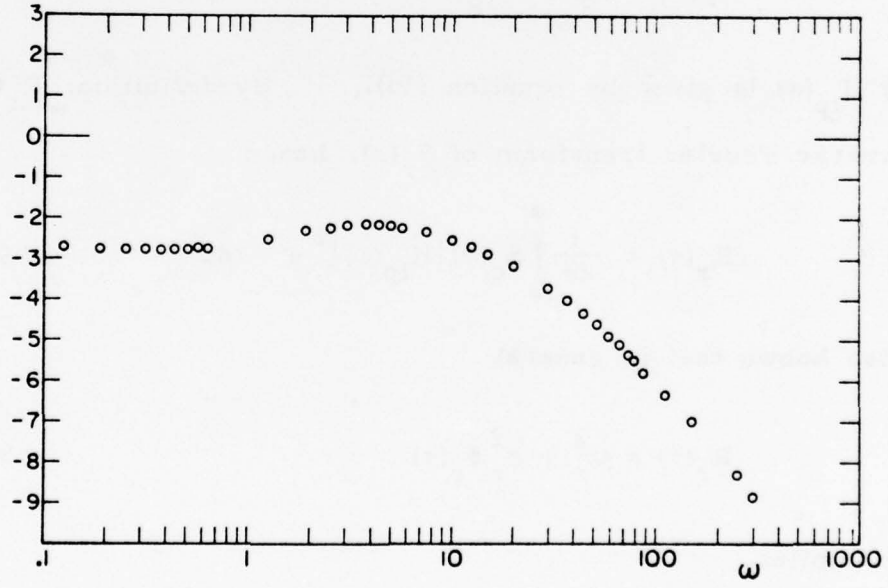


Figure 25. Output Power Spectrum of the Achromatic Model

Now from equation (74)

$$R_s(\tau) = \mu_s^2 + \sigma_s^2 \rho_r(\tau) \tag{100}$$

therefore

$$R_s(\tau) = \mu_s^2 + \frac{\sigma_s^2}{\sigma_r^2} R_r(\tau) - \frac{\sigma_s^2}{\sigma_r^2} \mu_r^2 \tag{101}$$

Taking the Fourier transform we get

$$S_s(\omega) = 2\pi \left[\mu_s^2 - \frac{\sigma_s^2 \mu_r^2}{\sigma_r^2} \right] \delta(\omega) + \frac{\sigma_s^2}{\sigma_r^2} \mathcal{F}\{R_r(\tau)\} \tag{102}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform operation. But $R_r(\tau)$ is given by equation (97) in terms of the inverse Fourier transform of $S_q(\omega)|H_{lp}(\omega)|^2$, hence

$$S_s(\omega) = 2\pi \left[\mu_s^2 - \frac{\sigma_s^2 \mu_r^2}{\sigma_r^2} \right] \delta(\omega) + \frac{\sigma_s^2}{\sigma_r^2} S_q(\omega) |H_{lp}(\omega)|^2 \tag{103}$$

Of course $S_t(\omega)$ follows directly from

$$S_t(\omega) = S_s(\omega) |H_{hp}(\omega)|^2 \tag{104}$$

where $H_{hp}(\omega)$ is given by equation (92). If we assume $R_q(\tau)$ to be first order Markov of parameter α then

$$S_t(\omega) = \left\{ 2\pi \left[\mu_s^2 - \frac{\sigma_s^2 \mu_r^2}{\sigma_r^2} \right] \delta(\omega) + \frac{\sigma_s^2}{\sigma_r^2} \frac{2\alpha}{(\alpha^2 + \omega^2)} \frac{.14^2}{(.49 + \omega^2)^2} \right\}$$

$$\times \left| \frac{\omega^2 + 10^{-4}}{4 \times 10^{-3} + .8\omega^2} \right|^2 \qquad (105)$$

Figure 25 was obtained by setting $\alpha = .1$ and using μ_r, μ_s, σ_r^2 and σ_r^2 (as determined from actual images) in equation (105).

In Section 3.5 we showed the pdf at the output of a logarithmic nonlinearity is Gaussian, given that the input pdf is lognormal. In addition, Figure 12 indicated that the lognormal assumption is valid for typical imagery and Figure 14 shows the pdf of the output to be strongly Gaussian. Furthermore, the reasonable assumption of Markov statistics at the input of the nonlinearity leads to an expression for the power spectrum $S_y(\omega)$. The band-pass filter in the achromatic HVS model has also been verified by several different experiments and it is given by equation (93). A review of the basic assumptions which led to the rate distortion function defined by the pair of equations (89) and (90) reveal that they have all been satisfied with the possible exception of stationarity. Thus, we see that the achromatic HVS models of Figures 20 and 21 enable us to apply the rate distortion equations to achromatic imagery. The rate distortion curves shown in Figure 26 were obtained by solving the parametric equations

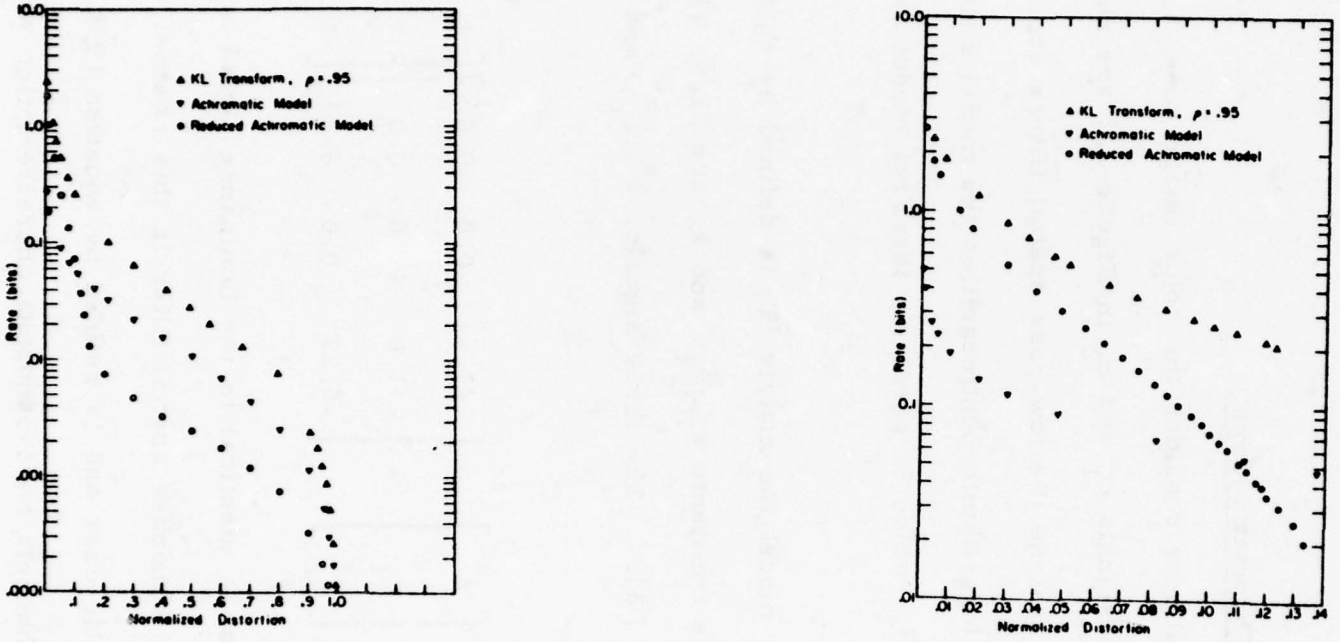


Figure 26. Rate Distortion Curves for the Achromatic Model

(89) and (90) for various values of distortion θ. This operation is sometimes referred to as the "water-filling" procedure.

5.2. The Chromatic Case

We will now consider the color image case, i.e., the chrominance signals c_1 and c_2 in Figure 2.2 are not both zero. If we again assume the low-pass spatial filters can be passed through the logarithmic nonlinearities the model reduces to that in Figure 27 which is precisely the Frei model for color vision [3].

In this model the matrix [T] is defined by equation (23) and the three constants k_1, k_2, and k_3 are 21.5, 41.0, and 6.27 respectively [3]. The three signals, ℓ^*, c_1^*, and c_2^* are therefore given by

$$\begin{bmatrix} \ell^* \\ c_1^* \\ c_2^* \end{bmatrix} = \begin{bmatrix} 21.5 & 0.0 & 0.0 \\ -41.0 & 41.0 & 0.0 \\ -6.27 & 0.0 & 6.27 \end{bmatrix} \begin{bmatrix} t_1^* \\ t_2^* \\ t_3^* \end{bmatrix} \qquad (106)$$

The ℓ^* signal is identical to the luminance signal of the achromatic case and the bandpass spatial filter in this channel is identical to the achromatic case and is defined by equation (93). The two chromatic channels have bandpass characteristics which peak at 4 cycles/degree for c_1 and 2 cycles/degree for c_2. These peak

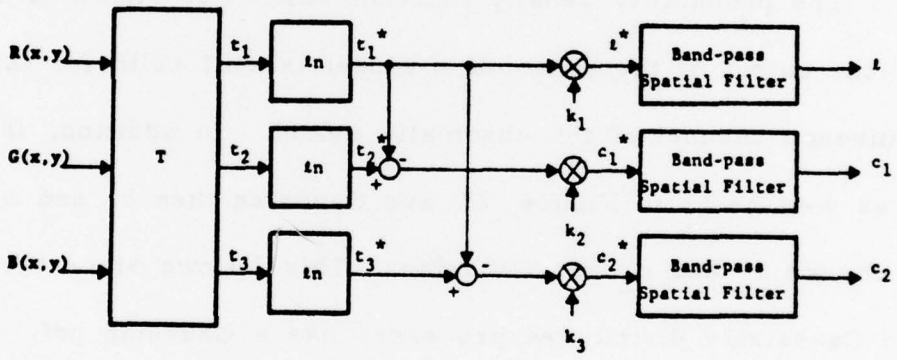


Figure 27. Simplified Chromatic HVS Model

frequency points (for ℓ, c_1, and c_2) have been established through psychophysical techniques by Faugeras [37, Figure 3.9]. The three bandpass filters may therefore be defined by

$$H_{\ell}(\omega) = 2.6[0.0192 + 0.018\omega] \exp[-(0.018\omega)^{1.1}]$$

$$H_{c_1}(\omega) = 2.6[0.0192 + 0.036\omega] \exp[-(0.036\omega)^{1.1}]$$

$$H_{c_2}(\omega) = 2.6[0.0192 + 0.072\omega] \exp[-(0.072\omega)^{1.1}] \qquad (107)$$

where the subscripts ℓ, c_1, and c_2 refer to the appropriate channel.

The probability density function which was shown to be valid for the output of the achromatic model is still valid for the luminance channel of the chromatic model. In addition, if t_2^* and t_3^* as well as t_1^* in Figure 27 are Gaussian then c_1^* and c_2^* and of course c_1 and c_2 are Gaussian. This follows since the sum of two Gaussianly distributed processes has a Gaussian pdf. Probability plots of ℓ, c_1, and c_2 for the Kodak GIRL are shown in Figure 28. The straight lines in these three plots indicate the underlying pdfs are strongly Gaussian.

In order to apply the rate distortion equations developed in Section 4.6, we need the output power spectra for ℓ, c_1, and c_2. We may again draw upon the results of Section 3.5 to establish that the processes at t_1^*, t_2^*, and t_3^* are first order Markov if the original inputs are Markov. Plots of the first 14 spatial correlation

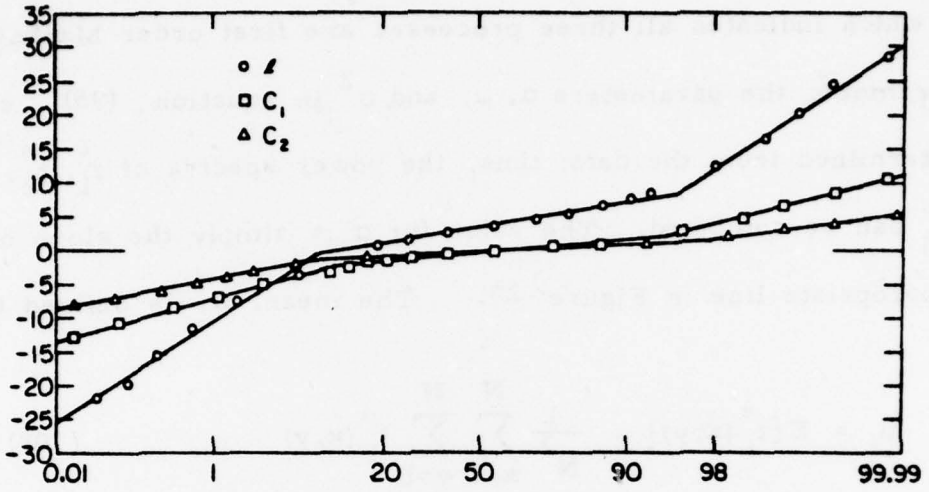


Figure 28. Linear-probability Plot of Histogram Data from $GIRL_{\ell}$, $GIRL_{c1}$, and $GIRL_{c2}$

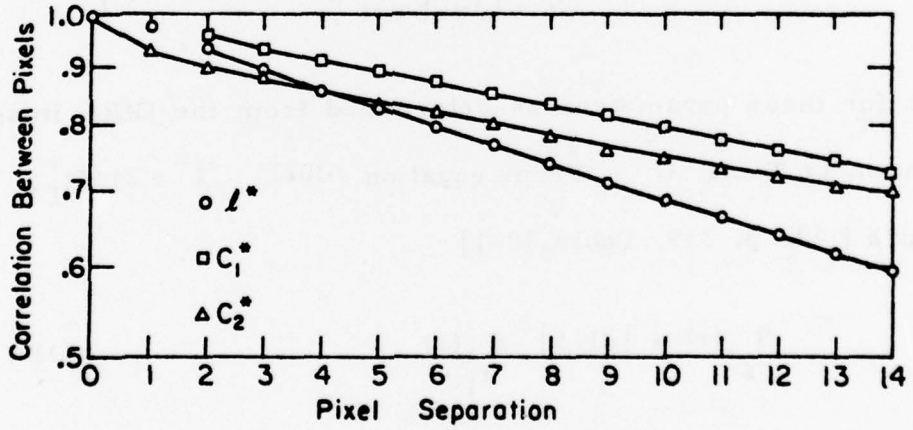


Figure 29. Adjacent Pixel Correlation Curves for $GIRL_{t_1^*}$, $GIRL_{t_2^*}$, and $GIRL_{t_3^*}$

coefficients computed in the t_1^*, t_2^*, and t_3^* planes are shown in Figure 29. They form three straight lines in the log-linear plots which indicates all three processes are first order Markov. Furthermore, the parameters α, μ, and σ^2 in equation (95) can be determined from the data; thus, the power spectra of t_1^*, t_2^*, and t_3^* can be computed. The value for α is simply the slope of the appropriate line in Figure 29. The mean, μ, is defined by

$$\mu_i = E\{t_i^*(x,y)\} = \frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N} t_i^*(x,y) \qquad (108)$$

where $i = 1, 2, 3$ and N is the width and length of the square image array. Similarly, the variance, σ^2, becomes

$$\sigma_i^2 = E\{[t_i^*(x,y) - \mu]^2\} = \frac{1}{N^2 - 1}\left\{\sum_{x=1}^{N}\sum_{y=1}^{N} t_i^{*2}(x,y) - N^2\mu_i^2\right\} \qquad (109)$$

Values for these parameters as determined from the GIRL image are shown in Table 1. From equation (106) $\ell^* = 21.5t_1^*$, therefore [39, p. 339, Table 10-1]

$$S_{\ell^*}(\omega) = |21.5|^2 S_{t_1^*}(\omega) \qquad (110)$$

Equation (106) also defines c_1^* and c_2^* as

$$c_1^* = -41t_1^* + 41t_2^* \qquad (111)$$

TABLE 1

STATISTICAL PARAMETERS FROM GIRL IMAGE

Parameter \ Color Coordinate	ℓ	c_1	c_2
α	0.0388	0.0228	0.024
μ	3.96	-0.228	-0.064
σ^2	18.26	3.90	1.26

and

$$c_2^* = -6.27\, t_1^* + 6.27\, t_3^* \tag{112}$$

Now the sum of two random variables,

$$z(t) = x(t) + y(t) \tag{113}$$

has a power spectrum defined by [39, p. 337]

$$S_{zz}(\omega) = S_{xx}(\omega) + S_{yy}(\omega) + S_{xy}(\omega) + S_{yx}(\omega) \tag{114}$$

Therefore,

$$S_{c_1^*}(\omega) = 41^2 S_{t_1^*}(\omega) + 41^2 S_{t_2^*}(\omega) - 41^2 S_{t_1^* t_2^*}(\omega) - 41^2 S_{t_2^* t_1^*}(\omega) \tag{115}$$

and

$$S_{c_2^*}(\omega) = 6.27^2 S_{t_1^*}(\omega) + 6.27^2 S_{t_3^*}(\omega) - 6.27^2 S_{t_1^* t_3^*}(\omega) - 6.27^2 S_{t_3^* t_1^*}(\omega) \tag{116}$$

define the power spectra of c_1^* and c_2^*. For the case of decorrelated color planes the cross-spectra are zero and equations (115) and (116) reduce to

$$S_{c_1^*}(\omega) = 41^2 [S_{t_1^*}(\omega) + S_{t_2^*}(\omega)] \tag{117}$$

and

$$S_{c_2^*}(\omega) = 6.27^2 [S_{t_1^*}(\omega) + S_{t_3^*}(\omega)] \tag{118}$$

Since we have expressions for $S_{\ell^*}(\omega)$, $S_{c_1^*}(\omega)$, and $S_{c_2^*}(\omega)$, the output power spectra are defined by

$$S_{\ell}(\omega) = S_{\ell^*}(\omega) \left| H_{\ell}(\omega) \right|^2$$

$$S_{c_1}(\omega) = S_{c_1^*}(\omega) \left| H_{c_1}(\omega) \right|^2$$

and

$$S_{c_2}(\omega) = S_{c_2^*}(\omega) \left| H_{c_2}(\omega) \right|^2 \tag{119}$$

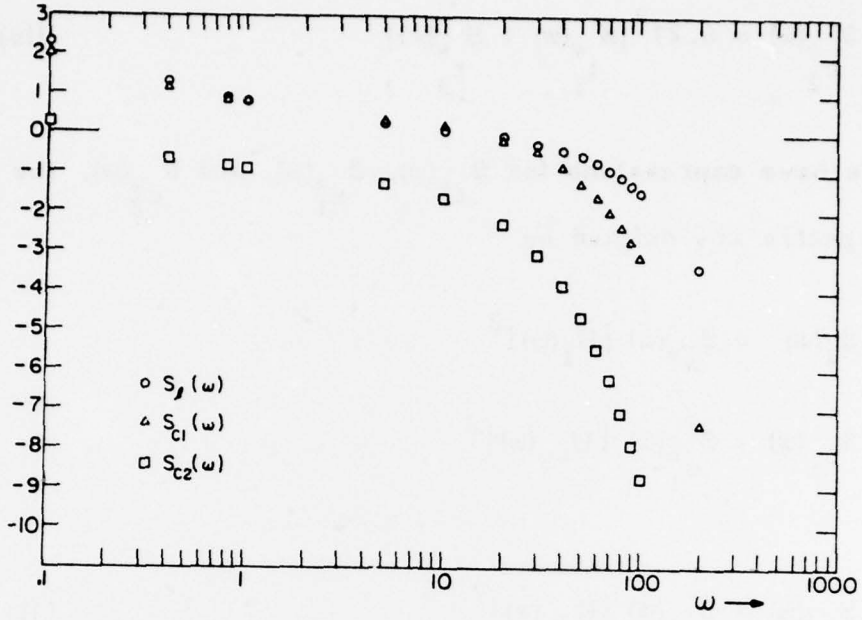
where $H_{\ell}(\omega)$, $H_{c_1}(\omega)$, and $H_{c_2}(\omega)$ are given by equation (107) Plots of $S_{\ell}(\omega)$, $S_{c_1}(\omega)$, and $S_{c_2}(\omega)$ are shown in Figure 30.

