USCIPI REPORT #1055

Stochastic Singular Value Decomposition Texture Measurement for Image Classification

by

Behnam Ashjari

February 1982

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 400
Los Angeles, CA 90089-2564 U.S.A.

ACKNOWLEDGEMENTS

Sincere gratitude is presented to the individuals who graciously took steps and accompanied me along the path leading to the accomplishment of this dissertation.

Dr. William K. Pratt, now of Vicom Systems, Incorporated, continued his guidance, as the chairman of my dissertation committee, after his leave from USC. His suggestions on the practical portion of the thesis and his advice throughout my entire graduate career have been truly helpful.

Appreciation is offered to the math member of my committee, Professor Theodore E. Harris, for a reading of the thesis and his encouraging comments on the theoretical work. The interest of the other member of the committee, Dr. A. A. Sawchuk, is also acknowledged.

While it is not possible to mention all the people who assisted in the preparation of this document, special thanks are given to Corinne Leslie for her rapid and accurate word processing of the manuscript; Ray Schmidt for a splendid photographic work; Paul Liles for providing computer consultation; Dr. Keith Price for his remarks on computer graphic devices; and Professor George A. Bekey for his assistance and cooperation.

Most of the applications and experiments of the research have been performed at the USC Image Processing Laboratory. The USCIPI is thanked for providing support and facility.

Cooperation and hospitality of members of the Image Processing

Group at Institut de Recherche d'Informatique et d'Automatique (now INRIA) Recquencourt, France, particularly, Drs. Olivier Faugeras and Andre Gagalowicz during a six month visit, at the onset of this project, is hereby sincerely thanked.

Production of a doctoral dissertation is a thought provoking experience: its research portion satisfies the instinctive childish curiousity, its organization and structure caresses the human sentiment for construction, and its writing rewards the student with strengthening of vocabulary and learning to communicate formally. The theoretical aspects (random matrices) and engineering applications (accurate and inexpensive feature extraction) of this work became particularly compelling to me.

Finally, the contribution of my family to this work has been significant. Deepest affection is expressed to them for their encouragement and love.

Behnam Ashjari

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

LIST OF DERIVATIONS

LIST OF DERIVATIONS (continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

ABSTRACT

The purpose of this study is to develop a highly accurate technique for feature extraction with low computationality requirements. The particular application explored in this work is image classification. Other possible applications include fast texture segmentation and the detection of foreign texture in a textural background.

The singular value decomposition (SVD), a technique of unitary matrix transformation, has been used for extracting features from a texture field. From a large (512x512) texture field with correlation among its pixels, small (32x32) sample matrices are randomly selected. The sample matrices (texture windows) are non-overlapping. Upon SVD transformation on each sample window, a set of (32) singular values are obtained. The singular values contain much of the information regarding correlation content of matrix elements and their interrelationships. In reality, the SVD reduces two-dimensional processing to one-dimensional, resulting in a substantial saving in computation. Feature selection is performed on the vector of singular values to further reduce its dimensionality to four or less. The reduced dimensional singular value vectors are then used for image classification.

The singular value decomposition is utilized in a stochastic context and the problem of textural feature extraction is approached from a statistical viewpoint. A connected theory based on stochastic SVD is developed for deriving probability functions of a

bidirectionally correlated random texture field and probability functions of its singular values.

A family of SVD textural features is introduced. The features' performance are evaluated, individually and in various combinatorial forms, in terms of their strength in texture classification. The classifier is Bayesian and its error criterion is the Bhattacharyya distance measure.

Experiments are performed on two types of textures: artificial and natural. For the first type, bidirectionally correlated artificial (computer generated) textures are used. This set of experiments provides an evaluation for the SVD features in a controlled environment. For the second type, four natural textures are used to provide a real-world evaluation of the features. The SVD features classify all of the textures with low probability of error. Computational requirements are significantly lower than previously developed textural feature extraction techniques.

PREFACE

The research component of this dissertation has been analysis of stochastic Singular Value Decomposition and investigating it on the real world problem of texture measurement. This document is a product of research performed at the Image Processing Institute, University of Southern California.

The dissertation is divided into four parts, including ten chapters and seven appendices. Part one provides the background and includes the first two chapters:

Chapter 1 presents an introduction to the dissertation.

Chapter 2 reviews some of the important related approaches in texture analysis and feature extraction. Advantages and disadvantages of existing techniques are discussed; and, the approach and methodology of this study is specified. Particularly, an attempt has been made to define the concepts of "texture," and "feature extraction."

Part two provides the theoretical framework and includes Chapters 3, 4, 5 and 6.

Chapter 3 provides the definition for the theory of singular value decomposition. It then proceeds with SVD perturbation theory and Jacobian of SVD. Appendices A through C support this chapter.

Chapter 4 introduces texture modeling, the stochastic model used in this work, and its justification. Appendices D and E compliment this chapter.

The next two chapters are devoted to a stochastic/statistical

treatment of the texture fields and their singular values.

Chapter 5 presents stochastic singular value decomposition. This chapter derives the probability density and distribution functions of random texture fields, the joint probability density function of the singular values of a random field, and finally, probability functions of the largest singular value of a random texture field. Appendix F supports this chapter.

Chapter 6 deals with texture multivariate statistical derivations via SVD.

Part three concentrates on application and experiments. Computer generation of artificial image textures and experiments with simulated (artificial) and natural textures are performed:

Chapter 7 presents SVD texture measurements and feature extraction.

Chapter 8 applies the mathematical developments of the previous chapters to the problem of image classification. Evalution of the SVD texture feature extraction method in terms of defined criteria is also the subject of this chapter.

Chapter 9 investigates the SVD technique for classification of natural textures. Appendix G compliments this chapter.

Chapter 10 discusses the applications of this work, and some concluding remarks and suggestion for further research are presented.

Part four includes a glossary of symbols and abbreviations, appendices A through G and a reference section.

CONVENTIONS

Throughout this thesis, the symbol [ ] is used for referencing

and the symbol ( ) is used for equation numbering. Figures are represented like equations except that they don't have the ( ). Derivations are represented with the chapter number followed by a point and then the derivation number. Sub-derivations are subsequently numbered after their pertinent derivation. For example, [42] means reference 42, (3-11) means equation 3-11. (D-7) means equation 7 in appendix D. 4-1 means Figure 4-1, and 4.2 means derivation 4.2. Derivation 4.2-1 is the first sub-derivation for derivation 4.2.

Throughout this work, the word "sample" with its mathematical statistics implications is used which essentially means that a set of several vectors are taken from a population. In engineering, a "sample" is usually referred to one vector from a population.

Matrices are represented with underlined capital and vectors with underlined small letters, e.g. $\underline{F}$ and $\underline{f}$ means matrix F and vector f. A list of selected symbols are presented in the glossary. Three separate lists for derivations; figures; and tables are presented after the table of contents.

PART ONE: BACKGROUND

Chapters 1 and 2 present the description of the problem, motivation to this work, and historical background. A survey on texture analysis and texture feature extraction techniques are the topics of this part.

CHAPTER 1

INTRODUCTION

During the past two decades, many classical signal processing techniques for processing time series and one-dimensional continuous time signals were modified for two-dimensional data, and image processing emerged as an independent field of study within electrical engineering. In the past few years, image analysis has developed into a significant area of research within image processing. Image analysis involves the extraction of useful information and measurements from an image field. For this purpose, it overlaps with pattern recognition and artificial intelligence. Because of the enormous amount of information in a picture, and the high level of correlation of pictorial data, extraction of "image features" is a necessary and important concept of image analysis. The term "feature" is commonly used to denote a particuluar combination of attributes or characteristics that play an intermediate role in the interpretation of an original picture. One type of feature commonly used in image analysis is texture. Texture will be defined in Chapter 2. There have been several studies on image texture feature extraction and various theories and techniques have been developed on the subject. The present study is concerned with a new technique.

A brief description of the problem is presented in Section 1.1. Section 1.2 presents the motivation and Section 1.3 follows with the historical background of the present work.

1.1 SVD Texture Processing

Singular Value Decomposition (SVD) is a numerical technique of matrix transformation, which has been used as a linear algebraic method for obtaining least square solutions for a set of homogeneous equations and matrix pseudo-inversion [1-2]. In image processing, SVD has been used for image restoration [3-4]. The SVD technique has also the promise of being applicable to image feature extraction [5-6].

The mathematical definition of the SVD is

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^{T} \qquad (1\text{-}1)$$

where, the matrix $\underline{F}$ is decomposed into three matrices $\underline{U}$, $\underline{S}$, and $\underline{V}$. Matrices $\underline{U}$ and $\underline{V}$ are unitary and $\underline{S}$ is a real diagonal matrix.

The number of non-zero singular values of a matrix is equal to its rank. The singular values display the rank in a quantitative form [17].

The singular values of a matrix can be considered as descriptors or features of the matrix elements and their inter-relationships. If a matrix is composed of randomly chosen real numbers, the singular values will tend toward equality. On the other hand, a highly structured matrix with a high degree of correlation and dependence will exhibit a few dominating singular values. However, not every matrix is suitable for SVD analysis. It will be amply discussed later that a highly structured matrix such as an image of man-made objects cannot effectively be analyzed with SVD. On the other hand,

a matrix of random values such as an image texture can efficiently be described by the SVD. Suitability of an image for SVD will be discussed in Chapter 7. Considering $\underline{F}$ as a texture field, the application of the SVD in feature extraction will be apparent from the following discussion.

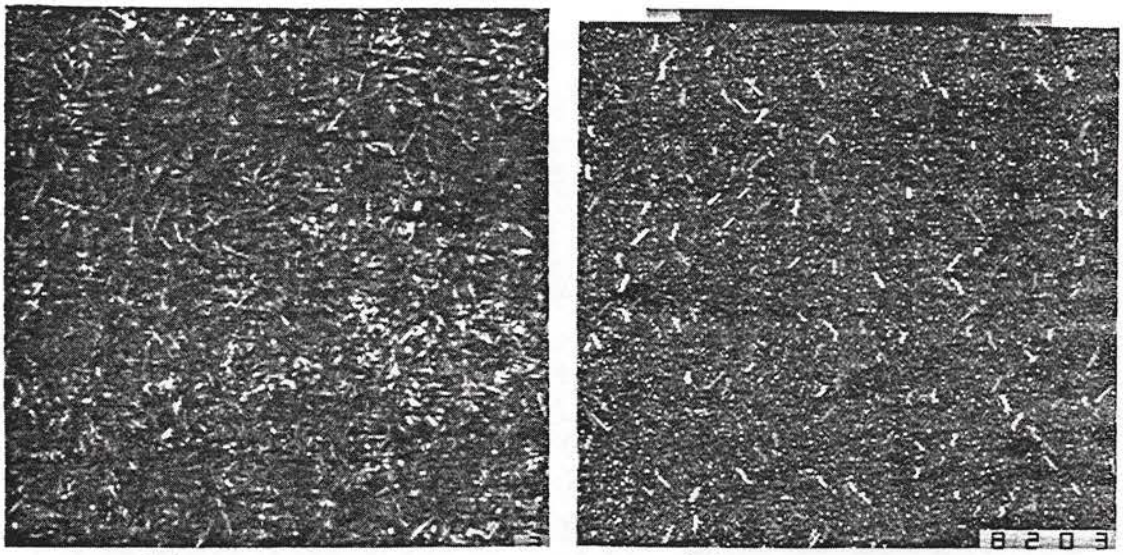
Since the elements of the matrix $\underline{F}$ are random variables, obviously, the elements of the matrices $\underline{U}$, $\underline{S}$, and $\underline{V}$ will also be random variables. In this work, it has been verified that the variation of the elements of the matrix $\underline{F}$ is mostly reflected in the singular values of $\underline{F}$ rather than in the elements of the $\underline{U}$ and $\underline{V}$ matrices. In other words, the singular values contain most of the essential information. A window taken from the texture image can be modeled as a random field whose essential information is preserved in a vector with singular values of the window as its components. It is obvious that a random vector whose components are the singular values can be used as a feature vector which serves a dual purpose: dimensionality reduction from N^2 to N, resulting in faster and less expensive feature selection and better performance in terms of classification accuracy. The subsequent chapters will cover the technique and its application in more detail.

1.2 Motivation

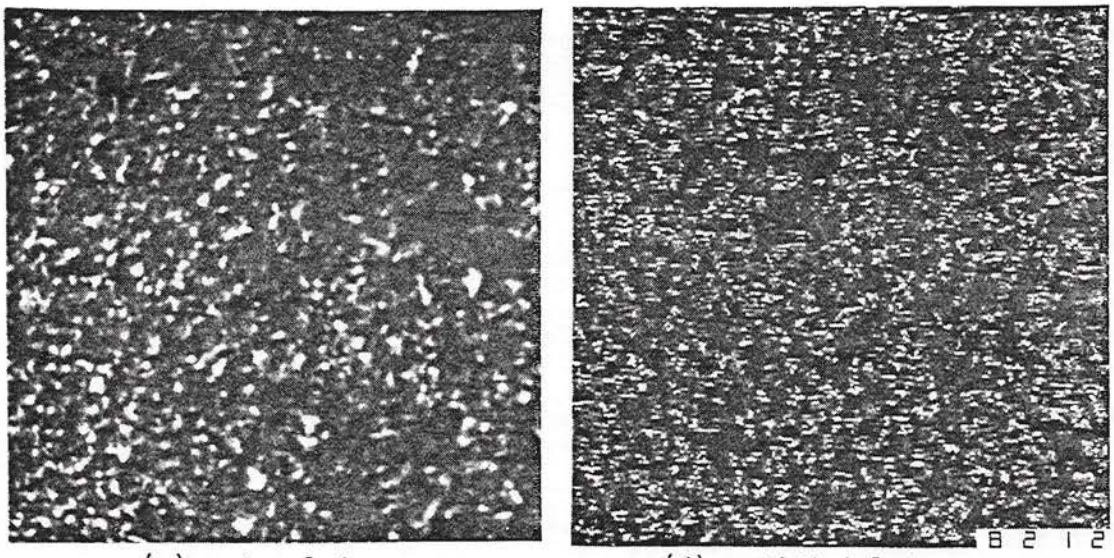
Since the advent of the digital computer, the application of numerical techniques has grown considerably for extracting useful information from large matrices such as pictures. SVD is one of the useful numerical techniques for dealing with large matrices. The primary experiments performed on natural and artificial grass and ivy

texture fields were encouraging [5]. The prominent result was that similar looking textures have similar ordered singular value distribution curves. Figure 1-1 contains examples of natural and artificial textures. The artificial textures were created by a random number generator with superimposed randomly placed spikes and blobs. Figures 1-2 and 1-3 depict their ordered singular value distributive curves. It was conjectured [Ibid] that the singular values of an uncorrelated texture field would be relatively equal and its distribution curve would be flat, and that of a highly correlated texture field would be relatively steep. The steepness could be attributed to the low rank of such a field, which would produce a few dominant singular values and the rest would be aproximately zero. Figure 1-4 shows the qualitative behavior associated with such a conjecture [Ibid].

A very realistic model for non-structural image texture, especially, for natural texture is a stochastic one. Unfortunately, there exists neither a stochastic nor a statistical treatment of the singular value decomposition. The present work is intended to serve two purposes: first, to develop a unified mathematical theory based on stochastic and statistical formulation of the singular value decomposition; and, second, to apply SVD in the physical world for devising an accurate, fast and, economical technique for image feature extraction. The results obtained in this dissertation have immediate engineering application in image processing and pattern recognition for classifying, recognizing, and segmenting two dimensional information.



(a) natural grass (b) artificial grass



(c) natural ivy (d) artifivial ivy

Figure 1-1. Natural and Artificial Grass and Ivy.

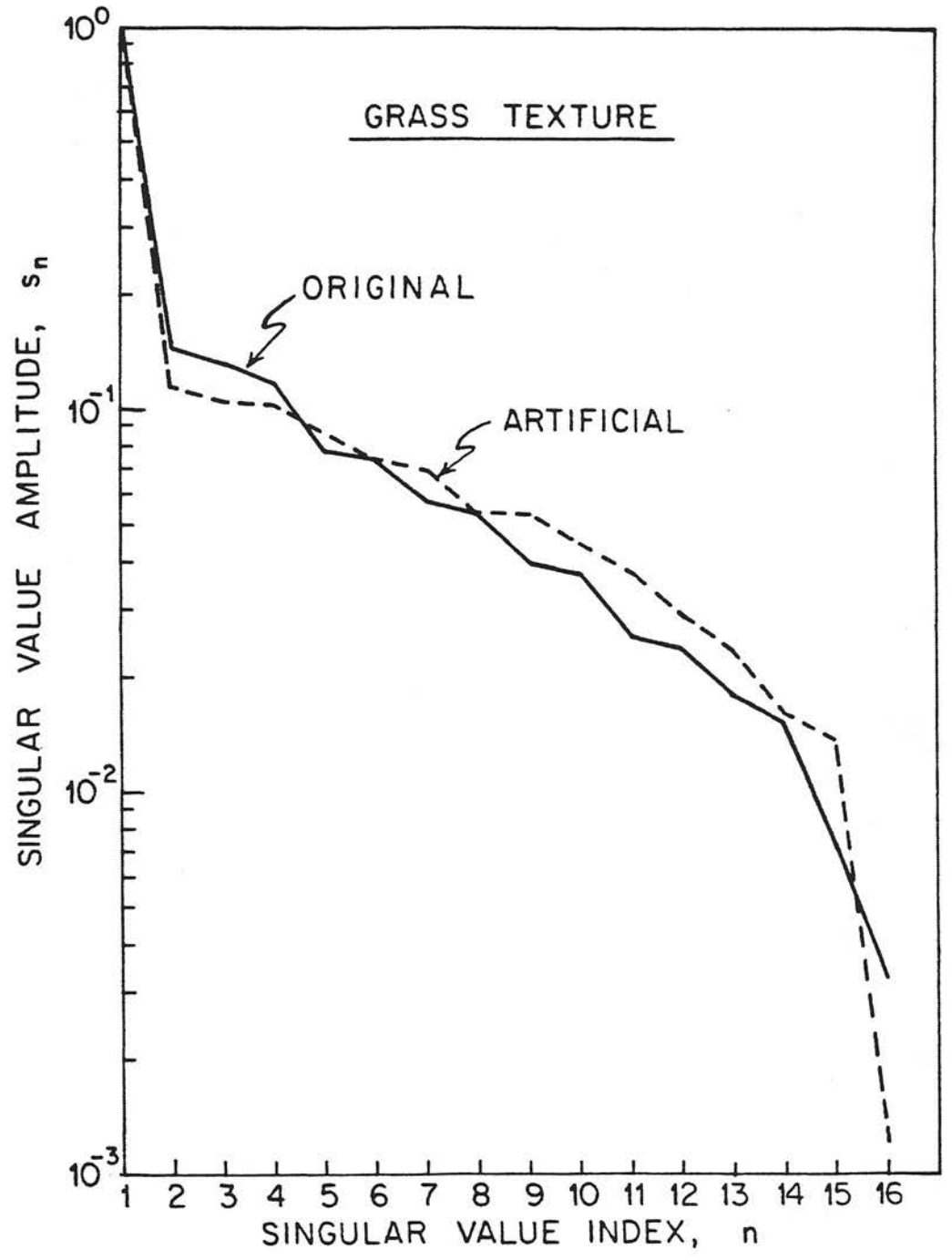


Figure 1-2. Singular Values of Natural and Artificial Grass Texture.

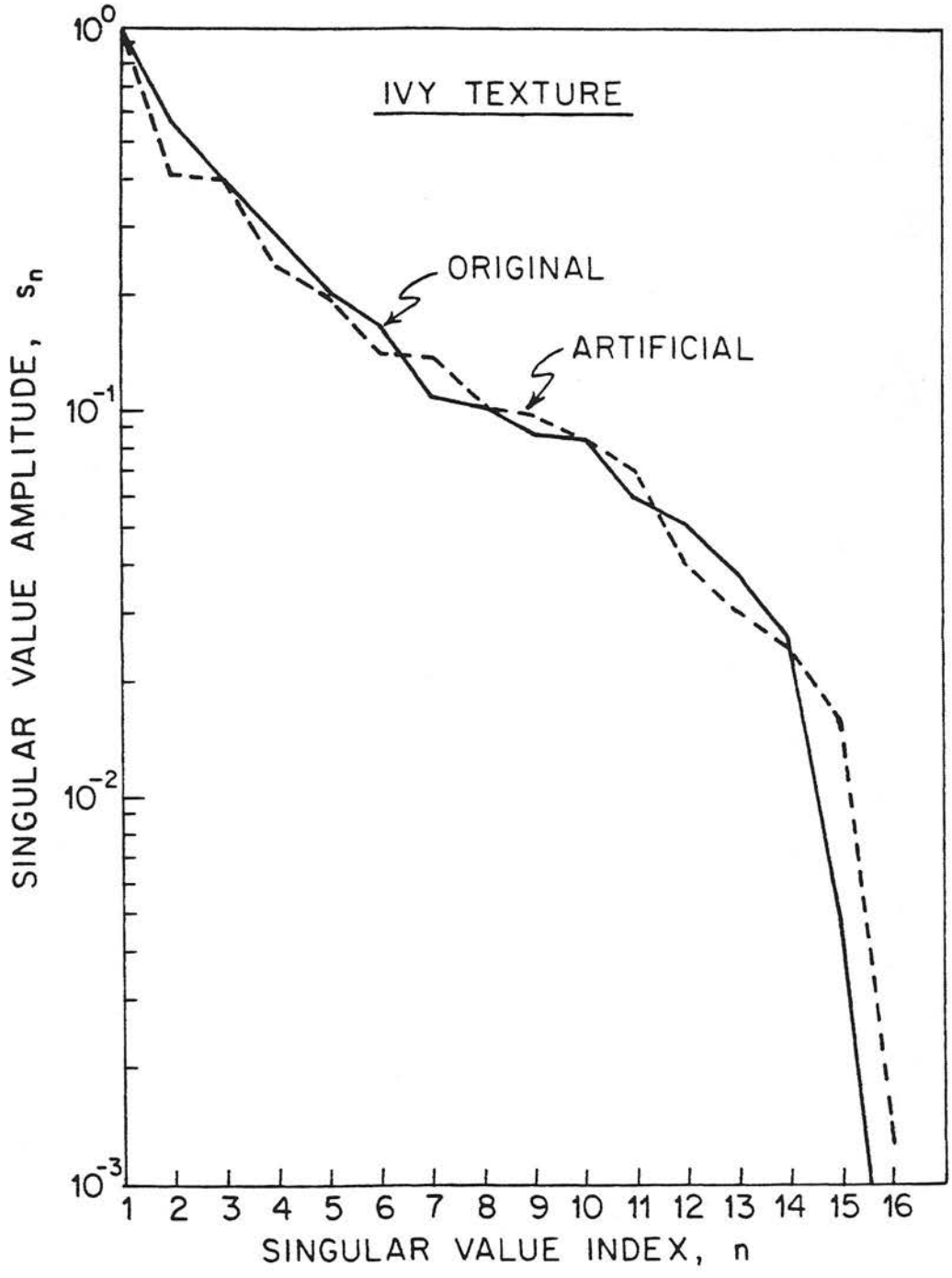


Figure 1-3. Singular Vaules of Natural and Artificial Ivy Texture.

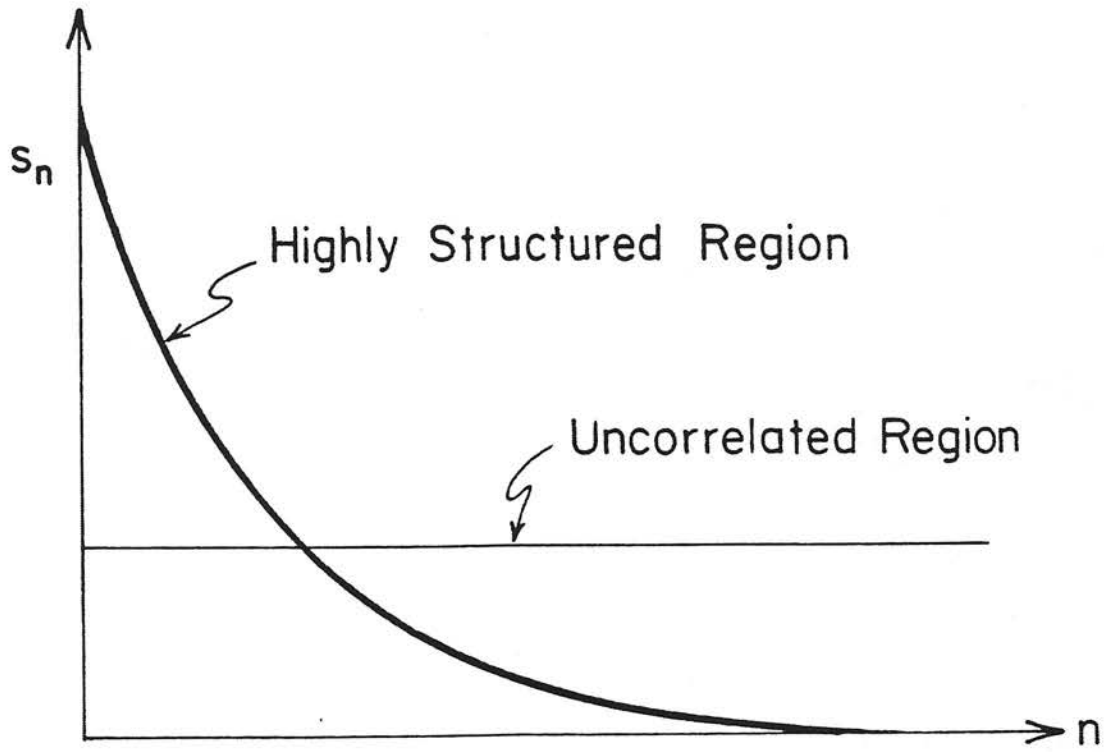


Figure 1-4. Example of Qualitative Singular Value Behavior of Image Regions.

1.3 A Brief Historical Review of Random Matrix Theory

Few works can be found concerning random matrices. From a historical standpoint, the theory involving samples from population with a single-variate, especially when the population was normal had been developed in the late 19th and early 20th centuries. The interest in obtaining the simultaneous distribution of the covariance matrix in samples from a multivariate population had grown in the scientific community since 1900. However, the extension from single-variate to multivariate developments was made rather slowly. In 1908, 'Student' considered the first four moments of samples taken from a population, assumed for simplicity to be normal, and applied Karl Pearson's method to infer the distribution of standard deviation [7].

In 1915, Ronald Fisher solved the problem for the bivariate case. This was the probability distribution function of a 2x2 sample covariance matrix of a set of independent observations from a 2-variate normal population [8](*). Fisher, later, extended his work to the 3-variate case. It was not until 1928 that John Wishart found the same distribution for a general multivariate case [9]. The Wishart distribution is the distribution of a k-dimensional sample covariance matrix in samples from a k-variate normal population. The sample covariance matrix is defined as follows:

(*) In engineering literature, sample covariance is sometimes referred to as experimental covariance.

$$\underline{X}\,\underline{X}^T = \sum_{i=1}^{n} (\underline{x}_i - \underline{\mu}_x)(\underline{x}_i - \underline{\mu}_x)^T \qquad (1\text{-}2)$$

where,

$$\underline{X} = [\underline{x}_1 - \underline{\mu}_x\ , \underline{x}_2 - \underline{\mu}_x, \ldots, \underline{x}_n - \underline{\mu}_x)T \qquad (1\text{-}3)$$

knowing that,

$$\underline{x}_i = N(\underline{\mu}_x, \underline{K}_c) \qquad \text{and} \qquad (1\text{-}4)$$

$$\underline{x}_i \perp\!\!\!\perp \underline{x}_j \quad \text{for } i \neq j \qquad (1\text{-}5)$$

In the r.h.s. of (1-2), replacing $\underline{\mu}_x$ by $\bar{x}$, the sample mean vector and multiplying the summation by 1/n or 1/(n-1), respectively, gives the biased or unbiased estimate to the population covariance matrix, and is called the sample covariance matrix. Essentially, all of the above mentioned solutions arose from classification problems of different multivariate normal observations. The next set of problems that needed to be tackled were the distribution of the roots of certain determinantal equations which were required for various statistical tests of classification accuracy. One of the most important problems was to obtain the distribution of latent roots (eigenvalues) of a sample covariance matrix. Obviously, a sample covariance matrix is a positive definite symmetric matrix whose elements are random variables that vary according to certain laws. A special case of this problem, namely the case of uncorrelated population covariance ($\underline{K}_c = \sigma^2 \underline{I}_k$) (see equation 1-4) was solved approximately simultaneously in 1939 by Fisher [10] in Britain, Girshick [11] Hsu [12], Mood [13], in the United States, and Roy [14] in India, although Mood's paper was not published until 12 years later.

In 1960, James [15] derived the distribution of the eigenvalues

of a sample covariance matrix for a general population covariance matrix. The latter distribution was made possible as a result of the discovery of zonal polynomials.

In 1959, Faddeev [16] solved a special form of a perturbation problem. He assumed that the perturbation values $[da_{ij}]$ of a deterministic matrix $\underline{A} = [a_{ij}]$ were uncorrelated with mean zero and standard deviation σ and solved for the covariance matrix of the perturbation characteristic roots $d\lambda_i$. Although Faddeev's derivation is interesting it has the following differences with our goal: (a) it is in terms of eigenvalues not singular values (the two are entirely different for a general matrix), (b) the covariance terms of $d\lambda_i$ are also functions of the eigenvectors (we want a relation which eliminates the eigenvector variations), and (c) Faddeev's derivations are for uncorrelated case (we want a relation for correlated variables).

With such a perspective, this project commenced; and as its results will show, it has the promise of providing singular value decomposition as a technique for treating multivariate statistical concepts. At the beginning of this research, the problem of deriving joint probability density functions of the singular values of a bidirectionally correlated random matrix interested us. This problem will be solved in Chapter 6 as a byproduct of stochastic singular value decomposition.

CHAPTER 2

TEXTURE AND TEXTURE FEATURE EXTRACTION

Although there has been much research on image texture, there seems to be no unified quantitative definition for texture. Because of the importance of texture, this chapter has been partly devoted to the review and survey of texture analysis approaches. It is hoped that this chapter will provide sufficient background for the remainder of the dissertation.

2.1 Texture

Pickett [19] defines texture as:

"The underlying requirement of a pattern to be considered as texture is that many similar elements with roughly the same dimensions be arrayed over a region." Pickett [19] continues: "The basic requirement or an optical pattern to be seen as texture is that there be a large number of elements (spatial variation in intensity or wavelength), each to some degree visible, and, on the whole, densely and evenly arrayed over the field of view....." Pickett continues that "The variation may be continuous fluctuations in the position of a line...at an edge or along a margin...or they may be discrete elements overlapped with some characteristic distribution of shape, size, color, shading, and orientation." Hawkins [20] gives a more elaborate definition of texture based upon three ingredients: "(1) some local 'order' is repeated over a region which is large in comparison to the order's size; (2) the order consists in the non-random arrangements of elementary parts; and (3) the parts are

roughly uniform entities having approximately the same dimensions everywhere within the textured region." The above statements, however, do not imply a universal, quantitatively defined, mathematical definition for texture. Pratt [21] explains the absence of a direct relation between quantitative measures and qualitative descriptions of textures: "Although these descriptions of texture seem perceptually reasonable, they do not immediately lead to simple quantitative textures measures in the sense that description of edge discontinuity leads to the quantitative definition of an edge in terms of its location, slope angle, and height." Lack of precise and scientific definition of the term texture arises for two reasons according to Muerle [22]: "Visual texture is a property of pattern scenes which has escaped precise definition so far. Part of the problem is that a very precise definition has not yet been required" and there is "difficulty of achieving a consensus of the views of diverse group of professional people who are using texture."

Visual texture is not an absolute and independent phenomenon. Its appearance must be considered relative to a variety of factors such as distance of view and perception. Based on the distance that texture is viewed from, examples can be noted as a tissue section looked at with an electron microscope or the image of a group of cell as a lump, a wheat scale under a microscope or a wheat field in an aerial photograph, concrete viewed from a few feet distance and an aerial image of an urban area. Pictorial examples of textures include the Brodatz' textures [23] such as the pictures of woven aluminum wire, reptile skin, grass lawn, raffia weave, ceramic coated

brick wall, beach sand, water, woven brass mesh, wool, etc. In pictures of a human face, hair and skin have textural properties. For a data-base on digitized texture images, one can refer to [24]. (Figures in Section 2.1.1 will show pictorial examples of various types of texture.)

2.1.1 Definition

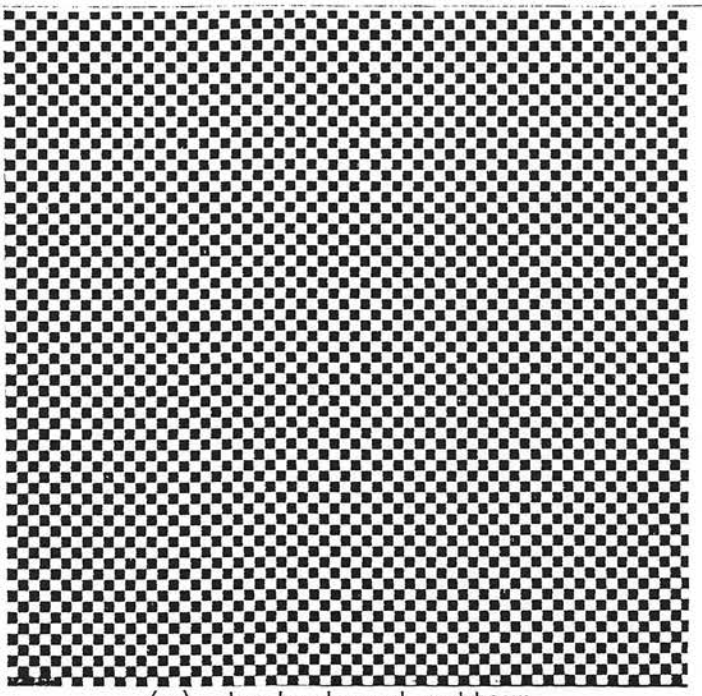
Texture can be categorized as either structural or non-structural. Structural texture can have a regular or irregular pattern. A non-structural texture is usually irregular. In the following discussion, these concepts will be formalized as an attempt to explain and define texture. Our categorization of texture arises mainly from a modeling and simulation point of view.

Structural Texture

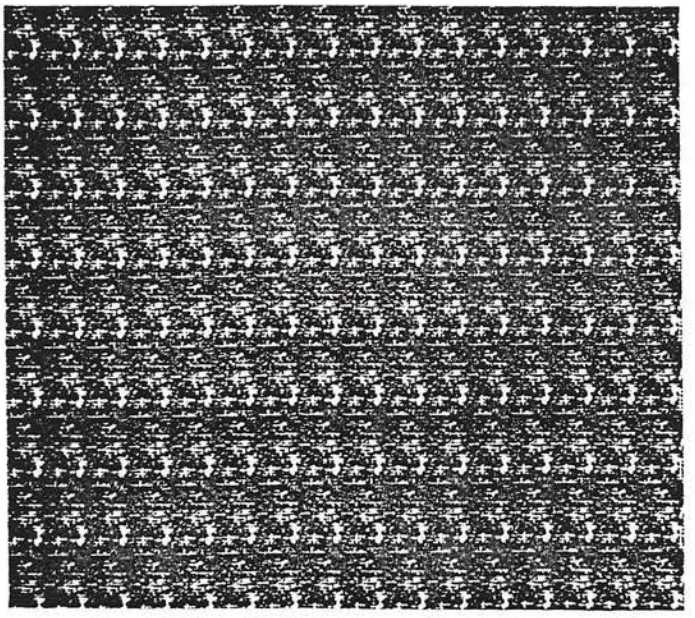
A structural texture is generally viewed as an image of organized elements. One of the requirements of texture to be categorized as structured is that its unit patterns must neither be too small to be unlaboriously countable nor so large that they are counted at a glance. By unit pattern, it is meant the smallest portion of texture whose repetition over a region reproduces the original structural pattern. In structural texture, there must be edge and discontinuity between neighboring unit patterns so that they are easily separated from each other by human eye. The unit patterns can either be identical or approximately identical. They can have a regular or irregular placement rule. For determining the identity of a unit pattern, a coordinate system which is invariant to the unit

patterns locations can be defined on attributes such as geometry, size, orientation, optical properties, phase, frequency and spatial relation to its neighboring elements. Thus, obtaining one point on the coordinate system for each unit pattern. We define subpattern as a portion of texture usually created by several unit patterns. The case of identical subpatterns exists when all such points fall over each other, while the case of approximately identical subpatterns arises when these points fall near each other resulting in a dense cluster. Generally, the less dense the points, the more irregular the appearance and vice versa. A regular structural texture can be uniquely reconstructed by a unit pattern and a simple placement rule. Examples of identical subpattern texture is given in Figure 2-1. In Figure 2-1(a) of the checker board pattern, every diametrically symmetric subpattern containing an equal number of black and white squares is capable of reproducing the original texture by its non-overlappig repetition over a region. The smallest of such subpatterns is a 2x2 square, which we call a unit pattern. In Fig. 2-1(a), there are two possible unit patterns which are in 90^{0} phase with each other w.r.t. orientation. Unit patterns may not necessarily be identical in their micro structure; but they may look so in their macro form. Figure 2-1(b) presents a VLSI network [154] whose subpatterns appear identical.

Approximately identical unit pattern texture arises when slight differences exists between corresponding values of the same attribute in unit patterns. By approximate it is meant that the differences (in units of pixels or gray tone) must be a fraction of the dynamic



(a) checkerboard pattern



(b) a VLSI network

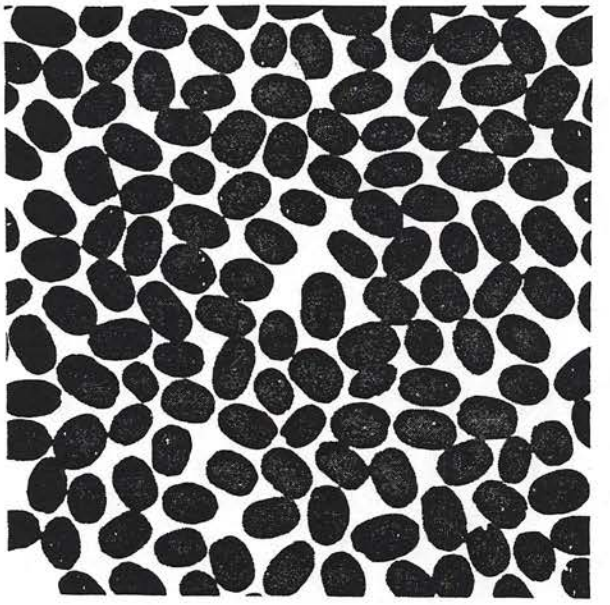
Figure 2-1. Example of Structural Texture with Identical Unit Patterns.

range of that particular attribute. An example of such a type of structural texture appears in Figure 2-2. Figure 2-2(a) shows the unit patterns having an elliptical shape with approximately the same dimensions. In addition, the placement of unit patterns differ in orientation. In most of the approximately-identical-unit-pattern structural textures, the variations in unit patterns and their placement is not regular and hence cannot be represented by a simple formula. On the other hand, there is usually some degree of randomness in such textural patterns. For example, in Figure 2-2(a) there seems to be a randomness in orientation θ of each elliptical unit pattern with θ being a uniformly distributed random variable between zero and π. Figure 2-2(b) shows an approximately identical unit-pattern to be constituted of a few polygons whose number of sides are randomly selected from an integer between 3 to 6. There is also randomness in the length and thickness of the sides.