The three curves of Figure 30 can now be used to compute a rate distortion curve. A slight modification of equations (89) and (90) is required to accommodate the three independent spectra; hence,

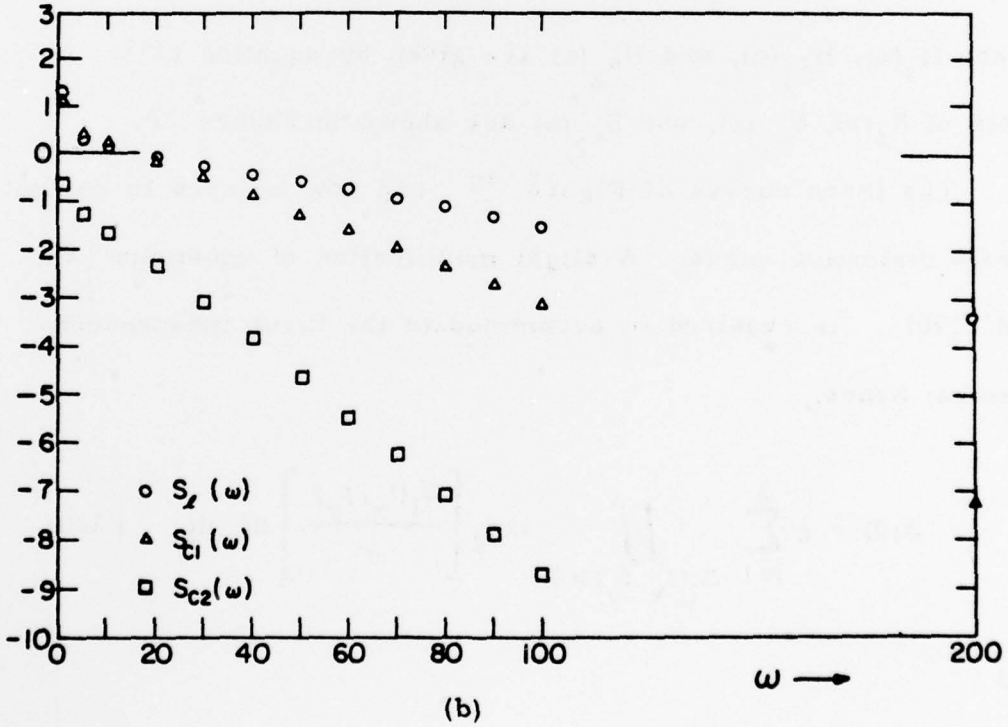
$$R(\theta) = \tfrac{1}{2} \sum_{i=1}^{3} \iint_{S_i(f_x, f_y) > \theta} \log_2 \left[\frac{S_i(f_x, f_y)}{\theta} \right] df_x \, df_y \tag{120}$$

and

$$d(\theta) = \sum_{i=1}^{3} \int_{-\infty}^{\infty} \min \left[S_i(f_x, f_y), \theta \right] df_x \, df_y \tag{121}$$



(a)



(b)

Figure 30. Power Spectrum Plots for ℓ, c_1, and c_2

where $i = 1, 2, 3$ refers to ℓ, c_1, and c_2 respectively. The resultant rate distortion curve obtained by using the parameters in Table 1 is shown in Figure 31.

In this section we have developed expressions for the output power spectra of our achromatic and chromatic HVS models. In addition, a set of parametric rate distortion equations based on mean square error and Gaussian pdf was used to obtain a set of curves for the theoretical coding performance of our HVS models. These curves can be used to evaluate the results of the coding experiments which will be detailed in the next two sections.

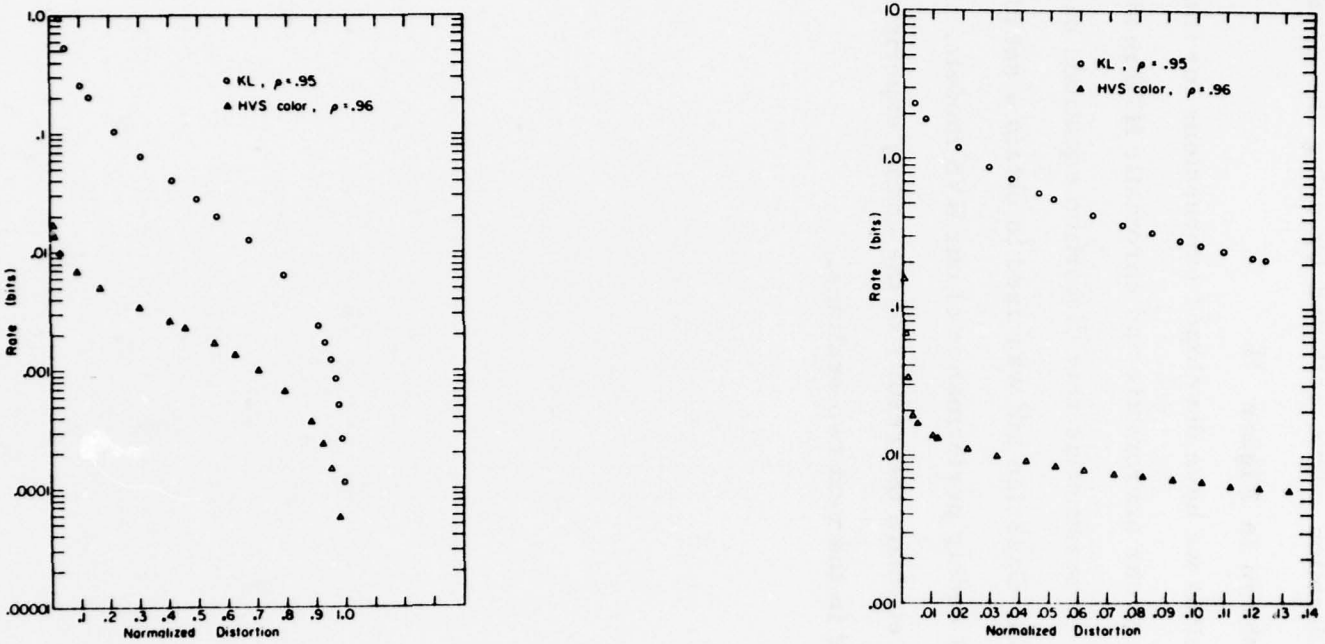


Figure 31. Rate Distortion Curve for the Chromatic Model

SECTION VI

ACHROMATIC CODING EXPERIMENTS

In this section the results of several coding experiments on black and white imagery will be presented. The initial experiments involve standard transform coding techniques and are included for comparative purposes. The later experiments make use of the achromatic model of the HVS developed in Section V.

6.1. Block Cosine Transform Coding

Block transform picture coding has been investigated by several researchers [11], [38], and [58] and we will not develop the theory here. Rather, the procedure as implemented, will be presented and the reader is referred to the references for the theoretical details.

The first step is to obtain a variance matrix for the picture to be coded. This matrix will be of the same block size as the subpicture size. The variance matrix is used for two purposes. The number of bits to be used to encode a particular transform coefficient will be proportional to the variance for that coefficient. Also, each coefficient will be normalized by its respective variance prior to being quantized with a Max quantizer [47]. We will

assume the picture data is first order Markov. A block Toeplitz array with the desired correlation, ρ, is generated as

$$\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \rho^2 & \\ \rho^2 & \rho & & & \\ \vdots & \rho^2 & & & \\ \rho^{N-1} & & & & 1 \end{bmatrix} \qquad (122)$$

where N is the width and length of the subpicture. For the case of spatially separable transforms, two of these arrays are used -- one for the row and one for the column statistics. They are both transformed (for this example by the DCT) yielding row and column covariance matrices. The diagonals of these two matrices are used to form a normalized variance matrix via an outer product expansion. Finally, assuming ergodic images, this matrix is multiplied by the spatial variance to obtain an unnormalized variance matrix for the transform domain.

The process used for determining the bit assignment was developed by Pratt [59]; and, for the case of Gaussian data, the algorithm is optimal. Basically the algorithm uses the Gaussian error function to decrement the largest variance of the array one bit at a time, until the total number of desired bits have been "spent." Each time a variance is decremented the bit value for

that location is incremented. When the process is completed the bitmap which has been generated will produce the minimum error if the data are Gaussian. If the desired average bit rate is B, and the subpicture size is N by N, then this procedure requires BN^2 passes thru N^2 data points. The computation involved grows quite rapidly and for $N > 32$ the cost versus optimality issue must be considered carefully. For this experiment $N = 8$ or 16; therefore, the computational time was not a major factor.

Once the variance matrix and bitmap are obtained, the picture is divided into subpictures of size $N \times N$ and a two dimensional cosine transform is performed on each subpicture. A reordered and scaled version of a 256 by 256 picture which was cosine transformed in 8×8 blocks was shown in Figure 17. The original picture (Figure 32) was a low noise version of the Kodak GIRL (note that she is facing the opposite direction from that usually seen. This is to aid in distinguishing this low noise version). The histograms and other statistical data discussed in Section 3.5 were obtained from this image. The vertical and horizontal correlation were nearly equal and a value for ρ of .96 was used to code this image. A 1 bit/pixel 8×8 bitmap is shown in Figure 33. The coded result for two block sizes is contained in Figure 34.

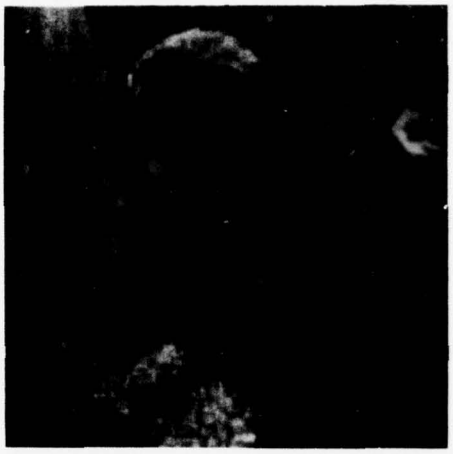
Close inspection of Figure 34 reveals one of the problems with this type of coder. When a subpicture contains a high contrast



Figure 32. Low-noise GIRL (original)

6	4	4	3	3	2	2	2
4	3	2	1	1	1	1	0
4	2	1	1	0	0	0	0
3	1	1	0	0	0	0	0
3	1	0	0	0	0	0	0
2	1	0	0	0	0	0	0
2	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0

Figure 33. A 1 bit/pixel Bitmap (8 x 8 block size, $\rho = .96$)



(a)



(b)

Figure 34. Two 1 bit/pixel Cosine Coded Images
a) 8 x 8 block size, NMSE = .39%
b) 16 x 16 block size, NMSE = .36%

edge, the D.C. value for that block's transform is coded with an error large enough to make the subpicture visible. This type of noise is very annoying to the viewer. Channel errors in the D.C. term also produce the same effect.

6.2. Full Image Cosine Transform Coding

This section covers the special case for N equal to the image width and number of lines, i.e., the subpicture size equals the size of the input image. Again, we will assume first order Markov statistics. Since N = 256 we will not be able to use the optimal Pratt bit assignment algorithm. For this experiment we will use the equation

$$b_{ij} = \left\lceil B + 2\log_{10}\sigma_{ij}^2 - \frac{2}{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\log_{10}\sigma_{k\ell}^2 \right\rceil \qquad (123)$$

where b_{ij} is the $ij^{\underline{th}}$ entry in the bitmap, B is the desired average bit rate, σ_{ij}^2 is the variance of the $ij^{\underline{th}}$ transform coefficient and $\lceil \cdot \rceil$ denotes integer part of. This algorithm is suboptimal due to the rounding operation $\lceil \cdot \rceil$ [59]. The σ_{ij}^2 are obtained from a 256 x 256 variance matrix computed as in the previous section. Because the variances become very small for large ij, we will use the fewest bits for these terms. A typical bitmap is shown in Figure 35. The white area in the upper left represents the maxi-



Figure 35. A 256 x 256 Cosine Domain Bitmap, $\rho = .96$



Figure 36. A 256 x 256 Cosine Coded Image
1 bit/pixel, NMSE = .24%

mum bit assignment which was nine in this case. The black area had zero bits assigned and the intermediate grays varied between one and eight. Two comments are in order. First, the upper left point is the "D.C." cosine coefficient and for this experiment this term was not coded, i.e., the bits allocated to this term were equal to the machine word size (36 bits). This "extravagence" represents an increase of $27/N^2$ or .000412 bits in the average bit rate. In doing this a stability in the coded image mean, which minimizes error and eliminates the need for scaling before viewing, is achieved. The second comment is in regards to the shape of the contours in the bitmap. They are hyperbolic with maximum number of bits assigned to the coefficients on the transform axes. An image coded to 1 bit/pixel in this manner is shown in Figure 36.

6.3. Full Image Fourier Transform Coding

The procedures discussed in the previous section can be implemented, with minor changes, in the Fourier transform domain. The major difference between the cosine and Fourier transforms is that the Fourier is complex. At first glance it would appear that we will double the number of coefficients which must be coded. However, due to the property of conjugate symmetry, which holds for the transform of pure real data (i.e., no imaginary part), this

is not the case. Thus, a 256 x 256 image transforms to a 256 x 129 complex Fourier plane. There are several ways to order the frequency coefficients in this plane. A common arrangement, and the one used in this work, is shown in Figure 37. In this diagram the D.C. term is located in the upper left corner. Frequency increases downward and to the right until the (0, 128) point is reached. The frequency decreases (on the right) from this point until the (0, -1) frequency is reached. The semicircles represent contours of constant radial frequency. Two 256 x 256 block Toeplitz matrices are Fourier transformed and the diagonal vectors are used to generate the desired 256 x 129 variance matrix and bitmap. A typical bitmap is shown in Figure 38. This bitmap readily illustrates the frequency symmetry. Note that the hyperbolic contours are still present.

The bitmap of Figure 38 is not complex. The complex Fourier coefficients are coded by Max quantizing the real and imaginary part of each coefficient to the corresponding rate in the bitmap. Therefore, twice the number of bits allocated to that location in the real bitmap are used and the final average bit rate is $2B/N^2$, where B is the total number of bits in the bitmap and, for the present example, $N = 256$. A picture which has been coded in this manner is shown in Figure 39.

6.4. Block Cosine Coding in the Perceptual Domain

Thus far we have considered the coding of the original image only. We will now consider the coding technique discussed in Section 6.1 as implemented on a preprocessed image. In particular, one processed with the achromatic model of the HVS as shown in Figure 21. The complete process is illustrated in Figure 40. A preprocessed image is shown in Figure 41. For this experiment the first order Markov assumption was still maintained and the Pratt bit assignment algorithm was used. The results for two block sizes and 1 bit/pixel are shown in Figure 42.

6.5. Full Image Cosine Coding in the Perceptual Domain

The full image techniques of Section 6.2 can be applied to the HVS preprocessed image also. The process is the same as that shown in Figure 40. The image shown in Figure 43 is a 1 bit per pixel result. The first order Markov assumption was used for this image and the bitmap was similar to that in Figure 35.

6.6. Full Image Fourier Transform Coding in the Perceptual Domain

In Sections 6.4 and 6.5 we considered the cosine coding of preprocessed imagery. The filtering process shown in Figure 40. is implemented in the Fourier domain, thus coding in the Fourier

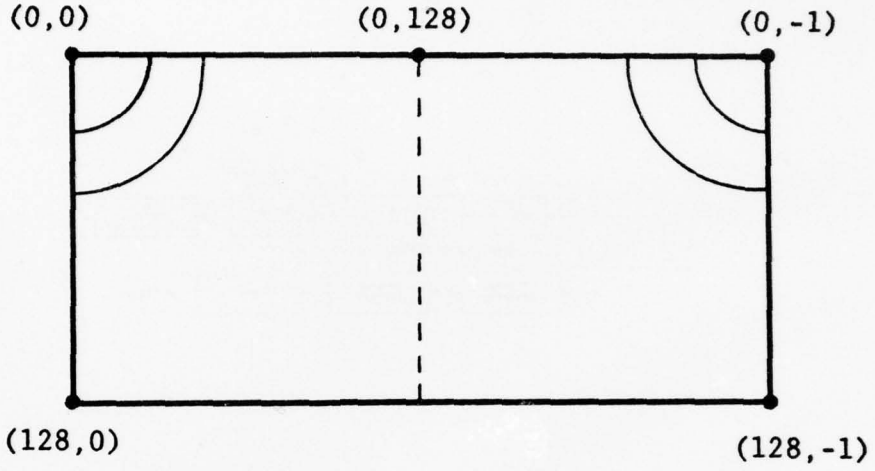


Figure 37. Fourier Domain Frequency Location

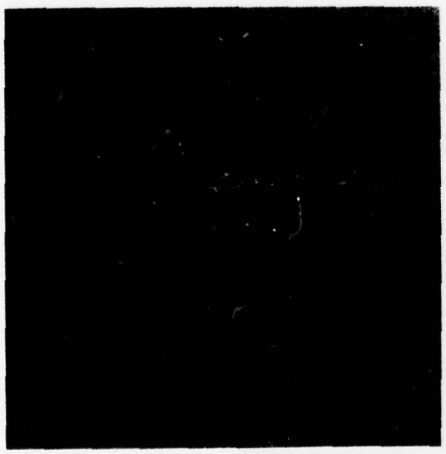


Figure 38. A 256 x 129 Fourier Domain Bitmap

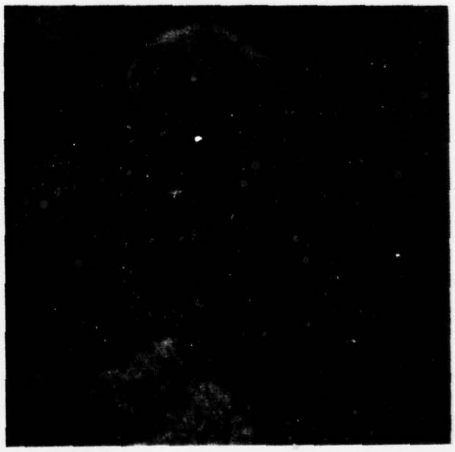


Figure 39. A 256 x 256 Fourier Coded Image 1 bit/pixel, NMSE = .23%

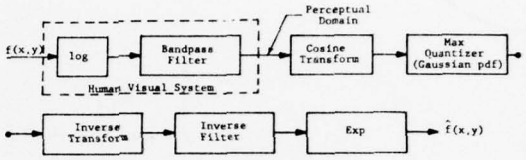


Figure 40. Psychovisual Cosine Coder



Figure 41. A HVS Preprocessed Image



(a)



(b)

Figure 42. Psychovisual Cosine Coded Images, 1 bit/pixel
a) 8 x 8 block size, NMSE = .57%
b) 16 x 16 block size, NMSE = .50%



Figure 43. A 256 x 256 Psychovisual Cosine Coded Image
1 bit/pixel, NMSE = .44%

domain is more expedient. The revised process is diagramed in Figure 44. The techniques for obtaining the variance matrix and bitmaps discussed in Section 6.3 were used to code the HVS pre-processed image. A bitmap similar to that of Figure 38 was obtained. A coded image is shown in Figure 45.

6.7. Perceptual Domain Power Spectrum Coding

The coding techniques discussed in previous sections had two things in common: a variance matrix was computed and first order Markov statistics were assumed. The Markov assumption is reasonable for the original image domain. In Sections 3.5 and 5.1 it was shown that the first order Markov assumption for the input to the achromatic HVS model led to a power spectrum equation of the form

$$S_z(\omega_r) = \left[\frac{2\alpha\sigma_y^2}{\alpha^2 + \omega_r^2} + 2\pi\mu_y^2\delta(\omega_r) \right] \left| H_{bp}(\omega_r) \right|^2 \qquad (124)$$

where $H_{bp}(\omega)$ is defined by equation (93). If we choose to not code the D.C. term then we need only consider

$$S_z(\omega_r) = \left[\frac{2\alpha\sigma_y^2}{\alpha^2 + \omega_r^2} \right] \left| H_{bp}(\omega_r) \right|^2 \qquad (125)$$

This equation defines the power for all $\omega > 0$ and can be used to

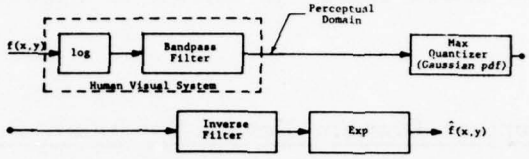


Figure 44. Psychovisual Fourier Coder

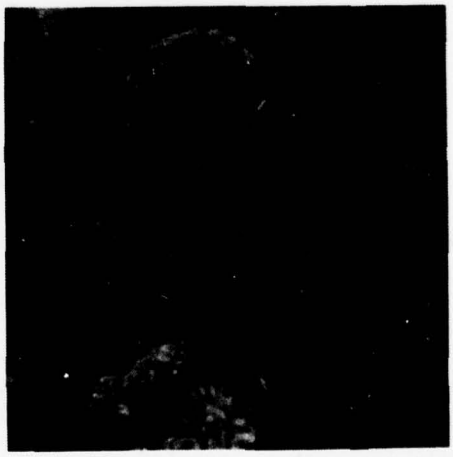


Figure 45. A 256 x 256 Psychovisual Coded Image
1 bit/pixel, NMSE = .26%

determine the variance and bit allocation for any $\omega_r > 0$. Thus, the modification of the Markov statistics which occurs due to the band-pass filter can be taken into consideration. Moreover, the generation of a 256×256 variance matrix is no longer required.

To obtain a bit assignment one merely solves equation (125) for a particular ω_r,

$$\omega_r = \omega_s \sqrt{i^2 + j^2} \tag{126}$$

where i and j are the indices of the Fourier coefficient to be coded and ω_s is the scale factor for conversion to radians/degree. The computed value from equation (125) , call it σ_{ij}^2, is used in

$$b_{ij} = \left\lceil \log_2 \gamma \sigma_{ij}^2 \right\rceil \tag{127}$$

to obtain the bit allocation for the ij$^{\underline{th}}$ coefficient. Equation can be rewritten as

$$b_{ij} = \left\lceil \log_2 \sigma_{ij}^2 + \log_2 \gamma \right\rceil \tag{128}$$

The factor, γ, is selected to yield the desired bit rate. From Figure 24 it can be seen that $\log_{10} \sigma_{ij}^2$ has a maximum value of approximately -2, therefore γ should be about 5×10^4 to obtain 9

bits for b_{ij} maximum. Experimentally, it was found that average rates of .1 to 1 bits/pixel required γ's of 8.9×10^4 to 9×10^5. The Fourier coefficient to be coded is normalized by σ_{ij}^2 and Max quantized to b_{ij} bits for the real part and b_{ij} bits for the imaginary part. Note that no storage other than that for the transformed image is required. A typical bit allocation is shown in Figure 46. The obvious difference between this bitmap and those in Figures 35 and 38 is that the contours are now semicircles of constant radial frequency. This characteristic shape is that of the isotropic filter function $H_{bp}(\omega)$. Thus, the coding technique is taking full advantage of the image filtering provided by $H_{bp}(\omega)$. Several coded images are shown in Figure 47.