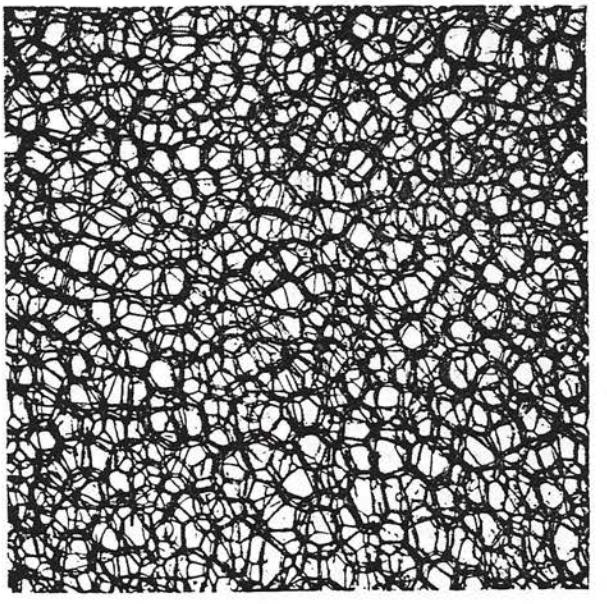
Non-Structural Texture

A homogeneous textural pattern appears non-structural if its unit patterns (defined above) are too small to be effortlessly visually countable and separable. In some cases, the unit patterns are as small as one pixel. One of the characteristics of a non-structural texture is that its percent duty cycle of repetition cannot readily be determined and is basically very low, while just the opposite is true for structural texture. Figure 2-3(a) shows beach sand as an example of non-structural texture.

A structural scene changes its characteristics when viewed from different distances. For example, the checker board pattern of



(a) coffee beans



(b) plastic bubbles

Figure 2-2. Examples of Structural Texture with Approximately Identical Unit Patterns.

Figure 2-1(a), if viewed from 4 to 5 feet distance with a 20/20 eye for at least five seconds, appears as a textural pattern of black and white stripes having 45^{o} angle ith vertical direction. This effect is because of the gestalt or grouping effect performed in the human visual system. The same texture viewed from 8 to 9 feet seems solid gray which is a non-structural scene. These observations show that in non-structural texture, the total effect of many unit patterns as a group on perceptual system is more important than properties of individual unit patterns.

A simple and regular non-structural texture can be modeled with a non-probabilistic formula. For example, for the texture created by putting black dots on intersection of every other two columns and rows of an otherwise white digital picture, the formula is

$$\frac{i \cdot j}{q} - \left[\frac{i \cdot j}{q} \right] \begin{cases} = 0 & \text{assign 0 (black)} \\ \neq 0 & \text{assign 255 (white)} \end{cases}$$

where I andj are row and column indices.

Three fundamental properties can be presented for non-structural texture: (a) In a digital textured image, the number of pixels is usually large in order to give enough repetition and outlook of the texture. This may contribute to the importance of the total (macro) effect of many pixels on the perceptual system in resemblence of the textural pattern. (b) Unit patterns are very small. Because of this property, the spatial relations and values of individual pixels are of concern throughout the texture. (c) Certain statistical

properties of the non-structural texture seem to closely relate to the appearance perceived by human eye. Among such statistics is correlation.

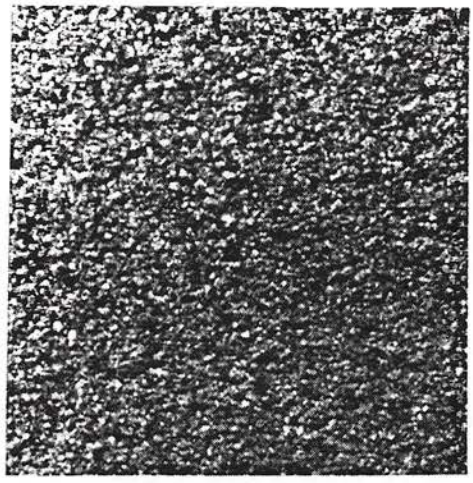
From the above properties, the following conclusions can be drawn: There is a large degree of freedom in choosing the pixel values to have the desired visual properties. Thus, for every pattern, there is a family of textures which can be perceived similarly. Such properties can be found in certain families of ergodic stochastic processes. By ergodic, we mean a stochastic process all of whose statistics (with probability 1) can be determined from a single member of the process. It is imperative to note that the term stochastic has nothing to do with scrambling and irregularities in textural pattern. A single texture can be modeled non-probabilistically, or if there is a random appearance in one or some of the attributes, a probabilistic factor can be used. Such a probabilistic factor must not be confused with the term stochastic process defined above. In stochastic notation, we deal with a class of textures in an ensemble sense. Depending on the generating model of stochastic process, the patterns in the class can be regular or irregular. For example, if the finite sequence of random variables representing the textures is normal and first order Markov with correlation factor of 0.96 in column direction and zero correlation along the rows, the pattern will be like that of Figure 2-3(d) which is obviously irregular. Another member of this family of stochastic processes can be generated using a different sequence of normal random variables with the same correlations as before. This pattern,

too, will look similar to Figure 2-3(d) and has the same statistical properties.

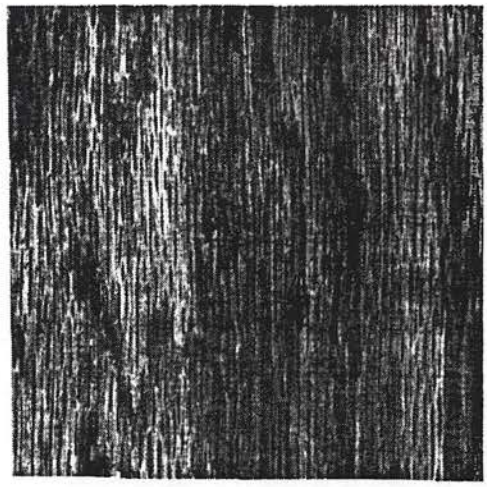
Thus, stochastic models are very important for characterizing non-structural textures. For a homogeneous stochastic texture, wide-sense stationarity seems quite a reasonable assumption and can be justified with the results obtained by Julesz et al. [31] and Pratt et al. [35]. Most aerial scenes are non-structural and can therefore be modeled stochastically.

In texture processing, there is a frequent referral to the term natural texture. Natural texture means a pattern whose scene has been created either by nature (grass and sand) or exists in every day life in a processes form (raffia weave and wool cloth). The pattern has then been photographed, digitized and transduced to numbers.

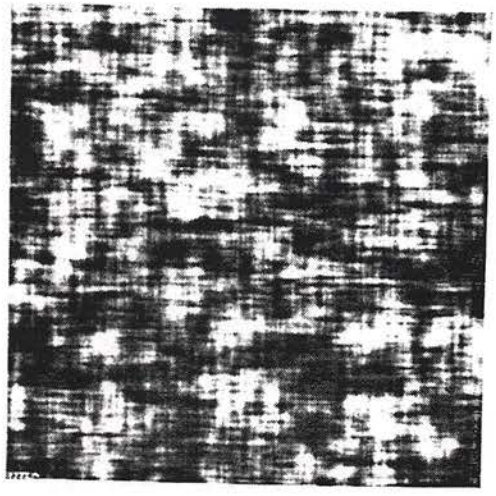
Artificial texture, on the other hand, refers to a pattern whose content is first generated by a machine in the form of numbers, line drawings, or light and then transduced to form an image such as Figures 2-3 (c and d). There are other types of textural patterns that do not logically fit into either natural or artificial categories. Such patterns are usually created by either industrial processes or by direct human intervention. Thus, we refer to the former as industrial natural textures and to the latter as man-made natural texture. An example of industrial texture is plastic bubbles in Figure 2-2(b) and one for the man-made texture is bricks organized in an orderly manner. Biological textures such as images of tissues can be categorized into natural texture patterns. It is hard to think of a natural texture whose pattern is regular. Barks on the



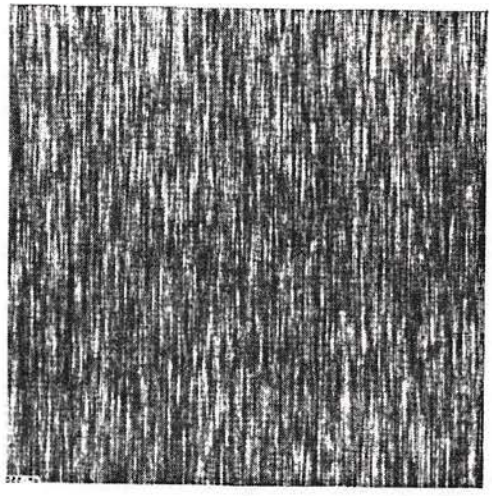
(a) natural and non-structural texture - beach sand



(b) natural texture - wood grain



(c) artificial, stochastic texture-bidirectional correlation $\rho_C = \rho_R = 0.96$



(d) artificial, stochastic texture - unidirectional correlation $\rho_C = 0.96$, $\rho_R = 0.0$

Figure 2-3. Examples of Non-Structural and Stochastic Textures.

trunk of a given tree (e.g. sequoia) having approximately the same age and on the same part of the world (e.g. California) have similar textural patterns. Photographed under similar lighting conditons with the same camera from the same distance (e.g. one foot) and going through a standard darkroom procedure, the bark images will look alike. Digitized with the same resolution, they look very close, although they are different when compared pixel by pixel. They form one class of textures which, like most of the natural textures, are best modeled by a certain family of stochastic processes. Figure 2-3(b) shows a natural wood grain. Such a texture is simulated by a normal, first order Markov model depicted in Figure 2-3(d). In this document, we are concerned with a stochastic non-structural view of texture and a statistical approach to feature extraction.

2.1.2 Texture Processing

Image texture can be studied at least around four topics:

i) Texture perception

ii) Texture synthesis

iii) Texture analysis

iv) Texture discrimination

For the first topic, more psychological/neurological research is needed on the eye and brain. Gibson [26] has shown the relation between texture cues and a sense of depth in perception. Metzger [25] first recognized the perception of texture in 1930, and since then this subject has provided stimuli for many scientists. Julesz [29-32], Pollack [33], Purks and Richards [34]; and Pratt, Faugeras, and Gagalowicz [35 and 141] have performed various experiments for

understanding the underlying statistical behavior of image texture. On the second topic, texture synthesis can be applicable in image coding where the coding process of synthesizable area of a segmented image becomes, potentially, more efficient than direct image coding. A method for generation of two-dimensional correlated texture fields will be introduced in Chapter 4. Recently, techniques have been developed to simulate some natural patterns using synthetic texture [142-143].

The third and fourth topics are concerned with performing feature extraction and useful measurement on textures for the purpose of image classification, segmentation and modeling. The last two topics of texture processing, namely, texture analysis and texture discrimination and the last two applications, namely, texture feature extraction and texture classification are the mainstream of the engineering application of this dissertation. Sections 2.2 and 2.3 present a brief survey on texture feature extraction.

2.2 Feature Extraction Theory

One of the most important topics of pattern classification is feature extraction(*). It is useful, for better understanding of this section, to define both terms, "feature" and "extraction".

(*)In the literature, this concept is alternatively referred to as feature selection.

Definition of Feature

Ledley [36] defines the term "feature" as follows: "The term feature is commonly used to denote a particular combination of attributes or characteristics that play an intermediate role in the interpretation of the original data, or picture in terms of the final pattern classification." Abdali [37] places the term feature into three related connotations: specialized measurements features, visual features, and pattern synthesizing features. In the first meaning, a set of measurements represent the pattern and each set of measurement is tested according to a "goodness" criteria resulting in a best combination of measurements. Features, in the second sense are used to denote some visual characteristic of patterns such as topological and geometrical. In the third connotation, the pattern is created from superposition of some unit patterns. In other words, features are subpatterns [38-39]. In texture processing, the last two interpretations of feature are useful for structural texture. Thus, because of the non-structural nature of this study, the type of features utilized in this document fit into the first category; i.e., a set of measurements on the texture field.

Extraction Defined

Feature extraction or selection is the process of mapping the original measurements into more effective and usually independent features. The mapping can be linear or non-linear and it, generally, must reduce the dimensionality. For the linear case, it is sufficient to find the coefficient of a linear function to maximize or minimize a criterion, or for the more complicated criteria, optim-

ization techniques can be applied to determine the coefficients [40].

In many applications, important features are nonlinear functions of the original measurements. Unfortunately, a reasonable general theory for nonlinear feature extraction is not available and analytical treatment of a nonlinear case is usually cumbersome. It can be said that most of the nonlinear feature extraction problems are solved heuristically, are, therefore, problem-specific. For more discussion on optimum feature extraction, see [41].

2.2.1 Evaluation of the Techniques

It was previously mentioned that features are tested according to some "goodness" criteria. The following are the criteria which are often used in pattern recognition.

i) Classification Evaluation Method:

This method involves measurement of classification error. There are two common approaches to this problem; namely, probability of error and separability.

Probability of Error

This criteria is popular because it is related to the Bayes classifier. The procedure is that the number of classification errors are counted experimentally. However, this criteria has several shortcomings. One major disadvantage is that an explicit mathematical expression is not available except for a very few special cases, making a theoretical analysis difficult. Even for the simplest one of these special cases, which is the normal distribution, the calculation of error requires numerical

integration, except for the equal covariance case. The second disadvantage is that this criteria requires the knowledge of the probability density function. When the distribution of samples in each class is not normal, the empirical derivation of the probability density function is tedious and especially erroneous when there are a small number of samples available [40].

Scatter Matrices for Discriminant Analysis

For multi-class classification problems, the features of interest maximize or optimize class separability. When there are only two classes, optimization reduces to maximization. In this type of classification problem, the goodness criteria is the separability function, which is independent of the coordinate system, but depends on the class distribution and the classifier to be used. Fukunaga [40] proposes five desirable conditions for separability criteria:

1) Monotonic relationship with the probability of error.
2) Monotonic relationship with the upper and lower bounds of the probability of error.
3) Invariance under one-to-one mappings.
4) Additive independent features.
5) Metric properties.

These criteria are simple, and the selection of optimum features is straightforward. The criteria are used for multi-class problems, and have explicit mathematical expressions. Scatter matrices consist of the within-class matrix, which shows the scatter of samples around their class expected vector and the between-class matrix, which shows the experimental covariance matrix of the class mean-vectors. One

well-known criterion of discriminant analysis is J = trace (between-class)/trace (within-class) [42]. For more types of criteria see [43].

One disadvantage of this type of feature selection is that as the number of classes increase, the criteria become more and more inaccurate.

ii) Figure of Merit Evaluation Method:

If more a priori information is available, such as the feature distribution for each pattern class, figure of merit can be used as an indirect measure for feature effectiveness [53]. This method has the advantage that it does not depend on any specific classification scheme, but it is related to bound on probability of misclassification. Some of the more common figure of merit criteria are divergence, entropy function, Bhattacharyya distance, and Chernoff bound. For more discussion on distance measures, see [44].

Bhattacharyya Distance(*)

B-distance is a scatter function of the conditional densities of features of two classes. It is monotonically related to the Chernoff bound of the probability of classification error using a Bayes classifier. For general cases, the B-distance and Chernoff bound criteria are not as practical as the probability of error because, first, they involve integration of the product of density functions, which could analytically be very difficult and, second, probability

(*) B-distance for simplicity.

of error has a direct physical meaning. Kailath [45] has derived explicit mathematical expressions for B-distance criteria using some well-known density functions. One of the density functions is the normal density of features for which the B-distance formula becomes substantially easier to work with, and it can be claimed that when there is a large number of samples from normal distributions, B-distance figure of merit is very efficient. For a more detailed discussion on this topic, see Chapter 7 where B-distance is explained as the criteria of evaluation in this study.

2.3 Some Important Texture Feature Extraction Techniques

Considerable research has been performed on the problem of texture analysis and various texture feature extraction techniques and theories have been developed which are of three forms: statistical, structural, and structural-statistical.

i) Statistical Texture Features:

This group of features is particularly useful for natural scenes and aerial images. It can be divided into four related categories whose differences are defined in terms of efficiency which usually means accuracy vs. cost. Measurements in this group can be made in spectral and spatial.

Spectral Texture Features

The most important class of features in this category is the extracts of the Fourier transform of image texture. The Fourier power spectrum in polar coordinate has a radial and angular distribution. For a wide-sense stationary field, the Fourier power spectrum is a real quantity and is equal to the magnitude square of the Fourier transform of the field. So, if ω_x and ω_y are the frequency coordinates of the Fourier transform, $F(\omega_x,\omega_y)$, r = radial $= \sqrt{\omega_x^2+\omega_y^2}$ and θ = angular = $\tan^{-1}(\omega_y/\omega_x)$. The radial distribution of a texture region gives a measure of coarseness. The finer the texture, the less correlation among its elements which, in turn,

results in a higher degree of variation in its autocorrelation function. It is well known that abrupt changes in a time function results in significant components in high values of its Fourier transform. Therefore, high variation in the autocorrelation function will result in the presence of significant components in high frequencies of the power spectrum. So the power spectrum's amplitude per frequency variation can provide a measure for texture fineness. The more power at high frequencies, the finer the texture. The angular distribution of the Fourier power spectrum is sensitive to the directionality of the texture. This concept is supported by the fact that angular distribution has high values concentrated around the perpendicular direction to the directionality of texture [48]. It shall be noted that the notion of directionality in texture plays an important role in texture processing because there is directionality in most of the natural textures such as mountains and agricultural fields. The frequency domain features have advantage over the spatial first order statistical features in terms of accuracy, but their performance is inferior to that of the spatial second order features to be discussed next. Such a disadvantage is in their high computational cost.

Spatial Texture Feature Based on First Order Statistical Measures

Examples of these types of features include image amplitude in terms of spatial values, luminance, tristimulus values, or other units. These features are often simple and practical. Mean and variance of an image are considered first order statistics. First

order histogram features are also among this group of features from which the mean, variance, skewness, and kurtosis of image can be derived [21]. Another in this category is gray-level difference method [153], which measures probability of differences between pairs of pixels at ends of a given displacement vector. First order techniques have advantages of simplicity and inexpensiveness, and their disadvantage is in their not taking arrangements of the pixels into consideration. To illustrate this shortcoming, one can create several completely different looking texture images with the same first order histograms.

Spatial Ad hoc Statistics

In this group, we include useful statistics which are neither first order nor second or higher order. An example is the gray-level run length method [48], which measures the length of consecutive pixels in the direction of a given vector, having the same gray level. A directional texture will have longer lines of consecutive pixels in its direction. In general, a coarser texture will have higher number of longer lines of equal pixel value. Another adhoc technique is a procedure using planar random walk for texture discrimination [148].

Spatial Texture Features on Second Order Statistics

Among this category, autocorrelation can provide a measure for coarseness based on the same reason given for Fourier power spectrum: The coarser a texture, the smoother its autocorrelation. Another use for the autocorrelation function is its ability to

measure periodicity of information in a picture: A spatially periodic texture will produce a periodic autocorrelation. An autocorrelation function can be computed by a direct method [21], or it can be obtained by frequency domain techniques. As another form of spatial domain feature, we can mention edge per unit area [49-50]. The finer the texture is, the higher the number of its edges will be per unit area.

The most important technique among spatial domain texture features is the family of second order joint density between all pixel pairs of a texture block, which is approximated by joint gray-level histrogram. Julesz et al. [29-32] and Pratt et al. [35] have shown that second order statistics are sufficient for human texture discrimination(*). Based on their results, second order gray-level statistics are considered very important for texture feature extraction theory and therefore, deserve some elaboration.

Joint Gray-Level Histogram Matrix(**)

Consider the (x,y) plane and a picture $f(x,y)$ defined over this plane. Let $\underline{\delta} = (\Delta x, \Delta y)$ be a vector in the plane. One can compute the joint probability density of any pair of gray-levels separated by the vector $\underline{\delta}$. For the discrete case, Δx and Δy are integers. By

---

(*)The strong sense of second order statistics is meant.

(**)In the literature, different names have been used for this concept. Some call it "gray-tone spatial dependence matrix," some have named it "gray-level co-occurrence matrix," and others have used "joint gray-scale histogram matrix."

counting the number of times that each pair of gray-levels occur at ends of $\underline{\delta}$, one can form an array of mxm where m is the number of gray levels present in the picture [51]. Therefore, a symmetric matrix is generated.

A simple set of separation vectors are considered to be (1,0), (1,1), (0,1), and (-1,1). This corresponds to the pairs of pixels with 0, or 180; 45 or 225; 90 or 270; and 135 or 315 degrees from the positive direction of the x-axis. As a consequence, there are four mxm matrices for each nxn sample window taken from the image, where n is the number of pixels in one row of the window, and m is the number of gray levels. If the texture is coarse, and $\underline{\delta}$ is small compared to the sizes of the texture element, the pairs of points at separation should usually have similar gray levels. This means that the high values in the joint gray level matrix should be concentrated on or near its main diagonal [48].

Let $P(i,j,\underline{\delta})$ be the $(i,j)^{th}$ element of the joint gray level histogram matrix of order m, which represents the number of occurrances of (i,j) at separation $\underline{\delta}$. Also, let S be a normalizing factor which is equal to the sum of all of the matrix entries. Then $P(i,j,\underline{\delta})/S$ is actually an approximation to the joint probability density function that gives the probability of the pair of gray levels (i,j) to occur at pairs of points separated by $\underline{\delta} = (\Delta x, \Delta y)$. As an example, a measure for correlation is

$$f_1 \triangleq \sum_{i=0}^{m} \sum_{j=0}^{m} (i \cdot j)(P(i,j,\underline{\delta})/S) \qquad (2\text{-}2)$$

where m is the number of possible gray levels.

Definition (2-2) is a discrete version of the two-dimensional continuous autocorrelation problem,

$$R_{vw}(\underline{\delta}) \triangleq E_{\underline{\delta}}\{vw\} = \int_{-\infty}^{\infty}\int vwf(v,w,\underline{\delta})dvdw \qquad (2\text{-}3)$$

where $f(v,w,\underline{\delta})$ is the joint probability density function of v and w with spatial separation vector $\underline{\delta}$. In formula (2-3), v and w are two random variables which can have any value and exist at any location within the assigned frame of boundaries on x-y plane as long as their locations are separated by the vector $\underline{\delta}$ (or $-\underline{\delta}$), and $f(v,s,\underline{\delta})$ being zero outside of the boundary.

f_1 from formula (2-2) can be considered a texture measure. Haralick et al. [51-52] have suggested 14 texture features based on the joint gray-level histograms which have proved to be effective in terms of classification accuracy. However, this family of texture features has its share of disadvantage. Its critics believe that two of the major problems with them are high computationality for obtaining the two-dimensional histograms, and poor accuracy for low contrast textures [53]. Davis [149] has studied an approach using generalized joint gray level histogram matrices for texture analysis. However, the generalized technique does not seem to improve the classification accuracy significantly while preserving its high computational cost.

ii) Structural Texture Feature:

For structural texture description, a formal language has been

developed by Garlucci [46]. This abstract linguistic technique for describing texture suggests a model where the primitives are line segments and polygons and their arrangements. Other structural texture features are based on artificial intelligence techniques such as finding topological features, e.g. connectivity in a pattern [47]. Davis [151] has investigated the structure of cellular textures. A methodology for feature extraction and reconstruction of structural texture has been proposed in [152]. All of these techniques deal with highly structured and usually man-made patterns of texture.

iii) Structural-Statistical Texture Features:

Features using a hierarchical mixture of structural and statistical attributes has been proposed [54]. The approach is structural in that it views texture as an organized phenomenon for which subpatterns and their spatial arrangements are taken into account. They are statistical in the sense that they extract local statistics from pixels. An example of this method is extraction of texture primitives and performing measurements to determine a rule of placement for them [150].

2.4 Summary

In this chapter, the notions of texture, feature, and extraction have been defined and explained, and various models for texture and notable techniques for its measurement have been examined. It was reasoned that because of the nature of variation in a non-structural textured pattern, specifically, that of a natural texture, a

statistical approach is presumed to be the most appropriate in our study. It was discussed that some advances has been made in extracting useful information from texture and the most famous of statistical texture features are the Haralick measures. However, Haralick features are computationally expensive, and do not much reduce the enormous redundancy in textural information. Some of the other techniques are algorithmic, i.e., they do not provide a mathematical insight into the nature of textural information. In the following chapters, based on a stochastic model for texture, we will attempt to present an accurate, fast, and inexpensive technique for texture modeling, measurement, feature extraction, and classification. Particularly, Chapter 7 is devoted to a family of SVD texture measures which provides good accuracy with low computationality. The evaluation of SVD measures will be performed in Chapters 8 and 9.

PART TWO: THEORY

This part includes Chapters 3, 4, 5, and 6. A model for image texture is established, and probabilistic and statistical developments on SVD are presented.

CHAPTER 3

DIFFERENTIAL SINGULAR VALUE DECOMPOSITION

In this chapter, we address a differential treatment of singular value decomposition. Such a treatment results in the Jacobian of SVD which will be used in subsequent chapters for derivation of p.d.f. and probably distributions of singular values. In Section 3.1, the notion and mathematical approach of singular value decomposition is defined. In Sections 3.2 and 3.3, differentiation of decompositions which are most useful for our later developments are presented.

3.1 SVD

Singular value decomposition is a powerful technique of matrix transformation. SVD has lately earned an important place in the fields or disciplines that deal with two dimensional fields. Among the areas of science that heavily utilize two-dimensional processing are digital image processing (for the obvious reason of two-dimensionality of image fields), and pattern recognition. In working with two-dimensional data, SVD can greatly simplify the task of extracting information from patterns. It can, also, provide us with a tool for obtaining analytical insight into the data. As it will be explained and proved in the subsequent sections of this chapter, SVD assists us in understanding many useful and important inter-relationships among elements of a texture field. In its special case, SVD analysis reduces to eigenvalue analysis. It is also possible to extend the time-invariant approach of this work to a time-invariant one for television applications. In this work,

efforts concentrated on the discrete time-invariant case which is sufficient for the classification.

3.1.1 Unique Definition and Notation

It is desired to obtain a precise definition for singular value decomposition. The reader may wonder why the adjective "precise" is used for "definition" since a mathematical definition is synonymous with preciseness or at least it should be. Section 3.1.3 on non-unique decomposition of singular values will shed light on why the word precise is used here.

Let $\underline{F}$ be a $\underline{k}\times\underline{n}$ ($k\leq n$) general complex matrix(*). It is possible to decompose $\underline{F}$ into a product of three matrices as

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T \qquad (3\text{-}1)$$

such that, $\underline{U}$ is a $\underline{k}\times\underline{k}$ unitary matrix, hence

$$\underline{U}^T\underline{U} = \underline{I}_k \qquad (3\text{-}2)$$

with the real part of the elements in the first row positive,

$\underline{S}$ is a $\underline{k}\times\underline{k}$ real diagonal matrix with its diagonal elements called

---

(*)Any development for $k\leq n$ can be extended to the case where $k>n$. However, the $k \leq n$ assumption has the advantage of making $\underline{F}\,\underline{F}^T$ a full rank matrix which is extensively used in this document.

singular values(*) of $\underline{F}$. They are non-negative and in descending order, where the number of non-zero singular values corresponds to the rank of $\underline{F}$. $\underline{V}$ is a nxk unitary matrix hence,

$$\underline{V}^T\underline{V} = \underline{I}_k \tag{3-3}$$

Then, the decomposition (3-1) of $\underline{F}$ is **unique.**

Remark: For derivations of Section 3.2 and on, without loss of generality, $\underline{F}$ will be considered real. This assumptions is based on the fact that the applications and experiments of this work are performed on digital pictures. Digital pictures are usually real valued arrays.

3.1.2 Linear Algebraic Properties of SVD

It can be proved that for

$$\underline{\Lambda} \triangleq \underline{S}^2 \tag{3-4}$$

we have,

$$\underline{F}\,\underline{F}^T = \underline{U}\,\underline{\Lambda}\,\underline{U}^T \tag{3-5}$$

and,

$$\underline{F}^T\underline{F} = \underline{V}\,\underline{\Lambda}\,\underline{V}^T \tag{3-6}$$

(*)In the literature, the name principal values has occasionally been used instead of singular values. The term outer product expansion also refers to SVD.

Remark: If F ix $\underline{kxn}$ ($k \leq n$) then the rank of $\underline{F}^T\underline{F}$ is $r \leq k$ which is not full. Thus, $\underline{V}$ has r nonzero eigenvalues.

Letting

$$\underline{\Lambda} \triangleq \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) \tag{3-7}$$

$$\underline{S} \triangleq \text{diag}(s_1, s_2, \ldots, s_k) \tag{3-8}$$

$$\underline{U} \triangleq [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k] \tag{3-9}$$

$$\underline{V} \triangleq [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k] \tag{3-10}$$

$\underline{u}_i \triangleq$ the ith eigenvector associated with $\underline{U}$ (3-11)

$\underline{v}_i \triangleq$ the ith eignevector associated with $\underline{V}$ (3-12)

r = rank of $\underline{F}$

Then, a hybrid form for SVD can be obtained as

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T = \sum_{i=1}^{r} s_i \underline{u}_i \underline{v}_i^T \tag{3-13}$$

or equivalently,

$$\underline{F} = \underline{U}\,\underline{\Lambda}^{1/2}\underline{V}^T = \sum_{i=1}^{r} \lambda_i^{1/2} u_i v_i^T \tag{3-14}$$

square root of λ_i can be positive or negative. However, by

agreement, the positive sign of $\lambda_i^{1/2}$ is taken as singular value s_i.

In Section 3.1.4, it will be explained how the relation (3-13) and (3-14) have found application in data compression.

Important Remark: Only for a non-negative definte, symmetric matrix are the singular values and eigenvalues equivalent. For this group of matrices $u_i = v_i$. For a general matrix, the two decomposition--singular value and eigenvalue--are entirely different.

3.1.3 Non-Unique Decomposition of Singular Values

Let $\underline{F}$ be a general complex $\underline{kxn}$ matrix. It is possible to decompose $\underline{F}$ into a product of three matrices as $\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T$ such that, $\underline{U}$ is a $\underline{kxk}$ unitary matrix, $\underline{S}$ is a $\underline{\mathbf{kxk}}$ real, diagonal matrix, and $\underline{V}$ is a $\underline{nxk}$ unitary matrix. It is, also, possible to decompose $\underline{F}$ as $\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T$ such that $\underline{U}$ is a $\underline{kxn}$ unitary, $\underline{S}$ is a $\underline{\mathbf{nxn}}$ real diagonal, and $\underline{V}$ is a nxn unitary matrix. A third method of decomposition is when $\underline{U}$ is $\underline{kxk}$ and unitary, $\underline{S}$ is $\underline{\mathbf{kxn}}$ and $\underline{V}$ is nxn and unitary. In this case

$$\underline{S} = \begin{array}{c} \begin{matrix} \overset{r}{} & \overset{n-r}{} \end{matrix} \\ \left[\begin{array}{c|c} \begin{matrix} s_1 & & \\ & \ddots & \\ & & s_r \end{matrix} & \underline{0} \\ \hline \underline{0} & \underline{0} \end{array}\right] \begin{matrix} r \\ \\ k-r \end{matrix} \end{array} \qquad (3\text{-}15)$$

The above decompositions are not unique. As an example, consider the first type of decomposition. For Proof of its non-uniqueness, the following remarks shall be presented:

Remark 1. Consider a k-dimensional column vector $\underline{e}_i$ such that its ith element is 1 and the rest of its elements 0 as $\underline{e}_i =$

$[0,0,\ldots,\underset{i^{TH}}{1},\ldots,0,0]^T$. Then, a kxk identity matrix can be represented as

$$\underline{I}_k = [\underline{e}_1,\underline{e}_2,\ldots,\underline{e}_i,\ldots,\underline{e}_j,\ldots,\underline{e}_k]. \qquad (3\text{-}16)$$

Upon interchanging the ith and jth column of I_k, a new matrix is obtained as

$$\underline{P}_{ij} = [\underline{e}_1,\underline{e}_2,\ldots,\underline{e}_j,\ldots,\underline{e}_i,\ldots,\underline{e}_k]. \qquad (3\text{-}17)$$

Remark 2. The matrix $\underline{P}_{ij}$ is called an elementary matrix of the first kind which represents the elementary operation of the first kind meaning interchanging ith and jth columns (rows) of a matrix which has k columns. In order to interchange the ith and jth column (row) of a matrix, P_{ij} must be post-multiplied (premultiplied) by the matrix.

Remark 3. The same matrix $\underline{P}_{ij}$ can be premultiplied to a matrix which has k rows and the ith and jth rows will be interchanged.

Remark 4. The second kind of elementary operation is multiplying any columns (row) by a nonzero constant c. The third kind is addition to any column (row) of any other column (row) multiplied by an arbitrary nonzero constant c.

Remark 5. The operations of second and third kind can be represented by their pertinent right (for operation on the columns) or left (for operation on the rows) elementary matrices. "Right" refers to the postmultiplication and "left" to the

premultiplication of an elementary matrix.

Remark 6. Any matrix can be considered equivalent to the product of a series of left and right elementary matrices of the first, second and third kinds [55-56].

Remark 7.
$$\underline{P}_{ij}^2 = \underline{I}_k \text{ and } \underline{P}_{ij} = \underline{P}_{ij}^T \tag{3-18}$$

Derivation 3.1: Let $\underline{F}$ be $\underline{kxn}$, $\underline{U}$ a $\underline{kxk}$ unitary, $\underline{S}$ a $\underline{kxk}$ diagonal, and $\underline{V}$ a $\underline{nxk}$ unitary. The decomposition $\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T$, in its general form, is **not unique.**

The proof of this derivation can be given in two parts, each of which is sufficient to prove the non-uniqueness of SVD.

First Part of the Proof:

Using Remark 7,

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T = \underline{U}\,\underline{P}_{ij}^2\,\underline{S}\,\underline{P}_{ij}^2\,V^T \tag{3-19}$$

But (3-19) is

$$F = \underline{U}(P_{ij}\underline{P}_{ij}^T)\underline{S}(\underline{P}_{ij}P_{ij}^T)V^T = (\underline{U}\,\underline{P}_{ij})(\underline{P}_{ij}^T\underline{S}\,P_{ij})(\underline{P}_{ij}^T\underline{V}^T) \tag{3-20}$$

$P_{ij}^T\underline{S}P_{ij}$ means the ith and jth diagonal elements are interchanged. $\underline{U}\,\underline{P}_{ij}$ means that the ith and jth columns of $\underline{U}$ and $\underline{P}_{ij}^T\underline{V}^T$ means that ith and jth row of $\underline{V}^T$ are interchanged too. The latter means the ith and jth columns of $\underline{V}$ must be interchanged. Due to many possible combinations of the interchanges, the decompositon is non-unique. This completes the first part of the proof.

Second part of the Proof:

Using equation (3-13), it is easily verified that multiplying the elements of the ith columns of $\underline{U}$ and $\underline{V}$ i.e. u_i and v_i by a minus sign does not alter $\underline{F}$. Even if the order of singular values may be set fixed, the combinatorial possibility of various negative multiplication for columns of $\underline{U}$ and $\underline{V}$ without altering $\underline{F}$ is 2^k, which is another indication of the non-uniqueness of SVD. Hence, the proof. Q.E.D.

The non-uniqueness of the decomposition $\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T$ can lead into analytical difficulties. Therefore, there is a need for a unique definition of SVD which was already given in Section 3.1.1. With this definition the first row of $\underline{U}$ is positive for real (orthogonal) $\underline{U}$ [and the real parts of the first row of $\underline{U}$ is positive for a complex (unitary) $\underline{U}$] which results in a unique decomposition. In order to make the first row of $\underline{U}$ positive, the following procedure shall be performed:

i) SVD is performed.

ii) The singular values are ordered descending-wise.

iii) Every column of $\underline{U}$ whose first component is negative is multiplied by (-1).

Obviously, according to the second part of the proof, the corresponding columns of $\underline{V}$ must also be multiplied by (-1). Hence, the first row of $\underline{U}$ will become positive. Such $\underline{U}$ is unique; and hence, SVD related to it will be unique as well.

3.1.4 A Perspective of the Literature

As a remark on the history of SVD, it should be mentioned that this concept was known to the mathematicians of the early 20th Century in the sense that they knew how to algebraically decompose a real general matrix into a diagonal matrix and two different orthogonal matrices. The elements of the diagonal matrix were called principal values as opposed to the presently used singular values.

The algebraic method becomes cumbersome for medium to large matrices. Since the advent of digital computers, numerical computation of the singular values of a matrix has been proposed through various methods: one method is to compute $\underline{F}\,\underline{F}^T$ and obtain a diagonalization of it. Taking the square root of the diagonalized $\underline{F}\,\underline{F}^T$ gives the singular values. However, this technique rapidly propagates the computer round-off error and, therefore, is not desirable. The computation of the singular values based on plane rotational methods has been proposed by Kogbetliantz [57], Hestenes [58], Forsythe and Henrici [59]. A QR algorithm was suggested by Kublanovskaja [60]. A combination of Housholder Transformation to bi-diagonalize $\underline{F}$ and then QR algorithm to obtain the singular values were developed by Golub and Reinsch [1]. Golub and Kahan have applied singular values in least square solutions and calculating the pseudo-inverse of a matrix [134]. The most recent approach for computation of singular values has been developed by Lawson and Hanson [61], and seems to be more advantageous than others. Moler [62] has elaborated on the conditioning of SVD matrix factorization. Applications of SVD in linear systems and linear algebra is

presented by Klema and Laub [63]. A SVD application in system theory of model reduction is discussed in [17] by Kung, and in filtering of two-dimensional signals by Lashgari, Silverman, and Abramatic [145] and by Lee [156]. Moler and Stewart [64] have recently used the PDQ factorization that requires less computation time and less computer memory for SVD representation of images. Based on the results in [4], Andrews and Patterson [130-132] have applied SVD to approximate photographs by truncation of negligible singular values, and thereby have shown SVD as an aid in image coding, data compression and image restoration. Huang and Narendra [5] corrupted an image by Gaussian noise and succeeded in restoring the original image through the use of 38 out of 50 singular values of the image plus noise field. The use of more or less number of singular values would decrease signal-to-noise. Sahasrabudhe and Vaidya [65] have discussed the relation between the behavior of singular values of an image with its autocorrelation. The above mentioned works have all been in deterministic domain. The present research is concerned with stochastic SVD.

3.2 SVD Perturbation Theory

$\underline{F}$ is $\underline{k}\text{x}\underline{n}$, and without loss of generality, we assume that ($k \leq n$) and will continue this assumption throughout the rest of the dissertation.(*) The objective is to derive a relationship between a perturbation of $\underline{F}$ and its SVD representation. Let $\underline{F}$ be perturbed as

---

(*)For image processing puposes, k is, usually equal to n, in other words, texture fields are usually square matrices.

$$\underline{F}+\underline{\Delta}(\underline{F}) = \begin{bmatrix} F_{11}+\Delta F_{11} & \cdots & F_{1n}+\Delta F_{1n} \\ \vdots & & \vdots \\ F_{k1}+\Delta F_{k1} & \cdots & F_{kn}+\Delta F_{kn} \end{bmatrix} \tag{3-21}$$

obviously $\underline{S}$ will change to $\underline{S}+\underline{\Delta}\,\underline{S}$ as

$$\underline{S}+\underline{\Delta}(\underline{S}) = \begin{bmatrix} s_1+\Delta s_1 & & 0 \\ & \ddots & \\ 0 & & s_k+\Delta s_k \end{bmatrix} \tag{3-22}$$

likewise $\underline{U}$ and $\underline{V}$ will be perturbed. We can see that if $\underline{\Delta}(\underline{F})\to\underline{0}$, then $\underline{\Delta}(\underline{S})\to\underline{0}$. It is desired to find a limiting relationship as

$$J_{\underline{F}}(\underline{U},\underline{S},\underline{V}) = \lim_{\Delta F_{ij}\to 0} \frac{\prod_{i,j} \Delta F_{ij}}{\left(\prod_{i} \Delta s_i\right)\left(\prod_{i,j} \Delta U_{ij}\right)\left(\prod_{i,j} \Delta V_{ij}\right)}. \tag{3-23}$$

where for $\prod_{i,j} \Delta U_{ij}$ and $\prod_{i,j} \Delta V_{ij}$, we must exclude the redundant or zero terms. (See Appendix C). In other words, it is desired to obtain a differential form for dF in terms of dU, dS, and dV.

3.2.1 Definitions for Differential Forms of Matrices

If

$$\underline{F} = [F_{ij}] = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1n} \\ F_{21} & F_{22} & \cdots & F_{2n} \\ \vdots & \vdots & & \vdots \\ F_{k1} & F_{k2} & \cdots & F_{kn} \end{bmatrix} \tag{3-24}$$

is a kxn real matrix, then the differential volume element of $\underline{F}$ is defined to be the product of all differentials on elements of $\underline{F}$.