Now that we have a closed form expression for variance and bit allocation it is possible to code any size transform we wish. In particular, a 512×512 image (which is analogous to a standard TV image) may be coded. The results are shown in Figure 48. As can be seen from Figure 48 bit rates on the order of 1/10 of that previously achieved can be obtained with this technique and the degradation with decreasing rates is "graceful." A comparison of the coded rates and their associated distortion with the curves in Figure 26 indicates these results are consistent with the rate distortion curves.



Figure 46. Perceptual Power Spectrum Bitmap



Figure 47. Perceptual Power Spectrum Coded Images (N = 256)
Upper left: Original
Upper right: 2 bits/pixel, NMSE = .08%
Lower left: 1 bit/pixel, NMSE = .18%
Lower right: .5 bit/pixel, NMSE = .42%



(a) (b)



(c) (d)

Figure 48. Perceptual Power Spectrum Coded Images (N = 512)
a) Original
b) .5 bit/pixel, NMSE = .28%
c) .35 bit/pixel, NMSE = .50%
d) .1 bit/pixel, NMSE = .72%

SECTION VII

COLOR CODING EXPERIMENTS

This chapter contains the results of several color coding experiments. As in Section VI, the initial experiments involve relatively standard techniques and are mainly for comparative purposes. The last section contains results obtained with the model of the HVS developed in Section II and illustrated in Figure 27.

7.1. Color Coordinate Transformations

It has been shown that transform coding in a color coordinate space, such as the YIQ space, is preferable to coding in RGB space [58]. Indeed, Pratt has considered the color coordinate transformation followed by a spatial transformation of each color plane as a three-dimensional transformation [50]. The optimum coding transformation would be a three-dimensional KLT which would completely decorrelate the $3N^2$ color-image components. The computational complexity involved in such an approach has been discussed previously. However, several color-coordinate conversions provide a large amount of energy compaction and some decorrelation and therefore approach the optimum KL expansion.

We will use four of these conversions; YIQ, Lab, Faugeras (or F-space), and Frei (or G-space). These color spaces were presented in Section 3.4. The color image which will be used for the N = 256 experiments is the Kodak color GIRL. Black and white versions of the various color planes of this image are compared in Figures 49, 50, and 51.

The energy content of the color planes in several coordinate spaces was computed and the results are shown in Table 2. In addition, the correlation between color planes was computed and these results are shown in Table 3. The KL entry in Table 2 is from Pratt [58]. From Table 2 we see that the HVS model which was developed in Section II and is approximated by the Frei model maximizes the energy compaction. The difference between the cube root and logarithmic nonlinearities is minimal. For the case of correlation, obviously KL is the best. As to which of the others is second best is questionable. The YIQ conversion is much lower between planes 1 and 2, however, the correlation between planes 1 and 3 is higher than Lab or G_{cube}. Although the correlation between planes 2 and 3 is lower for YIQ, this is considered to be of secondary importance since the energy compaction indicates the bulk of the bits to be used in coding should be allocated to plane 1. It should also be pointed out that the data of Tables 2 and 3 were obtained without any spatial filtering.



(a) L



(b) Y



(c) F_1



(d) G_1

Figure 49. The 256 x 256 Color GIRL Image (Luminance Planes)



(a) a



(b) I



(c) F_2



(d) G_2

Figure 50. The 256 x 256 Color GIRL Image (First Chrominance Planes)



(a) b



(b) Q



(c) F_3



(d) G_3

Figure 51. The 256 x 256 Color GIRL Image (Second Chrominance Planes)

TABLE 2

ENERGY PACKING

System \ Coordinates	1	2	3
RGB	48.26%	28.76%	22.98%
Lab	85.38%	9.46%	5.17%
YIQ	88.55%	10.07%	1.38%
Faugeraus	94.00%	5.33%	0.67%
Karhunen-Loeve	94.00%	5.10%	0.90%
Frei	98.42%	1.39%	0.19%
Frei (cube root)	98.64%	1.20%	0.16%

TABLE 3

CORRELATION

Coordinates / System	1 to 2	1 to 3	2 to 3
Karhunen-Loeve	0.000	0.000	0.000
RGB	0.771	0.682	0.913
Faugeraus	-0.480	-0.394	0.812
Frei	0.414	0.338	0.818
Frei (cube root)	0.368	0.270	0.770
Lab	-0.353	-0.270	0.646
YIQ	0.018	-0.315	0.527

Therefore, the tables represent the color coordinate conversion characteristics only.

7.2. Block Cosine Transform Coding

The block transform coding procedure used for color imagery is an extension of the techniques outlined in Section 6.1 for monochrome images. The process is similar to that of Pratt et al. [50] and is as follows:

(1) Model the row and column variance matrices of RGB as first-order Markov processes and compute the variances of the elements of the color coordinate space to be coded.

(2) Spatially transform the color planes with the desired transform, obtaining T_1, T_2, and T_3.

(3) Model the probability density of the "DC" term of T_1 as a Rayleigh density and all other terms as Gaussian densities with variances as computed in step (1).

(4) Distribute the total number of bits between the color planes by a ratio consistent with the energy packing and the optimum .625/.275/.1 ratio for YIQ as determined by Pratt et al. [50].

(5) Assign a number of bits to each transform coefficient according to the Pratt algorithm discussed in Section 6.1.

All of the above steps are straightforward with the possible exception of 4). The ratio .625/.275/.1 for YIQ was determined

through a lengthy experimental process by Pratt, et. al. [50]. This ratio apparently does not change within a class of imagery [50]. Bit assignment based on total energy has been shown to be an effective strategy, therefore, the YIQ ratio was adjusted to .7/.2/.1 for the G and G_{cube} spaces and to .6/.25/.15 for Lab space. The Faugeras space bits were distributed with the YIQ ratio. It is recognized that this somewhat heuristic allocation of bits is questionable, however it was not the intent of the present work to investigate the bit allocation for this type of coding procedure. The YIQ ratio is optimal and an optimal ratio for each of the other spaces could be determined by the lengthy process as outlined in [50]. Although this would only have to be done once for each class of imagery, it is still a serious disadvantage to this type of coding.

The GIRL picture was coded following the above procedure for 8x8 and 16x16 blocks at several bit rates. In this work when we refer to a bit rate for a color image we mean the <u>total</u> average rate per pixel. Thus, 1 bit/pixel for the G_{cube} coded image implies .7 bits/pixel to the G_{c1}-plane, .2 bits/pixel to the G_{c2}-plane, and .1 bits/pixel to the G_{c3}-plane. Figure 52 contains the 1 bit/pixel results for the 16x16 cosine coded YIQ, Lab, and G_{cube} spaces. To aid in judging the comparative quality of the results, the three images are displayed in conjunction with the

Figure 52. Cosine Coded 1 bit/pixel 16 x 16 Blocksize

Upper left: Original
Upper right: YIQ, NMSE: Red=.58%, Green=.98%, Blue=1.69%
Lower left: Lab, NMSE: Red=.49%, Green=.76%, Blue=1.03%
Lower right: G_{cube}, NMSE: Red=.58%, Green=.79%, Blue=1.03%

Figure 53. Cosine Coded and Fourier Coded (1 bit/pixel 256 x 256 Blocksize)

Upper half: Cosine Coded
Left: YIQ, NMSE: Red=.42%, Green=.73%, Blue=1.25%
Right: G_{cube}, NMSE: Red=.36%, Green=.52%, Blue=.85%

Lower half: Fourier Coded
Left: YIQ, NMSE: Red=.42%, Green=.76%, Blue=1.59%
Right: G_{cube}, NMSE: Red=.39%, Green=.52%, Blue=.86%



Figure 52. Cosine Coded 1 bit/pixel 16 x 16 Blocksize



Figure 53. Cosine Coded and Fourier Coded (1 bit/pixel 256 x 256 Blocksize)

original. This quadruplet was photographed and processed as an entity and any differences between quadrants is a result of the coding and not the reproduction process.

As can be seen, the coded images contain a large amount of random colored noise. In addition, the blocking errors which were noted in the black and white block coding section are apparent in the color images as well. These blocking errors are accompanied by a large number of very low pixel values (i.e., ~0 on the 0 to 255 scale used for displaying). The source of this noise becomes evident when viewing the black and white triplet of the coded YIQ space as shown in Figure 53. The effect is worse for 8 x 8 blocks than 16 x 16 blocks. A little reflection reveals the problem. For bit rates of ~1 bit/pixel an 8 x 8 block has 64 bits to distribute throughout the 8 x 8 cosine transform domain. Of these 64 bits, 8 to 12 are usually assigned to the DC term (depending on the correlation used in the Markov model). This still leaves enough bits to obtain low quantization errors in the important low frequency and mid-frequency harmonics, as evidenced in Figures 33 and 34. When the average rate is reduced to .1 bit/pixel, as in the Q-plane coding for example, we are left with 6 bits for the entire block! This is not enough bits for the DC term alone. For 16 x 16 blocks the problem is not as acute since we would have 25 bits to distribute, but they would have to be allocated over 256



Figure 54. Monochrome Display of YIQ Space
Cosine Coded 1 bit/pixel
8 x 8 Blocksize

Upper left: Original
Upper right: Y
Lower left: I
Lower right: Q

coefficients. However, the higher harmonics grow less and less important and zero bits assigned to these coefficients have little affect on the coded image; thus, as the subpicture size grows the problem becomes less significant.

7.3. Full Image Cosine Transform Coding

In the previous section it was noted that small block sizes place a large burden on color coding because of the low number of bits assigned to the chrominance planes. The best results we could hope to achieve would be for the case of block size equal to the image dimensions. A .1 bit/pixel allocation in the Q-plane would give 6553 bits to be allocated. If we allocate 36 bits to the single D.C. term only $\frac{1}{2}\%$ of the total bits have been used on DC and this gives no DC error. In the previous section, for 8×8 blocks, even if we allocated all bits to DC components we would have a minimum error of .23% in the coded DC terms.

A 256×256 block size was used to code the various color planes as discussed in Section 7.2. The large block size was the only variation in the coding procedure. As expected, the results were better than for 8×8 or 16×16 blocks. A large amount of random colored noise was still present, however, the noise associated with subpicture size was not present. This is evident in the two coded images shown in the upper half of Figure 54

7.4. Full Image Fourier Transform Coding

A full image Fourier coder was implemented as a step toward the psychovisual coding to be discussed in Section 7.5. No significant difference in the coding results was anticipated since the Fourier and cosine transform both approach the optimum KL in energy packing for N = 256. Indeed, the black and white results of Section 6.3 revealed no significant improvements over that of Section 6.2. The total bit allocation between planes which was specified in Section 7.2 was used. The major variation in the coding procedure was brought about by the complex Fourier plane and the symmetries which exist. The method used to assign bits within a plane and to quantize the complex coefficients was that of Section 6.3. Two coded images are shown in the lower half of Figure 54. As anticipated, no significant improvement over the cosine coder was noted. The slight differences which may be seen between the two halves of Figure 54 are due to the different color spaces and inter-plane bit assignments rather than intra-plane bit assignment and type of spatial transform.

7.5. Perceptual Domain Power Spectrum Coding

In the previous sections one important problem which was common to all of the coding accomplished was that of inter-plane bit assignment. This problem can be handled easily through an

extension of the black and white power spectrum coding discussed in Section 6.7. In that section bits were assigned by equation (127) which contained a factor, γ, used to vary the bit rate. For the color coding case we merely select γ for the desired <u>total</u> rate and keep it constant for the coding of all three color planes. Thus, the percentage of bits assigned between planes is determined by the color power spectrum equations of Section 5.2.

From equations (95), (110), (117), (118), and (119) we have

$$S_{\ell}(\omega_r) = 21.5^2 \left[\frac{2\alpha_1 \sigma_1^2}{\alpha_1^2 + \omega_r^2} \right] \left| H_{\ell}(\omega_r) \right|^2 \tag{129}$$

$$S_{c1}(\omega_r) = 41^2 \left[\frac{2\alpha_1 \sigma_1^2}{\alpha_1^2 + \omega_r^2} + \frac{2\alpha_2 \sigma_3^2}{\alpha_2^2 + \omega_r^2} \right] \left| H_{c1}(\omega_r) \right|^2 \tag{130}$$

and

$$S_{c2}(\omega_r) = 6.27^2 \left[\frac{2\alpha_1 \sigma_1^2}{\alpha_1^2 + \omega_r^2} + \frac{2\alpha_3 \sigma_3^2}{\alpha_3^2 + \omega_r^2} \right] \left| H_{c2}(\omega_r) \right|^2 \tag{131}$$

Now for the imagery which has been used during this research $\alpha_1 \approx \alpha_2 \approx \alpha_3$, thus,

$$S_{\ell}(\omega_r) \approx 21.5^2 \left[\frac{2\alpha \sigma_1^2}{\alpha^2 + \omega_r^2} \right] \left| H_{\ell}(\omega_r) \right|^2 \tag{132}$$

$$S_{c1}(\omega_r) \approx 41^2\left[\frac{2\alpha}{\alpha^2+\omega^2}\right][\sigma_1^2+\sigma_2^2]\left|H_{c1}(\omega_r)\right|^2 \tag{133}$$

and

$$S_{c2}(\omega_r) \approx 6.27^2\left[\frac{2\alpha}{\alpha^2+\omega^2}\right][\sigma_1^2+\sigma_3^2]\left|H_{c2}(\omega_r)\right|^2 \tag{134}$$

But σ_1^2 is σ_ℓ^2 and furthermore, for uncorrelated color planes $\sigma_1^2+\sigma_2^2 = \sigma_{c1}^2$ and $\sigma_1^2+\sigma_3^2 = \sigma_{c2}^2$. Thus,

$$S_\ell(\omega_r) \approx 21.5^2\left[\frac{2\alpha\,\sigma_\ell^2}{\alpha^2+\omega_r^2}\right]\left|H_\ell(\omega_r)\right|^2 \tag{135}$$

$$S_{c1}(\omega_r) \approx 41^2\left[\frac{2\alpha\,\sigma_{c1}^2}{\alpha^2+\omega_r^2}\right]\left|H_{c1}(\omega_r)\right|^2 \tag{136}$$

and

$$S_{c2}(\omega_r) \approx 6.27^2\left[\frac{2\alpha\sigma_{c2}^2}{\alpha^2+\omega_r^2}\right]\left|H_{c2}(\omega_r)\right|^2 \tag{137}$$

The astute reader will have noticed that the delta function in equation (95) has been dropped. This is justified by again, as in Section 6.7, by not coding the $\omega_r = 0$ terms.

The method used to code in Section 6.7 is extended to color

simply by using the appropriate equation for the G-plane being coded and assigning bits by equation (127) with γ constant for all three planes. It should be noted that this process is only valid for G-space since the power spectrum equations were developed for the parameters in the G-space conversion. Coding in another color coordinate system would require changes in equations (135) thru (137)

The bit assignment equation was solved for $\gamma = 4 \times 10^3$ and variances as determined by equations (135) thru (137). The bit distribution between planes was 1.3 bits/pixel for G_1, .62 bits/pixel for G_2 and .01 bits/pixel for G_3. The GIRL picture was coded, with these computed variances and bit assignments, in the perceptual domain. The resultant 256 x 256 image was viewed side by side with the original on a Comtal display. It was extremely difficult to tell them apart. Some viewers had to have the minor difference pointed out. The difference consisted of a slight low spatial frequency tinge. This artifact was thought to be a result of the extremely low bit rate in the G_3 plane. The G_3 plane was coded to .09 bits/pixel and used with the previous G_1 and G_2 coded planes to obtain a color image with an overall bit rate of 2 bits/pixel. The coded image was virtually indistinguishable from the original. Several other bit rates (i.e., different γ's) were used. Three of the resultant color images along with the original are

shown in Figure 55. The results represent bandwidth compressions of 12-1 to 45-1.

Just as for monochrome images, we would expect an improvement in this performance by increasing N to 512. This was most certainly the case. The color Kodak GIRL was not available in a 512x512 scan so another image of the same class was selected. The original of this image, ANN, is shown in Figure 56. This image was selected for two reasons, first the fine detail in the design on the sweater would test the resolution capabilities and second the large amount of pure white in the collar of the blouse should bring out any random color noise. The image was coded following the method detailed earlier with $\gamma = 2 \times 10^4$. The bit distribution was .75 bits/pixel for G_1, .22 bits/pixel for G_2, and .03 bits per pixel for G_3. The quality of this coded image was so high that an experienced observer mistook it for the original when viewing the image on the Comtal display. The NMSE for this image was red = .13, green = .14, and blue = .38. In order to obtain an image which was degraded enough to be apparent after reproduction several lower rates were coded. The images shown in the lower half of Figure 56 were coded at .5 and .25 bits/pixel.

To establish the utility of the coding technique, five more 512x512 color images were coded. These images represented a

Figure 55. Perceptual Power Spectrum Coded (N = 256)

Upper left: Original
Upper right: 2 bits/pixel, NMSE: Red=.10%, Green=.18%, Blue=.70%
Lower left: 1 bit/pixel, NMSE: Red=.20%, Green=.33%, Blue=.84%
Lower right: .5 bit/pixel, NMSE: Red=.43%, Green=.66%, Blue=1.1%

Figure 56. Perceptual Power Spectrum Coded ANN Image (N = 512)

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=.13%, Green=.14%, Blue=.38%
Lower left: .5 bit/pixel, NMSE: Red=.18%, Green=.19%, Blue=.45%
Lower right: .25 bit/pixel, NMSE: Red=.26%, Green=.27%, Blue=.56%



Figure 55. Perceptual Power Spectrum Coded (N = 256)



Figure 56. Perceptual Power Spectrum Coded ANN Image (N = 512)

wide variation in subject content. They were all coded at 1, .5, and .25 bits/pixel. The originals and coded results are shown in Figures 57 thru 61. The 1 and .5 bits/pixel versions of these images were all coded with the same correlation and variance parameters. They were computed from the ANN image. The .25 bits/pixel images were coded with the same correlation parameters and bitmap; however, the normalization prior to Max quantizing was performed with the spatial variance for the respective image.

A question of considerable interest is, where are the coding errors manifested within the reconstructed image? This question may be answered by computing a difference image. If one subtracts a coded image from the original image and scales and displays the result; the areas of maximum error become readily visible. Three such images and the original are shown in Figure 62. The colors in the difference image represent the errors in the red, green, and blue planes. As in the achromatic case, the chromatic coding results compare favorably with the predicted performance (see Figure 31).

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=.60%, Green=.58%, Blue=1.1%
Lower left: .5 bit/pixel, NMSE: Red=1.1%, Green=1.0%, Blue=1.4%
Lower right: .25 bit/pixel, NMSE: Red=1.3%, Green=1.6%, Blue=1.9%

Figure 57. Perceptual Power Spectrum Coded LAKE Image (N = 512)

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=.17%, Green=.19%, Blue=.36%
Lower left: .5 bit/pixel, NMSE: Red=.24%, Green=.30%, Blue=.38%
Lower right: .25 bit/pixel, NMSE: Red=.46%, Green=.57%, Blue=.47%

Figure 58. Perceptual Power Spectrum Coded F16 Image (N = 512)



Figure 57. Perceptual Power Spectrum Coded LAKE Image (N = 512)



Figure 58. Perceptual Power Spectrum Coded F16 Image (N = 512)

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=.57%, Green=.75%, Blue=1.2%
Lower left: .5 bit/pixel, NMSE: Red=1.0%, Green=1.4%, Blue=1.4%
Lower right: .25 bit/pixel, NMSE: Red=1.8%, Green=2.2%, Blue=2.0%

Figure 59. Perceptual Power Spectrum Coded BUILDING Image (N=512)

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=1.5%, Green=2.0%, Blue=3.2%
Lower left: .5 bit/pixel, NMSE: Red=1.9%, Green=2.7%, Blue=3.7%
Lower right: .25 bit/pixel, NMSE: Red=2.3%, Green=3.4%, Blue=4.3%

Figure 60. Perceptual Power Spectrum Coded BABOON Image (N=512)



Figure 59. Perceptual Power Spectrum Coded BUILDING Image (N = 512)



Figure 60. Perceptual Power Spectrum Coded BABOON Image (N = 512)

Upper left: Original
Upper right: 1 bit/pixel, NMSE: Red=.55%, Green=.44%, Blue=1.8%
Lower left: .5 bit/pixel, NMSE: Red=1.1%, Green=.77%, Blue=1.9%
Lower right: .25 bit/pixel, NMSE: Red=.90%, Green=1.1%, Blue=2.3%

Figure 61. Perceptual Power Spectrum Coded PEPPERS Image (N=512)

Figure 62. Difference Images from ANN Coding Results (See Figure 56)

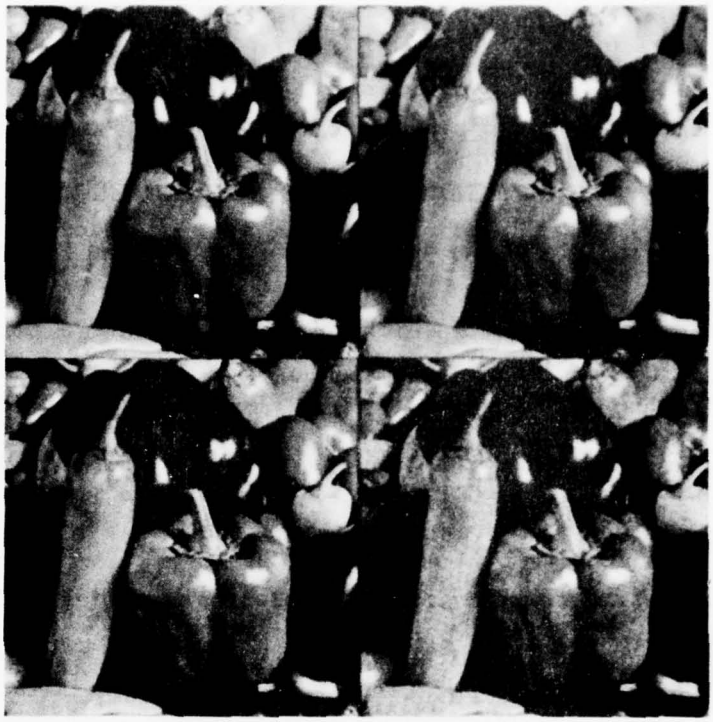


Figure 61. Perceptual Power Spectrum Coded PEPPERS Image (N = 512)



Figure 62. Difference Images from ANN Coding Results (See Figure 56)

SECTION VIII

IMAGE QUALITY MEASURES

A major problem which has plagued image processing has been the lack of an image quality measure which matches human subjective evaluation. Although several measures have been proposed and used, they usually suffer from one of two defects. They are either analytically non-tractable or they perform poorly against subjective evaluations. The next section contains a discussion of several quality measures which have been used. In Section 8.2 an image quality measure based on our model of the HVS is presented. A psychophysical paradigm, which was used to obtain subjective evaluations of two data bases (one monochrome and one color), will be described in Section 8.3. Then, in the last section, several image quality measures are compared to the subjective evaluation of the data bases.

8.1. Standard Image Quality Measures

One of the most commonly used quality or distortion measures is mean square error (MSE). For the case of an $N \times N$ discrete image, MSE may be defined as

$$\text{MSE} = \sum_{m=1}^{N} \sum_{n=1}^{N} [f(m,n) - \hat{f}(m,n)]^2 \qquad (138)$$

This particular distortion measure is attractive because it is tractable and a solution to the parametric rate distortion equations can be found for it. Unfortunately, MSE does not match human evaluation on many types of imagery. It is also possible to define a measure based on MSE and energy normalization [27]. We will call this measure normalized mean square error (NMSE), and for an N x N image,

$$\text{NMSE} = \frac{\sum_{m=1}^{N} \sum_{n=1}^{N} [f(m,n) - \hat{f}(m,n)]^2}{\sum_{m=1}^{N} \sum_{n=1}^{N} [f(m,n)]^2} \qquad (139)$$

Normalized mean square error performs somewhat better than MSE. It retains the analytic tractability and is easy to compute. For these reasons it has gained acceptance in some circles and therefore it has been used throughout the earlier chapters of this dissertation. It should be noted that NMSE can also be defined in the Fourier domain (FMSE) as,

$$FMSE = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} |F(m,n) - \hat{F}(m,n)|^2}{\sum_{m=1}^{M} \sum_{n=1}^{N} |F(m,n)|^2} \tag{140}$$

where N is the width and length of the original image and $M = N/2+1$ (recall the complex conjugate symmetry of the Fourier domain).

Another common measure is the normalized difference or normalized error (NE),

$$NE = \frac{\sum_{m=1}^{N} \sum_{n=1}^{N} |f(m,n) - \hat{f}(m,n)|}{\sum_{m=1}^{N} \sum_{n=1}^{N} |f(m,n)|} \tag{141}$$

This measure is particularly appealing because of its simplicity. The measure performs well for low intensity levels since incremental changes at low intensities are more noticeable than those at high intensities [29, p. 138]. The NE measure is not as easily manipulated as NMSE and for this reason it is not as popular as the latter.

Many attempts at defining image quality measures are based on some known property of the human visual system. One such measure, Laplacian mean square error (LMSE), is based upon the importance of edges to the human observer. This measure is

defined as [27],

$$\text{LMSE} = \frac{\sum_{m=2}^{N-1}\sum_{n=2}^{N-1}\left[G(m,n) - \hat{G}(m,n)\right]^2}{\sum_{m=2}^{N-1}\sum_{n=2}^{N-1}\left[G(m,n)\right]^2} \tag{142}$$

where

$$G(m,n) = f(m+1,n) + f(n-1,n) + f(m,n+1) + f(m,n-1) - 4f(m,n) \tag{143}$$

LMSE performs well for images which have been severely low-pass filtered. However, it is possible to generate severely degraded images with low spatial frequency noise which are measured "good" quality by LMSE. From equations (142) and (143) it can be seen that LMSE is not very tractable.

A similar measure can be obtained by retaining equation (142) and changing equation (143) to

$$\begin{aligned} G_s(m,n) = {} & \left|f(m+1,n-1) + 2f(m+1,n) + f(m+1,n+1)\right. \\ & \left. - f(m-1,n-1) - 2f(m-1,n) - f(m-1,n+1)\right| \\ & + \left|f(m-1,n+1) + 2f(m,n+1) + f(m+1,n+1)\right. \\ & \left. - f(m-1,n-1) - 2f(m,n-1) - f(m+1,n-1)\right| \end{aligned} \tag{144}$$

When G(m, n) in equation (142) is replaced by $G_s(m, n)$ we have a form of estimated gradient mean square error (GMSE). The function $G_s(m, n)$ is a Sobel operator defined on a 3 x 3 grid [60, pp. 271-272]. This measure produces some formidable analytic problems.

8.2. A Perceptual Image Quality Measure

The observant reader has no doubt noticed that GMSE and LMSE are simply NMSE computed in a transformed space. This approach to obtaining image quality measures is widely used since the actual distortion measure is based on mean square error and one merely selects an appropriate preprocessor. What better preprocessor could be selected than the HVS model we have developed? For the achromatic model depicted in Figure 21 we can define an achromatic perceptual mean square error ($PMSE_a$) as

$$PMSE_a = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[z(m,n) - \hat{z}(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[z(m,n)]^2} \qquad (145)$$

where z(m, n) and $\hat{z}(m, n)$ are given by

$$z(m,n) = \ell n[x(m,n)] \circledast h_{bp}(m,n)$$

and

$$\hat{z}(m,n) = \ell n[\hat{x}(m,n)] \circledast h_{bp}(m,n) \tag{146}$$

The function $h_{bp}(m,n)$ is simply the rectangular coordinate form of the point spread function equivalent of equation (93). This error criterion can also be defined in the Fourier domain and in this case we have

$$FPMSE_a = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[Z(m,n) - \hat{Z}(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[Z(m,n)]^2} \tag{147}$$

where

$$Z(m,n) = \mathcal{F}\{\ell n[x(m,n)]\} \cdot H_{bp}(m,n)$$

$$\hat{Z}(m,n) = \mathcal{F}\{\ell n[\hat{x}(m,n)]\} \cdot H_{bp}(m,n) \tag{148}$$

It should be evident that equations (145) and (147) are equivalent and therefore we will use the term achromatic perceptual mean square error for either case.

In a similar fashion, it is possible to define a chromatic perceptual mean square error ($PMSE_c$). In this case we simply compute the NMSE in the ℓ, c_1, and c_2 planes (see Figure 27). Hence

$$PMSE_c = NMSE_{\ell} + NMSE_{c1} + NMSE_{c2} \tag{149}$$

where

$$NMSE_{\ell} = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[\ell(m,n) - \hat{\ell}(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[\ell(m,n)]^2}$$

$$NMSE_{c1} = \frac{\sum_{m=1}^{N}\sum_{m=1}^{N}[c_1(m,n) - \hat{c}_1(m,n)]^2}{\sum_{m=1}^{N}\sum_{m=1}^{N}[c_1(m,n)]^2}$$

$$NMSE_{c2} = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[c_2(m,n) - \hat{c}_2(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[c_2(m,n)]^2} \tag{150}$$

The three expressions in equation (150) also have Fourier domain equivalents. Thus, the color counterparts of equation (148) become

$$L(m,n) = 21.5\,\mathcal{F}\{\ell n[t_1(m,n)]\}\,H_{\ell}(m,n) \tag{151}$$

$$C_1(m,n) = 41\,\mathcal{F}\{\ell n[t_2(m,n)/t_1(m,n)]\}\,H_{c1}(m,n) \tag{152}$$

and

$$C_2(m,n) = 6.27\,\mathcal{J}\{\ell n[t_3(m,n)/t_1(m,n)]\}\,H_{c2}(m,n) \qquad (153)$$

Of course, the coded versions $\hat{L}(m,n)$, $\hat{C}_1(m,n)$, and $\hat{C}_2(m,n)$ are similarly defined. The quantities $t_1(m,n)$, $t_2(m,n)$, and $t_3(m,n)$ are obtained from the RGB to T-space conversion defined by equation (23)

8.3. An Achromatic Subjective Image Quality Experiment

The ultimate image quality measure is a subjective evaluation. This type of measure is not without "pitfalls." Indeed, what is an important difference to one observer may go unnoticed by another. A reliable experimental result requires a large number of subjects of a "judicious" mix. They should be selected with the overall objective in mind. For example, to evaluate normal image viewing quality the observers should have a wide mix of background and experience. On the other hand, if one is developing a specialized image measure such as a texture measure, the observers should probably be familiar with this area. Since we are concerned with viewing quality, we will attempt to use unbiased observers with various backgrounds.

Another problem which is encountered is that of selecting the actual evaluation procedure. There are two general types of

subjective evaluation. In one, absolute evaluation, observers are shown an image and asked to rate it according to some pre-defined scale. The other, comparative evaluation, simply requires the observer to rank a set of images from best to worse. Extensive work has been done in the methods for scaling images, or the absolute evaluation method, particularly in evaluating television quality [61]. The rank ordering type of evaluation is more suitable to digital image processing and it is a fairly quick and easy test to perform. An additional favorable aspect is that it requires no training or familiarization tasks. The observers can be completely "naive" in this respect.

A convenient implementation of the comparative evaluation involves a type of "bubble sort." This method requires the observer to make a forced choice between two images. The chosen or best image is always retained for the next comparison until the set of images has been exhausted. The remaining image is ranked one and removed from the set. The procedure is repeated to find the second ranked image, etc., down to the $N^{\underline{th}}$ ranked image. This particular protocol has been used successfully by Mannos and Sakrison [7] and it is the evaluation technique selected for our experiments.

The monochrome data set was obtained by coding the 256 x 256 low noise scan of the GIRL picture. The image was coded to 2,

1.0, .75, and .5 bits/pixel with an 8x8 and a 16x16 block cosine coder. In addition, it was coded to 1, .5, .25, and .1 bits/pixel with the perceptual power spectrum coder. The twelve images were stored in digital form on a high speed disk. The images were displayed in pairs, diagonally opposite (i.e., quadrants 2 and 4), on a Comtal monitor. A PDP-11/40 was used to control the disk and accomplish transfer between the disk and monitor. The images were transferred sequentially to either quadrant. This enabled the rejected image (the worst image of a pair) to be replaced by the next image in the set. With this arrangement 66 pairings were required to order the entire set of 12 images.

The observer was seated in front of the monitor at a distance which gave a 6° viewing angle for a 256x256 image. This distance was selected to be consistent with the scaling factor which was used in the filter equations of the HVS processed images. The lighting within the viewing room was subdued and the average brightness of the display was approximately 15mL or 48 cd/m^2.

After all of the individual rankings were obtained an overall ranking for the group of observers was obtained by a weighted average. This average was defined by

$$R_i = \frac{1}{M} \sum_{j=1}^{M} n_{ij} \tag{154}$$

where M is the number of trials, R_i is the weighted average rank of the $i^{\underline{th}}$ image, and n_{ij} is the rank assigned to the $i^{\underline{th}}$ image during the $j^{\underline{th}}$ trial. Table 4 contains the final rankings for the achromatic data set.

8.4. A Chromatic Subjective Image Quality Experiment

The general methods outlined in the previous section were used to subjectively evaluate a set ten color images. In this case 30 256x256 image files were required. The color data set contained only ten images since the high speed disk could not store 12 color images. Since three times as much data was required for a complete image, the time required to display an image and the total time required to evaluate the entire data set was lengthened considerably.

The color image used for coding was the 256x256 color GIRL. The image was coded in the YIQ and Lab spaces with a block cosine coder. For the YIQ conversion, a 16x16 blocksize and rates of 2, 1, and .5 bits/pixel were used during coding. An 8x8 blocksize coder at the same rates was used to code the Lab-space. The image was also coded with the perceptual power spectrum coder at rates of 2, 1, .5, and .25 bits/pixel.

The above set of ten color images was subjectively evaluated by observers and the results were averaged by using equation

TABLE 4

SUBJECTIVE RANKINGS FOR ACHROMATIC IMAGE SET

Image	Blocksize	Rate (bits)	Subjective Rank
1	16	2	3
2	16	1	5
3	16	.75	7
4	16	.50	10
5	8	2	1
6	8	1	6
7	8	.75	8
8	8	.50	11
9	256	1	2
10	256	.50	4
11	256	.25	9
12	256	.10	12

(154). These average rankings are tabulated in Table 5.

8.5. Computed Image Quality Experiement

The two image data sets generated for the experiments of Sections 8.3 and 8.4 were ranked by the error between each of them and the original image. Minimum error was ranked one, second smallest two etc., until the largest error was ranked twelfth. The monochromatic error computations were performed with the equations defining $PMSE_a$ -- equations (145) and (146). For the color images the equations

$$NMSE_C = NMSE_R + NMSE_G + NMSE_B \qquad (155)$$

and

$$LMSE_C = LMSE_R + LMSE_G + LMSE_B \qquad (156)$$

along with equation (149) for $PMSE_c$ were used to rank the data set.

The results of these computations are shown in Tables 6 and 7. The subjective ranks have been included for comparative purposes. From Table 6, the correlation between PMSE and the subjective ranking of the achromatic image set is higher than that of NMSE and LMSE. For a data set size of 12 the confidence level of the correlations is greater than 99.9%. Thus, the PMSE is definitely the better distortion measure for this data set. It should

TABLE 5

SUBJECTIVE RANKINGS FOR CHROMATIC IMAGE SET

Image	Blocksize	Color Space	Rate (bits)	Subjective Rank
1	256	Gcube	2.00	1
2	256	Gcube	1.00	2
3	256	Gcube	.50	6
4	256	Gcube	.25	9
5	16	YIQ	2.00	4
6	16	YIQ	1.00	5
7	16	YIQ	.50	10
8	8	Lab	2.00	3
9	8	Lab	1.00	7
10	8	Lab	.50	8

TABLE 6

COMPUTED AND SUBJECTIVE RANKINGS FOR
ACHROMATIC IMAGE SET

Image	NMSE (%)	LMSE (%)	PMSE (%)	Subjective Rank (S.R.)
1	.27	60	3.2	3
2	.43	81	5.4	5
3	.51	89	6.1	7
4	.67	100	7.8	10
5	.28	64	3.0	1
6	.43	89	4.6	6
7	.57	100	6.1	8
8	.83	133	8.6	11
9	.26	75	1.2	2
10	.42	85	2.5	4
11	.73	93	5.0	9
12	1.55	99	9.1	12
Correlation to S.R.	.85	.84	.92	

be noted that this test was a severe one in the sense that three types of noise were contained in the images; Gaussian, 8 x 8 blocking errors, and 16 x 16 blocking errors. To the author's knowledge, comparable tests have not been performed. Previous subjective tests have dealt with a single type of noise (usually Gaussian).

The correlation results from the chromatic experiment are not as clear-cut as the achromatic case. The NMSE correlation to subjective rank is slightly higher than PMSE correlation to subjective rank (see Table 7). LMSE is definitely inferior to NMSE and PMSE. Four types of correlation were computed on the chromatic data set. The first was the standard correlation coefficient defined as the covariance divided by the product of the standard deviations,

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} \tag{157}$$

where x was the vector of actual errors measured and y was the subjective rank vector. Ranks were also assigned to the images according to minimum error under each measure. Equation (157) was then used to compute the correlations between rank orderings. The last two measures have been specifically developed for "ranked" data. The Spearman rank correlation coefficient is defined as

TABLE 7

COMPUTED AND SUBJECTIVE RANKINGS FOR CHROMATIC IMAGE SET

Image	NMSE		LMSE		PMSE		Subjective Rank (S. R.)
	Percent	Rank	Percent	Rank	Percent	Rank	
1	1.00	1	211	1	11.02	1	1
2	1.37	2	250	2	13.64	2	2
3	2.19	5	279	3	19.85	4	6
4	3.82	9	292	5	29.44	9	9
5	2.13	4	302	6	20.40	5	4
6	3.32	7	403	8	27.45	7	5
7	5.26	10	562	10	40.37	10	10
8	1.67	3	280	4	17.08	3	3
9	2.62	6	342	7	21.93	6	7
10	3.74	8	410	9	27.93	8	8
Correlation to S. R.	.92	.96	.74	.76	.90	.94	
Spearman Rank Correlation		.96		.76		.94	
Kendall Tau Coefficient		.91		.73		.87	

$$r_s = 1 - \frac{6 \sum_i D_i^2}{N(N^2 - 1)} \tag{158}$$

where the D_i are pairwise differences and N is the number of ranked images [62, pp. 245-249]. The Kendall tau coefficient or τ statistic is defined as [62, pp. 249-252]

$$\tau_K = \frac{\text{(number of agreements) - (number of disagreements)}}{\text{total number of pairs}} \tag{159}$$

The four types of correlation were computed on the chromatic results and they are tabulated in Table 7. It should be noted that the chromatic experiment was even more difficult than the achromatic experiment since three different color spaces were used in obtaining the image set.

SECTION IX

SUMMARY AND CONCLUSIONS

The primary thesis of this research was that models suitable for digital image processing -- and in particular color image bandwidth compression -- could be developed from the basic characteristics of the human visual system. This hypothesis has been demonstrated and the theoretical and practical implications are summarized in the next section. The conclusions which can be drawn from the results of this work are also discussed in Section 9.1. In the last section several recommended areas for future research are pointed out.

9.1 Summary and Conclusions

It has been demonstrated that simple mathematical models can be developed from the physiological and psychophysical traits of the HVS. These models were shown to be analytically tractable and expressions for their statistical characterization were obtained. Given a standard image model, the output power spectrum of an achromatic and a chromatic model were derived. These power spectrum expressions were used to code black and white and color imagery down to rates lower than that previously achieved. In

addition, the outputs of the models were shown to be statistically compatible with the basic assumptions necessary to obtain a solution for the parametric rate distortion equations. Those equations were solved and rate versus distortion curves which demonstrate the near optimality of the coding algorithm were presented.

The utility of these models as a preprocessor for image quality measurements was also demonstrated. It was shown that normalized mean square error is an effective distortion measure when applied to the preprocessed images. The combination of NMSE and the HVS preprocessor was referred to as perceptual mean square error (PMSE). A subjective evaluation of twelve monochrome and ten color images indicated that PMSE is a valid image quality measure.

One can conclude from this work that what has been conjectured previously is true. The HVS can be used to develop very effective preprocessors for image systems. Moreover, with a few simplifying assumptions, these models can be analyzed and efficient algorithms for image bandwidth compression and quality measurement can be obtained.

9.2 Recommended Future Work

Several areas which may be fruitful future research topics are apparent. One area of practical importance would be the application

of the techniques used to obtain the power spectrum equations for our HVS model to the YIQ and Lab color-coordinate spaces. If the analogous expressions are obtained, particularly for Lab space, bandwidth compressions similar to those obtained in this work should be possible.

Another area along these lines is to use the power spectrum equations tc code in the cosine domain. It is realized that the cosine domain is not a true frequency plane per se; however, there is reason to believe this approach would prove fruitful. A key ingredient of the successful coding in this dissertation has been the circularly symmetric bitmaps and this symmetry can be produced in any frequency or sequency domain with the appropriate "power spectrum" equation.

The basic algorithm can also be simplified by eliminating the filtering operation. Since the spatial filter is an integral part of the power spectrum expression and bit allocation is determined from this expression, a type of filtering is being performed in the quantization process. Of course, with the cosine transform, this would give a very simple algorithm with definite real time processing capabilities. The next step would be to study the block size problem. It could very well be that a 16 x 16 or 32 x 32 block cosine coder can be implemented with the power spectrum bit assignment technique.