Definition I. Differential Volume Element

$$dF \triangleq \prod_{i,j}^{k,n} dF_{ij} \tag{3-25}$$

where dF_{ij} is the differential element of the $(i,j)^{th}$ components of the matrix $\underline{F}$.

Definition II. Differential Volume Element for a Real Diagonal Matrix $\underline{S}$

If $\underline{S} = \text{diag}(s_1, s_2, \ldots, s_k)$, then the differential volume element of $\underline{S}$ is defined to be the product of all differentials on the diagonal elements.

$$dS \triangleq \prod_{i}^{k} ds_i$$

$$= ds_1 ds_2 \ldots ds_k \tag{3-26}$$

Definition III. Vector and Matrix of Differentials

If $\underline{a} = [a_1, a_2, \ldots, a_k]^T$ is a vector then,

$$\underline{d}(\underline{a}) \triangleq [da_1, da_2, \ldots, da_k]^T \tag{3-27}$$

and

$$\underline{d}(\underline{F}) \triangleq [dF_{ij}] \tag{3-28}$$

Note 1: $\underline{d}(\underline{F})$ of Definition (3-28) is a matrix of differential elements of $\underline{F}$, and hence

$$dF = \text{product of all elements in } \underline{d}(\underline{F}) \tag{3-29}$$

Note 2: (3-25) and (3-26) are scalars while (3-27) and (3-28) are arrays.

3.2.2 Mathematical Reasoning for SVD Jacobian

In this section, a more general variation on Jacobian will be presented which is different from the standard Jacobian in that it involves matrix relations.

The objective here is to determine a relation between dF and dUdSdV as

$$dF = [J_{\underline{F}}(\underline{U},\underline{S},\underline{V})]dUdSdV \qquad (3\text{-}30)$$

where $J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$ is Jacobian of SVD transformation, dF and dS are defined in (3-25) and (3-26) and dU and dV are defined in Appendix C.

By Jacobian, it is meant to describe a mathematical relation between the volume elements of $\underline{F}$ with those of its decomposed counterparts $\underline{U}$, $\underline{S}$, and $\underline{V}$. Notion of Jacobian in this case is depicted by $J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$.

The main question to be considered is the following: consider the space of elements of the matrix $\underline{F}$. This space is mapped into another space which contains elements of the matrices $\underline{U}$, $\underline{S}$, and $\underline{V}$. Then, one seeks to determine the behavior of dependent variables (elements of $\underline{F}$) as a function of the independent variables (elements of $\underline{U}$, $\underline{S}$, and $\underline{V}$) near a given point. To put this argument in mathematical terms, let us consider a point such as f_1 in the space of the elements of $\underline{F}$. The point f_1 has (kn) coordinates in space R^{kn} as

$$f_1 = (F_1(1,1),\ldots,F_1(k,n)) \tag{3-31}$$

The image of f_1 in the joint space of $\underline{U}$, $\underline{S}$, and $\underline{V}$ is a point called $P_{U_1S_1V_1}$ with (k^2+k+kn) coordinates in space R^{nk}. The basis for space of $\underline{U}$ is $\frac{1}{2}k(k-1)$ dimensional; for $\underline{S}$ is k-dimensional; and for $\underline{V}$ is $[nk - \frac{1}{2}k(k+1)]$-dimensional. Since spaces of $\underline{U}$, $\underline{S}$, and $\underline{V}$ are independent and disjoint, the sum total of their basis dimensions will add up to kn which is to be expected.

$$P_{U_1S_1V_1} = [(U_1(1,1),\ldots,U_1(k,k),S_1(1),\ldots,S_1(k),V_1(1,1),\ldots,V_1(k,n)] \tag{3-32}$$

one wishes to study $\underline{F}$ near $P_{U_1S_1V_1}$ given that $\underline{F}_1 = \underline{U}_1\underline{S}_1\underline{V}_1^T$.

In the physical world, the behavior of $\underline{F}$ near a sufficiently close distance of $P_{U_1S_1V_1}$ is linear provided that the variable elements of $\underline{U}_1$, $\underline{S}_1$, and $\underline{V}_1$ be restricted to a small variation. Because of this linear behavior in the close vicinity of $P_{U_1S_1V_1}$, the first partial derivatives at $P_{U_1S_1V_1}$ are needed to linearly approximate $\underline{F}$ at that point. As a result, second or higher order partial derivatives are considered to be zero.

Thus, it suffices to determine a mathematical relationship between the product of the first power of the differential elements of $\underline{F}$ with the product of those of $\underline{U}$, $\underline{S}$, and $\underline{V}$. In simpler terms, using the definition (3-25) and (3-26), it is desired to obtain a relation between differential volume elements of $\underline{F}$, $\underline{U}$, $\underline{S}$, and $\underline{V}$ or, notationally, between dF,and dUdSdV. This mathematical relation will

result in the [generalized] Jacobian $J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$.

3.2.3 Derivation of the Jacobian of SVD Transformation

For the derivation of the Jacobians of SVD, and spectral factorizations, three mathematical concepts will be used:

i. Jacobians of matrix transformations [66-67].

ii. Strategy like that of decomposition of a multivariate normal distribution into three independent distributions [68].

iii. Orthogonal functions in the hyperbolic space of matrices [99].

It must be emphasized that the results are not restricted to the case of normal samples and that they hold for samples from any distribution.

For image processing applications, $\underline{F}$ is always a real-valued matrix. Consider SVD in (3-1),

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T \qquad (3\text{-}33)$$

where, $\underline{F}$ is a real $\underline{kxn}$ ($k \leq n$) matrix, $\underline{U}$ is a real $\underline{kxk}$ orthonormal matrix, $\underline{S}$ is a real $\underline{kxk}$ diagonal matrix with non-negative elements, and $\underline{V}$ is a real $\underline{nxk}$ orthonormal matrix.

Applying differential operations on both sides of (3-33), we will have

$$\underline{d}(\underline{F}) = \underline{d}(\underline{U})\underline{S}\,\underline{V}^T + \underline{U}\,\underline{d}(\underline{S})\underline{V}^T + \underline{U}\,\underline{S}\,\underline{d}(\underline{V}^T) \qquad (3\text{-}34)$$

where, $\underline{d}(\underline{F})$ is defined in (3-28).

$\underline{V}$ is a $\underline{nxk}$ ($k \leq n$) orthonormal matrix. In order to form $\underline{V}$ into a square orthonormal matrix, we choose a nx(n-k) matrix $\underline{Q}$ such that the partitioned $[\underline{V}|\underline{Q}]$ is orthonormal. Therefore,

$$\underline{V}^T[\underline{V}|\underline{Q}] = [\underline{I}_k|\underline{0}] \tag{3-35}$$

Premultiplying $\underline{d}(\underline{F})$ of (3-30) by $\underline{U}^T$ and postmultiplying it by $[\underline{V}|\underline{Q}]$, we obtain

$$\underline{U}^T\underline{d}(\underline{F})[\underline{V}|\underline{Q}] = \underline{U}^T\underline{d}(\underline{U})\underline{S}\ \underline{V}^T[\underline{V}|\underline{Q}] + \underline{U}^T\underline{U}\ \underline{d}(\underline{S})v^T[\underline{V}|\underline{Q}] + \underline{U}^T\underline{U}\ \underline{S}\ \underline{d}(\underline{V}^T)[\underline{V}|\underline{Q}] \tag{3-36}$$

The reason for modifying the $\underline{nxk}$ $\underline{V}$ to a $\underline{nxn}$ matrix $[\underline{V}|\underline{Q}]$ is that, eventually, the determinant of [V|Q] shall be taken and for that matter it must be a square matrix [68].

Using Appendix A, Section 3.2.1, and the decompositon method of [68], we can modify (3-36) and state the following derivation:

Derivation 3.2: The Jacobian of SVD transformation is

$$J_{\underline{F}}(\underline{U},\underline{S},\underline{V}) = \text{abs}\left[\frac{2^{2k}\pi^{(\frac{k^2}{2}+\frac{kn}{2})}}{\Gamma_k(\frac{k}{2})\Gamma_k(\frac{n}{2})} \left(\prod_{i=1}^{k} s_i\right)^{n-k} \prod_{i<j}^{k} (s_i^2 - s_j^2) \right] \tag{3-37}$$

where, s_i is the i^{th} singular value and $\Gamma_k(n/2)$ is the k-variate gamma function of argument n/2 (a constant). Appendix B elaborates on the multivariate gamma function.

Remark: One can say that $J_{\underline{F}}(\underline{U},\underline{S},\underline{V}) = \dfrac{d(\underline{U}\ \underline{S}\ \underline{V}^T)}{\underline{dU}\ \underline{dS}\ \underline{dV}^T}$ where dF = $d(\underline{U}\ \underline{S}\ \underline{V}^T)$ is defined in (3-25), dS in (3-26), and dU and dV^T = dV are defined in Appendix C.

Proof:

Obtaining the volume elements of both sides of equation (3-36) and using [68, page 71] followed by definition (3-26), equations (C-14) and (C-23) of Appendix C, we obtain

$$dF = abs\left[|s|^{n-k} \prod_{i<j}^{k} (s_i^2 - s_j^2)\right] dS\tilde{d}(U)\tilde{d}(V) \qquad (3\text{-}38)$$

Using (C-35) and (C-38) of Appendix C on (3-38), and noting the fact that SVD of $\underline{F}$ is obtained with the consideration of uniqueness of decomposition, (C-41) is used

$$dF = 2^{-k}\left(\frac{2^k \pi^{\frac{k^2}{2}}}{\Gamma_k\left(\frac{k}{2}\right)}\right)\left(\frac{2^k \pi^{\frac{kn}{2}}}{\Gamma_k\left(\frac{n}{2}\right)}\right) abs\left[|\underline{s}|^{n-k} \prod_{i<j}^{k} (s_i^2 - s_j^2)\right] dS dU dV \qquad (3\text{-}39)$$

Explanation: The factor 2^{-k} at the beginning is because of having the first row of U positive. This is due to the fact that the $\underline{U}$ here is one of the 2^k combinationally possible $\underline{U}$'s without sign restriction.

$$|\underline{S}| = \prod_{i=1}^{k} s_i \qquad (3\text{-}40)$$

Thus, from (3-39) and (3-40)

$$dF = \underbrace{\text{abs}\left[\frac{2^k \pi^{\left(\frac{k^2}{2}+\frac{kn}{2}\right)}}{\Gamma_k\left(\frac{k}{2}\right)\Gamma_k\left(\frac{n}{2}\right)}\right]}_{\text{Constant}} \left(\prod_{i=1}^{k} s_i\right)^{n-k} \prod_{i<j}^{k} (s_i^2 - s_j^2)\,]\,dS\,dU\,dV \quad (3\text{-}41)$$

where, dU is the normalized invariant (Haar) measure on the orthogonal group O(k) and the differential form dV is the normalized invariant measure on the Stiefel Manifold $V_{k,n}$ of k-frames in Euclidean space R^n. Considering that

$$dF = J_F(\underline{U},\underline{S},\underline{V})\,dS\,dU\,dV \quad (3\text{-}42)$$

and comparing (3-42) with (3-41), the proof for (3-37) will be obtained. Q.E.D.

3.2.4 Analysis

Equations (3-37) of derivation 3.2 and (3-41) are always true regardless of whether singular values are ordered or not. If the singular values s_i, i = 1,...,k are arranged according to their descending order, then the "abs" sign of (3-37) can be removed. Equation (3-41) is of importance for our future derivations and analysis. Its significance is two-fold: one, being that $\underline{U}$ and $\underline{V}$ can be integrated out to result in the joint probability density function of the singular values of $\underline{F}$. And, two, (3-41) proves that, in SVD, the variations of $\underline{F}$ is a function of singular values of $\underline{F}$ rather than elements of $\underline{U}$ and $\underline{V}$. Formula (3-41) will be properly treated in Chapters 5 and 6 to obtain probability functions of the singular values and to develop stochastic SVD theory.

For the second significance, i.e. the property of singular values to carry most of the information of $\underline{F}$ and its elemental variations, the following theorem is presented:

Derivation 3.3: It is the singular values of $\underline{F}$ that determine the significant portion of the information regarding changes in $\underline{F}$.

Proof:

(3-37) of Derivation 3.2 shows that the generalized Jacobian for SVD transformation of $\underline{F}$ into $\underline{U}$, $\underline{S}$ and $\underline{V}$ is only a function of the singular values of $\underline{F}$ rather than a function of the elements of its unitary eigenmatrices $\underline{U}$ and $\underline{V}$. Therefore, one can attest that it is the singular values which determine the Jacobian and hence the significance of the change in elements of $\underline{F}$ with respect to the variations in $\underline{S}$, $\underline{U}$ and $\underline{V}$. **Q.E.D.**

Numerical Example:

As an example for application of the Derivation 3.3, consider the 2x2 matrix

$$\underline{F} = \begin{bmatrix} 5.000 & 9.000 \\ 2.000 & 7.000 \end{bmatrix}$$

Upon SVD decomposition of $\underline{F}$

$$\underline{F} = \underline{U}\,\underline{S}\,\underline{V}^T \qquad (3\text{-}43)$$

$$\underset{\underline{F}}{\begin{bmatrix} 5.000 & 9.000 \\ 2.000 & 7.000 \end{bmatrix}} = \underset{\underline{U}}{\begin{bmatrix} 0.819 & 0.574 \\ 0.574 & -0.819 \end{bmatrix}} \underset{\underline{S}}{\begin{bmatrix} 12.536 & 0 \\ 0 & 1.356 \end{bmatrix}} \underset{\underline{V}^T}{\begin{bmatrix} 0.418 & 0.908 \\ 0.418 & -0.418 \end{bmatrix}^T} \tag{3-44}$$

Upon incrementing $\underline{F}$ by $\underline{\Delta}(\underline{F}) = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}$

$$\underline{F}+\underline{\Delta}(\underline{F}) = \begin{bmatrix} 5.100 & 9.100 \\ 2.100 & 7.100 \end{bmatrix} \tag{3-45}$$

The SVD of (3-45) will be

$$\underline{F}+\underline{\Delta}(\underline{F}) = (\underline{U}+\underline{\Delta}(\underline{U}))(\underline{S}+\underline{\Delta}(\underline{S}))(\underline{V}^T+\underline{\Delta}(\underline{V}^T)) \tag{3-46}$$

$$\underset{\underline{F}+\underline{\Delta}(\underline{F})}{\begin{bmatrix} 5.100 & 9.100 \\ 2.100 & 7.100 \end{bmatrix}} = \underset{\approx\underline{U}+\Delta(\underline{U})}{\begin{bmatrix} 0.817 & 0.575 \\ 0.575 & -0.817 \end{bmatrix}} \underset{\approx\underline{S}+\Delta(\underline{S})}{\begin{bmatrix} 12.721 & 0 \\ 0 & 1.344 \end{bmatrix}} \underset{\approx\underline{V}+\Delta(\underline{V})}{\begin{bmatrix} 0.423 & 0.906 \\ 0.906 & -0.423 \end{bmatrix}^T} \tag{3-47}$$

Clearly,

$$\underline{\Delta}(\underline{U}) \cong \begin{bmatrix} -0.002 & 0.001 \\ 0.001 & 0.002 \end{bmatrix} \tag{3-48}$$

$$\underline{\Delta}(\underline{S}) \cong \begin{bmatrix} 0.185 & 0 \\ 0 & -0.012 \end{bmatrix} \tag{3-49}$$

$$\underline{\Delta}(\underline{V}) \cong \begin{bmatrix} 0.005 & -0.002 \\ -0.002 & -0.005 \end{bmatrix} \tag{3-50}$$

Using (3-37) of Derivation 3.2, the Jacobian of the SVD will be

$$J_{\underline{F}}(\underline{U},\underline{S},\underline{V}) = \frac{2^2 \pi^{\left(\frac{2^2}{2}+\frac{2x2}{2}\right)}}{\Gamma_2(1)\Gamma_2(1)} (s_1^2 - s_2^2)$$

$$= \frac{4\pi^4}{\pi^2} (157.161 - 1.839)$$

$$= 6131.867 \qquad (3\text{-}51)$$

where from Appendix B, $\Lambda_2(1) = \pi$. Using equations (C-14), (C-37) and (C-38) of Appendix C, ΔU which is the normalized (Haar) measure on the orthogonal group O(2) will be

$$\Delta U = \underline{U}_1^T \Delta(\underline{U}_1)$$

$$= [0.819 \quad 0.574] \begin{bmatrix} 0.001 \\ 0.002 \end{bmatrix} \qquad (3\text{-}52)$$

$$= 1.967 \times 10^{-3}$$

Using definition (3-22)

$$\Delta S = \Delta s_1 \Delta s_2$$

$$= 0.185 \times (-0.012) \qquad (3\text{-}53)$$

$$= -2.22 \times 10^{-3}$$

Since n=k=2, the Stiefel manifold for $\underline{V}_{2,2}$ will reduce to the

orthogonal group O(2).

$$\begin{aligned} \Delta V &= v_2^T \underline{\Delta}(\underline{v}_1) = -v_1^T \underline{\Delta}(\underline{v}_2) \\ &= [0.908 \quad -0.418] \begin{bmatrix} 0.005 \\ -0.002 \end{bmatrix} \\ &= 5.376 \times 10^{-3} \end{aligned} \tag{3-54}$$

Using the derived values of (3-51), (3-52), (3-53), and (3-54) in the r.h.s. of relation (3-42) of Derivation 3.2, one obtains

$$\begin{aligned} J_{\underline{F}}(U,S,V)\Delta U \Delta S \Delta V &= (6131.867)(1.967\times10^{-3}) \\ &\qquad (2.22\times10^{-3})(5.376\times10^{-3}) \\ &= 1.43 \times 10^{-4} \end{aligned} \tag{3-55}$$

The l.h.s. of (3-42) is dF which is the product of differentials in $\underline{\Delta}(\underline{F})$ as

$$\begin{aligned} \Delta F &= (0.1)(0.1)(0.1)(0.1) \\ &= 10^{-4} \end{aligned} \tag{3-56}$$

As it can be observed from comparison of (3-55) and (3-56), the two numbers are of the same order and are close. The slight difference between them has two reasons:

i) $\underline{\Delta}(\underline{U})$, $\underline{\Delta}(\underline{S})$, and $\underline{\Delta}(\underline{V})$ are approximations up to 3 digits after the decimal point.

ii) Relations (3-37), (3-41), and (3-42) are true at the limit when

$$\underline{\Delta}(\underline{F}) \to \underline{0}$$

$$\underline{\Delta}(\underline{U}) \to \underline{0}$$

$$\underline{\Delta}(\underline{S}) \to \underline{0}$$

$$\underline{\Delta}(\underline{V}) \to \underline{0}$$

From (3-51)-(3-55), three observations can be made:

Observation 1) Significance of the Singular Values:

The changes in $\underline{F}$ are mostly reflected in the Jacobian $J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$ which itself is a function of singular values. Table 3-1 shows that majority of variation in $\underline{F}$, is reflected in singular values of $\underline{F}$.

Observation 2) Significance of the Dominant Singular Value:

There are a pair of singular values for each of $\underline{F}$ and $\underline{F}+\underline{\Delta}(\underline{F})$. For both cases, the dominant singular value is much larger than the other one. This indicates that the dominant singular value can play an important role in information extraction.

Observation 3) Significance of Variation in the Dominant Singular Value:

As a result of variation in elements of $\underline{F}$, the dominant singular value reflects the largest change.

Table 3-2 presents the numerical results for observations 2 and 3.

3.2.5 Experimental Verification:

Derivations 3.2 and 3.3 are theoretical proofs of the significance of singular values of $\underline{F}$ in relation with the variation

Table 3-1. Significance of Singular Values and Jacobian of SVD.

SYMBOL	VALUE
ΔF	0.000100
ΔU	0.001967
ΔS	0.002220
ΔV	0.005376
$J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$	6131.867000

Table 3-2. Significance of the Dominant Singular Value and Its Variation.

	SINGULAR VALUES		CHANGE IN SINGULAR VALUES
	of $[\underline{F}]$	of $[\underline{F}+\Delta(\underline{F})]$	Δs_i AS A RESULT OF $\underline{\Delta}(\underline{F})$
FIRST SINGULAR VALUE (DOMINANT)	12.536	12.721	0.185
SECOND SINGULAR VALUE	1.356	1.344	-0.012

in $\underline{F}$. Figures 3-1 through 3-3 present the results of a set of experiments performed to verify this fact. In the experiments, artificial texture fields are generated. Then, matrices $\underline{U}$, $\underline{S}$, and $\underline{V}$ are obtained and changed to luminance amplitude information to be displayed by Tektronics 4014 graphic terminal. Any variation in $\underline{F}$ is followed by the variations in $\underline{U}$, $\underline{S}$, and $\underline{V}$. However, the ones for $\underline{U}$ and $\underline{V}$ are minimal while that of $\underline{S}$ is very noticeable. In Figure 3-1, the case of perfectly correlated and uncorrelated textures is shown. Figure 3-2 displays a Tektronix 4014 representation of luminance amplitude for $\rho = 0.5$ and $\rho = 0.6$ and shows the change in the singular values w.r.t. The correlation factor $\rho_C = \rho_R = \rho$. Figure 3-3 depicts the same for $\rho_C = \rho_R = \rho = 0.7$ and $\rho_C = \rho_R = \rho = 0.9$. The 3-dimensional representation of F, U, S, and V makes it easy to see any trend in the above matrices while numerical detection of changes in such large matrices is cumbersome to follow.

3.3 Differential Spectral Factorization

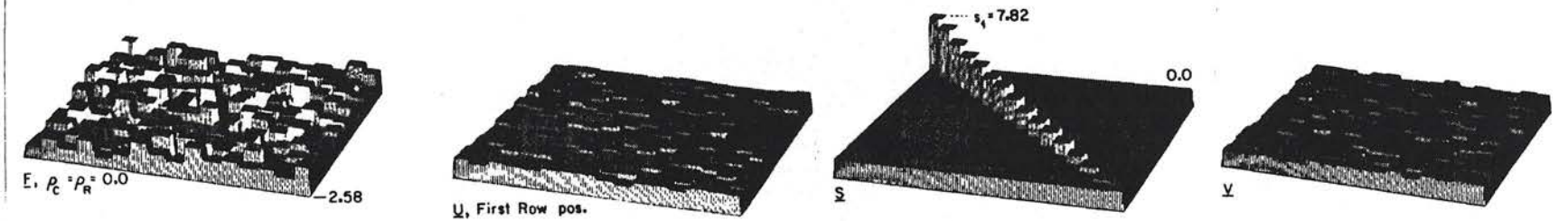
In this section, the procedure of Section 3.2.3 is utilized to obtain a Jacobian for spectral factorization transformation. Consider $\underline{F}$ as a real general $\underline{kxn}$ ($k<n$) matrix. Then, $\underline{F}\,\underline{F}^T$ is a $\underline{kxk}$ symmetric matrix.

It is known that any symmetric matrix can be spectrally factorized to its eigenvalues and eigenvector matrices. Hence,

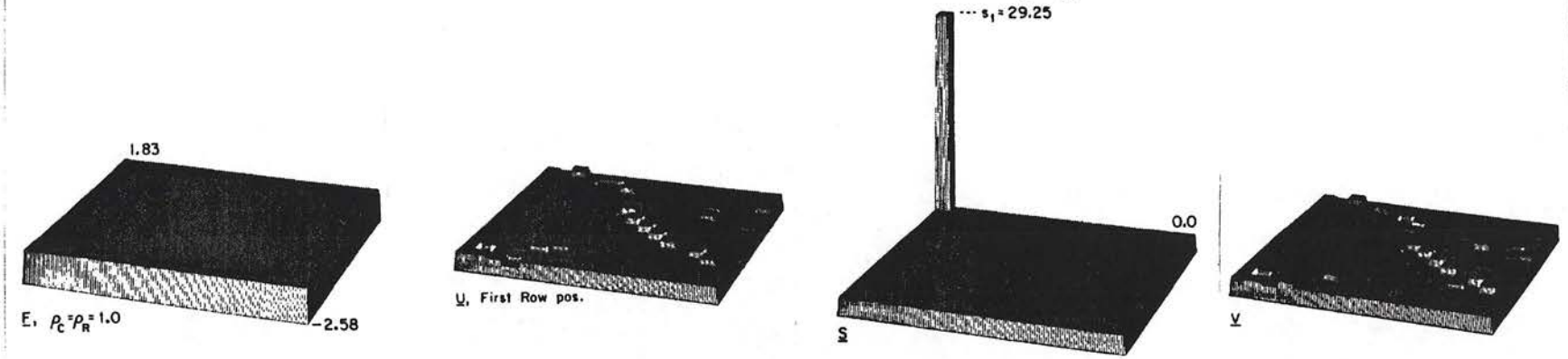
$$\underline{F}\,\underline{F}^T = \underline{U}\,\underline{\Lambda}\,\underline{U}^T \tag{3-57}$$

where,

$$\underline{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) \tag{3-58}$$

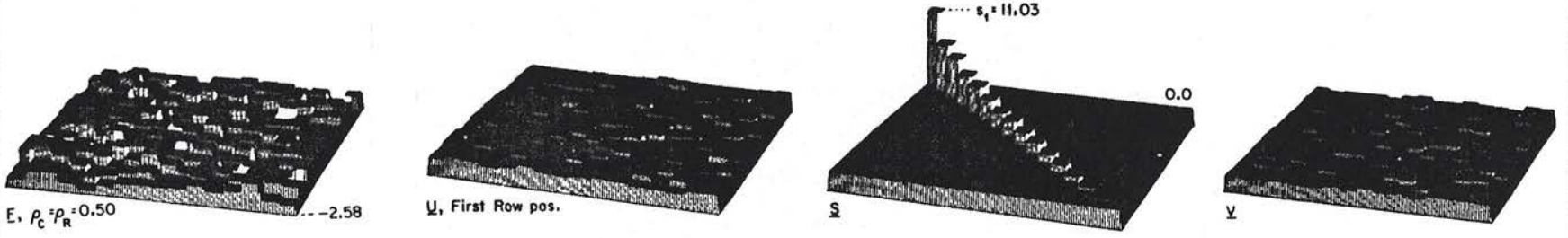


(a) perfectly uncorrelated texture $\underline{F}$, $\rho_C = \rho_R = 0.0$

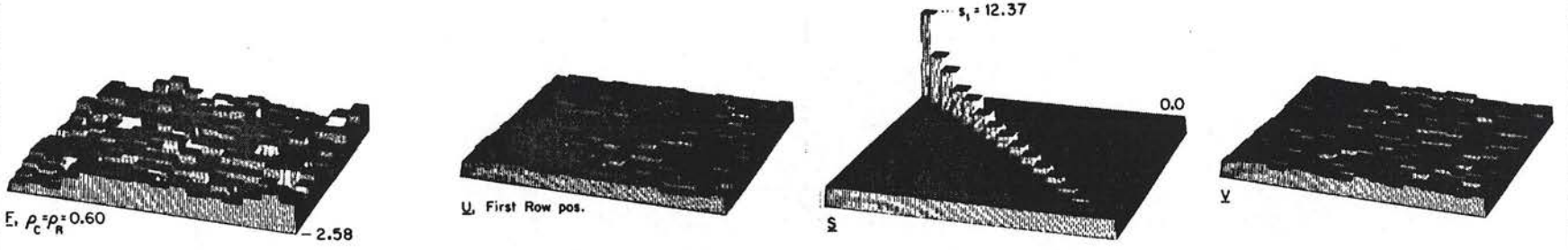


(b) perfectly correlated texture $\underline{F}$, $\rho_C = \rho_R = 1.0$

Figure 3-1. Luminance to Amplitude-Verification of Significance of the Singular Values.

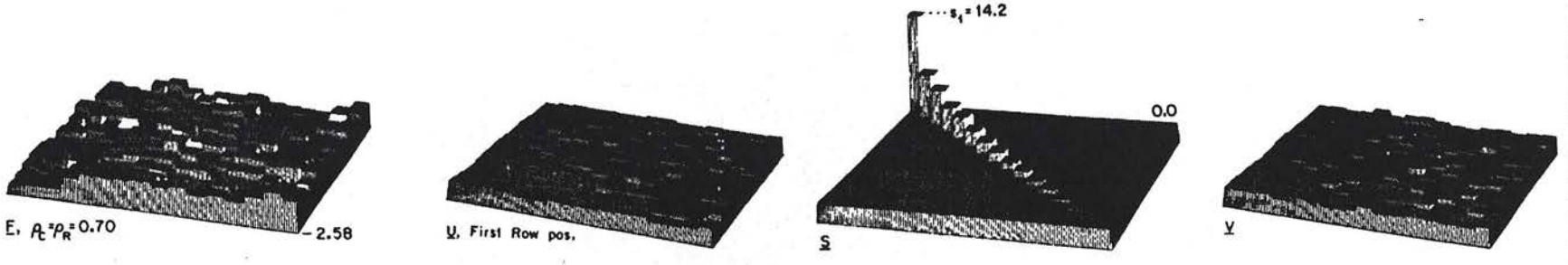


(a) texture $\underline{F}$, $\rho_C=\rho_R=0.5$

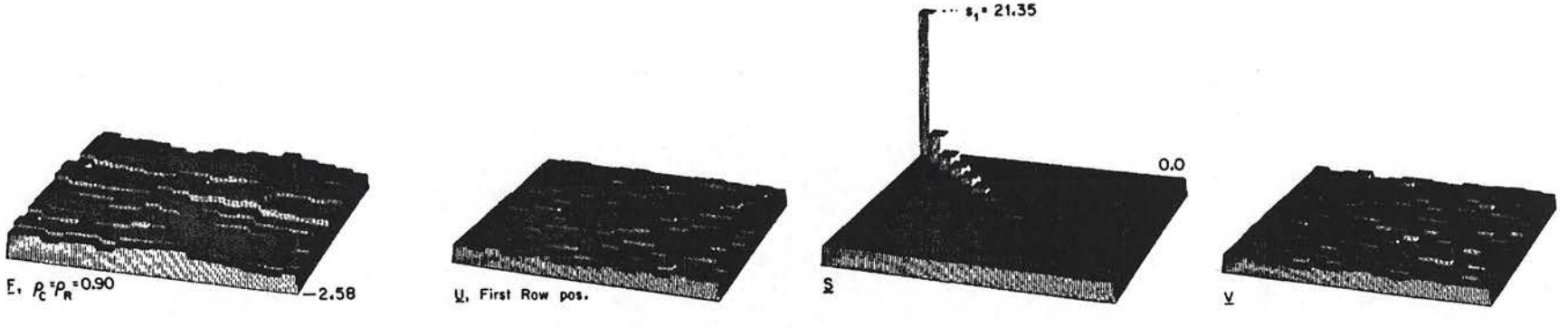


(b) texture $\underline{F}$, $\rho_C=\rho_R=0.6$

Figure 3-2. Luminance to Amplitude-Verification of Significance of the Singular Values.



(a) texture E, $\rho_C=\rho_R=0.7$



(b) texture E, $\rho_C=\rho_R=0.9$

Figure 3-3. Luminance to Amplitude-Verification of Significance of the Singular Values.

and, λ_i, $i = 1,\ldots,k$ is the i^{th} eigenvalue of $\underline{F}\,\underline{F}^T$. $\underline{U}$ is an orthogonal matrix of eigenvectors of $\underline{F}\,\underline{F}^T$ where

$$\underline{U} = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k] \tag{3-59}$$

u_i, $i = 1,\ldots,k$ is a k-dimensional column vector and

$$\underline{U}^T\underline{U} = \underline{I}_k \tag{3-60}$$

The objective, here, is to derive a relation for $d(FF^T)$ w.r.t. dU and $d\Lambda$ where,

$$d(FF^T) \triangleq \prod_{i<j}^{k} d(FF^T)_{ij} \tag{3-61}$$

$d\Lambda$ is defined in (3-26) as

$$d\Lambda \triangleq \prod_{i=1}^{k} d\lambda_i \tag{3-62}$$

and dU is the Haar measure on the orthogonal group $O(k)$ defined jointly by (C-14) and (C-38) of Appendix C.

3.3.1 Jacobian, $J_{\underline{F}\,\underline{F}^T}(\underline{U},\underline{\Lambda})$

In order to simplify reference to $\underline{F}\,\underline{F}^T$, an auxiliary matrix $\underline{\Phi}$ is used to denote it

$$\underline{\Phi} = \underline{F}\,\underline{F}^T \tag{3-63}$$

hence,

$$\underline{\Phi} = \underline{U}\,\underline{\Lambda}\,\underline{U}^T \tag{3-64}$$

differentiating (3-64),

$$\underline{d}(\underline{\Phi}) = \underline{d}(\underline{U})\underline{\Lambda}\,\underline{U}^T + \underline{U}\,\underline{d}(\underline{\Lambda})\underline{U}^T + \underline{U}\,\underline{\Lambda}\,\underline{d}(\underline{U}^T) \tag{3-65}$$

where, $\underline{d}(\underline{\Phi})$ is

$$\underline{d}(\underline{\Phi}) = [d\Phi_{ij}] \tag{3-66}$$

Pre- and post-multiplying both sides of equation (3-65) by $\underline{U}^T$ and $\underline{U}$

$$\underline{U}^T\underline{d}(\underline{\Phi})\underline{U} = \underline{U}^T\underline{d}(\underline{U})\underline{\Lambda} + \underline{d}(\underline{\Lambda}) + \underline{\Lambda}\,\underline{d}(\underline{U}^T)\underline{U} \tag{3-67}$$

According to Derivation A.6 of Appendix A

$$\underline{d}(\underline{U}^T)\underline{U} = -\underline{U}^T\underline{d}(\underline{U}) \tag{3-68}$$

Using (3-68) in (3-67)

$$\underline{U}^T\underline{d}(\underline{\Phi})\underline{U} = \underline{U}^T\underline{d}(\underline{U})\underline{\Lambda} + d(\underline{\Lambda}) - \underline{\Lambda}\,\underline{U}^T\underline{d}(\underline{U}) \tag{3-69}$$

The volume element on the l.h.s. of (3-69) is

$$\text{volume element}\{\underline{U}^T\underline{d}(\underline{\phi})\underline{U}\} = \text{abs}(|\underline{U}^T|^k|\underline{U}|^k) \prod_{i<j}^{k} d\phi_{ij}$$

$$= d\Phi \tag{3-70}$$

where, $d\Phi$ is defined in (3-61). Considering the r.h.s. of (3-69) represented by an auxiliary matrix $\underline{A}$.

$$\underline{A} = \underline{U}^T \underline{d}(\underline{U})\underline{\Lambda} + \underline{d}(\underline{\Lambda}) - \underline{\Lambda}\,\underline{U}^T\underline{d}(\underline{U}) \tag{3-71}$$

Clearly, $\underline{A}$ is a symmetric matrix of linear differential forms. Hence, the differential volume element of $\underline{A}$ will be product of the differential forms on the diagonal multiplied to that of the ones on the upper diagonal

$$\text{Volume Elements of } \underline{A} = \Big(\prod_{i=1}^{k} A_{ii}\Big)\Big(\prod_{i<j}^{k} A_{ij}\Big) \tag{3-72}$$

But

$$A_{ii} = \underline{u}_i^T \underline{d}(\underline{u}_i)\lambda_i + d\lambda_i - \lambda_i \underline{u}_i^T \underline{d}(\underline{u}_i) \tag{3-73}$$

Using Theorem A.5 of Appendix A

$$\underline{u}_i^T d(\underline{u}_i) = 0 \tag{3-74}$$

Thus

$$A_{ii} = d\lambda_i \tag{3-75}$$

Also,

$$\begin{aligned} A_{ij} &= \underline{u}_i^T \underline{d}(\underline{u}_j)\lambda_j - \lambda_j \underline{u}_i^T \underline{d}(\underline{u}_j) \\ &= \underline{u}_i^T \underline{d}(\underline{u}_j)(\lambda_j - \lambda_i) \end{aligned} \tag{3-76}$$

Using (3-75) and (3-76) in (3-72) and then equating the result with (3-70),

$$d\Phi = \text{abs}\Big[\prod_{i<j}^{k} (\lambda_i - \lambda_j)\Big] \prod_{i=1}^{k} d\lambda_i \prod_{i<j}^{k} \underline{u}_i \underline{d}(\underline{u}_j) \qquad (3\text{-}77)$$

Derivation 3.4: The Jacobian for spectral factorization is

$$J_{\underline{F}\,\underline{F}^T}(\underline{U},\underline{\Lambda}) = \text{abs}\Big[\frac{\pi^{k^2/2}}{\Gamma_k(k/2)} \prod_{i<j}^{k} (\lambda_i - \lambda_j)\Big] \qquad (3\text{-}78)$$

Proof: Using definition (3-62) and equation (C-14) of Appendix C in (3-77),

$$d\Phi = \text{abs}\Big[\prod_{i<j}^{k} (\lambda_i - \lambda_j)\Big] d\Lambda \tilde{d}U \qquad (3\text{-}79)$$

(C-38) of Appendix C is used in (3-79). Also, it is noted that for the sake of uniqueness of spectral factorization, $\underline{U}$ is restricted to have the elements of its first row positive. Then, (C-41) of Appendix C is also used in (3-79). Hence,

$$d\Phi = \text{abs}\Big[2^{-k} \Big(\frac{2^k \pi^{k^2/2}}{\Gamma_k(k/2)}\Big) \prod_{i<j}^{k} (\lambda_i - \lambda_j)\Big] d\Lambda dU \qquad (3\text{-}80)$$

But, $\underline{\Phi} = \underline{F}\,\underline{F}^T$ and

$$d(FF^T) = J_{\underline{F}\,\underline{F}^T}(\underline{U},\underline{\Lambda}) d\Lambda dU \qquad (3\text{-}81)$$

Comparison of (3-81) with (3-80), provides us with the proof for (3-78). **Q.E.D.**

Remark 1: (3-78) is always true regardless of the order of the

eigenvalues.

Remark 2: If the eigenvalues λ_i, $i = 1,\ldots,k$ are according to their descending order then the (abs) sign of (3-78) can be removed.

Remark 3: The Jacobian of (3-78) is of importance in the derivations of p.d.f. of the dominant singular value of a random texture field in Section 5.3.

Remark 4: Variations of the Jacobian (3-78) have been used in the mathematical statistics literature e.g. [122], although its formal derivation has not been located by the author.

3.4 Summary

In this chapter, a precise definition for singular value decomposition has been established, its various properties have been presented and a perspective from literature on the topic has been provided.

The proof of necessity for a unique decomposition is given and the notion of differential SVD is introduced. The concept of Jacobians for SVD and spectral factorization have been presented with its detailed proofs. An analysis and example is presented for the particularly important concept of Jacobian of SVD. The [generalized] Jacobian of SVD derived in this chapter holds regardless of the elemental distribution of $\underline{F}$ (texture field).

The next chapter has little relation with this one. However, together, they provide a framework for the forthcoming derivations of Chapters 5 and 6.

CHAPTER 4

STOCHASTIC TEXTURE MODELLING

An accurate mathematical model is essential for simulating a physical phenomenon. In this chapter, a model for image texture will be established which will be used throughout the rest of the dissertation. Let us assume that $\underline{F}$ is the texture field from which the singular values are extracted to be used as features for classification, recognition, or other applications. For stochastic modeling of $\underline{F}$, three mathematical concepts must be considered:

i) Lexicographic Transformation Technique

ii) Separable Covariance Matrix Concept

iii) Multivariate Normal Distribution

It is important to note that the first of the above concepts will assist in developing the second which is essential in our proof of the Toeplitz property for covariance matrices of our model. The above notions will be elaborated on in the following sections.