One very obvious area for further work is that of image quality measures. The subjective experiments performed during this research were very superficial. The results were certainly encouraging; however, many more images from several classes need to be processed before any definitive comparisons between PMSE, NMSE, GMSE, LMSE, or any other image quality measurement can be made.

APPENDIX A

SOME PHYSIOLOGICAL PROPERTIES OF THE HUMAN VISUAL SYSTEM

As pointed out in Section I , a primary goal of this research is to obtain quantifiable mathematical models of the HVS which are applicable to image coding. To work toward this end we must have some basic knowledge of the physiology of the HVS. The HVS models of Section II were developed based on the physiological properties discussed in this appendix. Before beginning let us consider what we mean by the human visual system. Throughout this research we will consider the eye, the optic tract, the lateral geniculate bodies and those portions of the striate (or visual) cortex which do not involve cognition to be the HVS.

A horizontal section of a right eye is shown in Figure A. 1. Light enters through the cornea and passes through the anterior chamber to the iris-lens structure. Upon exiting the lens, the light travels through the vitreous humor to the retina where it excites the photoreceptors which in turn convert these visible electromagnetic radiations to a type of frequency modulated signal. This electrical activity is passed via the optic nerve, through the optic chiasm to the lateral geniculate bodies (LGBs). From the LGB,

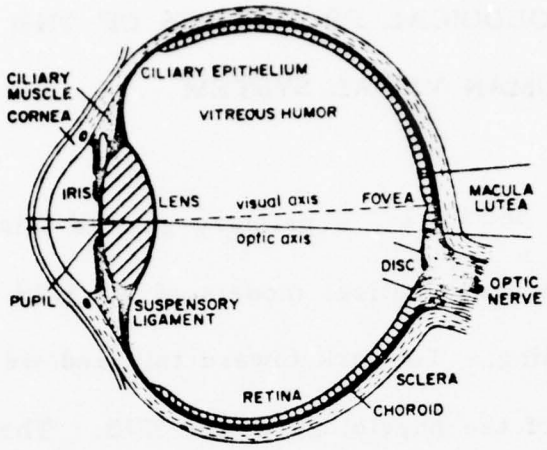


Figure A.1 Horizontal Section of Human Eye

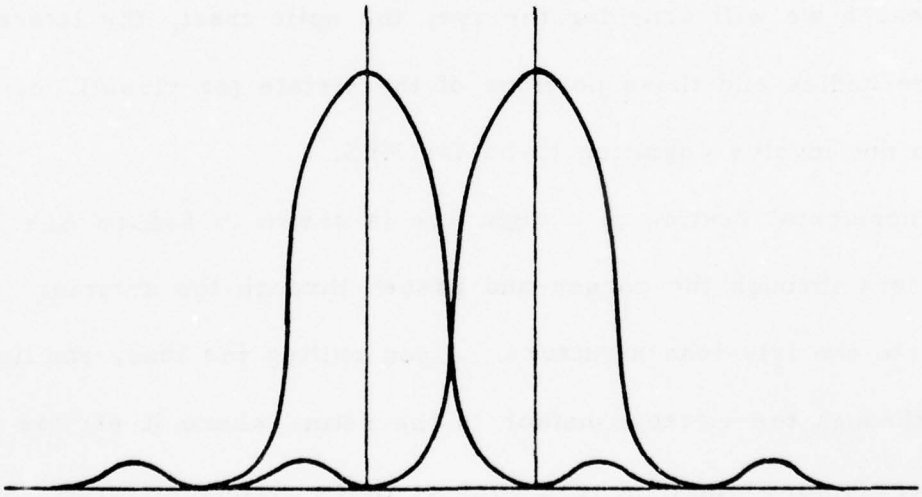


Figure A.2 Rayleigh Limit Diagram

the visual data passes to the occipital lobe region of the brain which contains the visual cortex. Through this pathway the original visual field of view is transmitted and mapped conformally onto area 17 of the striate cortex. Let us now consider Figure A.1 in detail.

A.1. The Ocular Optical System

The outer coat of the eye, the sclera, is protective in function [63]. The sclera (sometimes referred to as the "white" of the eye) is opaque except for the cornea, which is a transparent protuberance centered on the optical axis. The cornea has a refractive index of approximately 1.3771 and the aqueous humor (contained in the anterior chamber) has a refractive index of 1.3374 [22, p. 210]. The air-cornea-aqueous humor interface results in a lens power of 42 diopters which is approximately 2/3 of the total power of the eye. The remaining 1/3 is a result of the "crystalline" lens which has a refractive index of 1.42 [ibid]. Since the refractive index of the vitreous humor is 1.336 the differential index in the aqueous humor-lens-vitreous humor interface is lower than that of the corneal interfaces, hence a lower power. The crystalline lens is the most important element in the lens system however. This is because it is nonrigid and the shape and relative curvature of the two faces can be altered by the ciliary muscles. This action,

called accommodation, insures that the image is brought into focus at the retina, regardless of the distance of the object from the eye. The image which is finally formed on the retina is inverted (an upside down mirror image).

The process described in the previous paragraph produces a focused image on the retina, however, it does not control the intensity of this image. This is accomplished by a circular opening, the pupil, which is formed by the iris. The iris can adjust the diameter of the pupil from 2mm to 8mm (or an area variation of 16 to 1), thus, controlling the amount of light passing from the anterior chamber, through the lens, and into the vitreous chamber. The pigmented epithelium adjacent to the radial and circular muscles of the iris gives the eye its characteristic color (blue, green, or brown). Since aberrations in the dioptric system are the greatest in the periphery of the cornea and lens, pupillary constriction improves the quality of the image formed on the retina. Unfortunately, this action also decreases the resolution of the optical system through diffraction effects.

The resolving capability of any incoherent optical instrument is limited ultimately by the effects of diffraction [64, pp. 129-131]. The Rayleigh criterion of resolution states that two incoherent point sources are "barely resolved" by a diffraction-limited system when the bright central core of one Airy disk falls on the first dark

band of the other. This geometry is shown in Figure A.2 for a one-dimensional case. The minimum resolvable separation of the two point sources becomes

$$\delta = 1.22 \frac{\lambda}{d_0} \qquad (A,1)$$

where λ is the wavelength of the sources and d_0 is the diameter of the image-forming lens (i.e., the pupil diameter).

Riggs [65, pp. 333-334] has shown that visual acuity remains fairly constant as the pupil increases from 2.5mm to 5mm. This result indicates that within this range, the Rayleigh limit and optical aberration effects are balanced. The visual acuity of the total system involves other parameters however. We will revisit this subject in more detail later.

There is one type of aberration in the optical system of the eye which is measurable on axis, chromatic aberration. Since the refractive indexes of the ocular media are wavelength dependent, the optical power of the eye exhibits this dependence. If the image of a distant point source emitting all wavelengths is located on the optic axis and produces a focused image on the retina for a reference wavelength λ_0, then shorter wavelengths will image in front of the retina and longer wavelengths behind the retina. If the reference, λ_0, is set at the peak sensitivity wavelength for color sensitive photoreceptors (~578nm, a yellow), then variation in optic

power ranges from +.62 diopters for 750nm (red) to -2.63 diopters for 365nm (violet) [22, pp. 211-212]. As can be seen, the blue end of the spectrum has the largest variance in lens power. This implies there is more "defocus" for the blue end of the spectrum and hence less resolution.

In the previous paragraphs of this section we have briefly covered the optical system of the eye. This system is linear and, even though it is spatially and temporally variant and inhomogeneous, one can model the system quite accurately [1, p. 162], [66]. In the next element of the ocular system, the retina, we not only encounter complex inhomogeneities and interconnectivity patterns, but nonlinearities as well.

A.2 The Retina

The retina is a multi-layered structure which lines the interior of the rear wall of the eyeball. It extends about 100° on either side of the optic axis. The photoreceptors are located at the very back side of the retina. This means that light must pass through the optic nerve fibers, and other neural layers of the retina, before reaching the photoreceptors. This type of structure is referred to as an inverted eye. Figure A.3 illustrates the basic structure of the retina. The chorioid layer and pigment cells absorb any light which has passed through the entire system and has

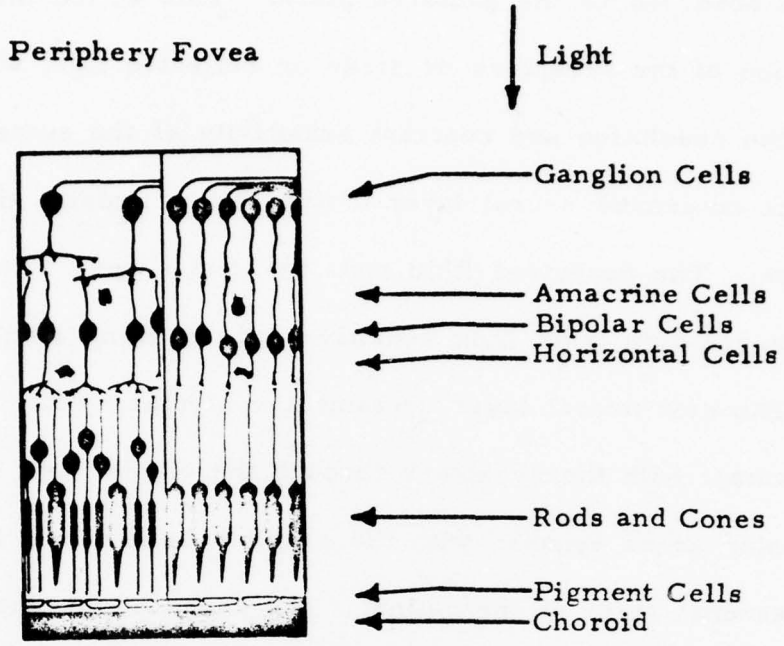


Figure A.3 Structure of the Human Retina

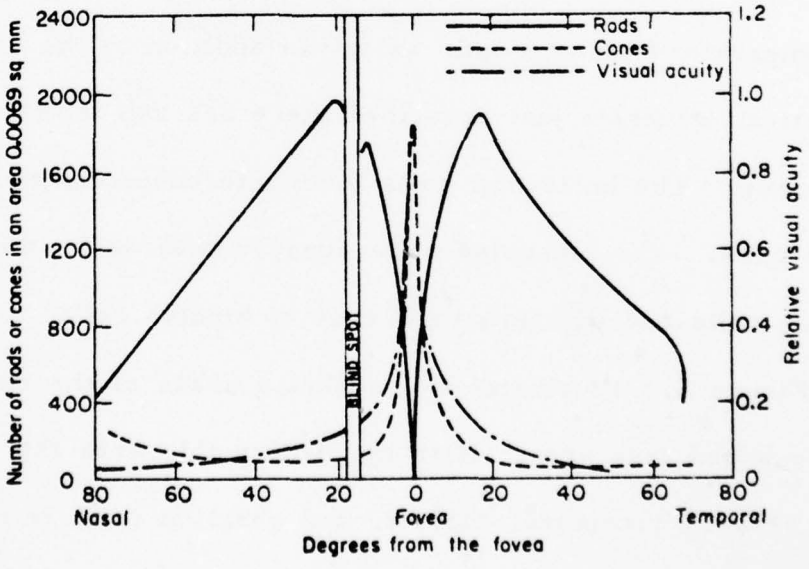


Figure A.4 Distribution of Rods and Cones

not been absorbed by the photoreceptors. This action minimizes stimulation of the receptors by stray or reflected light which would reduce the resolution and contrast sensitivity of the system.

The outermost neural layer of the retina contains the photoreceptors. The receptors (thin rod- or cone-shaped structures) are arranged with their light sensitive ends pointing away from the lens. The next neural layer contains the bipolar cells. These cells make contact with the receptors through the bipolar cell dendrites. The bipolar axons synapse with the ganglion cells which form the inner neuronal layer of the retina. The axons of the ganglion cells are gathered into the optic nerve at the optic disk which is located about 16° nasally from the optic axis. In this area there are no photoreceptors and a "blind-spot" results in the visual field located 16° temporally from the optic axis. In addition to the sequential or vertical structure just described there are two lateral systems of neurons. The horizontal cells form interconnections between receptor cells. The amacrine cells synapse with each other, with ganglion cells and with proximal ends of bipolar cells.

Figure A.3 illustrates two separate areas of the retina. One area is a rod free area. Note that within this area the correspondence between receptors, bipolar, and ganglion cells is one-to-one. The rod free area of the retina is a circular area of 500-600μ in diameter centered on the optic axis. This area is called the fovea

centralis and it subtends 1.7° to 2.0° of the visual field. There are 110,000 to 115,000 cones within this area. A smaller portion, the foveola (the inner 400μ diameter circle) is the most densely packed area and contains about 25,000 receptors. Outside the fovea centralis cone density begins to fall off rapidly and rod density begins to build up. A density profile for rods and cones and a relative acuity curve are shown in Figure A.4. There are approximately 6.5 million cones and 125 million rods in the retina. The optic nerve contains about 1 million ganglion axons. There is a one-to-one interconnectivity between ~100,000 of these ganglion cells and the cones in the fovea centralis. As a result, a 145 to 1 data reduction process must take place in connecting the remaining 131 million receptor outputs to 900,000 optic nerve channels. Thus, the relative acuity curve shown in Figure A.4 is a function of receptor density and interconnectivity (neural summation). Kabrisky [67, p. 18] has likened this arrangement to looking through a piece of frosted glass with a transparent spot in the center. We are not cognizant of the loss in acuity since the clear spot is always centered on where we are looking. If we consider the center-to-center spacing between cones in the fovea centralis (2 to 2.3μ), the corresponding subtended angle is 25 to 29 seconds. This is equivalent to approximately 60 cycles/degree subtended. As indicated previously, pupil diameters of 2.5mm to 5mm maintain relatively

constant acuity. Campbell and Gubisch [68] have shown that the optics of the eye produces the best image for a pupil diameter of 2.4mm. A recent paper by Synder and Miller [69] demonstrates that the theoretical optimum receptor packing with a 2.4mm pupil gives an angular spacing of 27.4 seconds; hence, the system appears to be consistent. Thus far we have considered only the basic anatomical arrangement of the photoreceptors within the retina. We will now discuss the functional relationships of these receptors.

The two types of receptors differ by more than their physical shape and size. The rods contain a purple pigment, rhodopsin, which has a peak spectral absorption at 505nm (within the green spectrum). When green light is absorbed several chemical reactions take place which convert the rhodopsin to retinene and a protein called scotopsin. If enough light is absorbed the retinene is further bleached to colorless vitamin A. Rhodopsin is continuously resynthesized from scotopsin and vitamin A or retinene. In complete darkness all of the scotopsin may be converted back to rhodopsin.

Proteins similar to scotopsin, photopsins, are found in cones. The cone pigments which produce photopsins are probably of three types. These pigments appear to absorb light maximally at 440nm, 535nm, and 565nm [70 , p. 330]. The actual pigments and proteins have yet to be completely isolated from the human retina. The spectral sensitivity curves shown in Figure A.5 were obtained

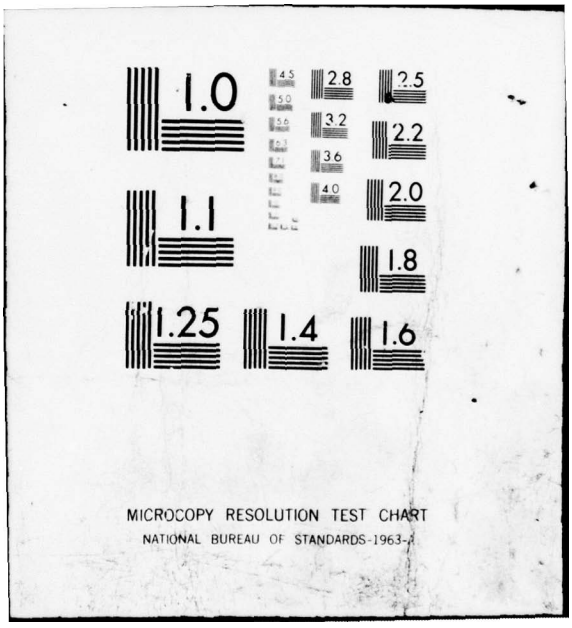
AD-A060 431 UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES IMAGE --ETC F/G 20/6
DIGITAL COLOR IMAGE COMPRESSION IN A PERCEPTUAL SPACE.(U)
JUL 78 H C ANDREWS, C F HALL F33615-77-C-1016

UNCLASSIFIED AFAL-TR-78-80 NL

3 OF 3
AD
A060431



END
DATE
FILMED
1-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

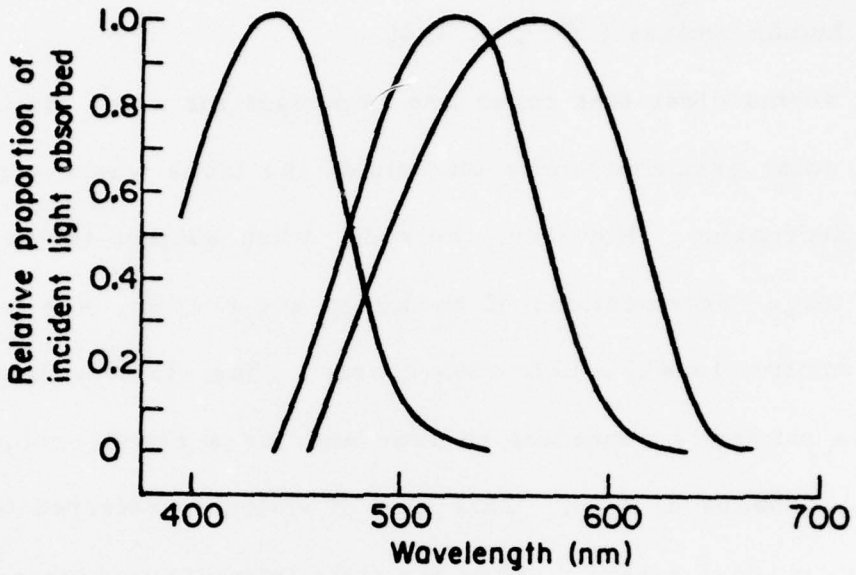


Figure A.5 Absorption Spectra for Three Types of Cones

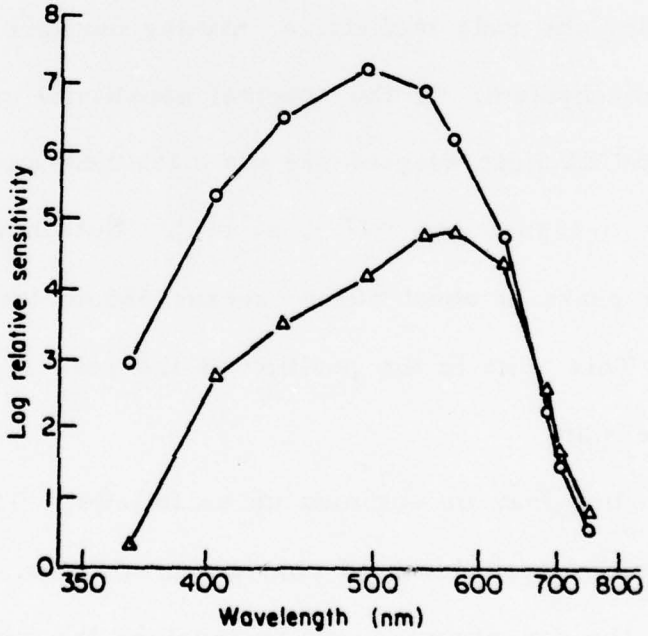


Figure A.6 Photopic and Scotopic Spectral Sensitivity Curves

with reflection densitometry measurements on single receptors from excised human retinas [70 , p. 332].

It seems clear that cones are important for color vision, and indeed, color sensitivity falls off outside the fovea where cone density is decreasing. However, the rods, when adapted to the dark so that large concentrations of rhodopsin are present, are much more sensitive to white light than cones. Thus, in dim light our vision is primarily dependent on rods and, as a result, colors appear as shades of gray. This type of vision is referred to as scotopic, or dark vision. When the light intensity is higher (as in daylight) the rhodopsin of the rods is almost entirely bleached out, thereby rendering the rods ineffective, making daylight (or photopic), vision a cone mechanism. If the spectral sensitivity curves for the dark adapted and daylight adapted eye are measured, one obtains curves similar to Figure A.6 [17 , p. 146]. Note how the scotopic (rod) curve peaks at about 505nm versus 555nm for the photopic (cone) curve. This shift in the position of the peak is referred to as the Purkinje shift.

The preceding may be summed up as follows. The retina is not a light sensitive transducer of constant properties. It contains two receptors: the day receptor, which involves the whole surface of the retina and functions at high luminous levels with a spectral sensitivity defined by the photopic curve shown in Figure A.6; and

the night receptor, which functions when the eye is dark adapted and is characterized by the scotopic curve in Figure A.6. The rods (night receptors) are almost completely absent from the fovea region where the cone (day receptor) density is highest. The cones appear to be totally responsible for color vision. This duality of the retina is sometimes referred to as the duplicity theory [71 , p. 387].

The minimum threshold for the rods appears to be one quantum of light whereas for the cones it is four or five quanta. Once the minimum threshold is exceeded the chemical processes previously mentioned take place. By some unknown mechanism the absorption of light and resultant chemical reaction produces an electrical response in the receptors that is transmitted to the bipolar cells. Unfortunately, it is not possible to monitor these signals at this point. The individual functions of the neuronal layers of the retina can only be conjectured. It is known that there is a nonlinear functional relationship between the nerve impulse output at the ganglion axons and the impinging light [1 , p. 163]. The functional form of this nonlinearity remains an issue [72]-[75] and [34]. debate centers on whether the functional form of the nonlinearity is logarithmic or a power law. The exponent range in the power law argument is usually .29 to .35 or approximately cube root. These two forms are nearly equal over a 1 to 100 range and the

logarithm is bound by the .29 and .33 curves out to about 600. In fact the difference between the .29 power curve and the logarithm curve at 1000 is less than 7% (see Figure A.7). The problem with this comparison is, "how should the data be scaled?" If one uses quanta of light to measure intensity then obviously we would be well beyond the range of close agreement. If we use trolands as our unit of measure (1 troland = 1 cd/m^2 illuminating a $1mm^2$ pupil area), then we are within the 1 to 100 range for most experimental data.

One of the primary results of the nonlinearity (regardless of the exact functional form) is the compression of the dynamic range of the input intensity. This results in a system which can handle light intensities over a range of 10 billion. Compared to the 16 to 1 area variation in pupil size we see that the main intensity compensation mechanism occurs in the photoreceptors. In fact, the pupilary response is transient in nature, always returning to approximately the same size after the photoreceptors have "adapted" to the change in illumination.

A.3. The Lateral Geniculate Bodies

The "coded" visual information exits the retina by propagating down the ganglion axons (the optic nerve) to the optic chiasm. At this point the optic nerves from both eyes decussate and the signals

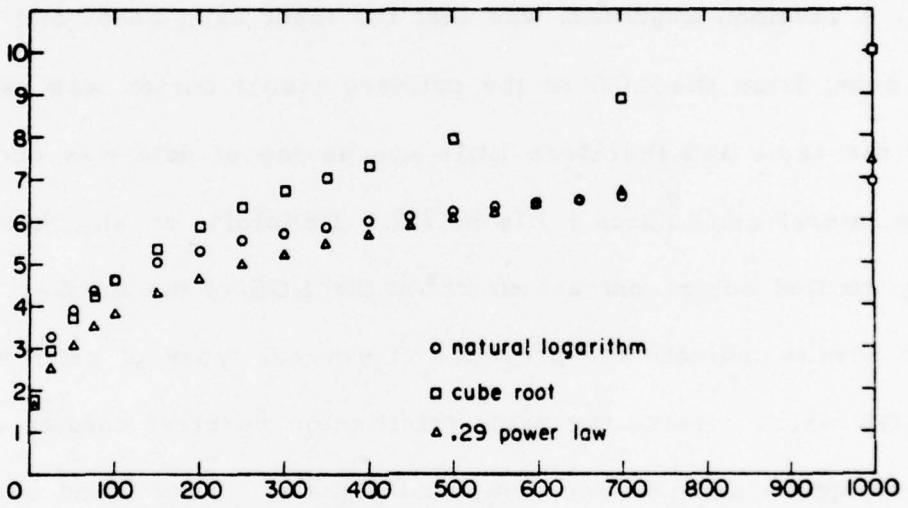


Figure A.7 Nonlinearity Comparison Curves

from the left half of the retina (w.r.t. optic axis) of both eyes proceed to the left lateral geniculate body (LGB). Similarly, the right half of the retina of both eyes provides signals to the right LGB. Since the retinae are stimulated by inverted images of the visual field, the left field maps to the right LGB and vice versa.

Until recently, the function of the LGB's was thought to be of minor consequence to the actual processing of the visual image itself. A common argument was that the input axon count and output axon count from the LGB to the primary visual cortex was essentially the same and therefore little processing of data was occurring in the lateral geniculates [67, p. 25]. DeValois et al., have recently studied color contrast effects in the LGB of the monkey [76]. Their results indicate the presence of several types of cells within the LGB which receive the basic tristimulus spectral outputs of the photoreceptors and produce compound signals. They found spectrally nonopponent cells which respond to all wavelengths with either an increase or a decrease in firing rate and spectrally opponent cells which respond with an increase in firing rate to some areas of the spectrum and a decrease to other areas. Four types of opponent cells were found: red excitatory and green inhibitory (+R - G), green excitatory and red inhibitory (+G - R), yellow excitatory and blue inhibitory (+Y - B), and blue excitatory and yellow inhibitory (+B - Y). The nonopponent cells appear to transmit brightness

information, whereas the opponent cells code the color information.

A.4. The Visual Cortex

The visual signals proceed from the LGB's to area 17 of the striate cortex which is located in the occipital lobe areas of the brain. The data appears to map conformally onto area 17. Investigation of the spectral sensitivity at this point indicates that the observed color opponent interaction is established at earlier levels of visual processing [77]. This finding indicates the spectral processing is occuring almost entirely within -- or prior to -- the LGB's. Several other neurological investigations of the primary visual cortex have been made which relate to the spatial content of the image.

The most noted experiments have been those of Hubel and Wiesel [78]-[81]. Early experiments by Kuffler [82] demonstrated the existence of concentric regions within the retinal mosaic which have on and off centers. These two types of structures produce a type of high-pass spatial filtering through lateral inhibition. Hubel and Wiesel found that at the cortical level there are "simple" cells which respond to spots of light on the retina anywhere within a long narrow rectangular area which is flanked by an inhibitory surround. Both "on" and "off" cells were found, including cells which responded to light-dark borders. The cell responses were

sharply selective to line orientation and to translational displacements of the stimulus. In addition to simple cells, complex cells were discovered. These cells appear to be located at the next level of processing. In these cells an appropriately oriented slit stimulus gives a response of about the same amplitude regardless of its position in the field. Pollen, et. al., [83] have suggested, based on their experimental work, that the complex structure of the striate cortex may be performing two-dimensional spatial decompositions of subdomains of the visual space. In a more recent publication Pollen and Taylor have shown that a Fourier decomposition of the spatial domain is consistent with Hubel and Wiesel's findings and they have pointed out several advantages of a system which performs such a decomposition [84].

The spectral and spatial decompositions of the visual field are by no means separable processes. Indeed, DeValois and Pease have demonstrated that whereas significant spatial processing of achromatic signals occurs at the retinal and LGB levels, comparable chromatic processing appears to occur at the cortical levels [24].

APPENDIX B

SOME PSYCHOPHYSICAL CHARACTERISTICS OF THE HUMAN VISUAL SYSTEM

In Appendix A we discussed the physiological facets of the visual system. One of the major problems of physiological studies is that they usually involve invasive techniques. That is, one inserts electrodes into an area of interest or exposes neuronal structure in vivo, etc. This type of research is not commonly performed on humans. Several animal species, from the Limulus (the horseshoe crab) to different varieties of monkey, have been used for these purposes. Although similarities in the basic structure of the HVS and certain animal visual systems certainly exist (particularly for higher primates), it is difficult to ascertain the detailed structure and interconnectivity of the HVS. Moreover, knowledge of the microstructure of a system (biological or otherwise) does not insure knowledge of function. In this regards, the sum of the parts is quite often exceeded by the whole. These problems are partially resolved by psychophysical techniques.

Boynton has defined visual psychophysics as, "an interdisciplinary area of scientific investigation relating the reactions of human observers to physically measurable aspects of the visual

environment in which they live" [85, p.8]. The key word is reactions and the basic thrust becomes that of studying the whole via input-output relationships. The mechanisms and/or organization which could produce these relationships may then be hypothesized. In this manner the two fields of study -- physiology and psychophysics -- complement one another.

B.1. A Fundamental Result

A recent paper by Campbell and Green readily demonstrates the "harmony" between visual psychophysics and physiology [86]. In this work a laser was used to image interference fringes onto the retina. By decreasing the contrast of the fringes with another source of light it was possible to determine the threshold of detection. This technique produces a measure of the resolving power of the retina-brain complex without prior modification by the optics of the eye. Measurements were then made of the visual resolution of "external" gratings (viewed from the face of an oscilloscope) whose intensity varied sinusoidally with distance across the gratings and which were imaged onto the retina by the optics of the eye. A comparison of the results yielded the modulation transfer function of the eye. Effects of pupil size and focus were measured and compared to the performance of an ideal optical system. The main results were; the retina-brain complex has a high frequency cutoff

and at every spatial frequency tested (2 to 40 cycles/degree) the optics decreased contrast sensitivity. The characteristics obtained for the optics was not in complete agreement with that obtained by Flamant in earlier work which did not use a psychophysical paradigm [87]. Flamant used a "double pass" technique in which a grating was focused on the retina and the reflected image analyzed. This technique does not require a response from the subject, however the grating passes through the optics twice and the reflective properties of the retina must be taken into account. Campbell and Gubisch then demonstrated that when the reflective properties of the retina are taken into account the two experimental techniques yield consistent results [68]. The modulation transfer functions of the eye for pupil diameters of 3mm and 6mm are shown in Figure B.1.

In the previous paragraph we discussed some psychophysical aspects of the dioptrics of the HVS. The main point is the dioptric system has been parameterized well enough that one can control, to an experimental degree of accuracy, the stimulus imaged upon the retina by a particular experimental apparatus. This is a prime precursor of a valid psychophysical experimental protocol. With this capability it is possible to study the retina-brain complex in detail. There are three main areas of interest in these studies (not necessarily independent); the spatial characteristics, the spectral characteristics, and the temporal characteristics. We will

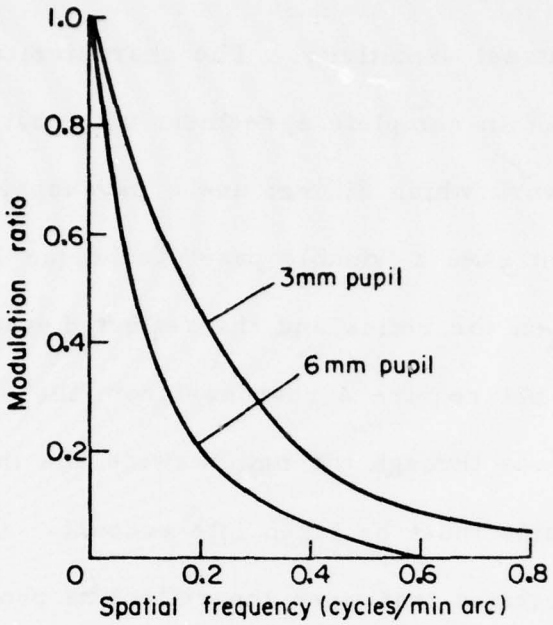


Figure B.1 Modulation Transfer Function of Ocular Media

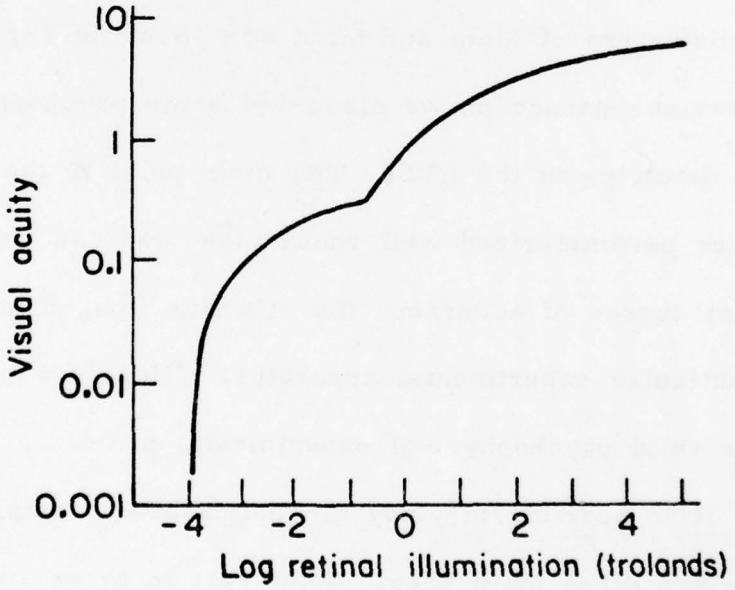


Figure B.2 Visual Acuity as a Function of Luminosity

begin with a discussion of the spatial characteristics.

B.2. Visual Acuity

One of the more important and misunderstood spatial characteristics of the HVS is visual acuity. Visual acuity is simply the capacity to discriminate the fine details of objects in the field of view. There are two reasons a trait so simply defined is misunderstood; firstly, there are several "types" of acuity tasks and secondly, for most tasks there is no single mechanism responsible for the response to the task. Acuity tasks may be grouped into the four classes; detection, recognition, resolution, and localization [65, p. 322].

The detection task merely involves stating whether an object is present in the visual field or not. This task has been used by some as a measure of the smallest objects which can be viewed by the HVS. This is misleading since the results of such paradigms cannot logically be separated from the absolute or differential sensitivity of the eye.

The task of recognition requires the subject to locate, describe, or name the object. The standard eye chart is an example of such a task. A common clinical object is the Landolt ring (a ring with a gap). The observer is asked to indicate the location of the gap. With high luminance backgrounds, gaps corresponding to

30 seconds of arc can be recognized. Intensity discrimination is not the limiting factor in Landolt ring acuity. Other factors, particularly foveal cone diameter and spacing, are important mechanisms affecting this type of test.

Resolution tasks require the observer to respond to a separation between elements of a pattern. The basic measurement becomes the minimum distance (between objects) which can be discriminated. Visual acuity is the reciprocal of the angular separation between two elements of the test pattern when the two elements are barely resolved. A favored pattern for this type of test is a grating of parallel light and dark stripes of equal widths. This type of object yields limits of one minute of arc. The resolution task is regarded as the most critical aspect of visual acuity. The results of such tests can be meaningfully related to the diffraction effects of the dioptrics and to the retinal mosaic.

The last type of acuity task, localization, depends on the discrimination of small displacements. An example of such a task is vernier acuity which is tested by using a broken, offset, straight line. The object becomes that of finding the minimum discernible lateral displacement of the two halves of the line. This type of task produces results which are similar to the detection of single black lines (2 to 4 seconds).

There are several factors which affect visual acuity. The

main ones are: (1) pupil size, (2) the dimensions of the retinal mosaic, (3) object intensity, (4) stimulus duration, (5) adaptive state of the photoreceptors, (6) eye movements, and (7) object contrast. The various tasks enumerated in previous paragraphs are affected differently by these factors. The effects of pupil size were discussed in Section A.1. In Section A.2 the limits imposed by the retinal mosaic were detailed. It was shown that these two factors limit visual acuity to approximately 30 seconds of arc.

Through personal observation, one can easily ascertain that while large objects are seen easily in dim light, small objects can be seen clearly only when the lighting is increased. This effect is primarily a function of scotopic versus photopic vision. Visual acuity is poorest at scotopic intensity levels where parafoveal or peripheral rod receptors predominate. For higher intensities (which exceed cone receptor thresholds) acuity rises steeply. As can be seen in Figure B.2, as intensity increases acuity rises to a maximum level which is maintained over a wide range of high intensities. As with other factors governing acuity, different data and interpretations are found for the different forms of acuity tasks, however, the basic relationship shown in Figure B.2 is maintained.

The effects of exposure time or stimulus duration have been studied by several researchers. These studies indicate that for the case of detection of bright disks on dark backgrounds, acuity is

proportional to the square root of exposure time. For bright line stimuli the proportionality is direct. No simple relationships appear for acuity versus time in the resolution tasks.

The state of adaptation of the photoreceptors is an important parameter in acuity tasks, particularly for the studies in stimulus duration. Craik found that in general, acuity is highest for conditions of near equal adapting and test luminances [89]. Prolonged dark adaptation is required to achieve scotopic vision, which is necessary for viewing objects at low intensity levels. Acuity is poor at these levels, but it is even poorer if adaptation is not complete. At high intensity levels the eye must be given prolonged adaptation to insure the cones are functioning most efficiently.

The eyes are never motionless, thus the retinal image must affect different receptors from one moment to the next. These motions could have three possible effects on visual acuity: (1) they may be so small acuity effects are precluded, (2) they may cause a "blurring" of the image, or (3) they may sharpen the image by "scanning" contours. Experimental evidence indicates that eye movement does not improve acuity and in some cases acuity is impaired by motion [88 , p. 178]. One of the more important characteristics of the HVS was discovered during these types of investigations. If the motion of the eye is completely counteracted, i.e., the image is stabilized on the retina, then the object fades out and

the field looks uniformly gray [17 , p. 405]. If the object is shifted or the intensity changed, it will reappear temporarily. A stabilized image which is illuminated once or twice a second remains visible [90 , p. 382]. It can be concluded that receptors which are continuously excited by the same stimulus cease to transmit information. If the receptors are excited intermittently, as during eye movement, then information is continuously transmitted. It appears then that eye motion is important for the maintenance of visibility but has little effect on actual resolution of objects if they are visible.

It has been found that for dark objects on bright backgrounds acuity is maximal for highest contrast between object and backgound [65 , p. 339]. Recent work with contrast gratings has produced a wealth of information and corresponding "theories" of vision. This area is discussed in detail in the following section.

B.3. Spatial Frequency Response Functions

So far, we have emphasized the standard techniques of visual acuity determination. In general, the spatial manipulation required to produce a criterion response confounds changes in the contrast and space parameters. For example, when two points are brought together the two light distribution peaks become closer and the absolute luminance of the trough increases. The latter effect reduces the contrast of the image. This situation is even more pronounced

when gratings of higher and higher spatial frequencies are considered. The contrast gets smaller and smaller, eventually becoming zero. One experiment which dissociates contrast and spatial separation is the interference fringe method of bypassing the dioptrics and creating a 100% contrast fringe on the retina. Another approach is to maintain a constant spatial pattern and vary only the contrast. These particular techniques are similar to the one dimensional frequency analyses performed on linear electrical networks. The system is subjected to a constant amplitude input sinusoid and the output amplitude and phase variations with frequency are determined. For linear systems (or systems operating in a linear range) this technique provides a complete characterization. In the space domain, where the input is periodically varying with distance, the system must be spatially invariant as well as linear. These two requirements cannot be over emphasized. The HVS does not satisfy either, however, in certain experimental procedures these conditions may be approached. In addition, the results of the experiments can be enlgihtening if one is cognizant of the limitations of the analysis; and, prediction of the system response to an arbitrary input is possible. For these reasons spatial frequency analysis of the human visual system has come into vogue recently [26 , p. 206].

The results of these experiments are usually conveyed in the form of contrast sensitivity functions or curves. These functions

characterize the ability of the visual system to transfer information at various spatial frequencies from stimulus input to output. Spatial frequency is usually expressed in cycles/degree. This convention relates different combinations of viewing distance and object size to the equivalent spatial frequency and hence, image size on the retina. Contrast sensitivity is defined as the reciprocal of percent threshold modulation (difference between peaks and troughs) required for the observer to distinguish the stimulus from a uniform field of equivalent luminance.

One of these experiments, that of Campbell and Green [86], quantified the dioptrics and it was discussed earlier. The results of this experiment indicate that as far as the high frequency characteristics are concerned, the dioptrics and the retina-brain complex yield curves which are of the same shape. The low-frequency portion of typical contrast sensitivity curves can only be attributable to the retina-brain complex however. The combined high- and low-pass characteristics produce an overall bandpass characteristic with a center frequency of approximately 5 cycles/degree (see Figure B.3). The high-frequency loss has been shown to be non-isotropic [91]. Gilbert and Fender have verified that the curves remain essentially unchanged for stabilized images [92]. The low frequency portion of the MTF has been found to be a function of luminance level [93]. The low-frequency attenuation also

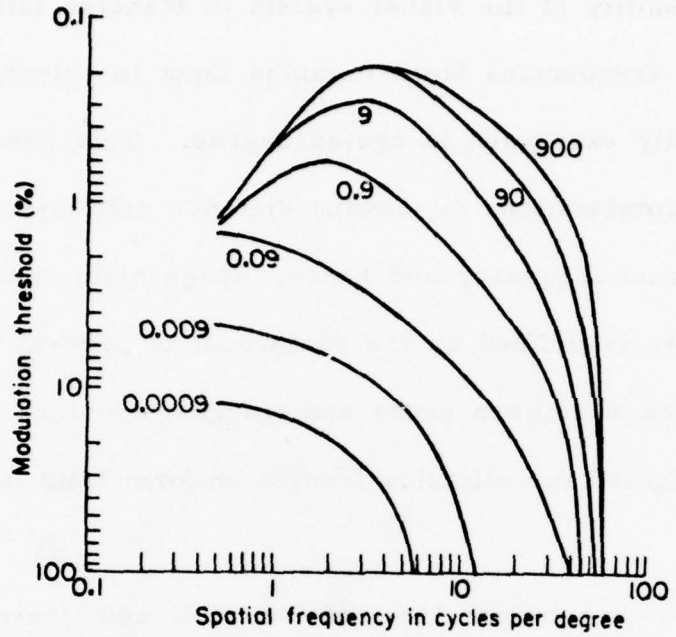


Figure B.3 Spatial Response as a Function of Mean Luminance in Trolands of HVS

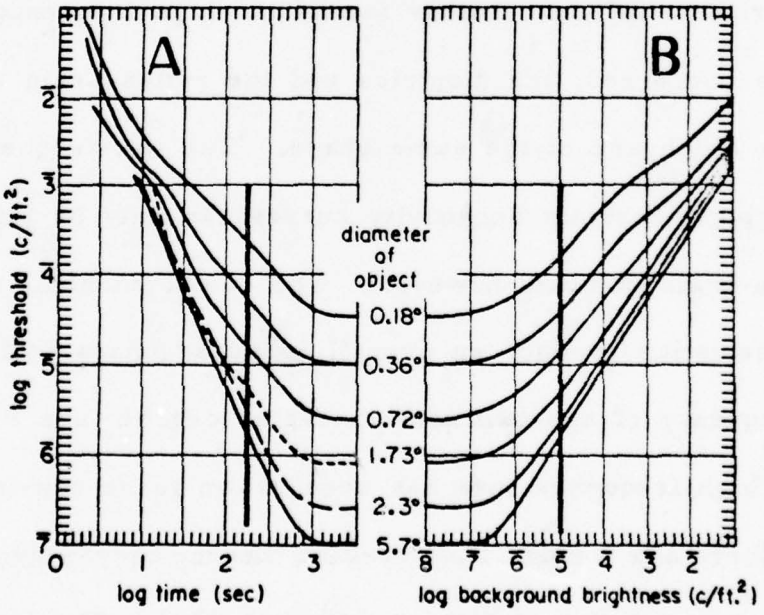


Figure B.4 Dark Adaptation and Equivalent Background

disappears with short exposure durations [94]. As Westheimer has pointed out, the high-frequency characteristics can be related to optical and anatomical limitations; however, the origin of the low-frequency traits is less clear [88, p. 182].