4.1 Lexicographic Transformation of an Image Field

In superposition operation applications, it is easier and computationally more feasible to work with vectors than matrices. Therefore, it is preferred to change the matrix $\underline{F}$ to a vector $\underline{f}$ without losing any information regarding the location and value of its elements. Such an operation is possible. The processing that must be performed on $\underline{F}$ can be done on its vector counterpart $\underline{f}$. The vector version of a matrix is nothing but the columns or rows of the matrix stacked on top of each other in an orderly manner. One form

of order, for example, is top-to-bottom, left-to-right scanning of columns. The operation of stacking the columns or rows of a matrix, in this document, is called lexicographic transformation(*). The mathematical technique to carry it out is the Kronecker product operations.(**)

Considering $\underline{F}$ as a texture field. Using a lexicographic transformation of $\underline{F} \rightarrow \underline{f}$, a vector $\underline{f}$ will be obtained. References [91] and [21] explain this type of transformation in mathematical terms and details.

4.2 Notion of Separability

The notion of separability of a covariance matrix has many practical advantages in image processing. Many types of texture images empirically display a separable property in their covariance matrix. The reason behind the separable property in visual texture can be explained in terms of texture's wide-sense stationarity. As an indirect outcome of [29-35], one can claim that second order statistics between various pairs of points with equal pairwise distance remains the same for a similar looking textured regions. In addition, the mean of every point throughout the texture field is the same and independent of location. These statistical properties of texture contribute to its wide-sense stationary behavior.

(*)Lexico for lexis a late Greek word meaning words or symbols and graphy for graphikos another Greek word meaning writing or forming in a specified way.

(**)In the literature, Kronecker product is sometimes referred to as direct product, and other times as tensor product.

The image array $\underline{F}$ will result in a lexicographic vector $\underline{f}$ for which the covariance matrix will be $\underline{K}_f$ where,

$$\underline{K}_f = \{(\underline{f}-\underline{\mu}_f)(\underline{f}-\underline{\mu}_f)^{*T}\} \qquad (4\text{-}1)$$

and $\underline{\mu}_f$ is the mean vector of $\underline{f}$.

For a wide-sense stationary $\underline{F}$, $\underline{K}_f$ is of block Toeplitz form [21]

$$\underline{K}_f = \begin{bmatrix} K_1 & K_2 & K_3 & \cdots & K_n \\ K_2^* & K_1 & K_2 & \cdots & K_{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ K_n^* & K_{n-1}^* & K_{n-2}^* & \cdots & K_1 \end{bmatrix} \qquad (4\text{-}2)$$

where $\underline{K}_f$ of equation (4-2) is called a **Hermitian** matrix(*). Although block Toeplitz forms do not directly imply separability, they can be easily represented by two separable Toeplitz forms. As a result, the wide-sense stationarity of image textures can simply be modeled by the separable connotation. Derivation 4.1 in the next section, will prove this conjecture.

4.2.1 Definition of Separable Covariance Matrices

The covariance function $K(s,t;s',t')$ of a continuous two dimensional field $F(s,t)$ is separable if $K(s,t;s',t') = K_C(s,s') \times K_R(t,t')$, where $K_C(.)$, $K_R(.)$ are the column and row covariance functions. Separability for the discrete field will be $K_{\underline{f}}(i,j,;i',j') = K_C(i,i')K_R(j,j')$ which means that the covariance matrix of the vector $\underline{f}$ can be expressed as the direct product of row and column covariance

---

(*)A Hermitian matrix is a matrix which is self-adjoint, where adjoint means conjugate transponse. A Hermitian matrix is the complex counterpart of a real symmetric matrix.

matrices. Appendix D elaborates on direct product relations. Using Definition (D-1), $\underline{K}_f$ can be represented as [21]

$$\underline{K}_f = \underline{K}_C \otimes \underline{K}_R = \begin{bmatrix} K_R(1,1)\underline{K}_C & K_R(1,2)\underline{K}_C & \cdots & K_R(1,n)\underline{K}_C \\ K_R(2,1)\underline{K}_C & K_R(2,2)\underline{K}_C & \cdots & K_R(2,n)\underline{K}_C \\ \vdots & \vdots & & \vdots \\ K_R(n,1)\underline{K}_C & K_R(n,2)\underline{K}_C & \cdots & K_R(n,n)\underline{K}_C \end{bmatrix} \qquad (4\text{-}3)$$

Remark: In equation (4-3), left direct product is used because $\underline{f}$, the image vector of lexicographic transformation of $\underline{F}$, is formed by column scanning of $\underline{F}$ in left to right, top to bottom order. If row scanning is used, the right direct product will be applicable. Throughout this chapter, left-direct product will be used consistently.

In equation (4-3), $\underline{K}_C$ is a kxk covariance matrix for columns of $\underline{F}$ and $\underline{K}_R$ is a nxn covariance matrix for rows of $\underline{F}$. The block Toeplitz form which was mentioned earlier as a result of the wide-sense stationarity of texture can be represented in separable form of (4-3) provided that both $\underline{K}_C$ and $\underline{K}_R$ matrices are, also, of Toeplitz forms.

Derivation 4.1: The block Hermitian Toeplitz form (4-2) can be represented by separable form of (4-3); for which the $\underline{K}_R$ is Hermitian Toeplitz and $\underline{K}_C$ is Hermitian.

Proof: We first prove that if (4-2) is separable, then $\underline{K}_C$ and $\underline{K}_R$ are Hermitian.

By comparing the $(i,j)^{th}$ and $(j,i)^{th}$ blocks of $\underline{K}_f$ in (4-2) with

the same in (4-3), we note that

$$\underline{K}_R(i,j)\underline{K}_C = [\underline{K}_R(j,i)\underline{K}_C]^{adj} \qquad (4\text{-}4)$$

where, adj means conjugate transpose. (4-4) leads to

$$\underline{K}_R(i,j)\underline{K}_C = \underline{K}_R^*(j,i)\underline{K}_C^{adj} \qquad (4\text{-}5)$$

where a * means the complex conjugate of a. (4-5) means that

$$\underline{K}_C = \underline{K}_C^{adj} \qquad (4\text{-}6)$$

hence $\underline{K}_C$ is Hermitian. Also from (4-5)

$$\underline{K}_R(i,j) = \underline{K}_R^*(j,i) \qquad (4\text{-}7)$$

means that

$$\underline{K}_R = \underline{K}_R^{adj} \qquad (4\text{-}8)$$

hence $\underline{K}_R$ is, also, Hermitian.

Now, the second part of the proof is the Toeplitz property of $\underline{K}_R$: Fixing an integer value for m such that $0 \leq m \leq n-1$. Then, comparing all $(i,i+m)^{th}$ blocks $(i = 1,\ldots,n-m)$ of $\underline{K}_f$ in (4-2) with each other shows that they all are represented by $\underline{K}_{m+1}$ matrix and therefore, are equal. Expecting the same equality between the

corresponding blocks of $\underline{K}_f$ in (4-3), results in all $K_R(i,i+m)\underline{K}_C$ for $i=1,\ldots,n-m$ to be equal. Hence

$$K_R(1,1+m) = K_R(2,2+m) = \ldots = K_R(n-m,n) \qquad (4\text{-}9)$$

Combining (4-9) and the Hermitian property of $\underline{K}_R$, we have

$$\underline{K}_R = \begin{bmatrix} K_R(1) & K_R(2) & K_R(3) & \ldots & K_R(n) \\ K_R^*(2) & K_R(1) & K_R(2) & \ldots & K_R(n-1) \\ \vdots & \vdots & \vdots & & \vdots \\ K_R^*(n) & K_R^*(n-1) & K_R^*(n-2) & \ldots & K_R(1) \end{bmatrix} \qquad (4\text{-}10)$$

which is a Toeplitz form. **Q.E.D.**

Derivation 4.1-1: Any block Hermitian Toeplitz matrix can be represented with two separable matrices for which one is Hermitian and the other is Hermitian and Toeplitz.

Remark: For the real cases, Hermitian will be replaced by symmetric.

Derivation 4.1-2: If the covariance of (4-2) is separable, then $\underline{K}_C$ and $\underline{K}_R$ of (4-3) are both of Toeplitz form.

Proof: $\underline{K}_f$ is a Hermitian matrix which can be represented both by left or right direct products depending on column or row scanning of the lexicographic transformations of $\underline{F} \rightarrow \underline{f}$. It was proved in Derivation 4.1 that $\underline{K}_C$ and $\underline{K}_R$ are both Hermitian and in addition, $\underline{K}_R$ is also of Toeplitz form. But Derivation 4.1 dealt with the left direct product. If we stack $\underline{F}$ row-wise, then $\underline{K}_f = \underline{K}_R \times \underline{K}_C$. As a

result, the proof of Derivation 4.1 can be applied again to obtain the result in a reverse manner meaning that this time $\underline{K}_C$ will be Hermitian Toeplitz and $\underline{K}_R$ will be Hermitian. Because $\underline{K}_f$ is the same for both cases and $\underline{K}_C$ and $\underline{K}_R$ are both Hermitian, then $\underline{K}_C$ and $\underline{K}_R$ must, also, be of Toeplitz form. Hence the proof. **Q.E.D.**

Derivation 4.1-3. If a wide-sense stationary field has separable property in its covariance, then the covariances along column and row direction are of Toeplitz form.

Derivation 4.1-4. The sufficient condition for a random field to be wide-sense stationary is that the covariances along columns and rows be separable and Toeplitz. This condition is, however, not necessary, i.e. it is possible to have wide-sense stationarity without having separable property.

Proof: If $\underline{K}_C$ and $\underline{K}_R$ are Toeplitz and separable then, $\underline{K}_f = \underline{K}_C \otimes \underline{K}_R$ is of block Toeplitz form which implies wide-sense stationarity.

Conclusions:

There are three concluding remarks:

1) The covariance matrix for a wide-sense stationary texture field is of block Toeplitz form.
2) A block Toeplitz form can be represented by a separable relation.
3) If a separable relation is used, then the separated covariances are, also, of Toeplitz form.

Remark: In order to perform experiments with our texture model in Chapters 8 and 9, first order Markov Toeplitz form will be used for $\underline{K}_C$ and $\underline{K}_R$ (although higher orders can also be, easily, used).

4.2.2 Two-Dimensional Separable Spectral Factorization

For establishing the model, the following two relations for Kronecker product operations shall be mentioned from Appendix D, followed by Derivation 4.2:

$$(\underline{A} \otimes \underline{B})(\underline{C} \otimes \underline{D}) \equiv \underline{A}\,\underline{C} \otimes \underline{B}\,\underline{D} \tag{4-11}$$

$$(\underline{A} \otimes \underline{B})^T \equiv \underline{A}^T \otimes \underline{B}^T \tag{4-12}$$

Derivation 4.2. If the covariance matrix $\underline{K}_f$ is separable, the spectral factorization of $\underline{K}_f$ will also be separable, i.e. if

$$\underline{K}_f = \underline{K}_C \otimes \underline{K}_R \tag{4-13}$$

Then, for spectral factorization of $\underline{K}_f$ as

$$\underline{K}_f = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T \tag{4-14}$$

$\underline{E}_f$ and $\underline{\Lambda}_f$ are also separable. Hence,

$$\underline{E}_f = \underline{E}_C \otimes \underline{E}_R \tag{4-15}$$

and

$$\underline{\Lambda}_f = \underline{\Lambda}_C \otimes \underline{\Lambda}_R \tag{4-16}$$

Proof: Upon spectrally factorizing $\underline{K}_C$ and $\underline{K}_R$

$$\underline{K}_C = \underline{E}_C \underline{\Lambda}_C \underline{E}_C^T \tag{4-17}$$

and,

$$\underline{K}_R = \underline{E}_R \underline{\Lambda}_R \underline{E}_R^T \tag{4-18}$$

The r.h.s. of (4-13) and (4-14) are equal. Therefore,

$$K_C \otimes \underline{K}_R = E_f \Lambda_f E_f^T \tag{4-19}$$

Using (4-17) and (4-18) in (4-19) we will have

$$\underline{E}_C \underline{\Lambda}_C \underline{E}_C^T \otimes \underline{E}_R \underline{\Lambda}_R \underline{E}_R^T = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T \tag{4-20}$$

but, using identity (4-11), we will change (4-20) to

$$(\underline{E}_C \underline{\Lambda}_C \otimes \underline{E}_R \underline{\Lambda}_R)(\underline{E}_C^T \otimes \underline{E}_R^T) = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T \tag{4-21}$$

reapplying the identity (4-11) on the l.h.s. of (4-21)

$$(E_C \otimes \underline{E}_R)(\underline{\Lambda}_C \otimes \Lambda_R)(\underline{E}_C^T \otimes \underline{E}_R^T) = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T \tag{4-22}$$

Using identity (4-12) on $E_c^T \otimes E_R^T$, we will have

$$(E_C \otimes E_R)(\Lambda_C \otimes \Lambda_R)(E_C \otimes E_R)^T = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T \tag{4-23}$$

hence, (4-15) and (4-16) [155]. **Q.E.D.**

4.3 A Practical Texture Model

The objective of this section is to form a mathematical model for texture field $\underline{F}$. The model must satisfy the requirements of the texture. There are five requirements for $\underline{F}$.

1) $\underline{F}$ is a matrix over the real field.
2) $\underline{F}$ has zero mean.
3) Covariance matrix of the elements of $\underline{F}$ is separable.
4) Elements of $\underline{F}$ are normally distributed.
5) $\underline{F}$ is a wide-sense stationary field.

Remark 1: The first of the above requirements is not an essential one for our theoretical developments. The derivations over the real field can easily be extended over the complex field. However, since the real life image texture fields are matrices of luminance values transduced to real numbers, the first assumption is of physical importance.

Remark 2: Requirement 2 is not essential either. Zero mean texture does not physically exist in nature because it implies having negative as well as positive values in a real texture field. However, considering ergodicity for image fields, it is possible to make a texture field zero mean by subtracting its experimental mean. In any event, the structure and inter-relationships between the pixels do not change, and even the covariance matrix of $\underline{F}$ remains the same. However, with zero mean, the analysis is simplified. It is mathematically possible to generalize the derivations in this document in order to develop and deal with non-central (non-zero

mean) cases.

Remark 3: Requirements 3 and 4 are essential in developing both the theoretical and experimental results.

Remark 4: The 5th requirement i.e. wide-sense stationarity is not essential and not needed for the theoretical developments of Chapters 5 and 6. However, in Chapters 8 and 9, wide-sense stationarity is assumed; and when combined with the third requirement along with the result of derivation 4.1-3, implies Toeplitz forms for the column and row covariances. In fact, for the experiments of the Chapters 8 and 9, first order Markov Toeplitz form will be used for $\underline{K}_C$ and $\underline{K}_R$.

Satisfying the 1st and 2nd Requirements

For satisfying the first two requirements, the following technique is used: Consider $\underline{f}$ as the lexicographic vector version of $\underline{F}$. $\underline{f}$ must have correlation among its elements. Noting that $\underline{K}_f$ can be factorized as (4-14), $\underline{f}$ with covariance $\underline{K}_f$ is generated from an uncorrelated process $\underline{m}$. From Appendix E

$$\underline{f} = \underline{E}_f \underline{\Lambda}_f^{1/2} \underline{m} \qquad (4\text{-}24)$$

where $\underline{m}$ is a white random process. It should be noted that mean of $\underline{f}$, i.e. $\mu_f = E\{\underline{f}\} = 0$.

For the third requirement which is separability, (4-15) and (4-16) will be used in (4-24)

$$\underline{f} = (\underline{E}_C \otimes \underline{E}_R)(\underline{\Lambda}_C^{1/2} \otimes \underline{\Lambda}_R^{1/2})\underline{m} \qquad (4\text{-}25)$$

Using identity (4-11) in equation (4-25)

$$\underline{f} = (\underline{E}_C \underline{\Lambda}_C^{1/2} \otimes \underline{E}_R \underline{\Lambda}_R^{1/2}) \underline{m} \qquad (4\text{-}26)$$

relation (D-12) can be rewritten from Appendix D, if, $\underline{f} = (\underline{A} \otimes \underline{B})\underline{m}$ where, $\underline{A}$ is a $\underline{kxk}$ and $\underline{B}$ is a $\underline{nxn}$ matrix. Then, for $\underline{F} \rightarrow \underline{f}$ and $\underline{M} \rightarrow \underline{m}$, we have

$$\underline{F} = \underline{A}\,\underline{M}\,\underline{B}^T \qquad (4\text{-}27)$$

Thus, using (4-26) and (4-27), $\underline{F}$ will be as

$$\underline{F} = (\underline{E}_C \underline{\Lambda}_C^{1/2}) \underline{M} (\underline{E}_R \underline{\Lambda}_R)^T \qquad (4\text{-}28)$$

In equation (4-28), $\underline{M}$ is a $\underline{kxn}$ random matrix whose elements are statistically orthonormal and have zero mean. Such a matrix can be called a white random matrix. Elements of matrix $\underline{F}$ are correlated with $\underline{K}_C$ as covariance matrix along the column and $\underline{K}_R$ as covariance matrices along the row directions where separable property holds as in (4-13).

Texture Field in a Multivariate Normal Sample

The 4th Requirement

The reason behind the normality requirement of the present texture model is a profound physical property:

Numerous independent random parameters are involved to form the visual appearance of a texture. Since these parameters are additive

along rows, $a_0 = \eta[1-\rho_R]$, $a_1 = \rho_R$, and $a_\omega = \sigma[1-\rho_R^2]^{1/2}$. After generation of a row, it will be correlated with the previous one along the column direction as $y_{n+1} = b_0+b_1y_n+b_w(x_{n+1}-\eta)/\sigma$ for $n = 1,\ldots,511$ where, $(x_{n+1}-\eta)/\sigma$ is a zero mean, unit variance normal random variable with x_{n+1} as the present pixel, y_n as the previous one and y_{n+1} as the one replacing x_{n+1} (y_{n+1} is correlated with y_n). As in the case for row direction, $b_0 = \eta[1-\rho_C]$, $b_1 = \rho_C$, $b_w = \sigma[1-\rho_C^2]^{1/2}$. In this procedure $\sigma = \sigma_C\sigma_R$ where σ_C^2 and σ_R^2 are the variance terms for the covariance matrix $\underline{K}_C$ and $\underline{K}_R$. The final matrix generated in this fashion has elements with mean η, variance σ^2, and every adjacent pair of pixels are correlated with $\sigma^2\rho_R$ correlation factor along the rows and ρ_C along the column, the covariance matrix $\underline{K}_f$ will be

$$\underline{K}_f = K_C \otimes \underline{K}_R = \sigma_C^2[\rho_C^{|i-j|}] \otimes \sigma_R^2[\rho_R^{|i-j|}] .$$

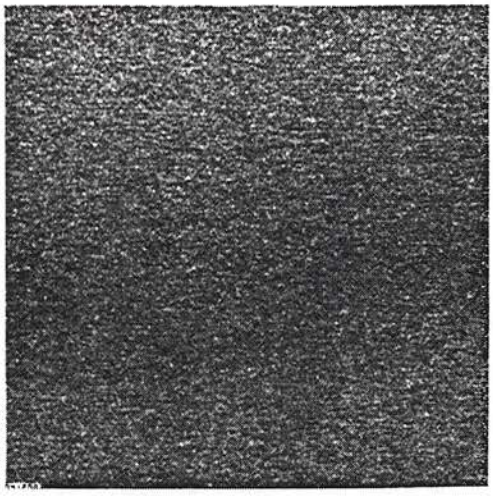
For our generations $\eta = 0.0$ and $\sigma_C = \sigma_R = 1.0$, hence

$$x_1 = \omega_0 = N(0,1) \tag{4-31}$$

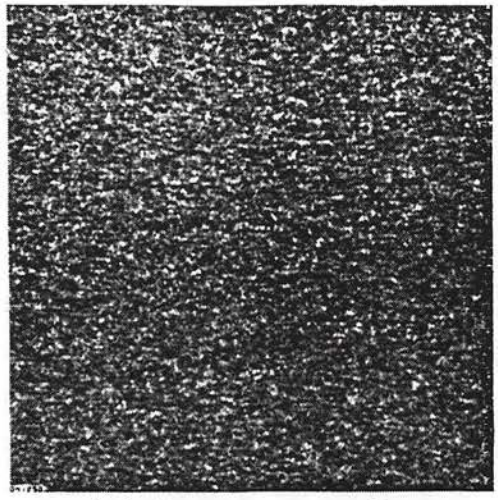
$$x_{n+1} = \rho_R x_n + (1-\rho_R^2)^{1/2}\omega_0 \qquad n = 1,\ldots,511 \tag{4-32}$$

$$y_{n+1} = \rho_C y_n + (1-\rho_C^2)^{1/2}x_{n+1} \qquad n = 1,\ldots,511 \tag{4-33}$$

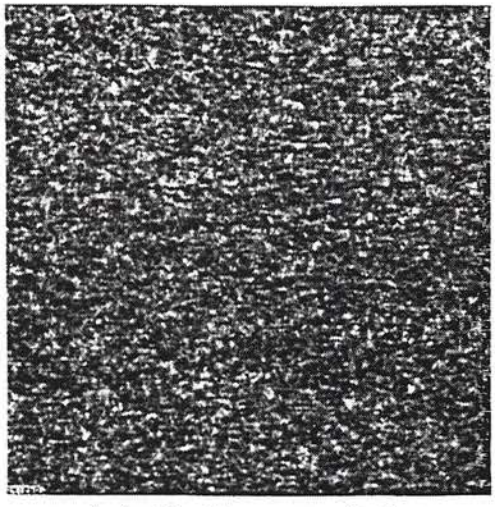
Figures 4-1 and 4-2 show examples of the stochastic textures that satisfy all 5 requirements of the model. In Figure 4-1 the textures are bidirectionally correlated with $\rho_C = \rho_R = \rho$ and in Figure 4-2, the textures are unidirectionally correlated with $\rho_C=\rho$ and $\rho_R = 0.0$ where in both figures ρ accepts values as 0.0 (perfectly



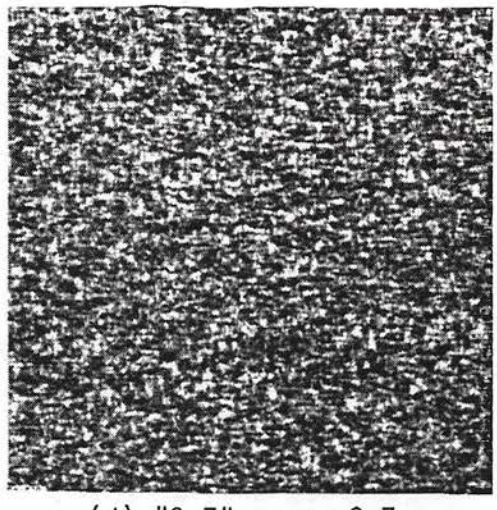
(a) "0.0", $\rho_C=\rho_R=0.0$



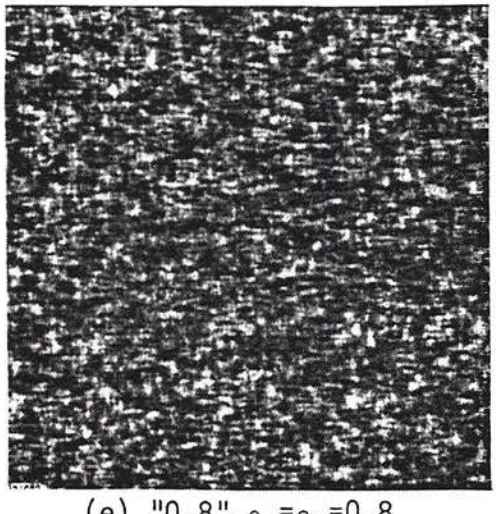
(b) "0.5", $\rho_C=\rho_R=0.5$



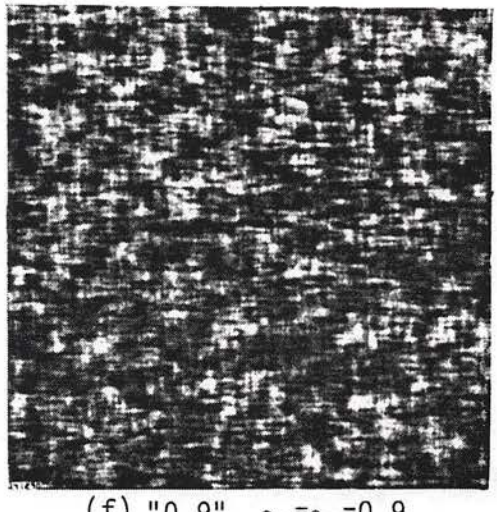
(c) "0.6", $\rho_C=\rho_R=0.6$



(d) "0.7", $\rho_C=\rho_R=0.7$

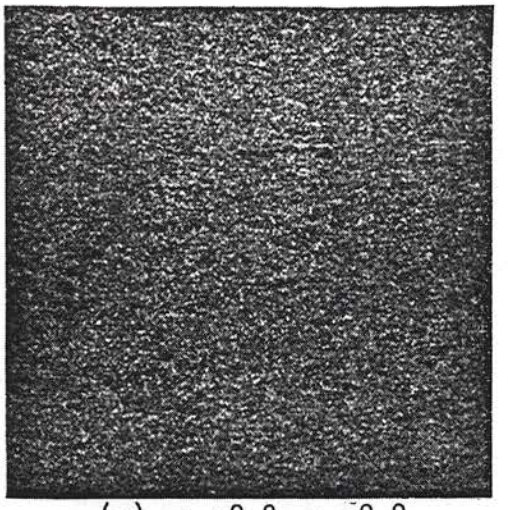


(e) "0.8", $\rho_C=\rho_R=0.8$

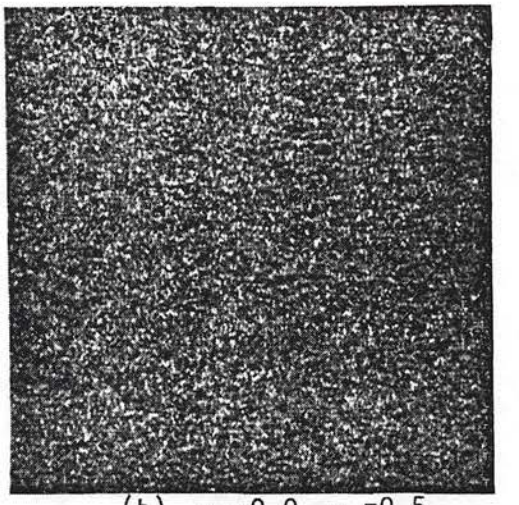


(f) "0.9", $\rho_C=\rho_R=0.9$

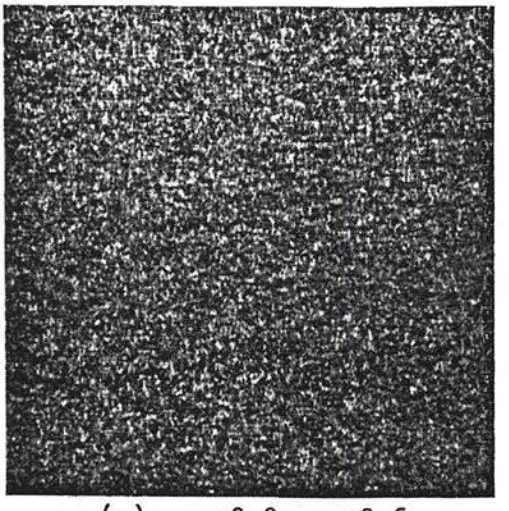
Figure 4-1. Stochastic Texture Models with Bidirectional Correlation.



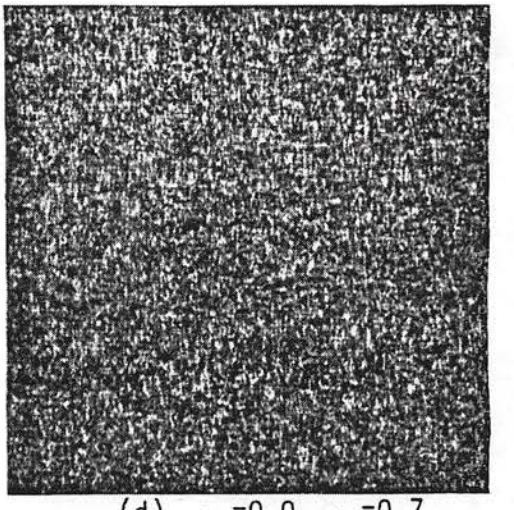
(a) ρ_R=0.0, ρ_C=0.0



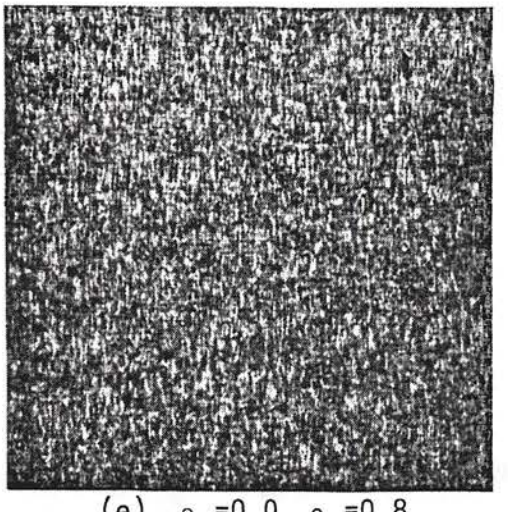
(b) ρ_R=0.0, ρ_C=0.5



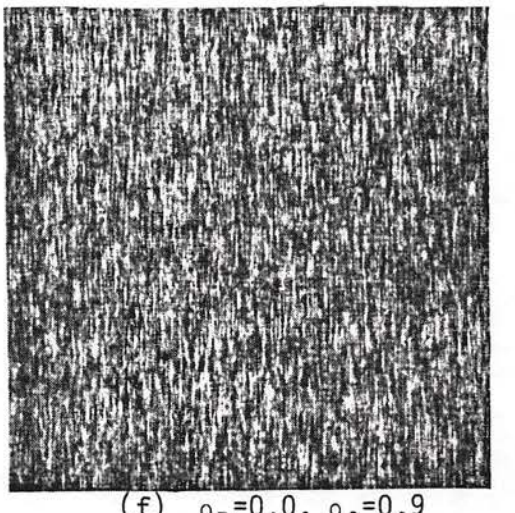
(c) ρ_R=0.0, ρ_C=0.6



(d) ρ_R=0.0, ρ_C=0.7



(e) ρ_R=0.0, ρ_C=0.8



(f) ρ_R=0.0, ρ_C=0.9

Figure 4-2. Stochastic Texture Models with Unidirectional Correlation.

uncorrelated), 0.5, 0.6, 0.7, 0.8, and 0.9.

4.4 Summary

The entire present chapter has been devoted to texture modeling. In this chapter, a stochastic texture model has been presented based on physical meaning and behavior of the many image textures.

The model to be used is concisely as follows: Modifying (4-28), one will have

$$\underline{F} = \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{M} \underline{\Lambda}_R^{1/2} \underline{E}_R^T \qquad (4\text{-}34)$$

where, $\underline{E}_C$ and $\underline{\Lambda}_C$ are defined by (4-17); and Λ_C and Λ_R are defined by (4-18). $\underline{M}$ is a white random matrix defined as in Appendix E.

Eq. (4-34) satisfies the first three requirements of the model; namely, real, zero mean, correlation with bidirectional and separable covariances. Using relation (4-30) in (4-29) and considering that $\underline{f}$ is knxl, then d = kn. Also, because of the second requirement $\underline{\mu}_f = \underline{0}$, therefore, (4-29) can be modified as:

$$\text{p.d.f.}(\underline{f}) = \frac{1}{(2\pi)^{kn/2} |\underline{K}_f|^{1/2}} \exp\{-\frac{1}{2} \text{tr}(\underline{K}_f^{-1} \underline{f}\, \underline{f}^T)\} \qquad (4\text{-}35)$$

(4-34) and (4-35) satisfy the first four requirements for the elements of $\underline{F}$. The last two equations provide the model for the theoretical works of Chapters 5 and 6. For real-world application and experiments, without loss of generality, wide-sense stationarity will also be considered for the model, as the fifth requirement.

CHAPTER 5

STOCHASTIC SINGULAR VALUE DECOMPOSITON

The derivations of Chapter 3, the model developed in Chapter 4 and the zonal polynomials from Appendix F are used in this chapter for obtaining probability functions of the texture field and its singular values.

5.1 Distribution of a Stochastic Texture Field(*)

Referring to the model for a stochastic texture field, [(4-34)-(4-35)], let $\underline{F}$ be a real stochastic texture field with zero mean, correlation in row and column directions, having a multivariate normal distribution. It is desired to derive the probability density and distribution function of $\underline{F}$. In order to establish enough background, the following section on mathematical preliminaries is presented.

5.1.1 Mathematical Preliminaries

Derivation 5.1: If $\underline{a}$ and $\underline{b}$ are vectors of the lexicographic transformation of $\underline{A}$ and $\underline{B}$, then

$$\mathrm{Tr}[\underline{a}\,\underline{b}^T] = \mathrm{Tr}[\underline{A}\,\underline{B}^T] \qquad (5\text{-}1)$$

where $\underline{A}$ and $\underline{B}$ are $\underline{kxn}$ matrices and $\underline{a}$ and $\underline{b}$ are $\underline{knx1}$ vectors.

(*) In the literature, a stochastic matrix or field usually means a matrix whose elements are probabilities of certain events [109-110]. In this document, however, by a stochastic or random matrix, it is meant a matrix whose elements are random variables.

(Note that, although $\underline{a}\,\underline{b}^T$ is a knxkn matrix and $\underline{A}\,\underline{B}^T$ is a kxk matrix, their traces are equal.)

Proof: $\underline{A}$ has n columns as $\underline{A} = [\underline{A}_1, \underline{A}_2, \ldots, \underline{A}_n]$ and $\underline{B}$ has n columns as $\underline{B} = [\underline{B}_1, \underline{B}_2, \ldots, \underline{B}_n]$. Then, using a hybrid representation

$$\underline{A}\,\underline{B}^T = \sum_{i=1}^{n} \underline{A}_i \underline{B}_i^T \tag{5-2}$$

Therefore

$$\mathrm{Tr}(\underline{A}\,\underline{B}^T) = \mathrm{Tr} \sum_{i=1}^{n} \underline{A}_i \underline{B}_i^T = \sum_{i=1}^{n} \mathrm{Tr}(\underline{A}_i \underline{B}_i^T) \tag{5-3}$$

$$\text{r.h.s. of (5-3)} = \mathrm{Tr}(\underline{A}_1 \underline{B}_1^T) + \mathrm{Tr}(\underline{A}_2 \underline{B}_2^T) + \ldots + \mathrm{Tr}(\underline{A}_n \underline{B}_n^T) \tag{5-4}$$

$$\text{r.h.s. of (5-4)} = \mathrm{Tr} \begin{bmatrix} \underline{A}_1 \underline{B}_1^T & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & & & & & & & & & \vdots \\ \vdots & & \underline{A}_2 \underline{B}_2^T & & & & & & & \vdots \\ \vdots & & \vdots & & & & & & & \vdots \\ \vdots & & \vdots & & & & & & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \underline{A}_n \underline{B}_n^T & \end{bmatrix}$$

$$= \mathrm{Tr} \begin{bmatrix} \underline{A}_2 \\ \underline{A}_2 \\ \vdots \\ \underline{A}_n \end{bmatrix} [\underline{B}_1^T, \underline{B}_2^T, \ldots, \underline{B}_n^T] \tag{5-5}$$

But

$$\underline{a}^T = [\underline{A}_1^T, \underline{A}_2^T, \ldots, \underline{A}_n^T] \quad \text{and} \quad \underline{b}^T = [\underline{B}_1^T, \underline{B}_2^T, \ldots, \underline{B}_n^T]$$

Thus,

$$\text{r.h.s. of (5-5)} = \mathrm{Tr}(\underline{a}\,\underline{b}^T)$$

Hence, (5-1). **Q.E.D.**

Derivation 5.2: If $\underline{f}$ is the vector of the lexicographic

transformation of $\underline{F}$, then

$$\underline{f}^T(\underline{A} \otimes \underline{B})\underline{f} = \mathrm{Tr}[(\underline{A} \otimes \underline{B})\underline{f}\ \underline{f}^T$$

$$= \mathrm{Tr}(\underline{A}\ \underline{F}\ \underline{B}^T\underline{F}^T) \qquad (5\text{-}6)$$

$$= \mathrm{Tr}(\underline{F}^T\underline{A}\ \underline{F}\ \underline{B}^T)$$

where $\underline{A}$ is a $\underline{kxk}$ and $\underline{B}$ is a $\underline{nxn}$ matrix.

Proof: The first equality of (5-6) is true according to its similarity with identity (4-30). The second and third equalities shall be proved as follows:

$$\mathrm{Tr}[(\underline{A} \otimes \underline{B})\underline{f}\ \underline{f}^T] = \mathrm{Tr}(\underline{I}_k\underline{A} \otimes \underline{I}_n\underline{B})\underline{f}\ \underline{f}^T], \qquad (5\text{-}7)$$

because $\underline{A} = \underline{I}_k\underline{A}$ and $\underline{B} = \underline{I}_n\underline{B}$ where $\underline{I}_k$ is identity matrix of order k.

Rewriting (D-7)

$$(\underline{I}_k\underline{A} \otimes \underline{I}_n\underline{B}) = (\underline{I}_k \otimes \underline{I}_n)(\underline{A} \otimes \underline{B}) \qquad (5\text{-}8)$$

Hence,

$$\text{r.h.s. of (5-7)} = \mathrm{Tr}[(\underline{I}_k \otimes \underline{I}_n)(\underline{A} \otimes \underline{B})\underline{f}\ \underline{f}^T] \qquad (5\text{-}9)$$

but, according to the circular property of the trace

$$\text{r.h.s. of (5-9)} = \mathrm{Tr}[\underline{f}^T(\underline{I}_k \otimes \underline{I}_n)(\underline{A} \otimes \underline{B})\underline{f}] \qquad (5\text{-}10)$$

$$\text{r.h.s. of (5-10)} = \mathrm{Tr}\{[(\underline{I}_k \otimes \underline{I}_n)\underline{f}]^T[(\underline{A} \otimes \underline{B})\underline{f}]\} \qquad (5\text{-}11)$$

Using (D-12), the lexicographic transformations of each vector inside

the brackets of the r.h.s. of (5-11) will be as follows:

$$(I_k \otimes I_n)f \;\sim\; \underline{I}_k \underline{F} \underline{I}_n \tag{5-12}$$

$$(\underline{A} \otimes \underline{B})\underline{f} \;\sim\; \underline{A} \underline{F} \underline{B} \tag{5-13}$$

Using (5-12) and (5-13) in the r.h.s. of (5-11) along with the result (5-1) of Derivation 5.1

$$\text{r.h.s. of (5-11)} \;=\; \mathrm{Tr}\{(\underline{I}_k \underline{F} \underline{I}_n^T)^T(\underline{A} \underline{F} \underline{B}^T)\} \tag{5-14}$$

Therefore,

$$\text{r.h.s. of (5-14)} \;=\; \mathrm{Tr}(\underline{F}^T\underline{A} \underline{F} \underline{B}^T)$$

$$= \mathrm{Tr}(AFB^TF^T) \tag{5-15}$$

Hence, the proof of (5-6). **Q.E.D.**

The following four definitions are necessary in order to consistently derive the functions of this section as well as the rest of the dissertation:

Definitions: For the random vector $\underline{f}$ and random matrix $\underline{F}$, we define

I. $g(\underline{f}) \triangleq$ p.d.f. of vector $\underline{f}$ (5-16)

II. $G(\underline{f}) \triangleq$ distribution function of vector $\underline{f}$ (5-17)

Then,

$$dG(\underline{f}) = g(\underline{f})df \tag{5-18}$$

Similarly

$$\text{III.} \quad g(\underline{F}) \triangleq \text{p.d.f. of matrix } \underline{F} \tag{5-19}$$

$$\text{IV.} \quad G(\underline{F}) \triangleq \text{distribution function of matrix } \underline{F} \tag{5-20}$$

Then,

$$dG(\underline{F}) = g(\underline{F})dF \tag{5-21}$$

where dF in (5-21) means the product of elements of $\underline{F}$, $\prod_{i,j} dF_{ij}$.

5.1.2 Probability Density Function

The following derivation results in the p.d.f. of a random texture field.