B.4. Lateral Inhibition

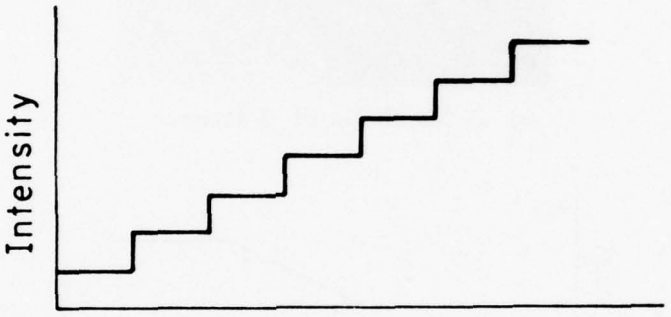
In Section A.4 we mentioned the experiments of Kuffler which demonstrated the existence of regions in the retina which have "on" and "off" centers. These types of regions can produce lateral inhibition which results in a low-frequency attenuation or high-pass filtering. Patel has established this fact through Fourier calculations [ibid]. The affects of adaptation and exposure duration on these receptive regions have been shown to be consistent with the elimination of the low-frequency effects [88, pp. 182-183]. The simple thesis that the low-frequency loss is due to lateral inhibition is not compatible with all observations however. For example, the results of two increment-threshold experiments are shown in Figure B.4. Note that in every case, as the diameter of the object increased the threshold decreased. If lateral inhibition is occuring in the HVS the threshold should begin to increase at some critical diameter. If the modulation threshold curves of Figure B.3 derive their low-frequency characteristic shape from lateral inhibition in the retina, reconciliation with the curves of Figure B.4 is necessary.

Several recent experimenters have questioned the validity of the low-frequency roll-off evidenced in most HVS MTF's [95]-[97]. The contention is that for the low-frequency gratings not enough cycles are within the visual field. Estevez and Cavonius [98] maintain that experiments of Hoekstra, McCann, and Savoy [95]-[97] caused illusory luminance gradients across the stimulus which resulted in a loss of sensitivity to mid-frequencies. They contend this mid-frequency loss has been misinterpreted as an absence of low-frequency attenuation. This particular issue is still unresolved; however, there are other experiments which indicate the presence of spatial interaction in the HVS.

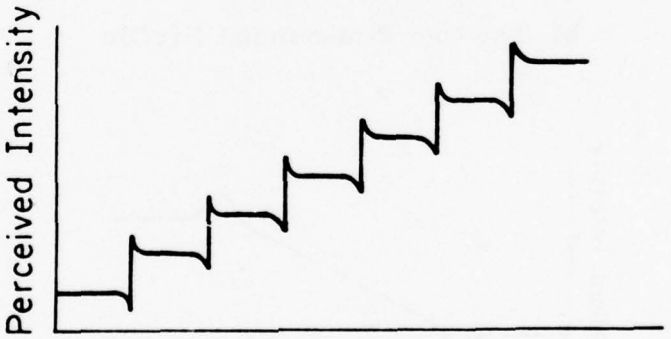
If there is no spatial interaction within the HVS, then the perceived brightness at any point in the visual field would be a function of the strength of excitation of the receptors lying under the retinal image of that specific point (the following discussion is based heavily upon Cornsweet's excellent presentation [17 , Ch. XI, pp. 268-310]). Several perceptual or psychophysical paradigms indicate this is not the case. A good example of this fact is demonstrated in Figure B.5. When the constant intensity step grey scale is viewed, a "scalloped" intensity pattern is perceived. Another common demonstration is the Mach band pattern shown in Figure B.6. In this case a dark and a light stripe appear to the left and right, respectively, of the center of the intensity gradient. These illustrations



a) A Uniform Grey Scale

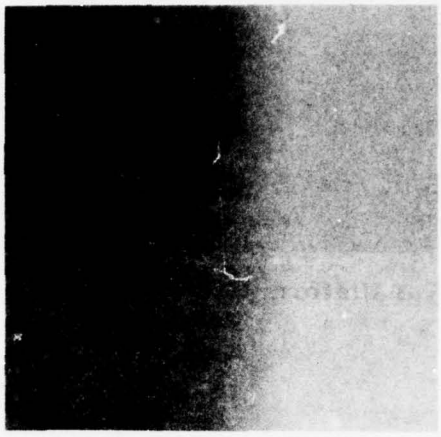


b) The One-dimensional Profile

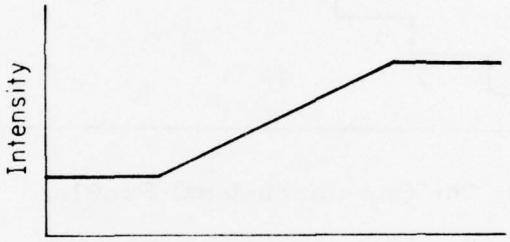


c) The Perceived One-dimensional Profile

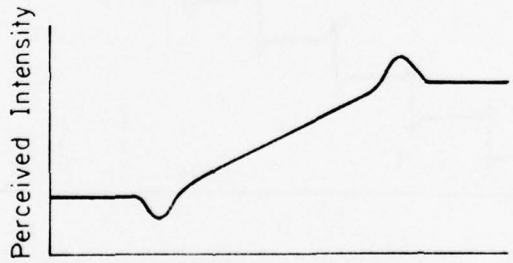
Figure B.5. Illustration of Lateral Inhibition Effects.



a) A Mach Band Pattern



b) The One-dimensional Profile



c) The Perceived One-dimensional Profile

Figure B.6. The Mach Band Effect.

indicate that perceived intensity is not a simple monotonic function of intensity stimulus. If one postulates the presence of lateral inhibition within the HVS and plots the outputs of a row of receptors being stimulated by a profile similar to that in Figure B.5b, an output similar to B.5c is obtained [17, pp. 303-304]. Thus, the hypothesis that lateral inhibition occurs within the HVS is consistent with these perceptual phenomena.

Although the previous paragraph indicates the presence of lateral inhibitory effects within the HVS, and hence high-pass filters, the experiments discussed do not quantify the filter parameters. The data from the sine-wave grating experiments could provide this parameterization if we assume the low-frequency portions of curves such as those shown in Figure B.3 are valid. A very significant work in this respect was performed by Mannos and Sakrison [7]. This work was primarily concerned with the efficient coding of images (as we are). Several subjective evaluation experiments were performed with images which were preprocessed, coded, and postprocessed with a model of the HVS which contained a bandpass filter. The filter parameters were varied for each set of experiments. The filter function which gave the best images (as judged subjectively) was very close to MTF curves obtained by various researchers via grating experiments [7, Figure 8, p. 535]. The primary difference being the peak frequency occured at 8 cycles per degree rather than the usual 5 to 6 cycles per

degree as obtained psychophysically. Thus, the low-frequency loss has been shown to be important perceptually.

The preceding discussions of the spatial characteristics of the HVS has shown them to be complex and not easily quantifiable. The lack of heterogeneity within the retinal structure cannot be overemphasized. This characteristic makes it extremely difficult to separate global and local characteristics of the system. Indeed, one of the main objections to grating paradigms is that they are global in nature. Many of the properties and traits we have discussed become relevant when modeling the HVS to perform perceptual, pattern recognition, or scene analysis tasks. For our purposes (system preprocessors) the global characteristics are the more pertinent characteristics.

B.5. Spectral Properties

Let us turn now to a second major area, the spectral characteristics of the HVS. The absorption spectra of the human visual photopigments were shown in Figure A.5. These curves were obtained through measurements on receptors in excised human retinas. The measurement technique used is very dependent on the adaptive state of the receptors. Obviously, human retinas are obtained under almost completely uncontrolled conditions and therefore the data is not totally reliable. Liebman has concluded that the data can be no

better than ±20-30nm and that published density curves cannot be regarded as indicative of what exists in the living eye [99 , p. 515]. To gain some true insight into the HVS spectral response we must once again turn to psychophysical experiments. First we will define some basic terms.

Colors have three main attributes; hue, saturation, and luminosity or brightness. Hue denotes the color appearance by name, e.g., red, orange, etc. It is the aspect of color which changes most strongly when the wavelength of the stimulus changes. Saturation refers to the purity of a hue or to which extent it appears to be diluted with white, grey, or black. The degree to which colors appear to emit more or less light is referred to as the luminosity or brightness of the color. The term luminosity is preferable since brightness of color means "colorfulness" to many people. The three attributes just defined can be used to describe any color. It should be noted that these are all subjective terms. In this sense color and wavelength of light are not synonymous. Indeed, several different combinations of wavelength may produce the same subjective color description. The visible band of electromagnetic radiation wavelengths extends from the short ultra violet rays below 397nm to the longer infrared heat waves above 723nm. The principle hues are: red, 647-723nm; orange, 585-647nm; yellow, 570-585nm; green, 521nm; blue, 480nm; indigo, 424-455nm; and violet, 397-424nm.

From Figure A.5 it can be seen that the three photoreceptor curves overlap. For example, a wavelength of 480nm would stimulate all three receptors. The difficulty that this situation generates in trying to design a reliable psychophysical paradigm is illustrated in Figure B.7. The curves of this figure are the result of a color naming experiment. The various wavelengths were presented to the subjects who responded with one of four hues; red, yellow, green, or blue. It can be seen that in the case of 580nm a variation of only ±40nm can shift the perceived response from green to blue to red. One way to eliminate some of the difficulties encountered in trying to measure responses of this trichromatic system is to select subjects with color vision difficiencies.

Some observers can only discriminate between wavelengths in restricted regions of the spectrum, and color-matching functions from them show that only two parameters are needed to describe their color vision. The simplest reason for this deficiency would be an absence of one of the three types of cones and this has been verified by using reflection densitometry. These dichromats are of three types: protanopes, who lack the 565nm cone; deuteranopes, who lack the 535nm cone; and the more rare tritanopes, who do not have the short-wavelength cones. It is known that blue light-absorbing cones are relatively sparse in the foveola [100 , p. 209]. Thus, blue lights imaged precisely in this area are confused with greens,

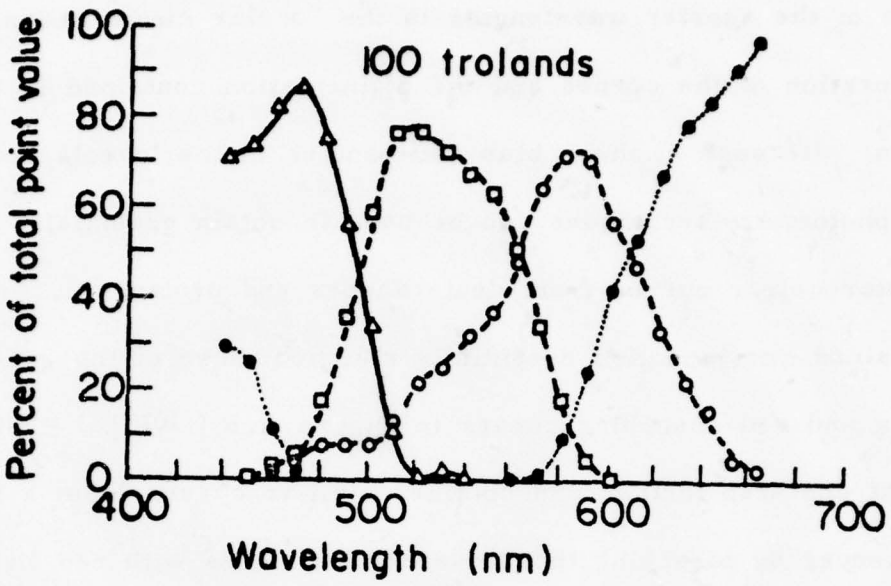


Figure B.7 Color Naming Functions

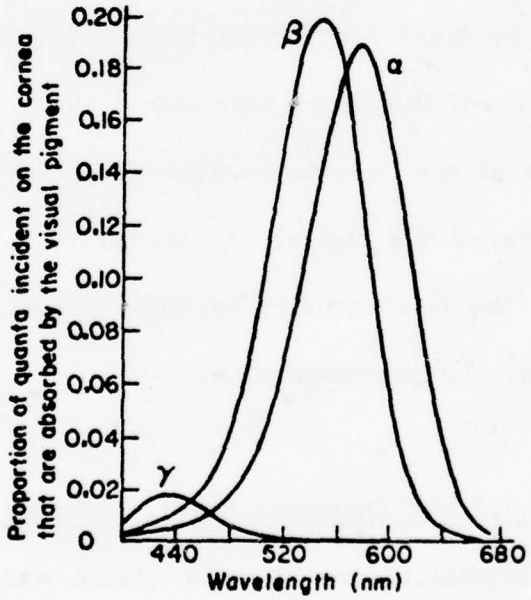


Figure B.8 Psychophysically Determined Spectral Response of the Photoreceptors (includes ocular media effects)

white, and yellows. This deficiency is compounded by the absorption of the shorter wavelengths in the ocular media caused by the coloration of the cornea and the pigmentation contained in the macula. Because of these blue deficiencies in the foveola, microspectrophotometry techniques can be used to obtain essentially single photoreceptor curves from deuteranopes and protanopes. Rushton obtained curves which essentially matched those of the green-absorbing and red-absorbing curves in Figure A.5 [101]. Rushton also went one step further and obtained similar curves from a normal observer by bleaching the red-absorbing cones with red light to obtain the green-absorbing curve and bleaching with blue-green light to obtain the red-absorbing curve. Figure B.8 contains three curves obtained by Wald which have been widely accepted as the absorption spectra of the three pigments [102]. These curves include the effects of the ocular media. When the difference in scaling is considered the curves of Figures A.5 and B.8 are quite similar. Thus, the trichromatic receptor theory is supported by both physiological and psychophysical data.

B.6. Trichromatic and Opponent Color Theories

The trichromatic theory of color vision was first postulated by an English chemist named Palmer in 1777 [103 , p. 56]. Twenty-five years later Young proposed the same theory of color

vision. Helmholtz brought Young's theory of "color sensations" forward in his Physiological Optics published in 1860. Because of this, the trichromatic theory is often referred to as the Young-Helmholtz theory of color vision. There is another theory for color vision which was proposed by Hering in 1820 [104 , p. 73], the so-called opponent theory of color vision. Hering was impressed by the existence of the five psychological sensations; red, yellow, green, blue, and white (recall the four hue curves of Figure B.7). In addition, the four basic hues seemed to operate in opposing pairs. Red and green seem to oppose in that there is no reddish-green color. Similarly, there are no yellowish-blues. Hering also assumed there must be a third black-white mechanism. This theory explained the existence of the five basic psychological primaries and the complementarity of negative after-images. For example, the after-image of a bright red stimulus seen against a white surface is green.

The two basic theories of vision, trichromatic and opponent, have generated much debate in the past 100 years. It now appears that both theories are correct. The experimental work of DeValois [76] has confirmed the existence of opponent cells in the LGB. Recent conjectures on the interconnectivity of the receptors and LGB cells demonstrate the compatibility of the two theories [90, p. 189], [104 , p. 76], and [105].

B.7. Luminosity and Color Constancy

So far, we have considered only the spectral aspects of perceived colors. The effects of luminance and contrast should also be considered when dealing with color vision. As the luminance of a colored stimulus is increased the apparent hue may undergo a change. Increasing luminance will shift reds and yellow-greens toward the yellows while blue-greens become bluer. This is the Bezold-Brucke effect, and it can be explained by using the opponent color theory. The red-green system simply has a lower threshold than the blue-yellow system. The appearance of a color is also altered by contrast phenomena.

If a constant luminance colored patch is viewed against a variable luminance white background, its appearance may change dramatically with changes in the background luminance. For example, an orange object will become brown with a high luminance background or a pastel orange with a low luminance surround. It is believed that lateral inhibition produces this effect and other similar effects [17, pp. 365-383]. If this is indeed the case, then Mach bands should occur in gradients of hue. Several researchers have investigated this phenomena and there is considerable disagreement as to whether "colored" Mach bands do indeed occur. Van Der Horst and Bouman maintain that they do not and hence, spatial inhibitory influences are lacking in the color mediating channels

[106]. On the other hand a recent paper by Green and Fast demonstrates that Mach bands similar to those which occur in achromatic luminance gradients also occur in constant hue luminance gradients [107]. However, the "Mach type" bands observed in hue gradients were not the type as predicted by lateral inhibition at the receptor level.

Spatial frequency contrast gratings of different hues have also been used in studying color vision [108]. Results of these studies verify the reduced sensitivity of the blue receptors (including the ocular media) and their scarcity. This later factor is evidenced by the reduction in resolution. The blue channel was found to peak at approximately 2 cycles/degree rather than 8 cycles/degree for red and green. In addition, the maximum frequency was between 10 and 20 cycles/degree, which represents an acuity decrease by a factor of 6. Of perhaps more importance is the fact that Green obtained low-frequency losses in all of his data, therefore implying that lateral inhibition is present. It becomes apparent that several spatial and spectral aspects of the HVS are inter-related and it may be some time before the true structure and nature of the system is known. To compound the problem these factors are also related to the temporal characteristics of the HVS.

B.8. Some Temporal Considerations

Since we are primarily concerned with "still" imagery, we will not discuss in detail the temporal aspects of the visual system. One of the most studied temporal characteristics is the response to flickering stimuli. At a given light intensity, a field is alternated between light and dark with increasing frequency until the flicker is no longer detected. That point is defined as the critical flicker frequency (CFF) for the particular stimulus conditions. One can obtain MTF's of the temporal system by varying the intensity of a field sinusoidally. The temporal MTF has been measured for a wide variety of stimulus and adaptation conditions [109]-[112]. At any mean level of luminance the system is maximally sensitive to frequencies between 5 and 25Hz (flicker free T.V. is scanned at 30Hz). Increased luminance shifts the high- and mid-frequency response to higher frequencies. The low-frequency portion of the curves is relatively insensitive to mean intensity changes and again, lateral inhibition may be their determinant [17 , pp. 410-416]. Some spatio-temporal and spectral-temporal effects are of more interest.

Tynon and Sekuler have found that sinusoidal gratings appear to be of higher spatial frequency when briefly flashed rather than presented for longer durations [113]. Other studies have shown that the contrast level for perceived flicker and that for which

the spatial structure of gratings becomes apparent occurs at two different thresholds [114]. These results have led some to posit that two temporal channels, one sustained and one transient in nature, exist in the HVS [114] and [115]. This proposition has been verified for the interconnections between the cat's retina and LGB [116]. The implications of these results are not clear at this time.

One of the more startling temporal phenomena is that of induced color. Colors may be perceived when a variety of stimulus patterns are illuminated intermittently with white light [90 , pp. 205-210], [104 , p. 152], and [117 , pp. 307-308]. These colors are commonly referred to as Fechner colors and they are usually demonstrated with a Benham's disc or top. The disc is rotated at about 5 to 10 rps and three colored rings of blue, green, and red appear. Such a disc is shown in Figure B.9. When rotated clockwise the lines denoted A appear blueish and those at C are reddish. A counterclockwise rotation interchanges the two colors. It has been suggested that the complex flickering patterns set up by the rotating disc produce time-varying activity in the optic nerve that is similar to the output of the photoreceptors when stimulated by colored lights [90, p. 207]. Several attempts have been made to produce subjective colors with stationary flickering lights. These experiments have been moderately successful.