Derivation 5.3: The probability density function of $\underline{F}$, namely, a real, random texture field in a multivariate normal sample, with bidirectional correlation is

$$g(\underline{F}) = \frac{1}{(2\pi)^{kn/2}|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} \exp\{-\tfrac{1}{2}\,\mathrm{Tr}(\underline{K}_C^{-1}\underline{F}\,\underline{K}_R^{-1}\underline{F}^T)\} \tag{5-22}$$

Proof: Using equation (4-35) of Chapter 4, p.d.f. of $\underline{f}$ is

$$g(\underline{f}) = \frac{1}{(2\pi)^{kn/2}|\underline{K}_f|^{1/2}} \exp\{-\tfrac{1}{2}\,\mathrm{Tr}(\underline{K}_f^{-1}\underline{f}\,\underline{f}^T)\} \tag{5-23}$$

For separable correlation in the texture's covariance, we replace $\underline{K}_f$ by $\underline{K}_C \otimes \underline{K}_R$ in equation (5-23). Thus

$$g(\underline{f}) = \frac{1}{(2\pi)^{kn/2}|\underline{K}_C \otimes \underline{K}_R|^{1/2}} \exp\{-\frac{1}{2} \mathrm{Tr}[(\underline{K}_C \otimes \underline{K}_R)^{-1}\underline{f}\,\underline{f}^T]\} \quad (5\text{-}24)$$

noting from (D-6) of Appendix D that

$$(A \otimes B)^{-1} \equiv A^{-1} \otimes B^{-1} \quad (5\text{-}25)$$

Using (5-25) followed by derivation 5.2 in the exponent of (5-24) and also applying (D-11) of Appendix D in the coefficient of (5-24), (5-22) will be derived. **Q.E.D.**

5.2 Joint Probability Functions of the Singular Values of a Random Field Texture in a Normal Sample with Bidirectional Correlation

In the derivations of the present section, it should be observed that $g(\underline{F})$ and $G(\underline{F})$ denote only the p.d.f. and distribution of $\underline{F}$ and not any general function of $\underline{F}$. The result (5-22) of Derivation 5.3 is used in deriving the joint probability functions of a singular values of a texture field. The essential idea is that in equation (5-22), $\underline{F}$ is replaced by $\underline{U}\,\underline{S}\,\underline{V}^T$, and $\underline{U}$ and $\underline{V}$ are integrated out. Hence, considering $g(\underline{S})$ as a joint p.d.f. of singular values,

$$g(\underline{S})d\underline{S} = \int_{\substack{\text{space of} \\ \underline{U} \text{ and } \underline{V}}} g(\underline{F})d\underline{F}\Big|_{\underline{F}=\underline{U}\,\underline{S}\,\underline{V}^T} \quad (5\text{-}26)$$

But,

$$dF = J_{\underline{F}}(\underline{U},\underline{S},\underline{V})dUdSdV \quad (5\text{-}27)$$

It was proven in equation (3-37) of Derivation 3.2 that $J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$ is

a function of only the singular values. Replacing $\underline{F}$ by its SVD components $\underline{U}\,\underline{S}\,\underline{V}^T$ in (5-22) and using (5-26)

$$g(\underline{S})dS = \frac{1}{(2\pi)^{kn/2}|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} J_{\underline{F}}(\underline{U},\underline{S},\underline{V})dS$$

$$\int_{V_{k,n}} \int_{O(k)} \exp\{-\frac{1}{2}\mathrm{Tr}(\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}\,\underline{U}^T)\}dUdV \tag{5-28}$$

where dU and dV are defined in Appendix C.

The exponential function in (5-28) can be represented in terms of its McLaurin Series expansion, where the McLaurin Series expansion for an arbitrary function f(x) is

$$f(x) = \sum_{P=0}^{\infty} \frac{x^P f^{(P)}(0)}{P!} \tag{5-29}$$

where, $f^{(P)}$ is the P^{th} derivative of f and it must exist.

Hence, considering $f(.) = \exp\{.\}$; knowing that the n^{th} derivative of an exponential function exists and is equal to itself, and that evaluated at zero results in unity, we derive

$$g(\underline{S}) = \frac{1}{(2\pi)^{kn/2}|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$$

$$\sum_{P=0}^{\infty} \frac{1}{P!} \int_{V_{n,k}} \int_{O(k)} \{\mathrm{Tr}(-\frac{1}{2}\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}\,\underline{U}^T)\}^P dUdV \tag{5-30}$$

The derivations hitherto have been based on probabilistic and group theoretic arguments. For the solution of equation (5-30), the

concept of zonal polynomials of the group representation enters into perspective.

Derivation 5.4: The joint p.d.f. of the singular values of a normal random texture field $\underline{F}$, with bidirectional correlation is

$$g(s_1,\ldots,s_k) = \frac{2^{k(1-\frac{n}{2})}\pi^{k^2/2}}{\Gamma_k(\frac{k}{2})\Gamma_k(\frac{n}{2})|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} \prod_{i<j}^{k}(s_i^2-s_j^2)\Big(\prod_{i=1}^{k} s_i\Big)^{n-k} \times$$

$$\sum_{P=0}^{\infty}\sum_{\rho}\frac{C_\rho(-\frac{1}{2}\underline{K}_C^{-1})C_\rho(\underline{K}_R^{-1})C_\rho(\underline{S}^2)}{P!C_\rho(\underline{I}_k)C_\rho(\underline{I}_n)} \qquad (5\text{-}31)$$

where, $\Gamma_k(n/2)$ is the multivariate gamma function (Appendix B) and $\underline{F}$ has been defined in Chapters 3 and 4.

Proof: Instead of $\{Tr(\cdot)\}^P$ in (5-30)(*), the equation (F-4) of Appendix F can be used to give

$$g(\underline{S}) = \frac{1}{(2\pi)^{kn/2}|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} J_{\underline{F}}(\underline{U},\underline{S},\underline{V})$$

$$\sum_{p=0}^{\infty}\frac{1}{P!}\sum_{\rho}\int_{V_{k,n}}\int_{O(k)} C_\rho(-\tfrac{1}{2}\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}\,\underline{U}^T)d\underline{U}d\underline{V} \qquad (5\text{-}32)$$

(*) Trace of a matrix is equal to the sum of its eigenvalues. Thus $\{Tr(\underline{A})\}$ is $(\lambda_1+\ldots+\lambda_k)^P = \sum_{\sum m_i=P}(P!/m_1!\ldots m_k!)\,\lambda_1^{m_1}\ldots\lambda_k^{m_k}$. But the latter polynomial is replaced by a simple one called zonal polynomial and shown as $C_\rho(\underline{A})$.

Using property (F-7), $\underline{U}^T$ can be brought to the beginning of the argument inside $C_p(.)$:

$$C_p(-\tfrac{1}{2}\underline{U}^T\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\,\underline{V}\,\underline{S}) = C_p(-\tfrac{1}{2}\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}\,\underline{U}^T) \quad (5\text{-}33)$$

Using (5-33), in (5-32) and then applying property (F-8),

$$\int_{O(k)} C_p(-\tfrac{1}{2}\underline{U}^T\underline{K}_C^{-1}\underline{U}\,\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S})d\underline{U} = \frac{C_p(-\frac{1}{2}\underline{K}_C^{-1})C_p(\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S})}{C_p(\underline{I}_k)} \quad (5\text{-}34)$$

noting that in (5-34), $\underline{U}$ is eliminated.

Taking $C_p(\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S})$ of (5-34) into consideration, and applying property (F-7) on it

$$C_p(\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}^2) = C_p(\underline{S}\,\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}) \quad (5\text{-}35)$$

Hence, noting (5-32)

$$g(\underline{S}) = (\text{other terms}) \int_{V_{k,n}} C_p(\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}^2)d\underline{V} \quad (5\text{-}36)$$

Applying property (F-8) on (5-36),

$$\int_{V_{k,n}} C_p(\underline{V}^T\underline{K}_R^{-1}\underline{V}\,\underline{S}^2)d\underline{V} = \frac{C_p(\underline{K}_R^{-1})C_p(\underline{S}^2)}{C_p(\underline{I}_n)} \quad (5\text{-}37)$$

Note that in (5-37) $\underline{V}$ is also eliminated. Denoting (5-37), (5-36), (5-34), (5-32), and (3-37), the joint p.d.f. of the singular values

of $\underline{F}$ will be as (5-31) and hence the proof of Derivation 5.4.

Q.E.D.

Equation (5-31) is an open form formula for p.d.f. of singular values. The zonal polynomials $C_\rho(.)$ can be recursively computed. See Appendix F for more insight into zonal functions.

5.2.1 P.d.f. for the Null Case

Derivation 5.4 gives the p.d.f. of the singular values in terms of zonal polynomials of the covariance matrices $\underline{K}_C$ and $\underline{K}_R$. An interesting result of the joint p.d.f. of singular values given by (5-31) is that by letting $\underline{K}_R = \underline{I}_n$ and $\underline{K}_C = \sigma^2\underline{I}_k$, the joint p.d.f. of the singular values of $\underline{F}$ will be obtained. This corresponds to the null case where there is no correlation among elements of $\underline{F}$. In such a case, using equation (F-4) and (F-10), $\sum_{\rho=0}^{\infty}\sum_{\rho}(\cdot)$ in (5-31) will reduce to $\exp\left[-\frac{1}{2\sigma^2}\,\mathrm{Tr}(\underline{S}^2)\right]$, and hence (5-31) will have a closed form given by (5-38)

$$g(s_1,\ldots,s_k) = \frac{2^{k(1-\frac{n}{2})}\pi^{k^2/2}}{\Gamma_k(\frac{k}{2})\Gamma_k(\frac{n}{2})\,\sigma^{kn}}\prod_{i<j}^{k}(s_i^2-s_j^2)\Big(\prod_{i=1}^{k}s_i\Big)^{n-k} \times \exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}s_i^2\Big\} \qquad (5\text{-}38)$$

(5-38) gives the joint p.d.f. of singular values of a kxn matrix $\underline{F}$ whose elements are independent normal random variables with mean zero and variance σ^2.

For a 2x2 $\underline{F}$, k=n=2, from (5-38), $g(s_1,s_2)$ will be

$$g(s_1,s_2) = \frac{1}{\sigma^4}(s_1^2-s_2^2)\exp(-\frac{s_1^2+s_2^2}{2\sigma^2}), \quad \infty > s_1 \geq s_2 \geq 0 \quad (5\text{-}39)$$

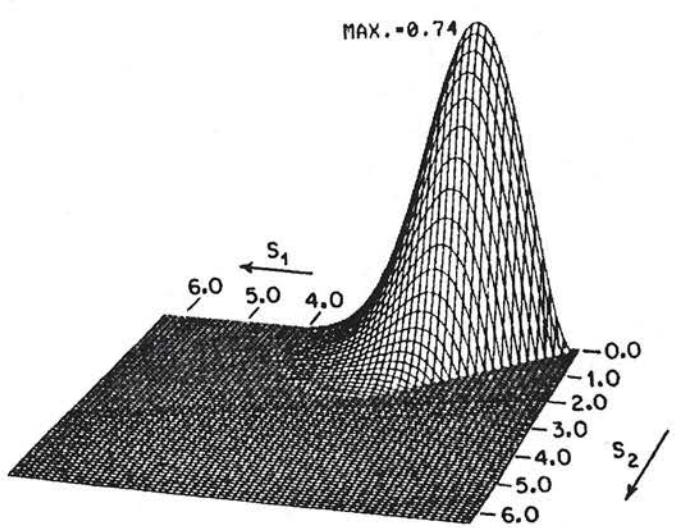
$g(s_1,s_2)$ given by (5-39) is displayed in Figure 5-1 for σ=1.0 and σ = 2.0. The maximum point happens at $s_1 = \sigma\sqrt{2}$, and $s_2 = 0$, where $g(\sigma\sqrt{2},0) = 2/(\sigma^2 e)$.

In (5-31), performing the transformation $\lambda_i = s_i^2$ for i = 1, ..., k, the joint p.d.f for the eigenvalues of a quadratic form will be obtained [123]. In (5-38), performing the transformation $\lambda_i = s_i^2$ for i = 1,...,k, the joint p.d.f. for the eigenvalues of a sample covariance matrix will be obtained for the case where each observation vector has uncorrelated components with identical variance σ^2. This distribution of eigenvalues is classic in multivariate statistical literature and has been simultaneously obtained by Fisher [10], Cirshick [11], Hsu [12], Mood [13], and Roy [14], and it is also presented in [81] and [117]. In (5-31), letting $\underline{K}_R = \underline{I}_n$, and performing the transformation $\lambda_i = s_i^2$, one obtains James' extension of the above mentioned classic formula for an arbitrary $\underline{K}_C$ [15].

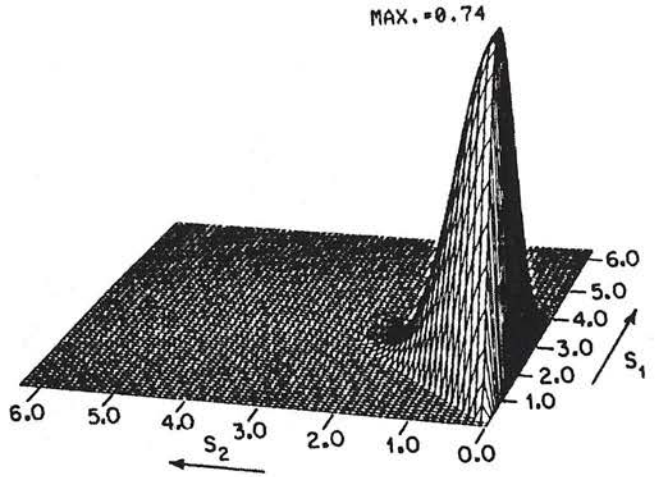
5.3 Probability Functions of the Dominant Singular Value

The largest singular value of a texture field plays a paramount role in SVD texture feature extraction for two reasons: (a) the majority of information is carried by the first few singular values particularly the dominant one, (b) utmost sensitivity of the dominant singular value to correlation among texture elements.

The cumulative distribution function of the largest characteristic root of a matrix in two to seven variate normal

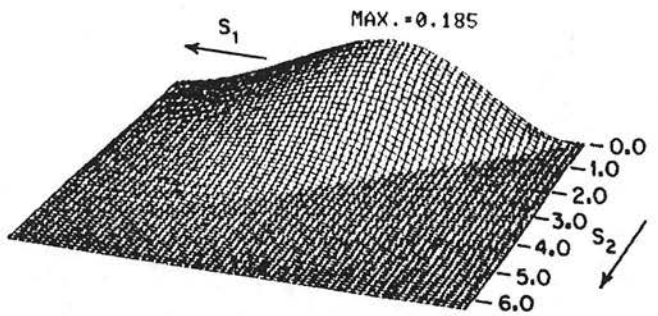


(a) σ = 1.0, front view

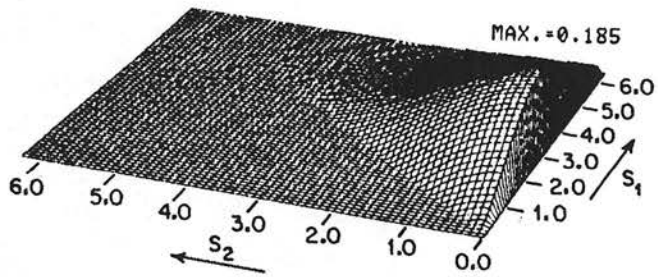


(b) σ = 1.0 side view

for (a) & (b) $g(\sqrt{2},0) = 0.74$



(c) σ = 2.0, front view



(d) σ = 2.0 side view

for (c) & (d) $g(2\sqrt{2},0) = 0.185$

Figure 5-1. Joint p.d.f. of singular values, $g(s_1,s_2)$ for a (uncorrelated) 2x2 texture field.

samples has been studied in [111]. The distribution of the largest eigenvalue of a sample covariance matrix in normal samples with population covariance matrix $\sigma^2 \underline{I}$ has been derived in [112] and the same for an arbitrary population covariance matrix has been derived in [113]. The following is a derivation for the p.d.f. of the dominant singular value of a random texture field with bidirectional correlation. The case of [113] is, in reality, a unidirectional quadratic form; while the case described below is for a bidirectional one.

Derivation 5.5: Let $\underline{\chi}$ be a diagonal matrix such that

$$\underline{\chi} = \begin{bmatrix} 1 & & & \underline{0} \\ & h_2^2 & & \\ & & h_3^2 \ddots h_k^2 & \\ \underline{0} & & & \end{bmatrix} \quad \text{and } 1 > h_2 > \ldots > h_k > 0$$

Then,

$$\int_{1>h_2>\ldots>h_k>0} \left(\prod_{i=2}^{k} h_i\right)^{2t-k} C_\rho(\underline{\chi}) \prod_{i=2}^{k} (1-h_i^2) \prod_{i<j} (h_i^2-h_j^2) \prod_{i=2}^{k} dh_i$$

$$= \frac{kt+P}{2^{k-1}} \frac{\Gamma_k(k/2)}{\pi^{k^2/2}} \frac{(t)_\rho}{(t+\frac{k+1}{2})_\rho} \frac{\Gamma_k(t)\Gamma_k(\frac{k+1}{2})}{\Gamma_k(t+\frac{k+1}{2})} C_\rho(\underline{I}_k) \qquad (5\text{-}40)$$

where,

$$(a)_\rho = \prod_{i=1}^{k} (a - \tfrac{1}{2}(i-1))_{p_i}$$

for $\rho = (P_1, P_2, \ldots, P_k)$, ρ is a partition of P into not more than k parts (see Appendix F). Also, as defined in Appendix B

and

$$\begin{cases} (x)_n = x(x+1)\ldots(x+n-1) \\ (x)_0 \triangleq 1 \end{cases}$$

Proof: Using Theorem 3 of [114] and modifying it, we will obtain

$$\int_{\underline{0}}^{\underline{I}} |\underline{\Phi}|^{(t-m)} C_{\rho}(\Phi) d\Phi = \frac{\Gamma_k(t)\,\Gamma_k(m)}{\Gamma_k(t+m)} \frac{(t)_{\rho}}{(t+m)_{\rho}} C_{\rho}(\underline{I}_k) \qquad (5\text{-}41)$$

where, $\underline{\Phi}$ is a positive definite symmetric matrix of order k. $\underline{\Phi}$ can be spectrally factorized as in (3-64); therefore $\underline{\Phi} = \underline{U}\,\underline{\Lambda}\,\underline{U}^T$.

The Jacobian (3-78) and the result of equation (3-80) are applied in (5-41). Also, letting m = (k+1)/2 and considering that if a positive definite matrix is varying between $\underline{0}$ and identity, its eigenvalues must vary in the space $1 > \lambda_1 > \lambda_2 > \ldots > \lambda_k > 0$. An integral equation will be obtained in terms of the eigenvalues. But, $\underline{U} \varepsilon O(k)$ and the integral of dU over the whole group is unity (Appendix C). Hence, performing the transformation $h_i^2 = \lambda_i/\lambda_1$ for i= 1,...,k on (5-41) and noting (F-10) of Appendix F for zonal polynomials, one will obtain (5-40) and hence the proof. **Q.E.D.**

P.d.f. of s_1, the Dominant Singular Value

Consider equation (5-31) which provides the joint p.d.f. of the singular values of the random texture field $\underline{F}$. In order to extract the p.d.f. of s_1, we need to integrate $g(s_1,s_2,\ldots,s_k)$ w.r.t. $s_2,\ldots,s_k$.

Derivation 5.6: The p.d.f. of the largest singular value of a random texture field is

$$g(s_1) = (\text{Constant}) \sum_{p=0}^{\infty} \sum_{\rho} \frac{(\frac{n}{2})_{\rho}}{(\frac{n+k+1}{2})_{\rho}} \frac{C_{\rho}(\underline{K}_C^{-1}) C_{\rho}(\underline{K}_R^{-1})}{C_{\rho}(\underline{I}_n)}$$

$$\times \left(-\frac{1}{2}\right)^p \frac{\frac{kn}{2}+p}{p!} s_1^{kn+2p-1} \qquad (5\text{-}42)$$

where,

$$\text{Constant} = \frac{1}{2^{(kn/2)-1}|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} \quad \frac{\Gamma_k(\frac{k+1}{2})}{\Gamma_k(\frac{n+k+1}{2})} \tag{5-43}$$

Proof: Consider equation (5-31) for the joint p.d.f. of $g(s_1,s_2,\ldots,s_k)$. Let $h_i \triangleq s_i/s_1$, hence $s_i = s_1 h_i$. Factoring all of the s_1's out, we will have

$$\begin{aligned} g(s_1,h_2,\ldots,h_k) = {} & \frac{2^{k(1-(n/2))}\pi^{k^2/2}}{\Gamma_k(\frac{k}{2})\Gamma_k(\frac{n}{2})\,|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} \\ & \cdot \prod_{i=2}^{k}(1-h_i^2) \prod_{i<j}^{k}(h_i^2-h_j^2)(s_1)^{2k(k-1)/2} \\ & \cdot \Big(\prod_{i=2}^{k} h_i\Big)^{n-k}(s_1)^{k(n-k)} \\ & \cdot \sum_{p=0}^{\infty} \sum_{\rho} \frac{C_\rho(-\frac{1}{2}\underline{K}_C^{-1})C_\rho(\underline{K}_R^{-1})C_\rho(\underline{X})(s_1)^{2p}}{p!\,C_\rho(\underline{I}_k)\,C_\rho(\underline{I}_n)} \\ & \cdot \Big(\prod_{i=2}^{k} dh_i\Big)(s_1)^{k-1}ds_1 \end{aligned} \tag{5-44}$$

where $\underline{X}$ has been defined in Derivation 5.5. Integrating (5-44) w.r.t. $h_2,\ldots,h_k$ yields

$$\begin{aligned} g(s_1)ds_1 = {} & \frac{2^{k(1-(n/2))}\pi^{k^2/2}}{\Gamma_k(\frac{k}{2})\Gamma_k(\frac{n}{2})|\underline{K}_C|^{n/2}|\underline{K}_R|^{k/2}} \sum_{p=0}^{\infty} \sum_{\rho} \frac{C_\rho(-\frac{1}{2}\underline{K}_C^{-1})C_\rho(\underline{K}_R^{-1})}{p!\,C_\rho(\underline{I}_k)C_\rho(\underline{I}_n)}\, s_1^{kn+2p-1}ds_1 \\ & \int_{1>h_2>\ldots>h_k>0} \Big(\prod_{i=2}^{k} h_i\Big)^{n-k} C_\rho(\underline{X}) \prod_{i<j}^{k}(h_i^2-h_j^2) \prod_{i=2}^{k}(dh_i) \end{aligned} \tag{5-45}$$

where,

$$\underline{X} = \begin{bmatrix} 1 & & & \underline{0} \\ & h_2^2 & & \\ & & \ddots & \\ \underline{0} & & & h_k^2 \end{bmatrix} \qquad (5\text{-}46)$$

Applying Derivation 5.5 on the integration of (5-45), we will obtain (5-42). Q.E.D.

Remark: The p.d.f. given by the Eq. (5-42) can be verified for a 1x1 matrix, which gives

$$g(s_1) = (2/\sqrt{2\pi\sigma^2}) \exp\left(\frac{-s_1^2}{2\sigma^2}\right) \text{ over the range } 0 \leq s_1 < \infty$$

This is logical because s_1 is equal to the absolute value of the 1x1 matrix.

5.4 Discussion on Accuracy and Convergence

The following steps have been taken into consideration to insure the accuracy of the p.d.f. of the largest singular value derived by Derivation 5.6:

1) The integration of (5-42) over the space $s_1 > 0$ is 1.
2) Considering $\lambda_1 = s_1^2$ and applying the Jacobian $\partial(s_1)/\partial(\lambda_1) = (2s_1)^{-1}$ we can derive the p.d.f. of the largest eigenvalue of $\underline{F}\,\underline{F}^T$. Then, letting $\underline{K}_R$ in the latter derivation be $\underline{I}_n$, i.e., no correlation among columns of $\underline{F}$, we will obtain Sugiyama's derivation, in [113], of the largest latent root of the sample covariance matrix and letting $\underline{K}_R = \underline{I}_n$ and $\underline{K}_C = \sigma^2\underline{I}_k$,

we will obtain the derivation in [112].

3) Convergence of $g(s_1)$ for $\underline{K}_R = \underline{I}_n$ can be shown with little difficulty: for $\underline{K}_R = \underline{I}_n$ (uncorrelated columns for $\underline{F}$) the following derivation gives the upper bound function for $g(s_1)$.

Derivation 5.7: For $\underline{K}_R = \underline{I}_n$, the p.d.f. of the dominant singular value of $\underline{F}$ is bounded as in (5-47)

$$g(s_1) < (\text{Const}) s_1^{kn-1} \left(\frac{kn}{2} + \frac{1}{2} s_1^2 \, \text{Tr}(\underline{K}_c^{-1})\right) \exp\left[\frac{1}{2} s_1^2 \, \text{Tr}(\underline{K}_c^{-1})\right] \quad (5\text{-}47)$$

where,

$$\text{Const} = \frac{1}{2^{\frac{kn}{2}-1} |\underline{K}_c|^{n/2}} \frac{\Gamma_k(\frac{k+1}{2})}{\Gamma_k(\frac{n+k+1}{2})}$$

Proof: Replacing $\underline{K}_R$ with $\underline{I}_n$ in (5-42)

$$g(s_1) = (\text{Const}) \, s_1^{kn-1} \sum_{p=0}^{\infty} \frac{\frac{kn}{2}+p}{p!} \left(\frac{1}{2} s_1^2\right)^p \sum_{\rho} \frac{(\frac{n}{2})_\rho}{(\frac{n+k+1}{2})_\rho} C_\rho(-K_C^{-1}) \quad (5\text{-}48)$$

Because terms before $\sum_{\rho}$ are positive, a bound can be established on the alternating part [122] as

$$\sum_{\rho} \frac{(\frac{n}{2})_\rho}{(\frac{n+k+1}{2})_\rho} C_\rho(-K_C^{-1}) < \left| \sum_{\rho} \frac{(\frac{n}{2})_\rho}{(\frac{n+k+1}{2})_\rho} C_\rho(-K_C^{-1}) \right|$$

$$< \sum_{\rho} \left| \frac{(\frac{n}{2})_\rho}{(\frac{n+k+1}{2})_\rho} C_\rho(-K_C^{-1}) \right|$$

$$= \sum_{\rho} \left| \frac{\left(\frac{n}{2}\right)_{\rho}}{\left(\frac{n+k+1}{2}\right)_{\rho}} \right| \left| C_{\rho}(-\underline{K}_C^{-1}) \right|$$

$$= \sum_{\rho} \frac{\left(\frac{n}{2}\right)_{\rho}}{\left(\frac{n+k+1}{2}\right)_{\rho}} \left| C_{\rho}(\underline{K}_C^{-1}) \right|$$

$$\leq \sum_{\rho} \left| C_{\rho}(\underline{K}_C^{-1}) \right| \qquad (5\text{-}49)$$

The last two steps in (5-49) were obtained because $C_{\rho}(-\underline{K}_C^{-1}) = (-1)^P C_{\rho}(\underline{K}_C^{-1})$ and

$$0 < \frac{\left(\frac{n}{2}\right)_{\rho}}{\left(\frac{n+k+1}{2}\right)_{\rho}} \leq 1. \qquad (5\text{-}49a)$$

According to, Remark 2, Appendix F, $C_{\rho}(\underline{K}_C^{-1}) > 0$ because $\underline{K}_C$ and hence $\underline{K}_C^{-1}$ are positive definite matrices. Thus, the last step in (5-49) will be equal to $\sum_{\rho} C_{\rho}(K_C^{-1})$. But, $\sum_{\rho} C_{\rho}(K_C^{-1})$ will be equal to $[Tr(K_C^{-1})]^P$. Hence (5-48) will lead to

$$g(s_1) \leq (Cont) s_1^{kn-1} \left[\sum_{P=0}^{\infty} \frac{kn/2}{P!} \left(\frac{1}{2} s_1^2\right)^P [Tr(K_C^{-1})]^P \right.$$
$$\left. + \sum_{P=1}^{\infty} \frac{1}{(P-1)!} \left(\frac{1}{2} s_1^2\right)^P [Tr(K_C^{-1})]^P \right] \qquad (5\text{-}50)$$

Using the McLaurin Series relation for exponential functions given by (5-29), (5-50) will lead to (5-47) and hence the proof. **Q.E.D.**

5.4.1 Application of the Bound

For $\underline{K}_C$, a kxk first order Markov covariance matrix,

$$\underline{K}_C = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \qquad (5\text{-}51)$$

where $0 \leq \rho \leq 1$. From $\underline{K}_C^{-1}$ given in [115], $\mathrm{Tr}(\underline{K}_C^{-1})$ will be obtained as

$$\mathrm{Tr}(\underline{K}_C^{-1}) = \frac{k+(k-2)\rho^2}{1-\rho^2} \qquad (5\text{-}52)$$

Therefore the bound presented by equation (5-47) for a kxn, texture field $\underline{F}$ with column covariance matrix $\underline{K}_C$ as given by (5-51), will be

$$g(s_1) \leq$$

$$\frac{1}{2^{\frac{kn}{2}-1}\,|\underline{K}_C^{-1}|^{n/2}} \frac{\Gamma_k\left(\frac{k+1}{2}\right)}{\Gamma_k\left(\frac{n+k+1}{2}\right)} s_1^{kn-1}\left[\frac{kn}{2}+\frac{1}{2}\, s_1^2\left(\frac{k+(k-2)\rho^2}{1-\rho^2}\right)\right]\exp\left(\frac{1}{2}s_1^2\left(\frac{k+(k-2)\rho^2}{1-\rho^2}\right)\right) \qquad (5\text{-}53)$$

5.5 A Model for $g(s_1)$ vs s_1

$g(s_1)$ given by (5-42) and (5-43) is an open form representation, which includes an infinite series. An infinite series including zonal polynomials is much like exp(x), which, if represented in a Taylor or McLaurin series expansion, its trailing terms are insignificant. For example, for $x\varepsilon(-1,1)$ in $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, we

have $\sum_{n=11}^{\infty} \frac{x^n}{n!} < 10^{-8}$ which means that with the first eleven terms of the power series, it is possible to numerically represent exp(x) up to eight digits of accuracy after the decimal point. It is also possible to speed up the convergence, by application of Remes single exchange algorithm [107], to a degree seven series while having the same accuracy. Zonal polynomials are of the same breed, but they have slow convergence. In order to obtain a graph for a general case of $g(s_1)$, the rate of convergence must be improved. Possible techniques for such enhancement have been explored. For example, convergence speed-up methods of Dahlquist and Bjorck [147] have been examined, but they seem to be unfit for zonal polynomials. Because of these problems associated with numerical evaluation of the open form for $g(s_1)$, an attempt has been made to develop a model for $g(s_1)$ that will provide some insight into its behavior.

The domain of the largest singular value (s_1) is the positive portion of the s_1 axis ($0 \leq s_1 < \infty$). It is evident that for $s_1 = 0$, $g(s_1) = 0$. Because the area under $g(s_1)$ is unity, and $g(s_1)$ is non-negative for all values of s_1, it can be concluded that $g(s_1)$ is convergent and that $g(s_1)$ asymptotically goes to zero. Rayleigh, Maxwell, or chi-square (of higher than two degrees of freedom) densities satisfy these conditions. Such models completely agree with the behavior of $g(s_1)$ for an uncorrelated 2x2 case. For a 2x2 matrix $\underline{F}$, a closed form for $g(s_1)$ can be analytically derived as a marginal density of $g(s_1,s_2)$ given by (5-39). Let $\underline{K}_C = \underline{K}_R = \sigma^2 \underline{I}_2$ be the column and row covariance matrices of dimension two. Then $g(s_1)$ is

$$g(s_1) = \frac{s_1}{\sigma^2} \exp\left(-\frac{s_1^2}{\sigma^2}\right) + \frac{\sqrt{2\pi}}{\sigma}\left(\frac{s_1^2}{\sigma^2} - 1\right)\exp\left(-\frac{s_1^2}{2\sigma^2}\right)\text{erf}\left(\frac{s_1}{\sigma}\right) \quad (5\text{-}54)$$

where, σ^2 = variance of the elements of $\underline{F}$, and

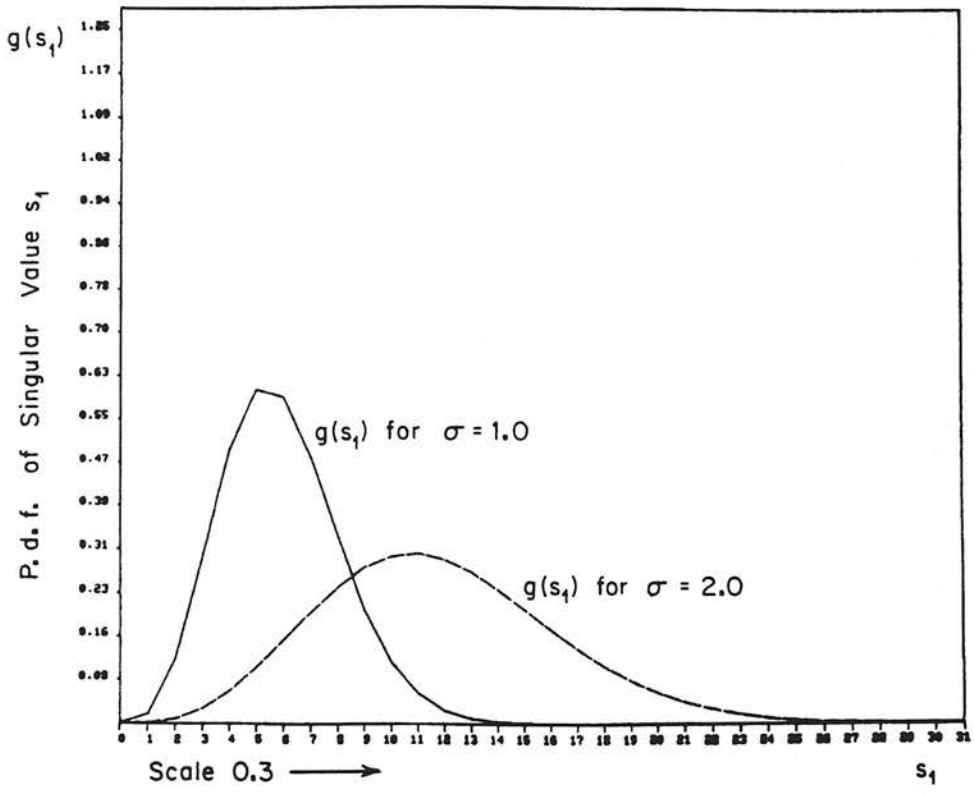
$$\text{erf}(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_0^y e^{-(x^2/2)} dx \quad (5\text{-}55)$$

It is possible to independently derive a closed form for $g(s_2)$, as another marginal density of $g(s_1,s_2)$ for which $\underline{K}_C = \underline{K}_R = \sigma^2 I_2$:

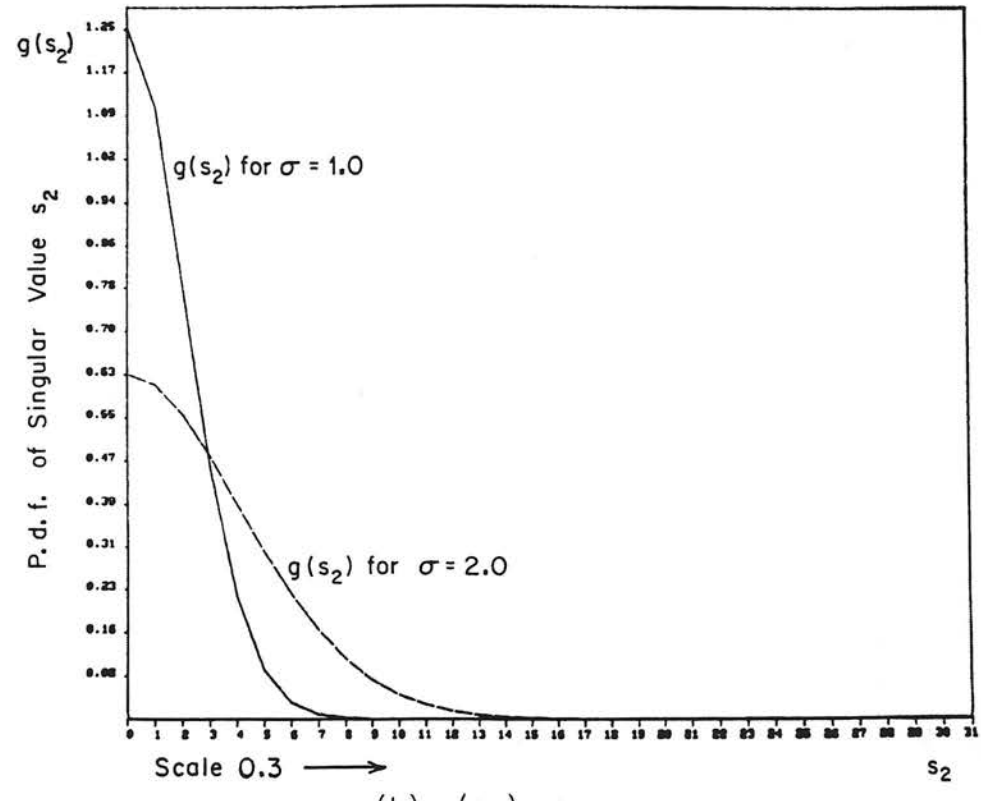
$$g(s_2) = \frac{s_2}{\sigma^2} \exp\left(-\frac{s_2^2}{\sigma^2}\right) + \frac{\sqrt{2\pi}}{\sigma}\left(\frac{s_2^2}{\sigma^2} - 1\right)\exp\left(-\frac{s_2^2}{2\sigma^2}\right)\text{erf}\left(\frac{s_2}{\sigma}\right)$$
$$- \left(\frac{1}{2}\right)\frac{\sqrt{2\pi}}{\sigma}\left(\frac{s_2^2}{\sigma^2} - 1\right)\exp\left(-\frac{s_2^2}{2\sigma^2}\right) \quad (5\text{-}56)$$

As can be seen, the mathematical representations of $g(s_1)$ and $g(s_2)$ are alike except for an extra term for $g(s_2)$. Plots of $g(s_1)$ and $g(s_2)$ are given in Figure 5-2. In Figure 5-2(a), $g(s_1)$ has been plotted for $\sigma = 1.0$ and $\sigma = 2.0$. The same values of σ have been plotted for $g(s_2)$ in Figure 5-2(b).

The plots of $g(s_1)$ shown in Figure 5-2(a) agree with the model of Rayleigh or Maxwell shape for a general order $g(s_1)$. In both cases of $\sigma = 1$ and $\sigma = 2$, the shapes are those desired, except for the case of $\underline{1x1}$ matrix which has its pdf of s_1 equal to twice the positive portion of a zero mean normal density with variance σ^2. The higher the σ, the wider and lower are $g(s_1)$ and $g(s_2)$. In fact, mathematically speaking, if σ is multiplied by n (in $g(s_1)$ and



(a) $g(s_1)$ vs s_1



(b) $g(s_2)$ vs s_2

Figure 5-2. Plots of $g(s_i)$ vs s_i for 2x2 Case.

$g(s_2)$), the mean and standard deviation of s_1 and s_2 will also multiply by n. The integration of $g(s_1)$ and $g(s_2)$ over the space of s_1 and s_2, respectively, will be unity in all cases. Table 5-1 numerically evaluates equation (5-54) to obtain integration of $g(s_1)$, and mean and standard deviation of s_1 with their respective computational errors for both cases $\sigma = 1.0$ and $\sigma = 2.0$. Table 5-2 shows the same for the second singular value using equation (5-56). One observation from Figure 5-2(b) is that it has its highest probability around zero. This behavior agrees with the space of variation $0 \leq s_2 \leq s_1 < \infty$.

For higher dimensional texture fields, a table of experimental moments and variance for the largest singular value of randomly generated textures can be obtained. Table 5-3 experimentally evaluates the largest singular value moments of a 32x32 texture field using 64 independent 32x32 sample texture fields for each correlation factor ρ. In Table 5-3, "ρ" denotes texture field whose correlation matrix is $\underline{K}_C = \underline{K}_R = [\rho^{|i-j|}]$. Figure 5-3 shows plots of the mean and standard deviation given by Table 5-3 as a function of ρ for the model of $g(s_1)$.

Based on the Table 5-3, Figure 5-2(a), 5-3, and the discussion presented in this section, the model of $g(s_1)$ for a general case is shown by Figure 5-4 as a function of s_1 w.r.t. different correlation factors.

5.6 Summary

In this chapter, using the derivations of Chapter 3 on differential SVD; and the texture model $\underline{F}$ of Chapter 4, we developed

Table 5-1. Numerical Evaluations for s_1.

σ	$\int g(s_1)ds_1$	Error	$E(s_1)$	Error	Standard Deviation of s_1	Error
1.0	1.00	3×10^{-6}	1.77	5×10^{-6}	0.65	2×10^{-6}
2.0	1.00	3×10^{-6}	3.54	2×10^{-6}	1.31	3×10^{-6}

Table 5-2. Numerical Evaluations for s_2.

σ	$\int g(s_2)ds_2$	Error	$E(s_2)$	Error	Standard Deviation of s_2	Error
1.0	1.00	5×10^{-6}	0.52	3×10^{-6}	0.40	2×10^{-6}
2.0	1.00	3×10^{-6}	1.04	1×10^{-6}	0.80	2×10^{-6}

Table 5-3. Experimental Evaluation of Mean, Second Moment, and Variance of the Largest Singular Value for 32x32.

TEXTURE	$E\{s_1\}$	$E\{s_1^2\}$	$Var\{s_1\}$
"0.0"	10.8060	116.9890	0.2191
"0.5"	16.4986	274.1390	1.9644
"0.6"	19.7199	393.2498	4.4426
"0.7"	24.7485	624.3566	12.0559
"0.8"	33.4432	1160.0092	42.2152
"0.9"	52.0764	2947.2329	239.0112

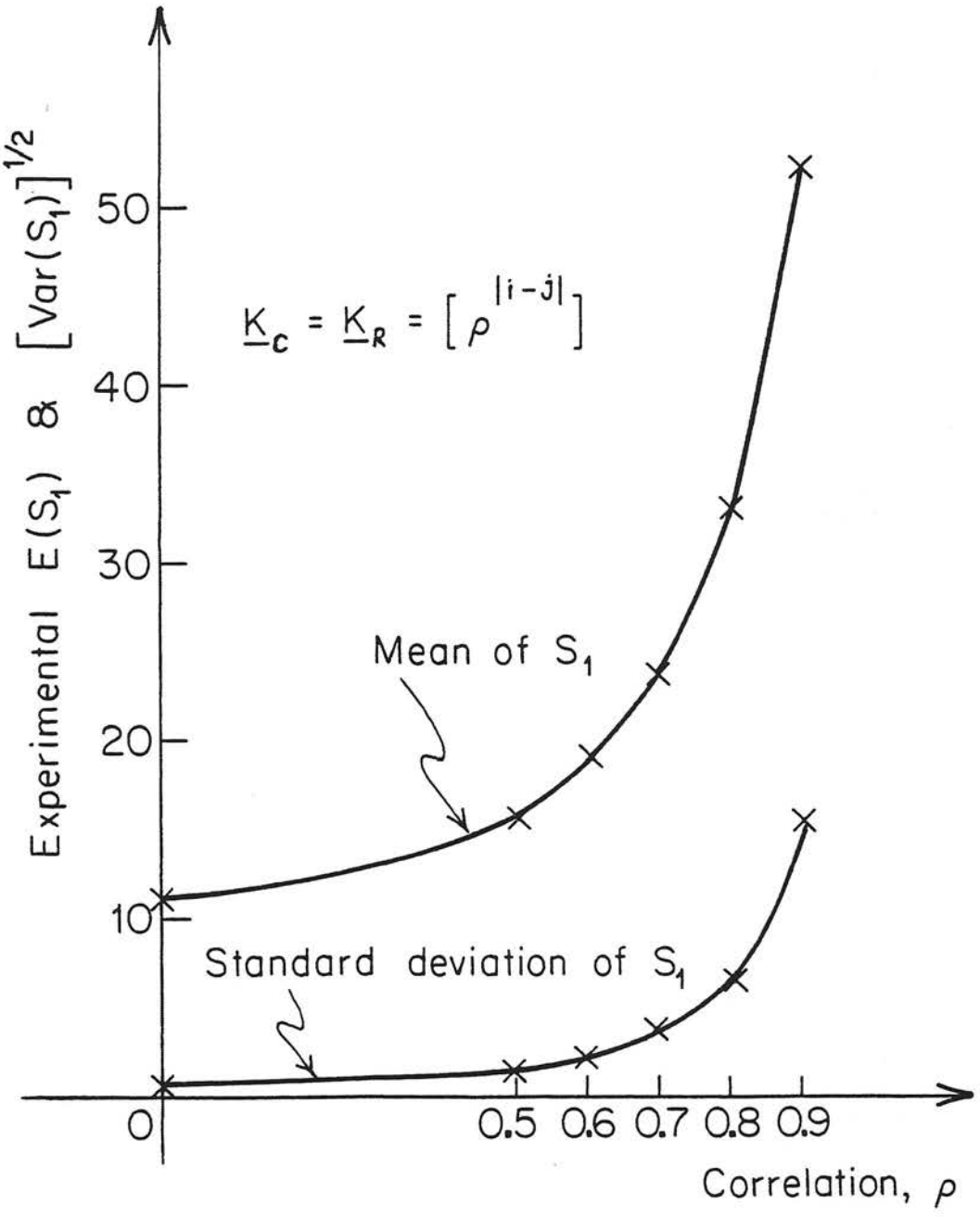


Figure 5-3. Plots of Mean and Standard Deviation of s_1 as a Function of ρ for 32x32 First Order Markov $\underline{K}_C$ and $\underline{K}_R$.

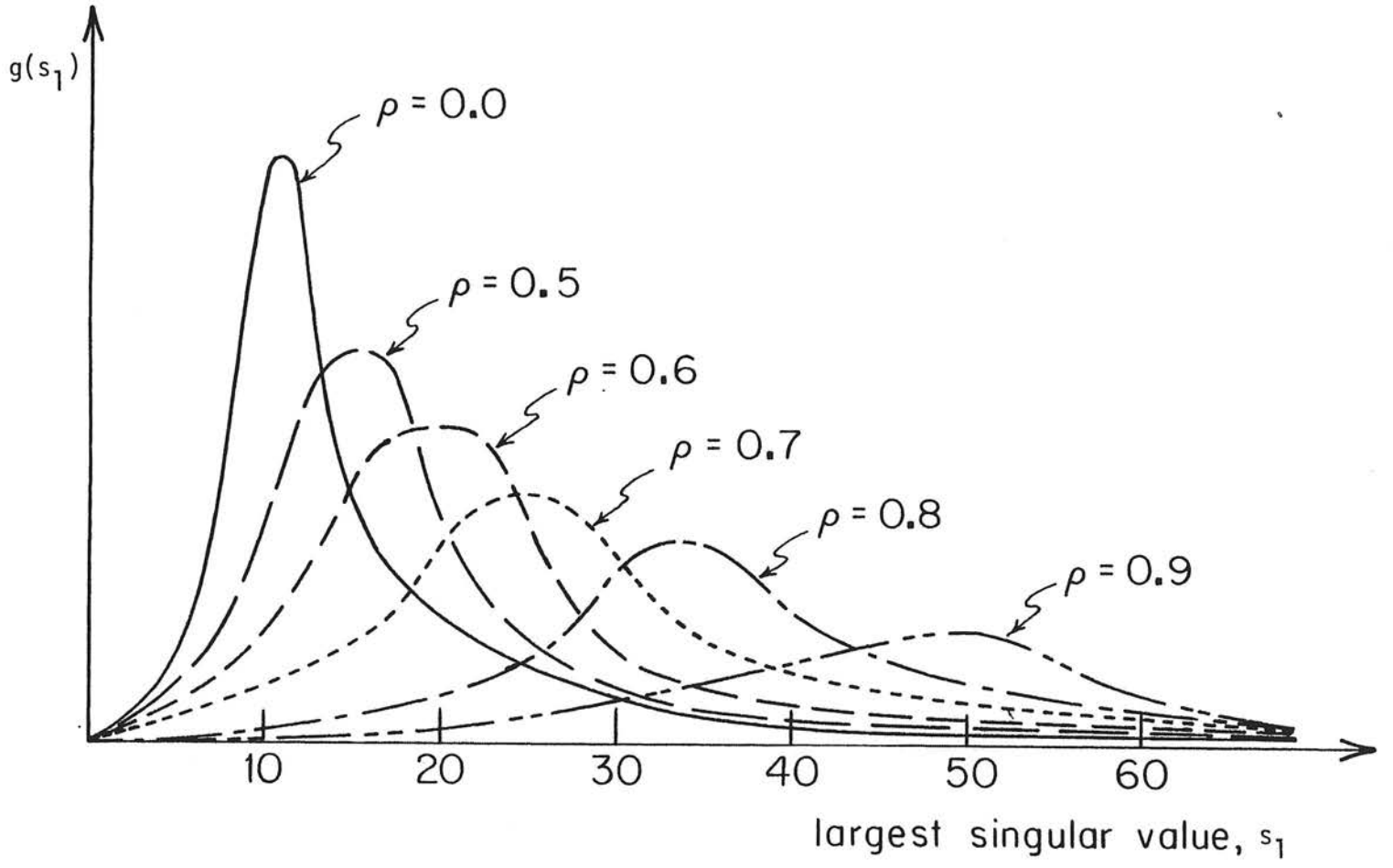


Figure 5-4. Model of $g(s_1)$ vs s_1 w.r.t. ρ.

the p.d.f. of a stochastic texture field. The joint p.d.f. of the singular values of $\underline{F}$ has been developed. From the latter derivation, the p.d.f. of the dominant singular values of $\underline{F}$ has been extracted, an upper bound for the p.d.f. of the dominant singular value has been derived, and a family of p.d.f. plots has been generated for various textural correlation factors.

Two of the interesting results of this chapter are that: a) p.d.f. and distribution function of the singular values of two stochastic texture fields are identical iff the mean and covariance matrices of the two are identical; and b) because the first and second order statistics determine the appearance of texture fields; similar looking textures have similar singular value distributions. In other words, the conjecture in [5] and [6] which was discussed by Figures 1-1 through 1-3 in Chapter 1 has been mathematically proved in the present chapter.

CHAPTER 6

MULTIVARIATE STATISTICAL DERIVATIONS FROM TEXTURE BY SVD

In this chapter, singular value decomposition will be used to derive multivariate relations.

6.1 Quadratic Form

Quadratic form is an important concept in multivariate analysis. The following presentation is a brief discussion, definition and derivation.

6.1.1 Definition

Let $\underline{X}$ and $\underline{Y}$ be $\underline{kxn}$ matrices whose n columns are independent k-variate random vectors from the same distribution d. Let $\underline{A}$ be a $\underline{nxn}$ positive definite symmetric matrix, and let $E[\underline{X}] = \underline{m}_X$ and $E[\underline{Y}] = \underline{m}_Y$. Then,

$$\underline{B} \triangleq (\underline{X}-\underline{m}_X)\underline{A}(\underline{Y}-\underline{m}_Y)^T \qquad (6\text{-}1)$$

where $\underline{B}$ is called a bilinear form in samples from distribution d. In (6-1), if $\underline{Y} = \underline{X}$ then

$$\underline{Q} = (\underline{X}-\underline{m}_X)\underline{A}(\underline{X}-\underline{m}_X)^T \qquad (6\text{-}2)$$

where $\underline{Q}$ is called a quadratic form in samples from distribution d.

For (6-2), $\underline{X}\ \underline{A}\ \underline{X}^T$ is called a <u>non-central</u> quadratic form and if $\underline{m}_X = \underline{0}$, i.e. the samples are taken from a zero mean population, $\underline{X}\ \underline{A}\ \underline{X}^T$ is a <u>central quadratic</u> form. Relation (6-2) itself is a

central quadratic form because if we let $\underline{Z} = \underline{X} - \underline{m}_X$, then, $E[\underline{Z}] = \underline{0}$, hence resulting in $\underline{Z}\,\underline{A}\,\underline{Z}^T$.

6.1.2 Formation From Texture Field $\underline{F}$

From relation (4-34),

$$\underline{F} = \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{M}\, \underline{\Lambda}_R^{1/2} \underline{E}_R^T \tag{6-3}$$

where $\underline{M}$ is a $\underline{kxn}$ random matrix obtained from a $\underline{knx1}$ white random process $\underline{m}$ (Appendix E). Hence, the elements of $\underline{M}$ are uncorrelated, zero mean, unit variance random variables. From (6-3), and the discussions of Chapter 4, it is evident that $\underline{F}$ is a $\underline{kxn}$ texture field with a separable covariance matrix $\underline{K}_C \otimes \underline{K}_R$ where $\underline{K}_C$ is a $\underline{kxk}$ covariance for columns and $\underline{K}_R$ is $\underline{nxn}$ for rows. $(\underline{E}_C, \underline{\Lambda}_C)$ and $(\underline{\Lambda}_R, \underline{E}_R)$ are the result of spectral factorizations of $\underline{K}_C$ and $\underline{K}_R$ (Section 4.2.2).

Note that if $\underline{K}_R$ is identity, i.e., if columns of $\underline{F}$ are independent,

$$\underline{X} \triangleq \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{M} \tag{6-4}$$

Forming $\underline{F}\,\underline{F}^T$:

$$\begin{aligned} \underline{F}\,\underline{F}^T &= \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{M}\, \underline{\Lambda}_R^{1/2} \underline{E}_R^T \underline{E}_R \underline{\Lambda}_R^{1/2} \underline{M}^T \underline{\Lambda}_C^{1/2} \underline{E}_C^T \\ &= \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{M}\, \underline{\Lambda}_R \underline{M}^T \underline{\Lambda}_C^{1/2} \underline{E}_C^T \\ &= \underline{X}\, \underline{\Lambda}_R \underline{X}^T \end{aligned} \tag{6-5}$$

Thus, $\underline{F}\,\underline{F}^T$ is a quadratic form. It can be proved that the p.d.f. of $\underline{X}\,\underline{K}_R\,\underline{X}^T$ is the same as $\underline{X}\,\underline{\Lambda}_R \underline{X}^T$ and that p.d.f. of the eigenvalues of both will also be the same. For $\underline{K}_R = \underline{\Lambda}_R = \underline{I}_N$, $\underline{F}\,\underline{F}^T = \underline{X}\,\underline{X}^T$ where $\underline{X}\,\underline{X}^T$

is a Wishart matrix [6]. The p.d.f. of $\underline{X}\,\underline{X}^T$ is given in [Ibid].

$$\underline{F} = \underline{X}\,\underline{\Lambda}_R^{1/2}\underline{E}_R^T \tag{6-6}$$

and if elements of $\underline{M}$, which generates $\underline{X}$ by relation (6-4), are independent and unit variance, and $\underline{M}$ has mean $\underline{\mathcal{M}}_M$ then

$$\begin{aligned} E[\underline{F}] &= E[\underline{X}]\underline{\Lambda}_R^{1/2}\underline{E}_R^T \\ \underline{\mathcal{M}}_F &= \underline{\mathcal{M}}_X\underline{\Lambda}_R^{1/2}\underline{E}_R^T \\ &= \underline{E}_C\underline{\Lambda}_C^{1/2}\underline{\mathcal{M}}_M\underline{\Lambda}_R^{1/2}\underline{E}_R^T \end{aligned} \tag{6-7}$$

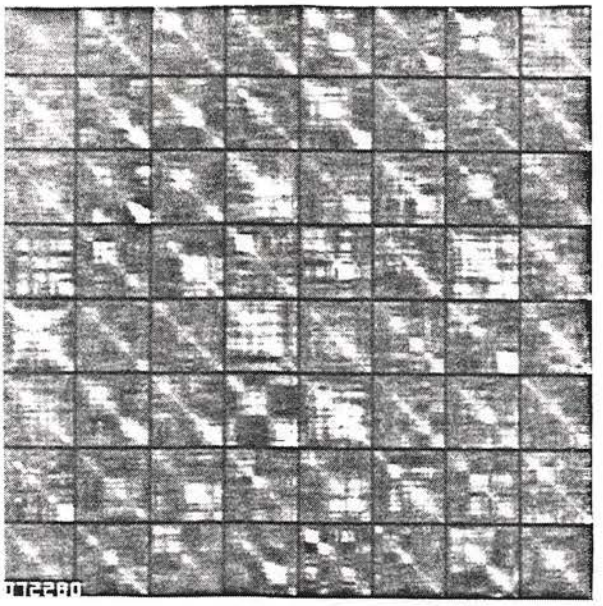
For non-zero $\underline{\mathcal{M}}_F$, $(\underline{F}-\underline{\mathcal{M}}_F)(\underline{F}-\underline{\mathcal{M}}_F)^T$ will be a central quadratic form, and $\underline{F}\,\underline{F}^T$ will be the non-central quadratic form. Figures 6-1 (a and b) and 6-2 (a and b) respectively show examples of central and non-central quadratic and Wishart forms.

Remark: If a certain mean $\underline{\mathcal{M}}_F$ is desired, a simple method is to generate a zero mean matrix as in (6-3) and then add $\underline{\mathcal{M}}_F$ to it. The rank of $\underline{\mathcal{M}}_F$ can be full, but it is usually one because columns of $\underline{X}$ usually have equal mean vectors.

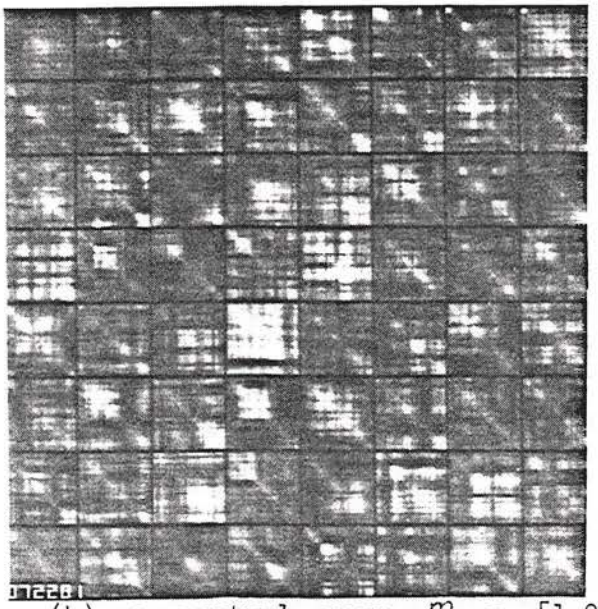
6.1.3 Moment

Let $\underline{F}$ be a separable kxn random texture field with bidirectional correlation.

Derivation 6.1: Mean matrix of non-central quadratic form $\underline{F}\,\underline{F}^T$ is $\underline{R}_C\,\mathrm{Tr}(\underline{K}_R)$, where $\underline{R}_C$ is correlation matrix along the columns and

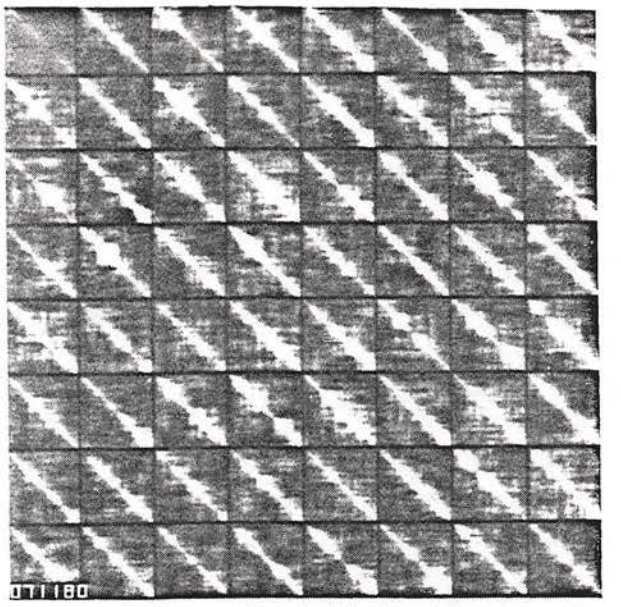


(a) central

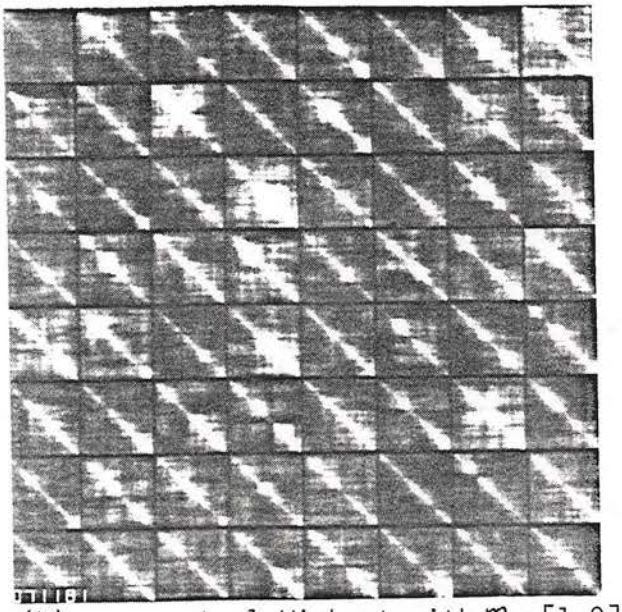


(b) non-central, mean = $\underline{m}_F$ = [1.0]

Figure 6-1. Sample 32x32 matrices from central and non-central $\underline{F}\,\underline{F}^T = \underline{X}\,\underline{\Lambda}_R\underline{X}^T$ with $\rho_C = \rho_R = 0.80$.



(a) central Wishart



(b) non-central Wishart with $\underline{m}_x$=[1.0]

Figure 6-2. Samples of 64 independent 32x32 matrices from the central and non-central Wishart distribution $W(\underline{X}\,\underline{X}^T, 32, 32, \underline{K}_C)$ where $\underline{\rho}_C = 0.80$.

$\underline{K}_R$ is covariance matrix along the row direction.

Proof: According to (6-5)

$$\underline{F}\,\underline{F}^T = \underline{X}\Lambda_R\underline{X}^T \tag{6-8}$$

where $\underline{X}$ is a matrix whose columns are independent and have correlation matrix $\underline{R}_C$ generated by $N(\underline{\mathcal{M}}_C, \underline{K}_C)$, and $\Lambda_R = \text{diag}[\lambda_R(1), \lambda_R(2), \ldots, \lambda_R(n)]$, then

$$\begin{aligned} E\{\underline{F}\,\underline{F}^T\} &= E\{\underline{X}\,\Lambda_R\underline{X}^T\} \\ &= E\{\sum_{i=1}^{k} \lambda_R(i) x_i x_i^T\} \\ &= \sum_{i=1}^{n} \lambda_R(i)(\underline{R}_C) \\ &= \underline{R}_C\, \text{Tr}(\Lambda_R) \\ &= \underline{R}_C\, \text{Tr}(\underline{K}_R) \end{aligned} \tag{6-9}$$

Q.E.D.

Derivation 6.2: Mean matrix of a central quadratic form $E(\underline{F}-\underline{\mathcal{M}}_F)(\underline{F}-\underline{\mathcal{M}}_F)^T$ is $\underline{K}_C\, \text{Tr}(\underline{K}_R)$, where $\underline{K}_C$ is the covariance matrix along column $\underline{F}$ where $E\{\underline{F}\} = \underline{\mathcal{M}}_F$.

Proof:

$$\begin{aligned} E\{(\underline{F}-\underline{\mathcal{M}}_F)(\underline{F}-\underline{\mathcal{M}}_F)^T\} &= E\{(\underline{X}-\underline{\mathcal{M}}_X)\Lambda_R(\underline{X}-\underline{\mathcal{M}}_X)^T \\ &= \sum_{i=1}^{n} \lambda_R(i) E\{(x_i-\underline{\mathcal{M}}_x)(x_i-\underline{\mathcal{M}}_x^T) \\ &= \underline{K}_C\, \text{Tr}(\underline{K}_R) \end{aligned} \tag{6-10}$$

where $\underline{\mathcal{M}}_C = E\{\underline{x}_i\}$, $\underline{X} = (\underline{x}_1, \ldots, \underline{x}_n)$, $E\{\underline{X}\} = \underline{m}_X$ and $x_i \perp\!\!\!\perp x_j$. **Q.E.D.**

Derivation 6.2-1: Mean of a non-central quadratic form is equal to the mean of central quadratic form plus quadratic form of the mean.

Proof: This is obvious because

$$E\{\underline{F}\,\underline{F}^T\} = E\{(\underline{F}-\underline{m}_F)(\underline{F}-\underline{m}_F)^T\} + \underline{m}_F\underline{m}_F^T \qquad (6\text{-}11)$$

Q.E.D.

Derivation 6.2-2: For $E\{\underline{F}\} = \underline{m}_F$ with rank one

$$\underline{m}_F\underline{m}_F^T = \underline{\mathcal{M}}_C\underline{\mathcal{M}}_C^T \, Tr(\underline{K}_R) \qquad (6\text{-}12)$$

where $\underline{\mathcal{M}}_C = E\{\underline{x}_i\}$ = mean vector of each column of $\underline{X}$.

Proof:

From the results of derivations 6.1, 6.2 and 6.2-1, and since $\underline{R}_C = (\underline{K}_C + \underline{\mathcal{M}}_C\underline{\mathcal{M}}_C^T)$

$$E\{\underline{F}\,\underline{F}^T\} = \underline{R}_C \, Tr(\underline{K}_R)$$

$$= \underline{K}_C \, Tr(\underline{K}_R) + \underline{\mathcal{M}}_C\underline{\mathcal{M}}_C^T \, Tr(\underline{K}_R) \qquad (6\text{-}13)$$

and hence the proof. **Q.E.D.**

Remark:

$$\underline{\mathcal{M}}_C = (\underline{E}_C\underline{\Lambda}_C^{1/2})\underline{\mathcal{M}}_m \qquad (6\text{-}14)$$

where, $\underline{\mathcal{M}}_m = (m_1, m_1, \ldots, \underset{\substack{\uparrow \\ k\text{-th element}}}{m_1}]^T$

Derivation 6.2-3: For $\underline{K}_R$ a first order Markov process with variance σ_R^2

$$E\{\underline{F}\,\underline{F}^T\} = n\sigma_R^2\underline{R}_C \tag{6-15}$$

Proof: The diagonal terms of $\underline{K}_R$ are all σ_R^2, hence $Tr(\underline{K}_R) = n\sigma_R^2$.

Q.E.D.

Derivation 6.2-4: If the texture field $\underline{F}$ is zero mean, then Derivation 6.2-3 will result in

$$E\{\underline{F}\,\underline{F}^T\} = n\sigma_R^2\underline{K}_C \tag{6-16}$$

Remark: $\underline{F}^T\underline{F}$ can also be called a quadratic since if $F^T = \underline{A}$, $\underline{F}^T\underline{F} = \underline{A}\,\underline{A}^T$.

Derivation 6.3:

$$E\{\underline{F}^T\underline{F}\} = \underline{R}_R\,Tr(\underline{K}_C) \tag{6-17}$$

$$E\{(\underline{F}-\underline{m}_F)^T(\underline{F}-\underline{m}_F)\} = \underline{K}_R\,Tr(\underline{K}_C) \tag{6-18}$$

where $\underline{R}_R$ and $\underline{K}_R$ are the correlation and covariance matrices along the rows.

Proof: We can define a nxk matrix $\underline{Y}$ whose k columns are uncorrelated and each column from $N(\underline{m}_R, \underline{K}_R)$

$$\underline{Y} = \underline{E}_R \underline{\Lambda}_R^{1/2} \underline{M}^T \tag{6-19}$$

hence,

$$\underline{Y}^T = \underline{M}\, \underline{\Lambda}_R^{1/2} \underline{E}_R^T \tag{6-20}$$

and therefore

$$\underline{F} = \underline{E}_C \underline{\Lambda}_C^{1/2} \underline{Y}^T \tag{6-21}$$

$$\underline{F}^T \underline{F} = \underline{Y}\, \underline{\Lambda}_C \underline{Y}^T \tag{6-22}$$

and the same proof as in derivations 6.1 and 6.2 will follow. **Q.E.D.**

Derivation 6.3-1:

$$E\{\underline{F}^T\, \underline{F}\} = \underline{R}_R\, \mathrm{Tr}(\underline{K}_C) = \underline{K}_R\, \mathrm{Tr}(\underline{K}_C) + \underline{m}_F^T \underline{m}_F \tag{6-23}$$

and for rank one $\underline{m}_F$

$$\underline{m}_F^T \underline{m}_F = \underline{\mathcal{M}}_R \underline{\mathcal{M}}_R^T\, \mathrm{Tr}(\underline{K}_C) \tag{6-24}$$

Derivation 6.3-2:

$$E\{\underline{U}\, \underline{\Lambda}\, \underline{U}^T\} = \underline{K}_C\, \mathrm{Tr}(\underline{K}_R) + \underline{m}_F \underline{m}_F^T = \underline{R}_C\, \mathrm{Tr}(\underline{K}_R) \tag{6-25}$$

and,

$$E\{\underline{V}\, \underline{\Lambda}\, \underline{V}^T\} = \underline{K}_R\, \mathrm{Tr}(\underline{K}_C) + \underline{m}_F^T \underline{m}_F = \underline{R}_R\, \mathrm{Tr}(\underline{K}_C) \tag{6-26}$$

where $\underline{\Lambda} = \underline{S}^2$.

6.2 Texture Energy

Based on SVD, some useful statistics can be obtained, which will be used in the application chapters.

Energy in each element of a random texture field $\underline{F}$ is equal to

the second moment of that element. Hence, if $\underline{F}$ is represented by its vector of lexicographic transformation $\underline{f}$, the diagonal elements of $E\{\underline{f}\,\underline{f}^T\}$ will contain the second moment of elements of $\underline{F}$ e.g. $E\{F^2(i,j)\}$.

For a separable $\underline{kxn}$ texture field $\underline{F}$ with mean matrix of rank one and using column scanning

Elemental Energy in $\underline{F}$ = Second Moment of the Corresponding Diagonal Term of $\underline{f}\,\underline{f}^T$. (6-27)

For a zero mean (central) $\underline{F}$ elemental energy is equal to

$$E\{F^2(i,j)\} = K_C(i,i)K_R(j,j) \tag{6-28}$$

Defining the matrix $\underline{V}_F$ for energy distribution of $\underline{F}$, $\underline{V}_F$ will be

$$\underline{V}_F = \underline{k}_C\underline{k}_R^T \tag{6-29}$$

where $\underline{k}_C$ and $\underline{k}_R$ are vectors of diagonal terms of $\underline{K}_C$ and $\underline{K}_R$ where $k_C = [K_C(1,1),\ldots,K_C(k,k)]^T$ and $k_R = [K_R(1,1),\ldots,K_R(n,n)]^T$. Of course, if $\underline{K}_C$ and $\underline{K}_R$ are first order Markov matrices, their diagonal terms for $\underline{K}_C$ and $\underline{K}_R$ will be equal to their variance coefficients σ_C^2 and σ_R^2. Hence

$$\underline{V}_F = \sigma_C^2\sigma_R^2 \begin{bmatrix} 1 & \ldots & 1 \\ 1 & \ldots & 1 \end{bmatrix} \tag{6-30}$$

where the matrix of 1's in (6-30) is a kxn matrix with all of its entries 1.

Total Energy:

The total energy in texture will be equal to the summation of energies of all elements of $\underline{F}$.

Derivation 6.4: The energy in a non-central random texture field is $Tr(\underline{R}_C)Tr(\underline{K}_R)$, and in a central one is, $Tr(\underline{K}_C)Tr(\underline{K}_R)$.

Proof:

$$\text{Texture Energy} = \sum_{i,j}^{k,n} E\{F^2(i,j)\} = Tr\, E\{\underline{f}\,\underline{f}^T\}$$

$$= E\{Tr(\underline{f}\,\underline{f}^T)\}$$

$$= E\{Tr(\underline{F}\,\underline{F}^T\}$$

$$= Tr\, E\{\underline{F}\,\underline{F}^T\}$$

$$= Tr(\underline{R}_C)Tr(\underline{K}_C) \qquad (6\text{-}31)$$

The equation (6-31) is consistant with equation (6-9). If $\underline{F}$ is central, $\underline{K}_C$ can replace $\underline{R}_C$ in (6-31). **Q.E.D.**

If $\underline{K}_C$ and $\underline{K}_R$ are first order Markov with variance coefficients σ_C^2 and σ_R^2, the energy will be $kn\sigma_C^2\sigma_R^2$. Equation (6-31) will be equal to the summation of all entries in $\underline{V}_F$.

Remark 1: For zero mean 512x512 artificial texture fields used in Chapters 4, 8, and 9, $\sigma_C^2 = \sigma_R^2 = 1.0$, hence the total energy = 512 x 512 = 262144.00, which is also equal to the texture variance.

Remark 2: Dividing energy texture element by standard deviation 512 will result in a unit variance texture field.

6.2.1 Moment of the Sum of Squares of Singular Values:

Since

$$\text{Texture's Total Energy} = E\{\text{Tr}\ \underline{F}\ \underline{F}^T\}$$

$$= E\{\text{Tr}\ \Lambda\} \tag{6-32}$$

$$E\left(\sum_{i=1}^{k} s_i^2\right) = \text{Tr}(\underline{R}_C)\text{Tr}(\underline{K}_C)$$

$$= \text{energy in texture} \tag{6-33}$$

6.3 Invariance of SVD to Unitary Transformation

A property of the SVD is its invariance to unitary transforms [6]. Consider a unitary transformation

$$\underline{Z} = \underline{L}_C^T \underline{F}\ \underline{L}_R \tag{6-34}$$

where,

$$L_C L_C^{*T} = \underline{I}_k \quad \text{and} \quad L_R L_R^{*T} = \underline{I}_n \ .$$

It is easily demonstrated that the characteristic equations of $\underline{Z}\ \underline{Z}^T$ and $\underline{F}\ \underline{F}^T$ are identical, and hence, the singular values of $\underline{F}$ and $\underline{Z}$ are the same. The eigenvector matrices are related by

$$\underline{U}_Z = \underline{L}_C^T \underline{U} \tag{6-35}$$

$$\underline{V}_Z = \underline{L}_R^T \underline{V} \tag{6-36}$$

As a consequence of this property, it is observed that the singular

values of an image block and its two-dimensional Fourier and Hadamard Transform, for example, are identical. Certainly, the structure of such arrays is substantially different.

Based on the above argument, if we form a random array $\underline{W}$

$$\underline{W} = \underline{E}_C^T \underline{F}\, \underline{E}_R \tag{6-37}$$

where $\underline{F}$ is a separable, zero mean correlated random texture field (Chapter 4) with covariance matrices $\underline{K}_C$ and $\underline{K}_R$. Upon spectral factorization of $\underline{K}_C$ and $\underline{K}_R$

$$\underline{\Lambda}_C = \underline{E}_C^T \underline{K}_C \underline{E}_C \tag{6-38}$$

$$\underline{\Lambda}_R = \underline{E}_R^T \underline{K}_R \underline{E}_R \tag{6-39}$$

E_C, $\underline{E}_R$, Λ_C, Λ_R, are eigenvector and eigenvalue matrices of $\underline{K}_C$ and $\underline{K}_R$, as noted. It can easily be proved that the covariance matrices of $\underline{W}$ along column and row are $\underline{\Lambda}_C$ and $\underline{\Lambda}_R$, and hence, $\underline{W}$ is an uncorrelated array. From (6-35) and (6-36), it has been established that the singular values of an uncorrelated array $\underline{W}$ and correlated array $\underline{F}$ are identical. This may seem to indicate that the SVD is useless for characterizing the structure of correlated arrays. But, this is not so. The spatial correlation information of $\underline{F}$ has simply been converted to another form--the variance distribution of $\underline{W}$. To show this, let $\underline{V}_W$ be a $\underline{k}\underline{x}\underline{n}$ matrix of variance terms of $\underline{W}$ in spatial correspondence with $\underline{W}$ [6]. Then as in (6-29)

$$\underline{V}_W = \underline{\gamma}_C \underline{\gamma}_R^T \tag{6-40}$$

where the vector $\underline{\gamma}_C$ and $\underline{\gamma}_R$ contain the diagonal terms of $\underline{\Lambda}_C$ and $\underline{\Lambda}_R$, respectively. For some well-defined processes, such as Markov processes, $\underline{V}_W$ can be determined analytically. The texture energy in $\underline{W}$ is $\mathrm{Tr}(\underline{\Lambda}_C)\mathrm{Tr}(\underline{\Lambda}_R)$, which is equivalent to that of $\underline{F}$, which is $\mathrm{Tr}(\underline{K}_C)\mathrm{Tr}(\underline{K}_R)$. This is logical because unitary transforms are energy preserving operations.

To summarize, if $\underline{F}$ is wide-sense stationary with separable covariance matrix, then the moments of $\underline{F}$ can be found indirectly from the moments of $\underline{W}$, which is an uncorrelated array with energy distribution given by equation (6-40) [6].

6.4 Moments of the Product of Singular Values

Before presenting Derivations 6.5 and 6.6, it is helpful to introduce the concept of generalized variance.

Definition:

A covariance matrix and its determinant are two of the multivariate analogs of univariate variance σ^2. The determinant of the covariance matrix is referred to as the generalized variance providing a scalar measure of variance for a multivariate process.

The unbiased sample covariance is well defined. Its determinant is defined by Wilks [171] as the sample generalized variance

$$|\underline{C}| = \left|\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})(y_i-\bar{y})^T\right| \qquad (6\text{-}41)$$

where $\bar{y}$ is the sample mean. The asymptotic distribution and moments

of $|\underline{C}|$ are given by Anderson in [81, page 171]. In the following derivation, the moments of $|\underline{C}|$ given in [81] are modified to give the moments of the determinant of the Wishart matrix, $\underline{X}\,\underline{X}^T$.

Derivation 6.5: If $\underline{X}$ is a matrix of n independent k-variates from $N(\underline{0},\underline{K}_C)$, then the h^{th} moment of $|\underline{X}\,\underline{X}^T|$ is

$$E\{|\underline{X}\,\underline{X}^T|^h\} = 2^{kh}|\underline{K}_C|^h \frac{\Gamma_k(\frac{n}{2}+h)}{\Gamma_k(\frac{n}{2})} \tag{6-42}$$

Proof: If $\underline{C}$ is formed by N observations from a K-variate normal population $N(\underline{\mu}_C,\underline{K}_C)$. The determinant of $\underline{C}$ will be

$$|\underline{C}| = \frac{1}{(N-1)^k}\left|\sum_{i=1}^{N}(y_i-\bar{y})(y_i-\bar{y})^T\right| \tag{6-43}$$

An orthogonal transformation can be made to eliminate the sample mean. Hence, it can be said that $|\underline{C}|$ is formed by N-1 independent observations from $N(0,\underline{K}_C)$. But

$$\underline{X}\,\underline{X}^T = \sum_{i=1}^{n} x_i x_i^T \tag{6-44}$$

where x_i is from $N(0,\underline{K}_C)$. Hence, in the derivation of [81], for sample generalized variance, we compensate for $1/(N-1)^k$ of (6-43) and replace n by N-1 to obtain the result for determinant of $\underline{X}\,\underline{X}^T$. Then, using Appendix B, the derivation of [81] will be changed from the product of single variate gamma functions to multivariate gamma functions, and (6-42) will be obtained. Q.E.D.

Derivation 6.6: The m^{th} moment of the product of the singular values of a zero mean kxk texture field $\underline{F}$ in a normal sample is

$$E\{(\prod_{i=1}^{k} s_i)^m\} = 2^{\frac{km}{2}} |\underline{K}_C|^{\frac{m}{2}} |\underline{K}_R|^{\frac{m}{2}} \frac{\Gamma_k(\frac{k+m}{2})}{\Gamma_k(\frac{k}{2})} \qquad (6\text{-}45)$$

where $\Gamma_k(.)$ is k-variate gamma function of (.) (Appendix B).