Young has proposed the color channels of the HVS are sensi-

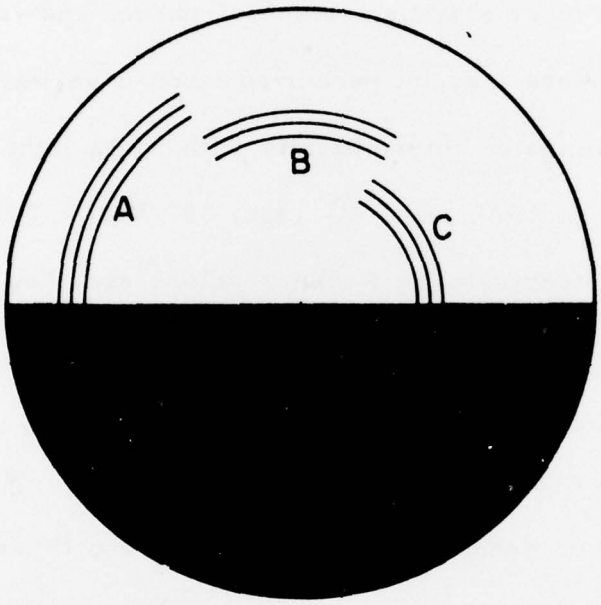


Figure B.9 Benham's Disc

tive to stimulus temporal phase information [118]. He has tested this hypothesis by stimulating the eye with electrical impulses which were compatible with Benham's disc signals convolved with the temporal impulse response of the HVS. The results indicate the relative phase relationships of temporal signals is the most important stimulus variable. The question of the exact physiological nature of the hypothesized phase signals, or the manner in which they are encoded and decoded to produce color sensations, remains unanswered however.

REFERENCES

1. C.F. Hall and E.L. Hall, "A Nonlinear Model for the Spatial Characteristics of the Human Visual System," IEEE Trans. Systems, Man and Cybernetics, vol. SMC-7, No. 3, pp. 161-170, March 1977.

2. A.N. Netravali and B. Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking," Proc. IEEE, vol. 65, No. 4, pp. 536-548, April 1977.

3. W. Frei, "Rate-distortion Coding Simulation for Color Images," Proc. SPIE, vol. 87, pp. 197-203, August 1976.

4. J.O. Limb, "Visual Perception Applied to the Encoding of Pictures," Proc. SPIE, vol. 87, pp. 80-87, August 1976.

5. T.G. Stockham, Jr., "An Overview of Human Observer Characteristics and their Effect on Image Transmission and Display," Proc. SPIE, vol. 66, pp. 2-4, August 1975.

6. F. Kretz, "Subjectively Optimal Quantization of Pictures," IEEE Trans. on Comm., vol. COM-23, No. 11, pp. 1288-1292, November 1975.

7. J.L. Mannos and D.J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," IEEE Trans. Info. Th., vol. IT-20, No. 4, pp. 525-536, July 1974.

8. T. Fukinuki, "Optimization of D-PCM for TV Signals with Consideration of Visual Property," IEEE Trans. on Comm., vol. COM-22, No. 6, pp. 821-826, June 1974.

9. T.S. Huang, et. al., "Image Processing," Proc. IEEE, vol. 59, No. 11, pp. 1586-1609, November 1971.

10. T.G. Stockham, Jr., "Image Processing in the Context of a Visual Model," Proc. IEEE, vol. 60, No. 7, pp. 828-842, July 1972.

11. P.A. Wintz, "Transform Picture Coding," Proc. IEEE, vol. 60, No. 1, pp. 809-820, July 1972.

12. T.G. Stockham, Jr., "Intra-frame Encoding for Monochrome Images by Means of a Psychophysical Model Based on Nonlinear Filtering of Multiplied Signals," Proc. 1967 Symp. on Picture Bandwidth Compression, T.S. Huang and O.J. Tretialk, Eds., New York: Gordon and Breach, 1972, pp. 415-442.

13. W.F. Schreiber, "Picture Coding," Proc. IEEE, vol. 55, No. 3, pp. 320-330, March 1967.

14. J.O. Limb, "Source-receiver Encoding of Television Signals," Proc. IEEE, vol. 55, No. 3, pp. 364-379, March 1967.

15. D.E. Pearson, "A Realistic Model for Visual Communication Systems, "Proc. IEEE, vol. 55, No. 3, pp. 380-389, March 1967.

16. J.L. Brown, "The Structure of the Visual System," in Vision and Visual Perception, C.H. Graham, Ed., New York: Wiley, 1965.

17. T.N. Cornsweet, Visual Perception, New York: Academic Press, 1970.

18. D. Jameson, "Theoretical Issues of Color Vision," in Handbook of Sensory Physiology, Vol. VII/4, D. Jameson and L. M. Hurvich, Ed., New York: Springer-Verlag, 1972.

19. K. Motokawa, Physiology of Color and Pattern Vision, New York: Springer-Verlag, 1970.

20. G. Westheimer and F. W. Campbell, "Light Distribution in the Image Formed by the Living Human Eye," JOSA, vol. 52, No. 9, pp. 1040-1045, September 1962.

21. W. Frei, "Modelling Color Vision for Psychovisual Image Processing," USCEE Report 459, pp. 112-122, 1973.

22. G. Wyszecki and W.S. Stiles, Color Science, New York: John Wiley, 1967.

23. C. H. Graham, "Color Mixture and Color Systems," in Vision and Visual Perception, C.H. Graham Ed., New York: Wiley, 1965.

24. R. L. DeValois and P.L. Pease, "Contours and Contrast: Responses of Monkey Lateral Geniculate Nucleus Cells to Luminance and Color Figures," Science, vol. 171, No. 3972, pp. 694-696, February 1971.

25. R.N. Haber, Contemporary Theory and Research in Visual Perception, New York: Holt, Rinehart and Winston, 1968.

26. R. Sekuler, "Spatial Vision," Annual Review of Psychology, vol. 25, M.R. Rosenzweig and L.W. Porter, Eds., Palo Alto, California: Annual Reviews Inc., 1974.

27. W.K. Pratt, Digital Image Processing, New York: John Wiley & Sons, Inc., 1978.

28. H.C. Andrews and B.R. Hunt, Digital Image Restoration, Englewood Cliffs: Prentice-Hall, 1977.

29. H.C. Andrews, Computer Techniques in Image Processing, New York: Academic Press, 1970.

30. H. C. Andrews, "Two-dimensional Transforms," in Picture Processing and Digital Filtering, T.S. Huang, Ed., New York: Springer-Verlag, 1975.

31. N. Ahmed, T. Natarajan, and K.R. Rao, "Discrete Cosine Transform," IEEE Trans. on Computers, vol. C-23, No. 1, pp. 90-93, January 1974.

32. A.K. Jain, "Some New Techniques in Image Processing," in Image Science Mathematics, C.O. Wilde and E. Barrett, Eds., North Hollywood, California: Western Periodicals Company, 1977, pp. 201-223.

33. L.G. Glasser, et. al., "Cube-root Color Coordinate System," JOSA, vol. 48, No. 10, pp. 736-740, October 1958.

34. S.S. Stevens, Psychophysics, New York: Wiley, 1975.

35. D.B. Judd, "Standard Response Functions for Protanopic and Deuteranopic Vision," JOSA, vol. 35, No. 3, pp. 199-221, March 1945.

36. W. Frei, "A New Model of Color Vision and Some Practical Implications," in Semiannual Technical Report, Report No. 530, Image Processing Institute, University of Southern California, March 1974, pp. 128-143.

37. O.D. Faugeras, "Digital Color Image Processing and Psychophysics Within the Framework of a Human Visual Model," UTEC-CSc-77-029, University of Utah, Salt Lake, June 1976.

38. A. Habibi and P.A. Wintz, "Image Coding by Linear Transformation and Block Quantization," IEEE Trans. Commun. Tech., COM-19, No. 1, pp. 50-62, February 1971.

39. A. Papoulis, Probability, Random Variables, and Stochastic Processes, New York: McGraw-Hill, 1965.

40. J. Aitchison and J.A.C. Brown, The Lognormal Distribution, Cambridge, Massachusetts: University Press, 1957.

41. C.E. Shannon, A Mathematical Theory of Communication, Urbana: University of Illinois Press, 1949.

42. T. Berger, Rate Distortion Theory, Englewood Cliffs: Prentice-Hall, 1971.

43. L.G. Roberts, "Picture Coding Using Pseudo-random Noise," IRE Trans. on Info. Th., vol. IT-8, No. 2, pp. 145-154, February 1962.

44. T.S. Huang, et.al., "Design Considerations in PCM Transmission of Low-resolution Monochrome Still Pictures," Proc. IEEE, vol. 55, No. 3, pp. 331-335, March 1967.

45. A.A. Sawchuk, Quantization Contour Elimination in PCM Television Using Edge Detection, B.S. Thesis, Dept. of Elect. Engr., M.I.T., Cambridge, June 1966.

46. P.F. Panter and W. Dite, "Quantization Distortion in Pulsecount Modulation with Nonuniform Spacing of Levels." IRE Proc., vol. 39, No. 1, pp. 44-48, January 1951.

47. J. Max, "Quantizing for Minimum Distortion," IRE Trans. Info. Th., vol. IT-6, No. 1, pp. 7-12, March 1960.

48. A. Habibi and G.S. Robinson, "A survey of Digital Picture Coding," Computer, No. 5, pp. 22-34, May 1974.

49. J.T. Kajiya, "Group Representations and the Modeling of Visual Perception," in Image Science Mathematics, C.O. Wilde and E. Barrett, Eds., North Hollywood: Western Periodicals Co., 1977, pp. 67-70.

50. W.K. Pratt, et. al., "Slant Transform Image Coding," IEEE Trans. on Commun., vol. COM-22, No. 8, pp. 1075-1093, August 1974.

51. R. M. Haralick and K. Shanmugam, "Comparative Study of a Discrete Linear Basis for Image Data Compression," IEEE Trans. on Systems, Man, Cybernetics, vol. SMC-4, No. 1, pp. 16-27, January 1974.

52. C. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," in IRE National Conv. Rec., part 4, pp. 142-164, March 1959.

53. L.D. Davisson, "Rate-distortion Theory and Application," Proc. IEEE, vol. 60, No. 7, pp. 800-808, July 1972.

54. R.G. Gallager, Information Theory and Reliable Communication, New York: John Wiley and Sons, 1968.

55. J.J.Y. Huan and P.M. Shultheiss, "Block Quantization of Correlated Gaussian Random Variables," IEEE Trans. Commun. Syst., vol. CS-11, No. 3, pp. 289-296, September 1963.

56. D.J. Sakrison and V.R. Algazi, "Comparison of Line-by-line and Two-dimensional Encoding of Random Images," IEEE Trans. Info. Theory, vol. IT-17, No. 4, pp. 386-398, July 1971.

57. C.F. Hall, "Image Filtering Based on Psychovisual Characteristics of the Human Visual System," Semiannual Technical Report, Report No. 740, Image Processing Institute, University of Southern California, March 1977, pp. 79-88.

58. W.K. Pratt, "Spatial Transform Coding of Color Images," IEEE Trans. on Commun. Tech., vol. COM-19, No. 6, pp. 980-992, December 1971.

59. W.K. Pratt, "Block Quantization Bit Assignment," submitted to IEEE Trans. on Info. Th.

60. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, New York: John Wiley & Sons, 1973.

61. D.E. Pearson, "Methods for Scaling Television Picture Quality: A Survey," in Picture Bandwidth Compression, T.S. Huang and O.J. Tretiak, Eds., New York: Gordon and Breach, 1972.

62. W. L. Hays and R. L. Winkler, Statistics, vol. II, New York: Holt, Rinehart and Winston, 1970.

63. G.L. Walls, The Vertebrate Eye and Its Adaptive Radiation, Bloomfield Hills, Michigan: The Cranbrook Press, 1942.

64. J.W. Goodman, Introduction to Fourier Optics, San Francisco: McGraw-Hill, 1968.

65. L.A. Riggs, "Visual Acuity," in Vision and Visual Perception, C.H. Graham, Ed., New York: Wiley, 1965.

66. G. Westheimer, "Optical Properties of Vertebrate Eyes," in Handbook of Sensory Physiology, vol. VII/2, M.G.F. Fourtes, Ed., New York: Springer-Verlag, 1972.

67. M. Kabrisky, A Proposed Model for Visual Information Processing in the Human Brain, Urbana: University of Illinois Press, 1966.

68. F.W. Campbell and R.W. Gubisch, "Optical Quality of the Human Eye," J. Physiology, vol. 186, No. 3, pp. 558-578, October 1966.

69. A.W. Snyder and W.H. Miller, "Photoreceptor Diameter and Spacing for Highest Resolving Power," JOSA, vol. 67, No. 5, pp. 696-698, May 1977.

70. I. Abramov and J. Gordon, "Vision," in Handbook of Perception, vol. III, E.C. Carterett and M.P. Friedman, Eds., New York: Academic Press, 1973.

71. D.J. Aidley, The Physiology of Excitable Cells, Cambridge: The University Press, 1971.

72. W.A.H. Rushton, "Peripherial Coding in the Nervous System," in Sensory Communication, W.A. Rosenblith, Ed., Cambridge, Mass.: MIT Press, 1961, pp. 169-181.

73. S.S. Stevens, "Neural Events and the Psychophysical Law," Science, vol. 170, No. 3962, pp. 1043-1050, December 1970.

74. S.S. Stevens, "The Psychophysics of Sensory Function," in Sensory Communication, W.A. Rosenblith, Ed., Cambridge, Mass.: MIT Press, 1961, pp. 1-33.

75. W.R. Uttal, The Psychobiology of Sensory Coding, New York: Harper and Row, 1973.

76. R.L. DeValois, et. al., "Analysis of Response Patterns of LGN Cells," JOSA , vol. 56, No. 7, pp. 966-967, July 1966.

77. P. Padmos and D.V. Norren, "Increment Spectral Sensitivity and Color Discrimination in the Primate, Studied by Means of Graded Potentials from the Striate Cortex," Vision Research, vol. 15, No. 10, pp. 1103-1113, October 1975.

78. D.H. Hubel and T.N. Wiesel, "Receptive Fields of Single Neurones in the Cat's Striate Cortex," J. Physiology (London), vol. 148, No. 3, pp. 574-591, October 1959.

79. D.H. Hubel and T.N. Wiesel, "Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex," J. Physiology (London), vol. 160, No. 1, pp. 106-154, January 1962.

80. D.H. Hubel and T.N. Wiesel, "Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat," J. Neurophysiology, vol. 28, No. 2, pp. 229-289, March 1965.

81. D. H. Hubel and T.N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," J. Physiology (London), vol. 195, No. 1, pp. 215-243, March 1968.

82. S.W. Kuffler, "Discharge Patterns and Functional Organization of Mammalian Retina," J. Neurophysiology, vol. 16, No. 1, pp. 37-68, January 1953.

83. D.A. Pollen, et.al., "How Does the Striate Cortex Begin the Reconstruction of the Visual World?" Science, vol. 173, No. 3991, pp. 74-77, July 1971.

84. D.A. Pollen and J.H. Taylor, "The Striate Cortex and the Spatial Analysis of Visual Space," in The Neurosciences Third Study Program, F. Worden and F.O. Schmitt, Eds., Cambridge, Mass.: MIT Press, 1973, Ch. 21, pp. 239-247.

85. R.M. Boynton, "The Psychophysics of Vision," in Contemporary Theory and Research in Visual Perception, R.N. Haber, Ed., New York: Holt, Rinehart and Winston, 1968.

86. F.W. Campbell and D.G. Green, "Optical and Retinal Factors Affecting Visual Resolution," J. Physiology (London), vol. 181, No. 3, pp 576-593, December 1965.

87. F. Flamant, "E'tude de la Re'partition Delumiere Dans L'image Retinienne D'unne Fente," Rev. Opt. (The'or. Instrum.), vol. 34, pp. 433, 459, 1955.

88. G. Westheimer, "Visual Acuity and Spatial Modulation Thresholds," in Handbook of Sensory Physiology, vol. VII/4, D. Jameson and L.M. Hurvich, Eds., New York: Springer-Verlag, 1972.

89. K.J.W. Craik, "The Effect of Adaptation Upon Visual Acuity," British Journal of Psychology, vol. 29, Part 3, pp. 252-266, January 1939.

90. L. Kaufman, Sight and Mind, New York: Oxford University Press, 1974.

91. F. W. Campbell, et. a., "The Effect of Orientation on the Visual Resolution of Gratings," J. Physiology (London), vol. 187, No. 2, pp. 427-436, November 1966.

92. D.S. Gilbert and D.H. Fender, "Contrast Thresholds Measured with Stabilized and Non-stabilized Sine Wave Gratings," Optica Acta, vol. 16, No. 2, pp. 191-204, March 1969.

93. F.L. VanNes and M.A. Bouman, "Spatial Modulation Transfer in the Human Eye," JOSA, vol. 57, No. 3, pp. 401-406, March 1967.

94. J. Nachmias, "Effect of Exposure Duration on Visual Contrast Sensitivity with Square Wave Gratings," JOSA , vol. 57, No. 3, pp. 421-427, March 1967.

95. J.J. McCann, et. al., "Visibility of Continuous Luminance Gradients," Vision Research, vol. 14, No. 10, pp. 917-927, October 1974.

96. J. Hoekstra, et. al., "The Influence of the Number of Cycles Upon the Visual Contrast Threshold for Spatial Sine Patterns," Vision Research, vol. 14, No. 6, pp. 365-368, June 1974.

97. R.L. Savoy and J.H. McCann, "Visibility of Low-spatial-frequency Sine-wave Targets: Dependence on Number of Cycles," JOSA, vol. 65, No. 3, pp. 343-350, March 1975.

98. O. Este'vez and C.R. Cavonius, "Low-frequency Attenuation in the Detection of Gratings: Sorting out the Artifacts," Vision Research, vol. 16, No. 5, pp. 497-500, 1976.

99. P.A. Liebman, "Microspectrophotometry of Photoreceptors," in Handbook of Sensory Physiology, Vol. VII/1, H.J.A. Dartnall, Ed., New York: Springer-Verlag, 1972.

100. Y. LeGrand, Light, Colour and Vision, New York: John Wiley & Sons, 1957.

101. W.A.H. Rushton, "Visual Pigments in Man," Sci. Amer., vol. 207, No. 5, pp. 120-132, November 1962.

102. G. Wald, "The Receptors for Human Color Vision," Science, vol. 145, No. 3636, pp. 1007-1017, September 1964.

103. D.L. MacAdam, "Color Essays," JOSA, vol. 65, No. 5, pp. 483-493, May 1975.

104. C.A. Padgham and J.E. Saunders, The Perception of Light and Color, New York: Academic Press, 1975.

105. L.M. Hurvich and D. Jameson, "An Opponent Process Theory of Color Vision," Psychological Review, vol. 64, No. 6, pp. 384-404, November 1957.

106. G.J.C. Van Der Horst and M.A. Bouman, "On Searching for "Mach Band Type" Phenomena in Color Vision," Vision Research, vol. 7, Nos. 11/12, pp. 1027-1029, November 1967.

107. D.G. Green and M.B. Fast, "On the Appearance of Mach Bands in Gradients of Varying Color," Vision Research, vol. 11, No. 10, pp. 1147-1155, October 1971.

108. D.G. Green, "The Constrast Sensitivity of the Colour Mechanisms of the Human Eye," J. Physiology, vol. 196, No. 2, pp. 415-429, May 1968.

109. D. H. Kelly, "Visual Responses to Time-dependent Stimuli. I. Amplitude Sensitivity Measurement," JOSA, vol. 51, No. 4, pp. 422-429, April 1961.

110. D.H. Kelly, "Flicker Thresholds," Proc. of Symp. on Information Processing in Sight Sensory Systems, Pasadena, California, November 1965, pp. 162-176.

111. D.H. Kelly, "Pattern Detection and the Two-dimensional Fourier Transform: Flickering Checkerboards and Chromatic Mechanisms," Vision Research, vol. 16, No. 3, pp 277-287, 1976.

112. H. DeLange Den, "Research into the Dynamic Nature of the Human Fovea-cortex Systems with Intermittent and Modulated Light. I. Attenuation Characteristics with White and Colored Light," JOSA, vol. 48, No. 11, pp. 777-784, November 1958.

113. P. Tynan and R. Sekuler, "Perceived Spatial Frequency Varies with Stimulus Duration," JOSA, vol. 64, No. 9, pp. 1251-1255, September 1974.

114. J.J. Kulikowski and D.J. Tolhurst, "Psychophysical Evidence for Sustained and Transient Detectors in Human Vision," J. Physiol. (London), vol. 232, No. 1, pp. 149-162, July 1973.

115. D.J. Tolhurst, "Sustained and Transient Channels in Human Vision," Vision Research, vol. 15, No. 10, pp. 1151-1155, October 1975.

116. B.G. Cleland, et. al., "Sustained and Trasient Neurones in the Cat's Retina and Lateral Geniculate Nucleus," J. Physiol. vol. 217, No. 2, pp. 473-496, September 1971.

117. J.L. Brown, "Flicker and Intermittent Stimulation," in Vision and Visual Perception, C.H. Graham, Eds., New York: Wiley, 1965.

118. R.A. Young, "Some Observations on Temporal Coding of Color Vision; Psychophysical Results," Vision Research, vol. 17, No.8, pp. 957-965, 1977.

*U.S.Government Printing Office: 1978 — 757-080/119