Proof: For a central quadratic form,

$$\underline{F}\,\underline{F}^T = \underline{X}\underline{\Lambda}_R\underline{X}^T \qquad (6\text{-}46)$$

where $\underline{X}$ has been defined in Section 6.1.2, and $\underline{\Lambda}_R$ is the eigenvalue matrix of $\underline{K}_R$, thus

$$|\underline{F}\,\underline{F}^T| = |\underline{X}\,\underline{\Lambda}_R\underline{X}^T| \qquad (6\text{-}47)$$

and for a square $\underline{F}$

$$|\underline{F}\,\underline{F}^T| = |\underline{\Lambda}_R|\,|\underline{X}\,\underline{X}^T|$$

$$= |\underline{K}_R|\,|\underline{X}\,\underline{X}^T| \qquad (6\text{-}48)$$

The l.h.s. of (6-48) is equal to $|\underline{U}\,\underline{\Lambda}\,\underline{U}^T|$, which itself is equal to $|\underline{\Lambda}| = (\prod_{i=1}^{k} \lambda_i) = (\prod_{i=1}^{k} s_i^2)$, where k is the dimension of $\underline{F}$. Hence,

$$(\prod_{i=1}^{k} s_i) = |K_R|^{1/2}\,|XX^T|^{1/2} \qquad (6\text{-}49)$$

Therefore

$$E\{(\prod_{i=1}^{k} s_i)^m\} = |K_R|^{m/2} E\{|XX^T|^{m/2}\} \qquad (6\text{-}50)$$

Because $\underline{F}$ is considered as square, then in (6-42) of Derivation 6.5, we let n=k and h = m/2. Putting the outcome in (6-48), one obtains (6-45) and hence the proof. **Q.E.D.**

Derivation 6.6-1: Second moment of product of singular values.

In (6-45), letting m=2

$$E\{(\prod_{i=1}^{k} s_i^2)\} = |\underline{K}_C|\ |\underline{K}_R|\ K! \qquad (6\text{-}51)$$

From (6-51), we can say that

$$E\{|\underline{F}\ \underline{F}^T|\} = |\underline{K}_C|\ |\underline{K}_R|\ k! \qquad (6\text{-}51a)$$

Derivation 6.6-2: h^{th} moment of the random determinant of $|FF^T|$.

$$E\{|\underline{F}\ \underline{F}^T|^h\} = 2^{kh}|\underline{K}_C|^h|\underline{K}_R|^h \frac{\Gamma_k(\frac{k}{2}+h)}{\Gamma_k(\frac{k}{2})} \qquad (6\text{-}52)$$

Derivation 6.6-3: From Derivation 6.5, all moments of a chi-square r.v. can be obained by letting k=1 and $\underline{K}_C = \sigma^2$. Hence, $y = \chi^2 = x_1^2+x_2^2+\ldots+x_n^2$

$$E\{(y)^h\} = 2^h\sigma^{2h}\frac{\Gamma(\frac{n}{2}+h)}{\Gamma(\frac{n}{2})}$$

$$= 2^h\sigma^{2h}(\tfrac{n}{2})(\tfrac{n}{2}+1),\ldots,(\tfrac{n}{2}+h-1) \qquad (6\text{-}53)$$

In (6-53), let h=1

$$E\{y\} = n\sigma^2 \tag{6-54}$$

and h=2

$$E\{y^2\} = n(n+2)\sigma^4 \tag{6-55}$$

Chi statistics: For h = 1/2

$$E\{x\} = 2^{1/2} \frac{\Gamma(\frac{n}{2} + \frac{1}{2})}{\Gamma(\frac{n}{2})} \sigma \tag{6-56}$$

In (6-56), letting n=2

$$E\{x\} = \frac{\sqrt{2}\pi}{2} \sigma \tag{6-57}$$

For the univariate quadratic form $q^2 = f_1^2 + f_2^2 + \ldots + f_n^2$, letting $\underline{F} = [f_1, f_2, \ldots, f_n]$, the h-th moment of q^2 and q can be derived using derivation 6.6-2 which will be that of (6-53)-(6-57) multiplied by the corresponding $|\underline{K}_R|^h$.

6.5 Stochastic Perturbation

A basic question is: If elements of a matrix $\underline{F}$ vary, in a random fashion, according to certain distribution laws, and just to make the problem more challenging let us say that there are correlation amongst elements of $\underline{F}$, then how and according to what distribution laws do the singular value of $\underline{F}$ vary?

In Chapter 5, from the p.d.f. of $\underline{F}$, we derived the p.d.f. of $\underline{S}$, i.e. the joint p.d.f.'s of the singular values of $\underline{F}$. Analytical extraction of the moments of individual singular values for a general order matrix is cumbersome. The p.d.f. of the largest singular

values was derived in the last chapter, and numerical evaluation of the moments of s_1 followed. Moments of s_1 are important because they play the most important role in image feature extraction. Such property could be attributed to the high degree of correlation in pictorial information, which reflects most of the energy in a picture in its first singular value. The following section gives analytical derivation of the moments of S.V.'s for a special case.

6.5.1 Individual Moments of Singular Values for a Special Case

Derivation 6.7: Moments of the singular values for an uncorrelated 2x2 Gaussian matrix are as follows:

$$E(s_1) = \sqrt{\pi}\,\sigma \tag{6-58}$$

$$E(s_2) = (1-\frac{\sqrt{2}}{2})\sqrt{\pi}\,\sigma \tag{6-59}$$

$$E(s_1^2) = (2+\frac{\pi}{2})\,\sigma^2 \tag{6-60}$$

$$E(s_2^2) = (2-\frac{\pi}{2})\,\sigma^2 \tag{6-61}$$

$$Var(s_1) = (2-\frac{\pi}{2})\,\sigma^2 \tag{6-62}$$

$$Var(s_2) = (2-2\pi+\sqrt{2}\pi)\,\sigma^2 \tag{6-63}$$

$$E(s_1 s_2) = \sigma^2 \tag{6-64}$$

$$E\{(s_1-E(s_1))(s_2-E(s_2))\} = \sigma^2(1-\pi+\frac{\sqrt{2}}{2}\pi) \tag{6-65}$$

Proof: Even for the apparently simple 2x2 case, analytical derivation of moments involve numerous calculations and integration by parts. Therefore, only the method is presented here.

Consider a 2x2 matrix $\underline{F}$ and let its columns be uncorrelated bivariate normal random vectors from the distribution $N(\underline{0},\sigma^2\underline{I}_2)$. Hence, $\underline{K}_R = \underline{I}_2$ and $\underline{K}_C = \sigma^2\underline{I}_2$. Then from (5-39) the joint p.d.f. of the singular values will be

$$g(s_1,s_2) = \frac{1}{\sigma^4}(s_1^2-s_2^2)\exp(-\frac{s_1^2+s_2^2}{2\sigma^2}) \qquad (6\text{-}65)$$

(See Figure 5-1 and Section 5.2.1 in Chapter 5 for related discussions.) Since $\infty > s_1 > s_2 > 0$, the n^{th} moment of s_i will be

$$E(s_i^n) = \int_0^\infty \int_{s_2}^\infty s_i^n g(s_1,s_2)ds_1 ds_2 \qquad (6\text{-}66)$$

for i = 1,2. The mathematical procedure is long and therefore eliminated. **Q.E.D.**

Based on (6-58)-(6-65), the approximate values of moments are shown in Table 6-1.

Remark: As it can be seen from Table 6-1, the moments of the second singular value are considerably less than those of the first one. From the same table, the covariance matrix of s_1 and s_2 can be obtained as

$$Cov(s_1,s_2) = \sigma^2 \begin{bmatrix} 0.429 & \frac{0.159}{2} \\ \frac{0.159}{2} & 0.159 \end{bmatrix} \qquad (6\text{-}67)$$

Table 6-1. Moments of S.V. for a 2x2 Case.

	MEAN	SECOND MOMENT	VARIANCE	CORRELATION	COVARIANCE
i	$E(s_i)$	$E(s_i^2)$	$Var(s_i)$	$E(s_1 s_2)$	$Cov(s_1, s_2)$
1	1.772σ	$3.57\sigma^2$	$0.429\sigma^2$	σ^2	$0.0795\sigma^2$
2	0.519σ	$0.429\sigma^2$	$0.159\sigma^2$		

6.6 Summary

SVD provides a useful tool for dealing with complicated multivariate problems. This chapter presented some concepts in texture energy distribution, and moments of singular values. Stochastic perturbation theory was discussed and using SVD, moments of the random determinant of a quadratic form and moment of product of singular values were derived. In the preceding section, we derived the covariance matrix of the singular values of a 2x2 random matrix where elements are normally distributed, uncorrelated and zero mean, and possess standard deviation σ.

One result of this derivation is the fact that although element of $\underline{F}$ are uncorrelated and zero mean, its singular values are correlated and non-zero mean. The ideal thing would be to analytically derive the covariance matrix of the singular values for a general case. For the most general case, we have got as far as:

i) Analytical derivation of the joint p.d.f. of the singular values - $g(\underline{S})$.

ii) Analytical extraction of the p.d.f. of the largest singular value - $g(s_1)$.

iii) Analytical derivation of a bound for the p.d.f. of largest singular value.

iv) Analytical and numerical derivation and graphical representation of $g(s_1)$ for a special case.

v) Analytical derivation of the moments of s_1 for a special case.

case.

vi) Empirical approximation of $E(s_1)$, $E(s_1^2)$, and $Var(s_1)$ for a general case.

PART THREE: APPLICATION

Attention is directed largely towards engineering applications of SVD in image feature extractions. Chapters 7, 8, 9 and 10 are included in this part.

CHAPTER 7

TEXTURE MEASUREMENT

Haralick et al. [51-52] have developed a family of texture measures based on two-dimensional histograms of textural images. Such measures provide up to 85% classification accuracy, but require a large amount of computation. In this chapter, a family of texture measures based on SVD are introduced which possess better performance. To be more specific, SVD texture measures require much less computation, but have higher classification accuracy than Haralick's measures. In particular, for low contrast textures for which Haralick's measures have accuracy limitation [53], SVD measures perform well. What follows is an introduction to this set of measures. Its evaluation will be performed by Bhattacharyya distance criteria in the next two chapters.

7.1 Deterministic Properties

Deterministic properties of SVD can be demonstrated by the following examples. Consider the rotation of a pattern. A NxN matrix with ones along one of its columns and zeros elsewhere. This matrix has one singular value. Now, consider another NxN matrix with ones along its main diagonal and zeros elsewhere. This second matrix has N singular values.

$$\begin{bmatrix} & 1 & \\ \underline{0} & \vdots & \underline{0} \\ & 1 & \end{bmatrix}_N \qquad \begin{bmatrix} 1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & 1 \end{bmatrix}_N$$

One Singular Value — N Singular Values

The rotation of the first matrix has affected the number of singular values. Consider another example

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & & 1 & 0 & \dots & 1 & 0 \end{bmatrix} \qquad (7\text{-}1)$$

The r.h.s. of (7-1) has one singular value because it is formed by one outer product of two vectors [5,6]. Consider the checkerboard matrix

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{bmatrix} \qquad (7\text{-}2)$$

The r.h.s. of (7-2) has two singular values because it is formed by the summation of two independent outer product vectors. (7-1) has alternating vertical stripes of zeros and ones, and (7-2) has alternating diagonals of zeros and ones. Thus, the two matrices are viewed as rotation of each other. However, their number of singular values differ.

For periodicity consideration of SVD, consider, again the matrix of (7-1) with alternating column of ones and zeros and compare it with (7-3)

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & & 1 \end{bmatrix} \qquad (7\text{-}3)$$

r.h.s. of (7-3) is a constant amplitude array with 100% duty cycle period and has the same number of non-zero singular values as the striped matrix of (7-1). Clearly, SVD is not a measure of periodic structure [5,6].

7.2 Stochastic Behaviour of Singular Values

It has been demonstrated that in the deterministic domain, the singular values of many pictures decrease rapidly with increasing index [130-132], and SVD is a deterministically optimal transform for energy compaction [2]. It turns out that singular values decrease quite rapidly for stochastic images as well. This section

illustrates such a fact. A set of experiments have been performed on artificially generated stochastic texture fields.

To generate a random matrix of order 16 whose elements are normally distributed with column covariance matrix of $\underline{K}_C$ and row covariance matrix of K_R with separable property, we consider the relation (4-34) developed in Chapter 4, $\underline{F} = (\underline{E}_C \underline{\Lambda}_C^{1/2}) \underline{M} (\underline{E}_R \underline{\Lambda}_R^{1/2})^T$ where $\underline{M}$ is a $\underline{16x16}$ white random matrix. Also, $\underline{K}_C = \underline{E}_C \underline{\Lambda}_C \underline{E}_C^T$ and $\underline{K}_R = \underline{E}_R \underline{\Lambda}_R \underline{E}_R^T$ as discussed in Chapter 4. Letting

$$\underline{K}_C = \underline{K}_R = \begin{bmatrix} 1 & \rho & \cdots\cdots & \rho^{15} \\ \rho & 1 & \cdots\cdots & \rho^{14} \\ \vdots & \vdots & & \vdots \\ \rho^{15} & \rho^{14} & \cdots & 1 \end{bmatrix} \qquad (7\text{-}4)$$

we will obtain a matrix $\underline{F}$ whose singular value distribution is a function of ρ. For, ρ = 0.0, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, we generate seven $\underline{16x16}$ matrices and seven singular value curves. Tabulation of the singular values is presented in Table 7-1. Figures 7-1 through 7-7, relate the singular value amplitudes to correlation factor. Figure 7-1 is a direct plot of this relationship. Figure 7-2 has the normalized version of the curves of Figure 7-1 in which all SV's in each curve are divided by the sum of singular values. Figure 7-3 shows the scaled version of the curves of 7-1 in which all SV's in each curve are divided by the largest singular value of that curve to give the first element a value one and the rest less than one. Figure 7-3 is particularly interesting because it displays the systematic increase in steepness as correlation increases. Figure

Table 7-1. Tabulation of the Singular Value Vectors w.r.t. ρ.

ρ	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}
0.0	7.827	6.996	5.780	4.969	4.810	4.052	3.857	3.105	2.544	2.212	2.03	1.809	1.107	0.886	0.375	0.028
0.5	11.027	7.052	5.896	4.352	4.040	3.173	2.421	2.243	1.814	1.481	1.158	1.076	0.728	0.585	0.261	0.020
0.6	12.366	6.787	5.640	4.026	3.735	2.727	2.056	1.821	1.464	1.214	0.929	0.841	0.596	0.462	0.212	0.017
0.7	14.199	6.460	4.979	3.569	3.181	2.199	1.628	1.363	1.099	0.919	0.696	0.613	0.448	0.338	0.160	0.012
0.8	16.917	5.779	3.855	2.858	2.352	1.563	1.121	0.899	0.726	0.609	0.460	0.395	0.295	0.218	0.105	0.008
0.9	21.348	4.119	2.221	1.714	1.265	0.818	0.569	0.441	0.355	0.299	0.226	0.190	0.143	0.105	0.051	0.004
1.0	29.247	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

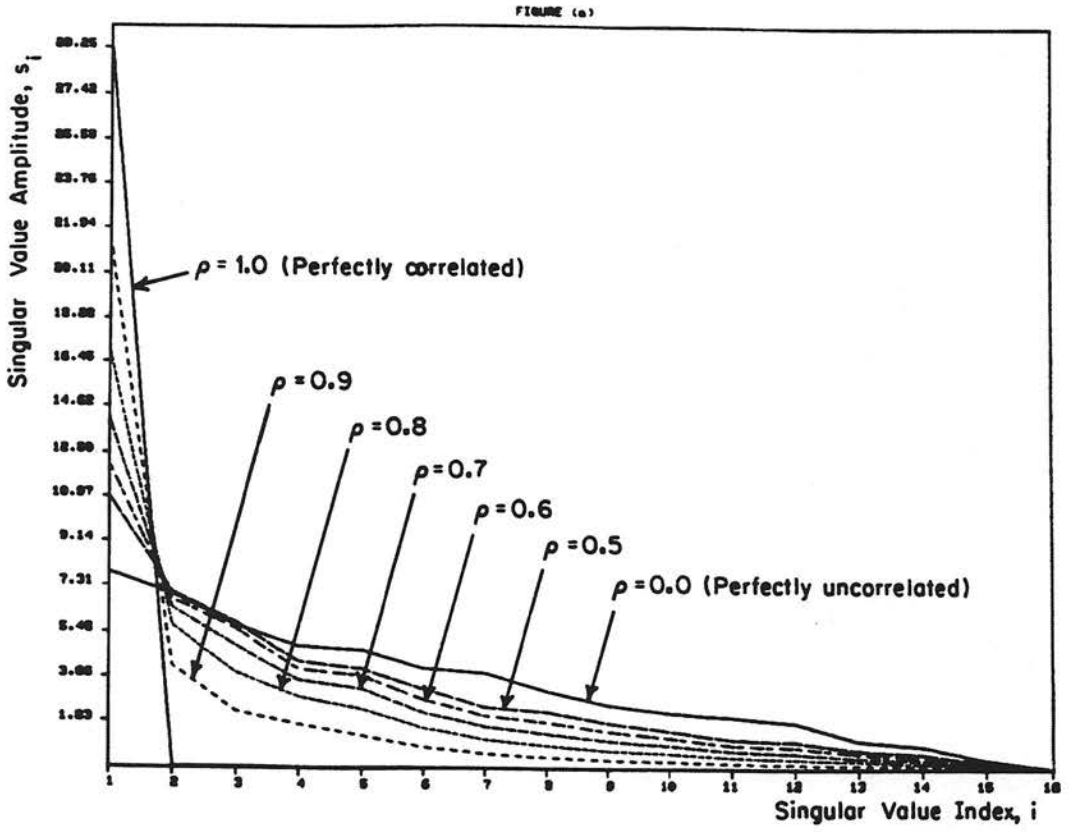


Figure 7-1. Singular Values Curves vs Index w.r.t. ρ.

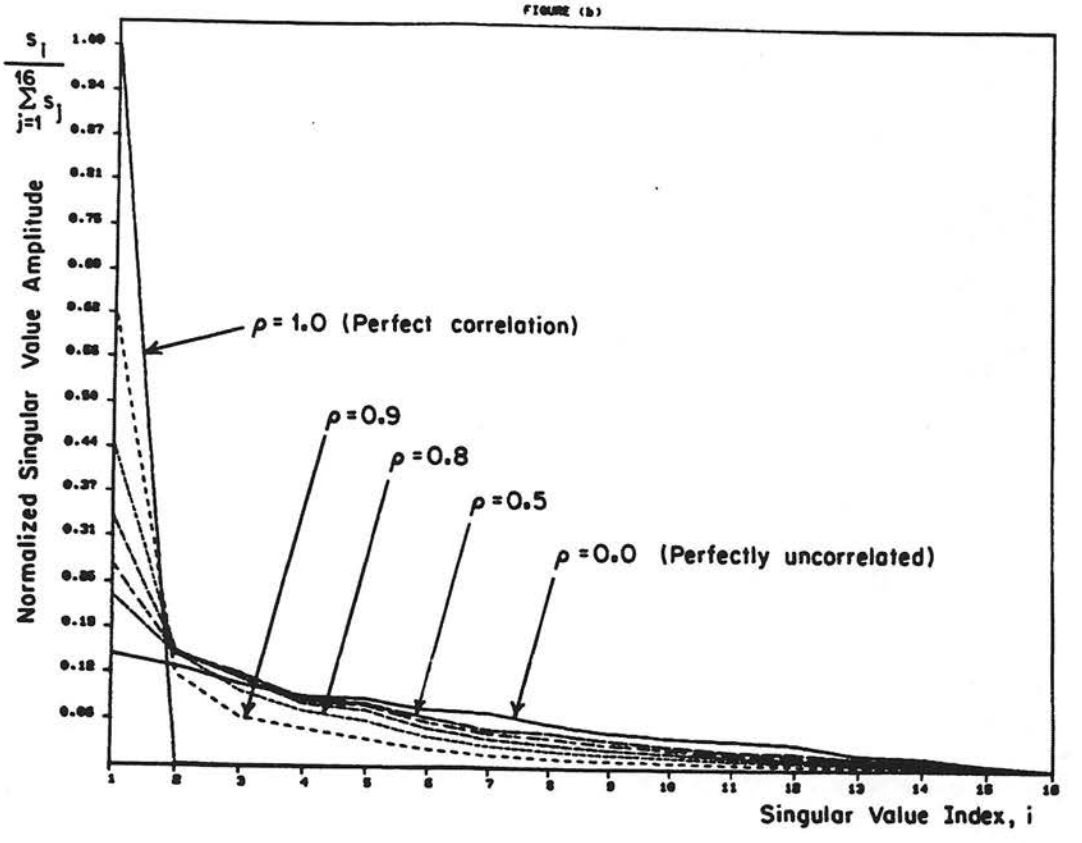


Figure 7-2. Normalized SV Curves vs Index w.r.t. ρ.

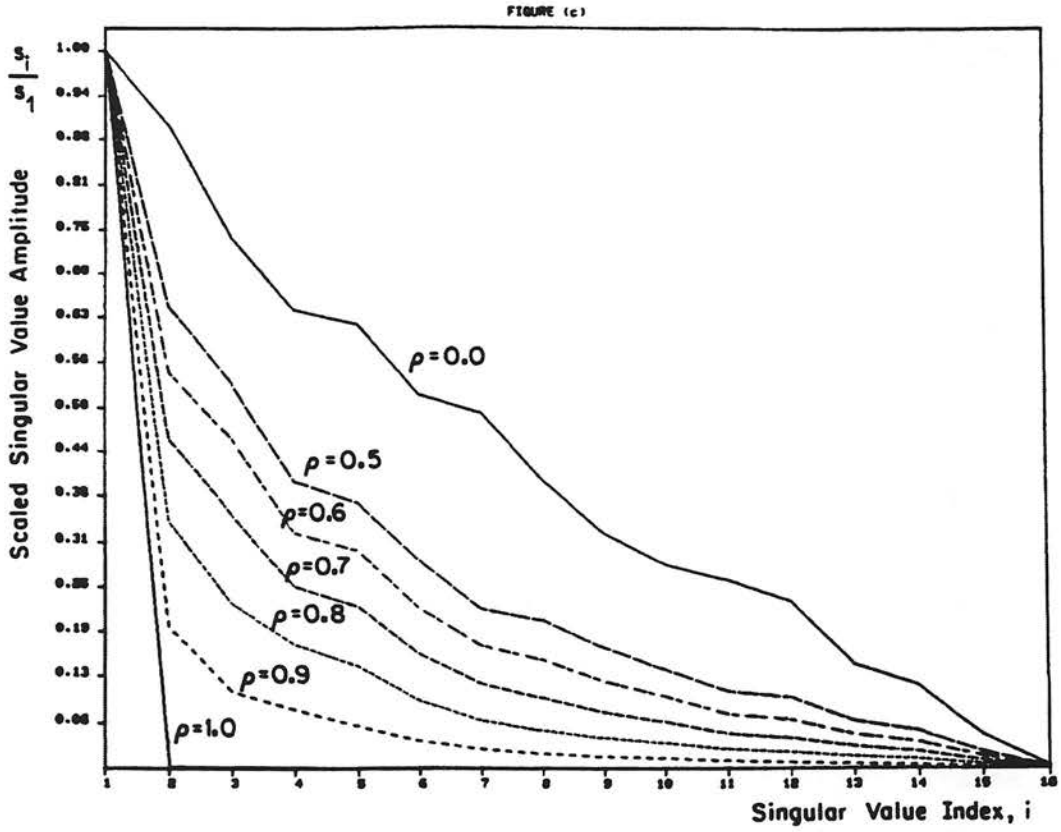


Figure 7-3. Scaled SV Curves vs Index w.r.t. ρ.

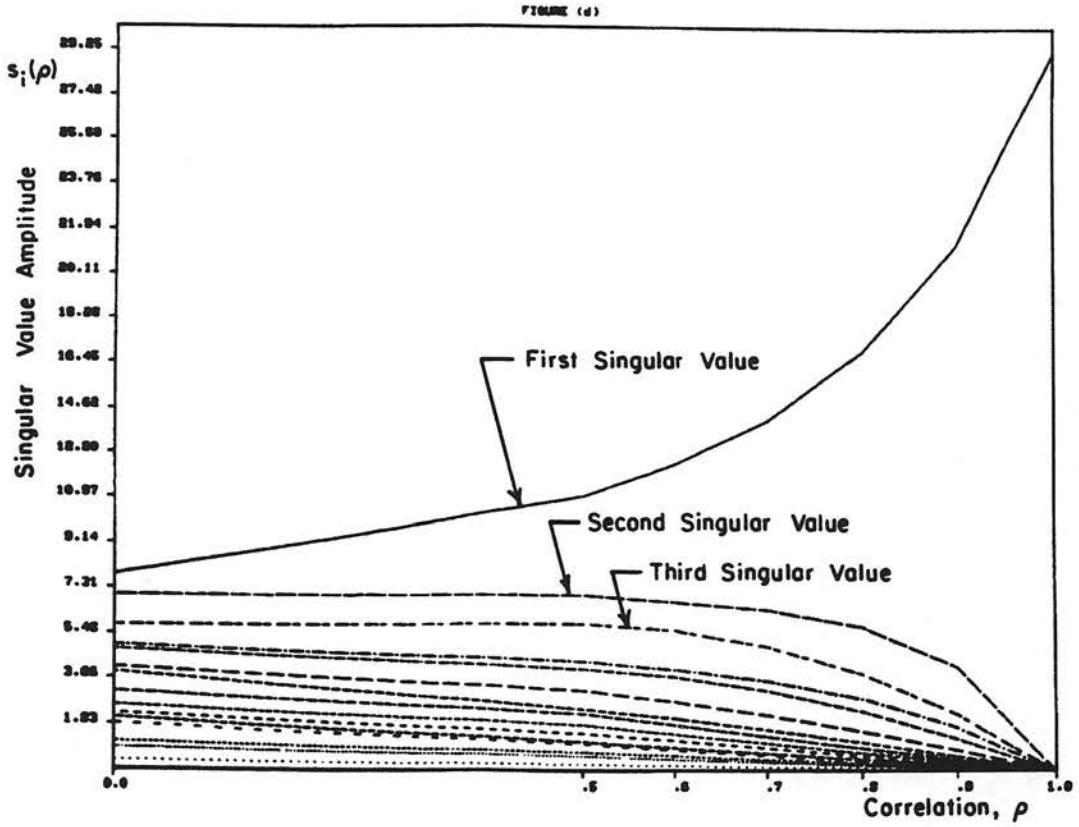


Figure 7-4. SV Curves vs ρ w.r.t. Index.

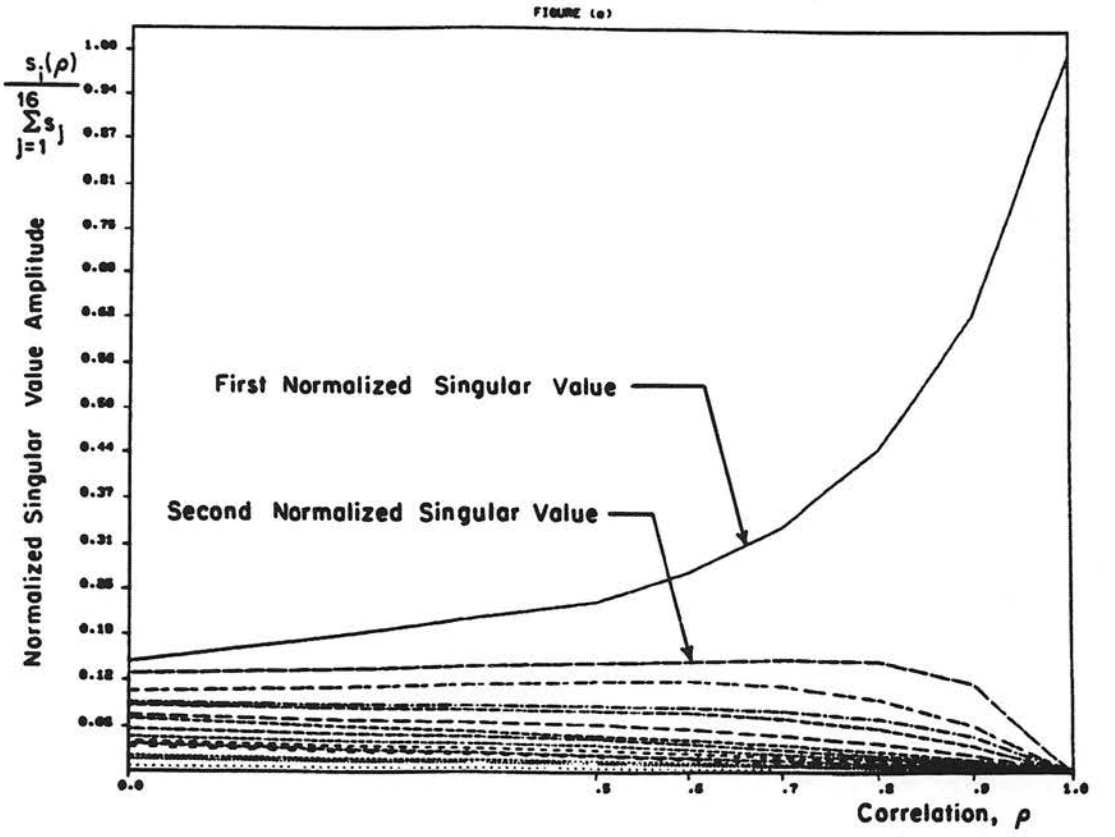


Figure 7-5. Normalized Curves vs ρ w.r.t. Index.

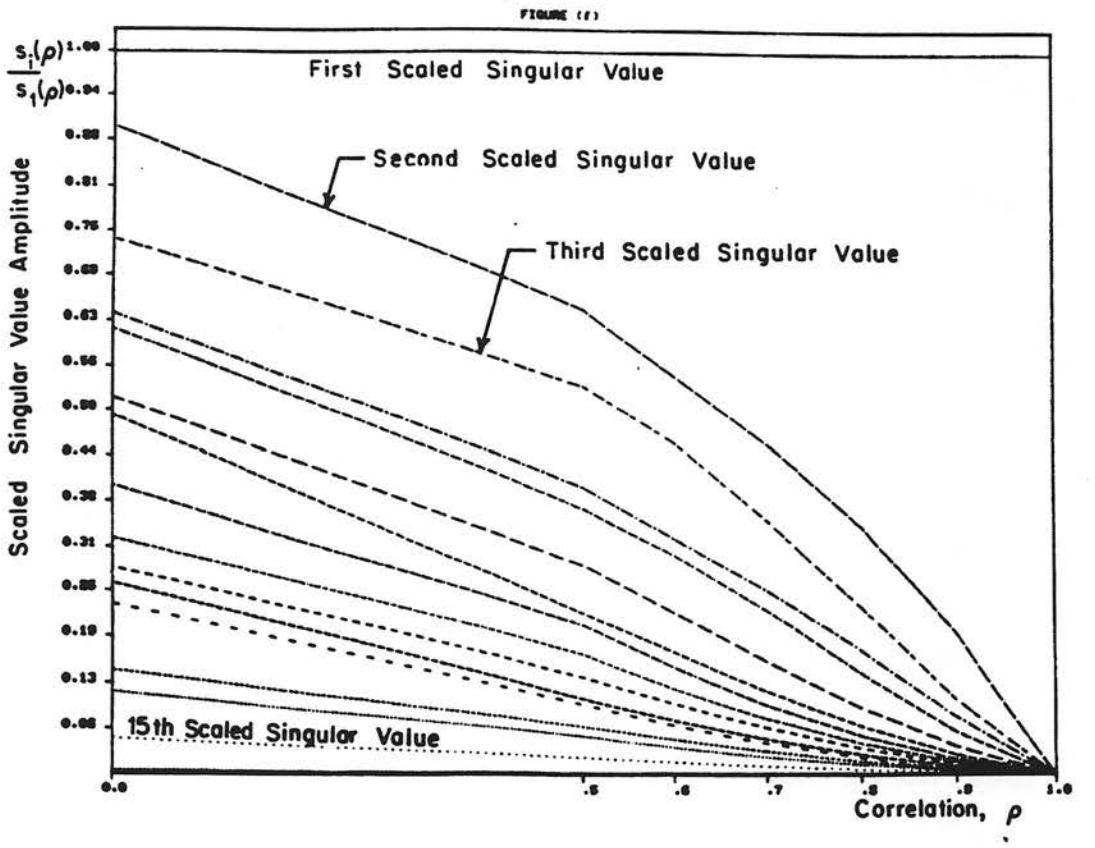


Figure 7-6. Scaled Curves vs ρ w.r.t. Index.

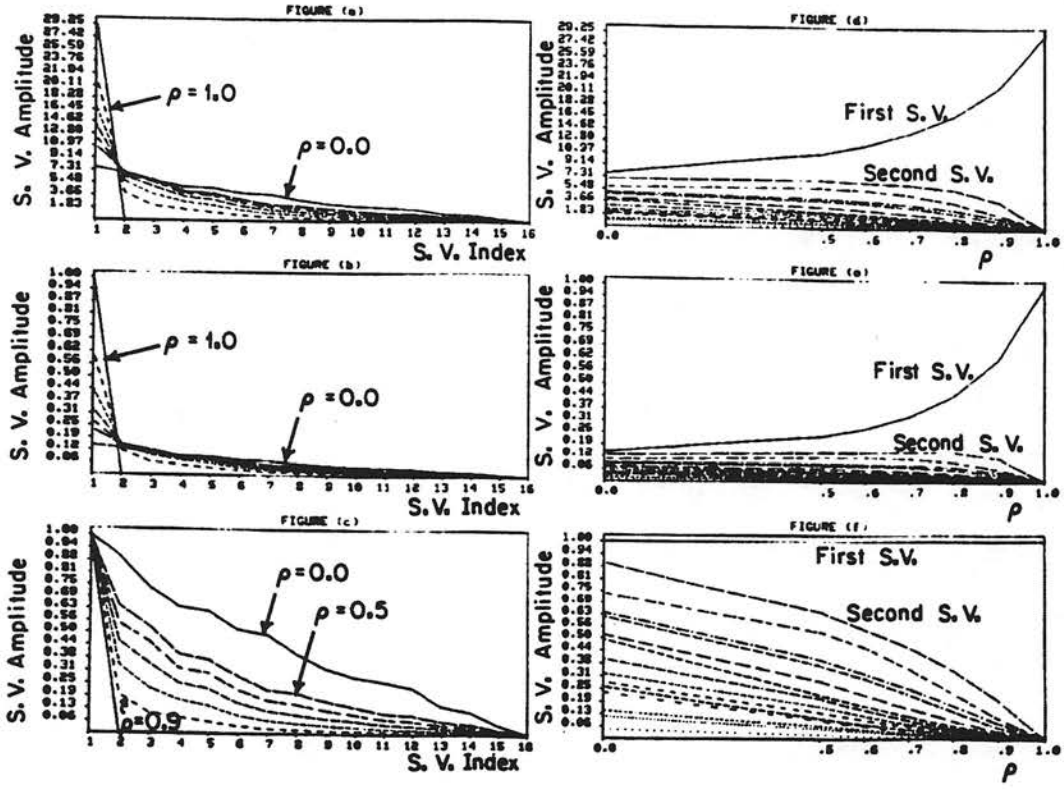


Figure 7-7. Graphical Comparison of SV Curves.

7-4 depicts the monotonic increase in the first (largest) singular value as correlation increases. The same graph shows a fall off in 2nd to 16th singular value. Taking the $\log_e$ of the curves in Figure 7-4 shows a linear increase for the largest SV. Hence, the largest SV increases exponentially. The normalized counterpart of Figure 7-4 is Figure 7-5. Figure 7-6 shows the scaled version of Figure 7-5 where it clearly shows the systematic decrease in the ratio of SV's w.r.t. the largest one as correlation increases. Figure 7-7 provides a graphical comparison of Figures 7-1 through 7-4 in which a and its counterpart d, b and e; and c and f are compared. Figures 7-1 through 7-6 show that the singular value curves can be used as a separating means in texture classification (see also Figures 3-1 through 3-3).

7.3 Significance of the Largest SV

In matrix theory, a norm can be specified for a general matrix $\underline{F}$ such that it is equal to the Euclidean length of the largest vector $\underline{y}$, where $\underline{y}=\underline{F}\ \underline{x}$ and $\underline{x}$ is of unit length [129]. We designate such a norm as

$$\|\underline{F}\|_2 = \|\underline{U}\ \underline{S}\ \underline{V}^T\|_2 = \|S\|_2 = s_1 \qquad (7\text{-}5)$$

$\underline{U}$ and $\underline{V}$ are unitary matrices and hence preserve the norm, and s_1 is the largest singular value of $\underline{F}$. The unit vector $\underline{x}$ which provides such $\underline{y}$ is the eigenvector v_1 corresponding to the largest singular value s_1. Therefore $\|F\|_2$ can be expressed as

$$\|F\|_2 = \sup_{\underline{x}\neq\underline{0}} \frac{\|\underline{F}\ \underline{x}\|}{\|\underline{x}\|} \qquad (7\text{-}6)$$

Since $\underline{F}$ is a continuous operator, sup becomes max, and hence

$$\|F\|_2 = \max_{\underline{x} \neq \underline{0}} \frac{\|\underline{F}\ \underline{x}\|}{\|\underline{x}\|} = s_1 \qquad (7\text{-}7)$$

where s_1 is the sup norm corresponding to ℓ_2-norm on the vector space of a general matrix $\underline{F}$. Another norm can be defined as the Euclidean norm of $\underline{F}$ as given by

$$\|\underline{F}\|_E = (\mathrm{Tr}\ \underline{F}\ \underline{F}^T)^{1/2} = (\sum_{i=1}^{k} \lambda_i)^{1/2} = (\sum_{i=1}^{k} s_i^2)^{1/2} \qquad (7\text{-}8)$$

Andrews et al. [130-132] have used the square of $\|\underline{F}\|_E$ in their experiments. For textural images, the $\|F\|_2$ norm performs well, and its simplicity encourages such use. It is obvious that the higher the correlation in a texture field, the lower its rank, and hence, the steeper its singular value curves; resulting in more significance attributed to s_1. Thus, in low contrast texture fields, the largest singular values is of primary importance for automatic discrimination.

A different significance can be attributed to s_1. The number of zero crossing--change of sign in components--of eigenvector v_1 corresponding to s_1 is zero which means that all components of v_1 have positive or negative values. The number of zero crossings in eigenvectors v_i linearly increases with the index [133, p.165]. Another significant behavior of s_1, is in its normalized form: Consider a perfectly correlated picture, e.g. the picture of a clear sky. A kxk sample window extracted from such a field has one

singular value. Hence, the normalized version will be (1,0,...,0). Consider the opposite, i.e. a picture with perfectly uncorrelated pixel values. In its ideal form, all singular values for such a picture will be equal, and hence, their normalized version will be $(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ Thus, the largest SV in any picture always varies as

$$\frac{1}{k} < \frac{s_1}{\sum_{i=1}^{k} s_i} \leq 1 \qquad (7\text{-}9)$$

and the rest vary as

$$\frac{1}{k} \geq \frac{s_j}{\sum_{i=1}^{k} s_i} \geq 0 \qquad \text{for } j = 2,3,\ldots,k \quad (7\text{-}10)$$

7.4 A Family of Texture Features

In order to compare singular values of several texture fields, the fields are standardized. Standardization simply means equalizing the experimental mean and variance. In our experiments, the texture fields are made zero mean and unit variance. Based on their stochastic properties, the singular values $(s_1,s_2,\ldots,s_k)$ provide measures or features of such texture fields. s_1 alone provides the majority of information about the texture. A better set of features is the normalized version of the singular values such as

$$\left(\frac{s_1}{s_1+s_2+\ldots+s_k}, \frac{s_2}{s_1+s_2+\ldots+s_k}, \ldots, \frac{s_k}{s_1+s_2+\ldots+s_k}\right)$$

The following family of measures can be considered as means for texture feature extraction.

7.4.1 Vector Features

Several vector features, which can be extracted from a texture window, are listed below:

I. $$\underline{z}_1 = (s_1, s_2, \ldots, s_k)^T \tag{7-11}$$

II. Normalized Singular Value Histogram

$$\underline{z}_2 = \left(\frac{s_1}{\sum_i s_i}, \frac{s_2}{\sum_i s_i}, \ldots, \frac{s_k}{\sum_i s_i} \right)^T \tag{7-12}$$

III. Normalized Length

$$\underline{z}_3 = \left(\frac{s_1}{(\sum_i s_i^2)^{1/2}}, \frac{s_2}{(\sum_i s_i^2)^{1/2}}, \ldots, \frac{s_k}{(\sum_i s_i^2)^{1/2}} \right)^T \tag{7-13}$$

IV. Energy in Individual Singular Values

$$\underline{z}_4 = (s_1^2, s_2^2, \ldots, s_k^2)^T \tag{7-14}$$

V. Normalized Energy

$$\underline{z}_5 = \left(\frac{s_1^2}{\sum_{i=1}^{k} s_i^2}, \frac{s_2^2}{\sum_{i=1}^{k} s_i^2}, \ldots, \frac{s_k^2}{\sum_{i=1}^{k} s_i^2} \right)^T \tag{7-15}$$

Remark: $\|F\|_E = (\sum_i s_i^2)^{1/2}$ defined in Section 7.3 is the Euclidean norm of the lexicographic vector of $\underline{F}$.

Non-linear Transformation for Dimensionality Reduction

Non-linear transformations can be made on the vector features $\underline{z}_1$ through $\underline{z}_5$ to produce effective scalar features. The non-linear mappings are as follows: A first order histogram h(i) of k levels can be represented rather accurately by its first four moments [21]. The four moments, mean, deviation, skewness, and kurtosis can represent a first-order histogram. z_1 through z_5 can be considered as first-order histograms. Thus, letting h(i) be the $\underline{i^{th}}$ component of any of the feature vectors $\underline{z}_1$ through $\underline{z}_5$, we can derive the following four moments of h(i): let $M_0 = \sum_{i=1}^{k} h(i)$, then

Mean - $$M_1 = \sum_{i=1}^{k} ih(i) \qquad (7\text{-}16)$$

Deviation - $$M_2 = [\sum_{i=1}^{k} (i-\frac{M_1}{M_0})^2 h(i)]^{1/2} \qquad (7\text{-}17)$$

Skewness - $$M_3 = \frac{1}{M_2^3} \sum_{i=1}^{k} (i-\frac{M_1}{M_0})^3 h(i) \qquad (7\text{-}18)$$

Kurtosis - $$M_4 = \frac{1}{M_2^4} \sum_{i=1}^{k} (i-\frac{M_1}{M_0})^4 h(i) - \frac{3}{M_0} \qquad (7\text{-}19)$$

The factor $3/M_0$ in the kurtosis makes the kurtosis of a Gaussian histogram zero in the limit. For normalized vector features $\underline{z}_2$ and $\underline{z}_5$ where $\sum_{i=1}^{k} h(i) = 1$, h(i) is analogous to a discrete probability density function.

Skewness, a measure of symmetry of a histogram, and kurtosis, are usually high in singular value feature vectors $\underline{z}_1$ and $\underline{z}_4$. Moments M_1 through M_4 provide a good prototype vector to represent either of $\underline{z}_1$ through $\underline{z}_5$. The moment of (7-16) to (7-19) will be equivalent to those explained in [21,p.472] provided that they are premultiplied by the matrix $\text{diag}(M_0^{-1}, M_0^{-1/2}, M_0^{1/2}, M_0)$.

VI. $$\underline{z}_6 = (M_1, M_2, M_3, M_4)^T \qquad (7\text{-}20)$$

In evaluation and classification, instead of using, for example the 32-dimensional feature vector $\underline{z}_1$ through $\underline{z}_5$, we use the 4-dimensional $\underline{z}_6$ with almost the same degree of accuracy, but less computational requirements. In the experiments of chapters 8 and 9, the unnormalized version of the four moments are used such that M_0^{-1}, $M_0^{-1/2}$, $M_0^{1/2}$, M_0 are not multiplied to M_1, M_2, M_3, M_4. This is because it is intended to preserve the information related to M_0 in $\underline{z}_6$. However, The results are obviously equivalent in SV vectors $\underline{z}_2$ and $\underline{z}_5$.

Scalar Features

VII. Entropy

From normalized vectors $\underline{z}_2$ and $\underline{z}_5$, one can extract a useful scalar feature, namely the entropy [21]. Letting h(i) equal the ith component of the vector $\underline{z}_2$ or $\underline{z}_5$

$$z_7 = -\sum_{i=1}^{k} h(i) \log_2[h(i)] \qquad (7\text{-}21)$$

where using L'hospital rule, one can prove that

$$\lim_{p \to 0} p \log p = 0 \qquad (7\text{-}22)$$

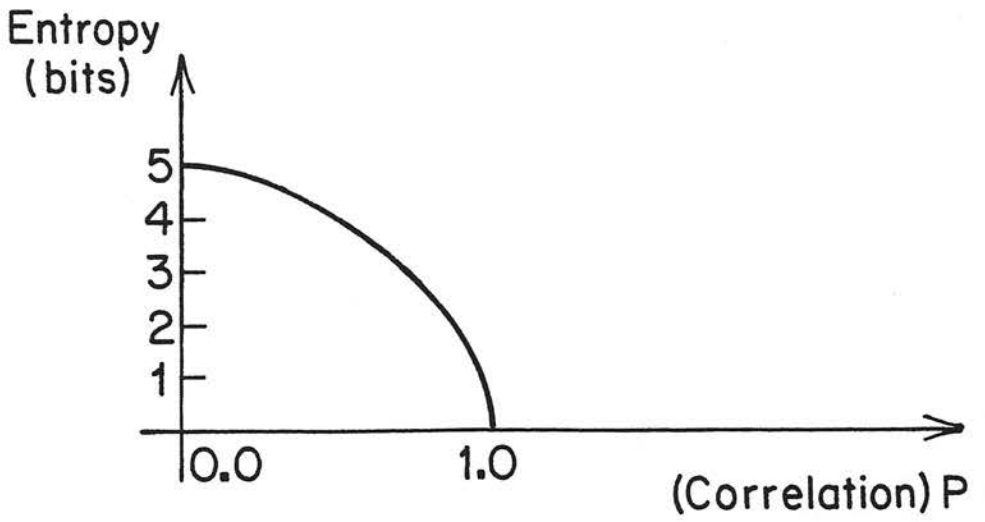


Figure 7-8. Singular Value Entropy for a 32x32 Texture Field.

Bound on the Entropy

Applying (7-9) and (7-10), it is easily proved that

$$\log_2^k > z_7 \geq 0 \qquad (7\text{-}23)$$

where k is the dimension of the texture field. For a 32x32 window, the entropy and its average are bounded by zero and five. From (7-23), the lower the correlation in texture, the higher the entropy. Hence, z_7 provides an additional measure on the correlation of a texture field. Figure 7-8 shows the entropy vs correlation factor ρ, where ρ is the parameter in the Toeplitz matrices $\underline{K_C} = \underline{K_R}$ given by (7-4). As shown in the figure, the singular value entropy is maximum when the correlation in texture is minimum, in fact, zero.

VIII. Energy of $\underline{z_1}$ and $\underline{z_2}$

Letting h(i) be the i^{th} component of any of the feature vectors $\underline{z_1}$ through $\underline{z_3}$,

$$z_8 = \sum_{i=1}^{k} [h(i)]^2 \tag{7-24}$$

$$\text{for } \underline{z}_1, \quad z_8 = \sum_{i=1}^{k} s_i^2 = \text{Tr } \underline{F}\,\underline{F}^T = \|F\|_E^2 \tag{7-25}$$

$$\text{for } \underline{z}_2, \quad z_8 = \frac{\sum_{i=1}^{k} s_i^2}{\left(\sum_{i=1}^{k} s_i\right)^2} < 1 \tag{7-26}$$

$$\text{for } \underline{z}_3, \quad z_8 = 1 \text{ for all textures.} \tag{7-27}$$

Therefore, z_8 derived from $\underline{z}_3$ does not provide a measure, but z_8 associated with $\underline{z}_1$ and $\underline{z}_2$ provides a measure.

IX. Product of Singular Values

$$z_9 = \prod_{i=1}^{k} s_i \tag{7-28}$$

For textures with high correlation or low contrast, z_9 is not a good choice because the small singular values are very close to zero, and their multiplication will result in a very small number and could lead to erroneous results.

X. The Largest Singular Value as a Scalar Feature

$$z_{10} = s_1 \tag{7-29}$$

largest component of normalized histogram:

$$z_{11} = \frac{s_1}{\sum_{i=1}^{k} s_i} \tag{7-30}$$

largest component of normalized length:

$$z_{12} = \frac{s_1}{\left(\sum_{i=1}^{k} s_i^2\right)^{1/2}} \tag{7-31}$$

Energy in the largest component of normalized SV histogram:

$$z_{12} = \frac{s_1}{\left(\sum_{i=1}^{k} s_i^2\right)^{1/2}} \tag{7-32}$$

largest energy ratio:

$$z_{14} = \frac{s_1^2}{\left(\sum_{i=1}^{k} s_i^2\right)} \tag{7-33}$$

Entropy Content of the Largest S.V. for z_{11} and z_{14}

Let $b = z_{11}$ or z_{14}, then

$$z_{15} = -b \log_2 b \tag{7-34}$$

XI. Measure on the Conditioning of a Texture Field

$$z_{16} = \frac{s_1}{s_k} \tag{7-35}$$

where $1 \leq z_{16} < \infty$.

For a texture, whose correlation content is high, the smallest

For a texture, whose correlation content is high, the smallest singular value is usually zero. Thus, z_{16} measure may not be a good choice for this type of textures.

Any combination of scalar features z_7 through z_{15} can be used to form a feature vector $\underline{z}_v$ for a given texture field. $\underline{z}_v$ can be combined with $\underline{z}_6$ to produce a feature vector to represent the texture. In a highly correlated texture, any of the features z_{10} through z_{14} related to the largest singular value can be utilized as an effective feature by itself.

Statistical Considerations

All derivations and experiments in this document are concerned with statistical cases. The mean of SV's, $E\left(\frac{s_1}{\sum_i s_i}, \ldots, \frac{s_k}{\sum_i s_i}\right)$ can be approximated as

$$E\left(\frac{s_j}{\sum_i s_i}\right) \approx \frac{E(s_j)}{E\left(\sum_i s_i\right)} \tag{7-36}$$

or, for z_5,

$$E\,\frac{s_j^2}{\sum_i s_i^2} \cong \frac{E(s_j^2)}{E\left(\sum_i s_i^2\right)} = \frac{Es_j^2}{Tr(\underline{K}_C)Tr(\underline{K}_R)} \tag{7-37}$$

Analytical extraction of moments of singular values is very difficult. Evaluation is possible. The m^{th} moment of $z_9 = \prod_{i=1}^{k} s_i$ has been given by equations (6-45), and its special case $E\{z_9^2\}$ has been given by (6-51) as $|\underline{K}_C|\,|\underline{K}_R|k!$. Hence, different values for correlation along the rows and columns will produce different moments for $\underline{z}_9$, which is an indication that the product of

singular values, can be used as a measure of texture correlation. An observation is that the second moment of z_9 assumes its maximum value, i.e. k! when $\rho_C=\rho_R=0.0$, which means the perfectly uncorrelated case. z_9 is at its minimum at zero when correlation ρ_C or ρ_R reaches its maximum--i.e. one--and the changes are monotonic. Dividing $E\{z_9^2\}$ by (k!) gives a normalized measure for correlation in terms of z_9. In a slightly more complicated case, one can derive the same type of normalized measures for the first moment $E\{z_9\} = E\{\prod_{i=1}^{k} s_i\}$. To be more specific, from relation (6-45), $E\{\prod_{i=1}^{k} s_i\} = 2^{\frac{k}{2}}|K_C|^{1/2}|K_R|^{1/2} \frac{\Gamma_k(\frac{k+1}{2})}{\Gamma_k(\frac{k}{2})}$. The dynamic range of $E\{z_9\}$ in this case is between zero and $2^{k/2} \frac{\Gamma_k(\frac{k+1}{2})}{\Gamma_k(\frac{k}{2})}$ for ρ_C and ρ_R varying from one to zero. Thus, dividing the moment $E\{z_9\}$ by $2^{\frac{k}{2}} \frac{\Gamma_k(\frac{k+1}{2})}{\Gamma_k(\frac{k}{2})}$, which is a well defined value, will provide a normalized measure of correlation.

7.5 Evaluation of the Features

We have determined six vector features and ten scalar ones based on singular values. We certainly would like to know which one of these features are most effective. There are three tasks which must be taken into account: one, determining which feature is the best; the second is testing the best feature on a variety of texture fields and comparing the results with each other; and third is to compare our results with those of Haralick et al. [51-52]. For the latter task the criteria must be related to the probability of error in

classification. For all of the three above mentioned tasks, we need a criteria of evalution suitable for texture processing. Of course, if all features in a combined form be used, the best result will be obtained. However, much of the information in features is redundant, and hence, it would not be economical to use them all. It should also be mentioned that this technique can best be compared with Haralick's if the same texture fields are used for comparison. In Chapters 8 and 9, of this document, we will perform the experiments on a set of artificial and natural textures and leave the close comparison based on the same texture fields for the future.

7.5.1 Methods

Four criteria have been considered for evaluating the texture features: (a) separability function based on scatter measurements [40]; (b) the Wald Sequential Test [135]; (c) divergence [136]; (d) Chernoff Bound to the probability of error and B-distance [137]. B-distance has been chosen as the best method suited for evaluation of SV features.

Chernoff Bound and B-distance

The probability of error in classification is upper bounded by the Chernoff bound. Consider the two-class case

$$\epsilon < P_1^{1-s} P_2^{s} \exp\{-b(s)\} \qquad (7\text{-}38)$$

where ϵ is the Bayes probability of error, P_1 and P_2 are apriori probabilities of the two classes, and $0 \leq s \leq 1$. For a certain s, the least upper bound for ϵ is obtained. In (7-38),

$$b(s) = -\ln \int_{\text{space of } \underline{x}} P(\underline{x}|\text{class 1})^{1-s} \, P(\underline{x}|\text{class 2})^{s} \, dx \qquad (7\text{-}39)$$

letting $s = 1/2$, gives the Bhattacharyya distance as $B(\text{class}_1, \text{class}_2) = b(1/2)$. Using a Bayes classifier, B-distance is monotonically related to the Chernoff bound where Chernoff bound is the least upper bound to the probability of classification error [137].

For two classes with samples from normal populations $N(\underline{M}_1, \underline{C}_1)$ and $N(\underline{M}_2, \underline{C}_2)$ B-distance is

$$B(\text{class}_1, \text{class}_2) = \frac{1}{8} (\underline{M}_1 - \underline{M}_2)^T \left[\frac{(\underline{C}_1 + \underline{C}_2)}{2}\right]^{-1} (\underline{M}_1 - \underline{M}_2) + \frac{1}{2} \ln \frac{\left|\frac{1}{2}(\underline{C}_1 + \underline{C}_2)\right|}{|\underline{C}_1|^{1/2} |\underline{C}_2|^{1/2}} \qquad (7\text{-}40)$$

where, $\underline{M}_1$ and $\underline{M}_2$ are the mean vectors and $\underline{C}_1$ and $\underline{C}_2$ are the covariance matrices of classes one and two.

The metric properties of B-distance have been proved in the statistical literature [40]. B-distance provides a measure for determining the best feature. It also gives a separability measure of various texture fields. It is related to the probability of error; thus, it provides a comparison between the SVD technique and Haralick's or other feature extraction techniques in terms of Bayes probability of classification accuracy. Its disadvantage is in its two-class restriction. However, the multi-class problem can always be divided into class pairs for B-distance computations. Another disadvantage is that the analytical derivation of B-distance for a

general density is difficult. It is, nevertheless, straightforward for normal densities. Our use of formula (7-40) for normal densities is justified based on the fact that the histograms of M_1, M_2, M_3 and M_4 exhibit normal behavior. Thus, (7-40) can be used as a genuine criteria for experiments of Chapters 8 and 9.

For equally likely a priori probability, the probability of error is related to B-distance as

$$\epsilon < \frac{1}{2} \exp(-B) \tag{7-41}$$

where ϵ is the Bayes probability of error and B is the B-distance. The lower bound to the probability of error is also related to the Chernoff bound. In general, the lower bound will be

$$\epsilon > \frac{1}{2} - \frac{1}{2} \left[1 - \exp(-2B)\right]^{1/2} \tag{7-42}$$

and for small errors, the lower bound can be approximated by

$$\epsilon > \frac{1}{4} \exp(-2B) \tag{7-43}$$

Table 7-2 gives the lower and upper bounds to the probability of error in terms of the B-distance for equally likely a priori probabilities.

7.6 Summary

In this chapter, a family of texture features has been introduced. The method of feature evaluation will be the

Table 7-2. Lower and Upper Bounds to Probability of Error in Classification.

$B(CLASS_1, CLASS_2)$	ERROR LOWER BOUND	ERROR UPPER BOUND
0.0	0.50	0.50
0.1	0.28	0.45
0.2	0.20	0.40
0.3	0.16	0.37
0.4	0.12	0.33
0.5	0.10	0.30
0.6	0.08	0.27
0.7	0.06	0.24
0.8	0.05	0.22
0.9	0.04	0.20
1.0	3.38×10^{-2}	1.84×10^{-1}
2.0	4.58×10^{-3}	6.77×10^{-2}
3.0	6.15×10^{-4}	2.48×10^{-2}
4.0	8.39×10^{-5}	9.16×10^{-3}
5.0	1.13×10^{-5}	3.37×10^{-3}
6.0	1.53×10^{-6}	1.24×10^{-3}
7.0	≈ 0	4.46×10^{-4}
8.0	≈ 0	1.68×10^{-4}
9.0	≈ 0	6.17×10^{-5}
10.0	≈ 0	2.27×10^{-5}
11.0	≈ 0	8.35×10^{-6}
12.0	≈ 0	3.07×10^{-6}
20.0	≈ 0	≈ 0
40.0	≈ 0	≈ 0

Bhattacharyya distance figure of merit, which provides the bound to probability of error for a Bayes classifier. B-distance gives a quantitative view of the separability of texture fields.

CHAPTER 8

CLASSIFICATION BY SVD

EXPERIMENTS ON ARTIFICIAL TEXTURE

In this chapter, we address the problem of supervised classification of artificial textures with some of the SVD features. Evaluation of the features is performed by the Bhattacharyya distance figure of merit. Experiments are performed on textures with controlled correlations, and texture features used in the process are the first four moments of the vector z_2, and scalar features z_7, z_{10}, z_{11}, and z_{15} defined in Chapter 7. In the next chapter, experiments are performed on all of the SVD features. Detection of one texture embedded in another texture is also the topic of discussion in this chapter.

8.1 Artificial Textures

A model was developed for texture in Chapter 4. The model is required to be zero-mean, real, separable, normally distributed, but not necessarily stationary for the theoretical developments of Chapters 5 and 6. For experiments of this chapter, this model will be utilized. Given the nature of the separable covariance matrix of the texture field $\underline{F}$, e.g. first, second, higher-order Markov, or non-Markovian, it is possible to generate a texture field in a sequential manner so that each pixel is a linear combination of its previous ones added to a noise factor [35] and [141]. Such a type of synthesis can be used for simulation of texture fields with separable N^{th} order Markov [138-140] covariance matrices. In the following

experiments on artificial texture, a wide-sense stationary first order Markov process $[\rho^{|i-j|}]$ is used for covariance matrices $\underline{K}_C$ and $\underline{K}_R$. A texture window is defined as a small square portion of textured image. Nonoverlapping sample windows are randomly selected from various locations of image and SVD is performed on them.

The Best Window Size

Several experiments have been performed to determine the best window size. The choices for a window size are limited since: a) it cannot be too large because of computational limitations, b) texture windows cannot be too small because they will be non-representative of texture, and also, they will increase the number of sample windows resulting, again, in high computationality.

For standard 512x512 or 256x256 picture sizes, we have experimentally determined that 32x32 window size is the best for our experiments.

Generation

The objective is to generate 512x512 texture fields which are zero mean, normally distributed, having separable first order Markovian covariance matrices $\underline{K}_C = \underline{K}_R = [\rho^{|i-j|}]$. Such texture fields are 512x512 real valued 36-bit per pixel fields (36 bits requirement of DEC10 computer). For visualizing purposes, we perform a linear transformation on each texture field to obtain 8-bit (integer) pictures with a dynamic range of 0 to 255. The linear transformation is

$$\frac{F_{ij}-F_{min}}{F_{max}-F_{min}}(255) \qquad \text{for } i,j = 1,\ldots,512 \quad (8\text{-}1)$$

where F_{ij} is the ij^{th} pixel of $\underline{F}$, F_{min} and F_{max} are the minimum and maximum real values of all textures in the set. Figure 8-1 shows the same texture fields for ρ = 0.5, 0.6, 0.7, and 0.9. Among four texture fields of Figure 8-1, F_{max} = 4.6124967 and F_{min} = -4.4934617.

Definition

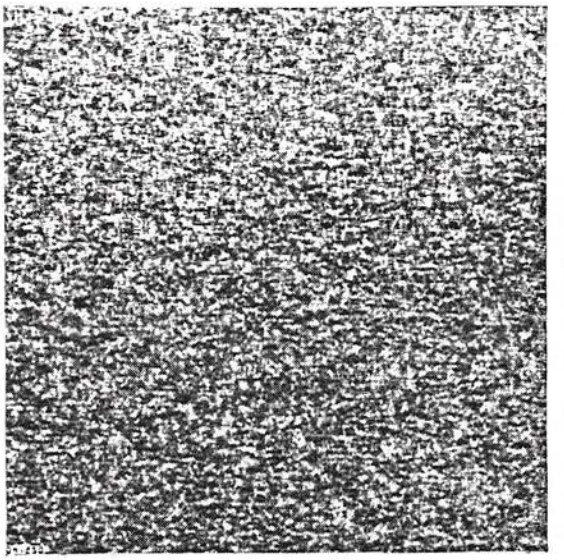
"0.5" means the texture shown in Figure 8-1(a) which represents a texture with $\underline{K_C} = \underline{K_R} = (0.5)^{|i-j|}$. Likewise for other values of ρ we have "0.6," "0.7," and "0.9."

Procedure

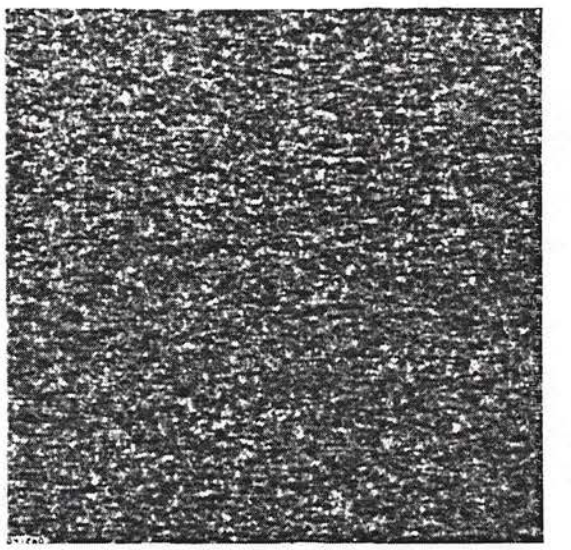
From texture fields "0.5," "0.6," "0.7," and "0.9," N non-overlapping sample windows are randomly extracted to produce N 32-dimensional singular value vectors $\underline{z_1}$ (see Chapter 7, relation (7-11) for definition of $\underline{z_1}$). From each $\underline{z_1}$, its normalized vector $\underline{z_2}$ is formed. There are two sets of experiments performed on $\underline{z_1}$ and $\underline{z_2}$:

(a) From each $\underline{z_2}$, the first four moments are computed to give the 4-dimensional vector $\underline{z_6}$ where $z_6 = (M_1,M_2,M_3,M_4)^T$. Various experiments on combinations of M_1, M_2, M_3, M_4 are performed to determine the effectivity of each combination.

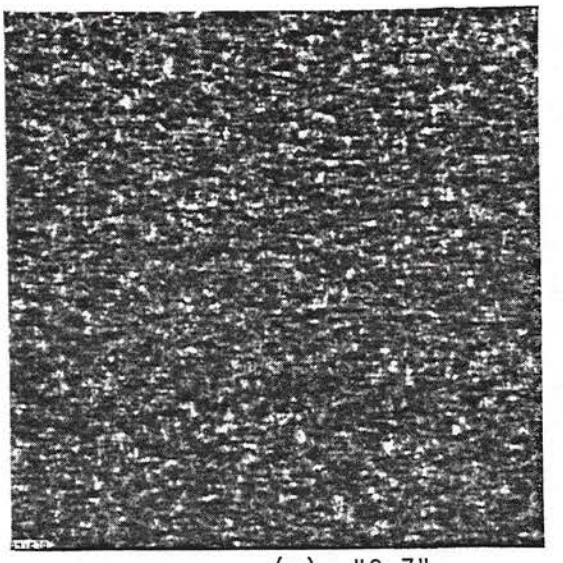
(b) The second set of experiments consists of finding scalar feature z_{10} from each $\underline{z_1}$, and scalar features z_7, z_{11}, z_{15} from each $\underline{z_2}$; and then various combinatorial possibilities of four features z_7, z_{10}, z_{11}, and z_{15} are tested.



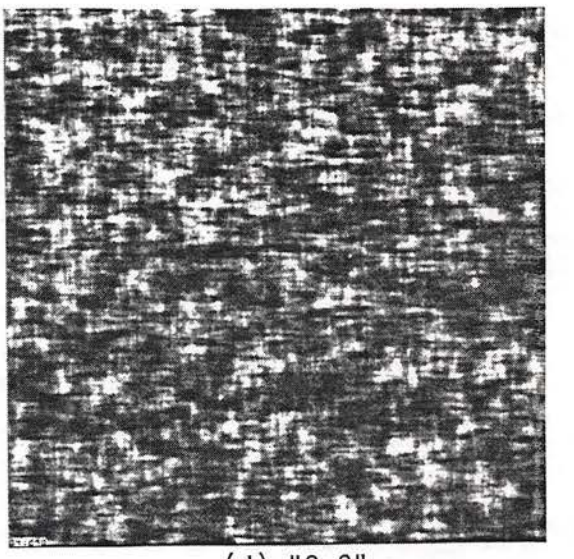
(a) "0.5"



(b) "0.6"



(c) "0.7"



(d) "0.9"

Figure 8-1. Artificial Texture Fields with Various Correlation Factors (512x512).

The criterion of evaluation is B-distance and the number of samples extracted are the same for all experiments making it easy to compare them.

Figure 8-2 gives the block diagram for the evaluator of part (a). Part (b) is essentially performed the same way except that instead of the four moments M_1, M_2, M_3, and M_4, we have z_7, z_{10}, z_{11}, and z_{15}.

Remark: Each 32x32 texture field has mean zero. Its second moment is $tr(\underline{K}_C \otimes \underline{K}_R) = tr(\underline{K}_C) \times tr(\underline{K}_R)$. For a Toeplitz matrix of 512x512, the trace is equal to 512. Hence, the second moment becomes

$$\text{Second Moment} = 512 \times 512 = 262144.0$$

The only question is what the value of N (the number of extracted sample windows) should be. To determine N, experiments are performed with all four features M_1, M_2, M_3, M_4 used. 32 samples are used for the first experiment, 64 for the second, 128 for the third, and 196 for the fourth one:

Number of Sample Windows

Table 8-1 shows the results of the B-distance vs various number of samples using all four moments. Figure 8-3 displays the semi-log scale of B-distance for various number of samples. From Table 8-1 and Figure 8-3 the following observations can be made:

Observation 1: It can be observed, from Table 8-1, that the higher $|\rho_\ell - \rho_m|$, the greater the B-distance, where ρ_ℓ, is the correlation factor in $\underline{K}_C = \underline{K}_R\ [\rho^{|i-j|}]$ for the ℓ^{th} class. Basically the more different the correlation content of a texture field pair, the higher

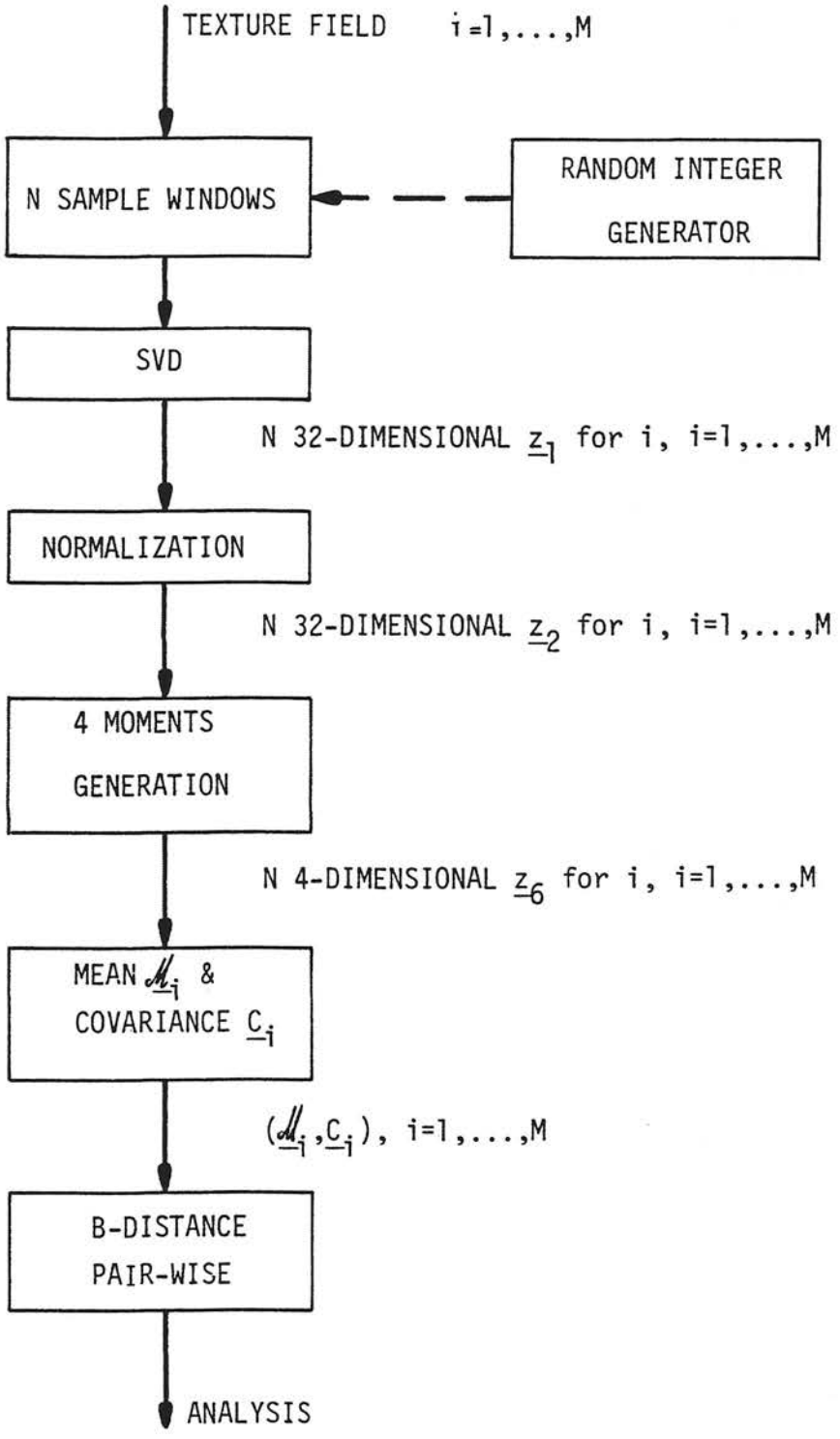


Figure 8-2. Block Diagram of the Evaluator.

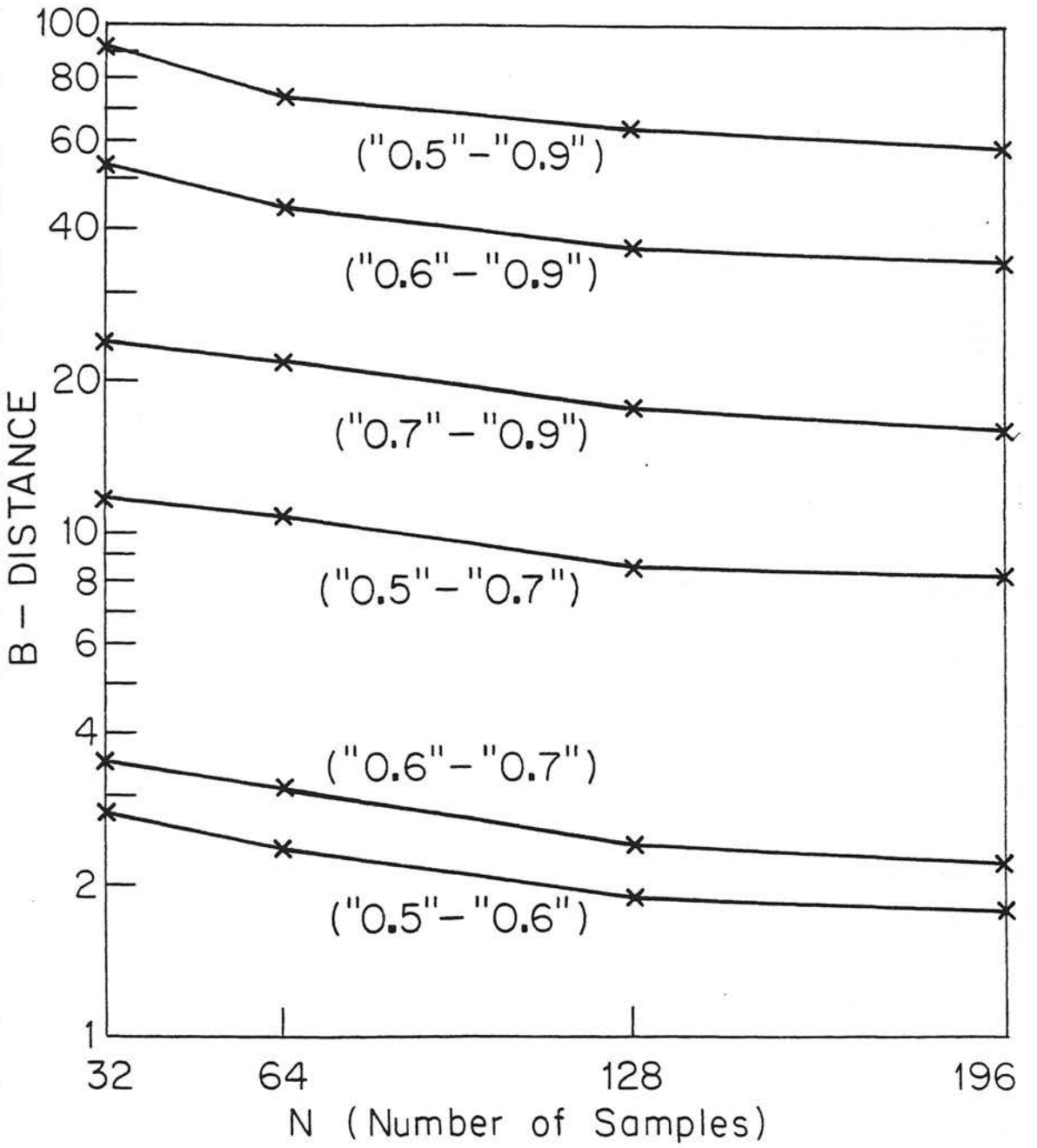


Figure 8-3. Semi-Log Scale of B-Distance vs Number of Samples for Artificial Textures.

Table 8-1. B-Distance for Various Numbers of 32x32 Sample Windows Extracted From Artificial Texture Using M_1, M_2, M_3, M_4.

ARTIFICIAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - (M_1,M_2,M_3,M_4)			
		32 SAMP.	64 SAMP.	128 SAMP.	196 SAMP.
"0.5"	"0.6"	2.76	2.35	1.91	1.81
"0.5"	"0.7"	11.89	10.89	8.69	8.10
"0.5"	"0.9"	91.89	72.17	64.20	59.97
"0.6"	"0.7"	3.50	3.14	2.43	2.27
"0.6"	"0.9"	53.78	45.11	38.15	35.32
"0.7"	"0.9"	23.92	22.44	18.13	16.78

Table 8-2. Bayesian Probability of Classification Accuracy in % Using M_1,M_2,M_3,M_4.

ARTIFICIAL TEXTURE PAIR		64 SAMPLES - (M_1, M_2, M_3, M_4)		
		B-DISTANCE	CLASSIFICATION ACCURACY	
			LOWER BOUND	UPPER BOUND
"0.5"	"0.6"	2.35	95.2	99.8
"0.5"	"0.7"	10.89	100.0	100.0
"0.5"	"0.9"	72.17	100.0	100.0
"0.6"	"0.7"	3.14	97.9	100.0
"0.6"	"0.9"	45.11	100.0	100.0
"0.7	"0.9"	22.44	100.0	100.0

the B-distance.

Observation 2: $|\rho_\ell - \rho_m|$ is not the primary factor which determines the B-distance. For example, |0.9-0.7| and |0.7-0.5| are both equal to 0.2; however, the B-distance for ("0.9," "0.7") pairs are considerably larger than those for ("0.7," "0.5") pair. The conclusion is that ρ_ℓ and ρ_m play individual roles in the B-distance computation. But, for equal absolute value of difference, the higher the $\max(\rho_\ell, \rho_m)$, the higher the B-distance.

Observation 3: There is a systematic decrease in B-distance values with increasing the number of samples.

Using 64 samples seems sufficient for producing the same order of classification accuracy as 196. Figure 8-4 displays 64 non-overlapping sample windows.

8.1.1 Moment Features of z_2

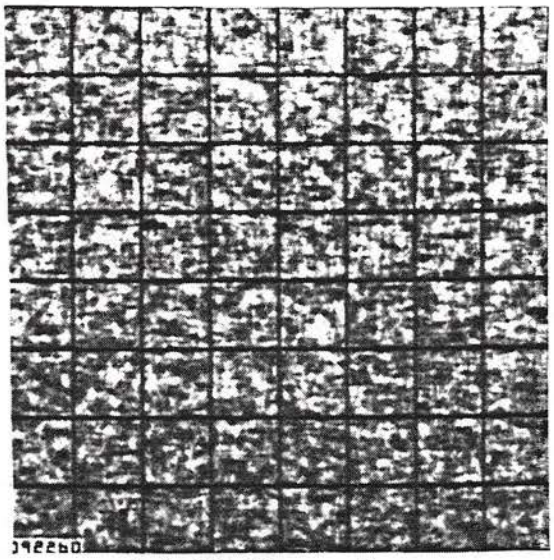
Using artificial textures "0.5," "0.6," "0.7," and "0.9," and applying various combinations of M_1, M_2, M_3, M_4, we obtain the following results:

Combination of M_1, M_2, M_3, M_4 Taken Four at a Time

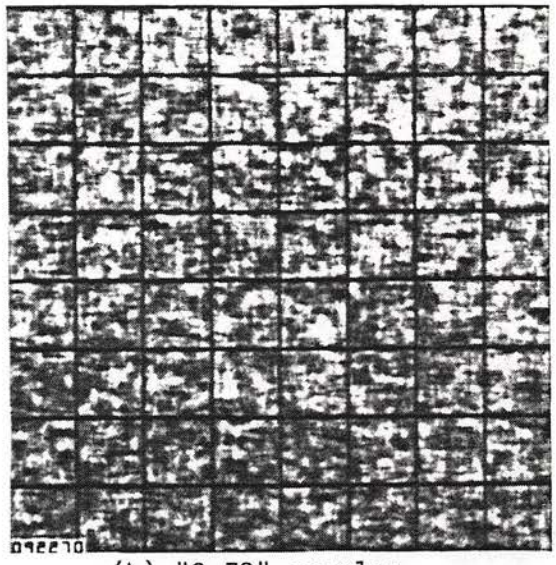
Table 8-2 shows the B-distances and classification accuracy using all four features.(*)

Taken Three at a Time

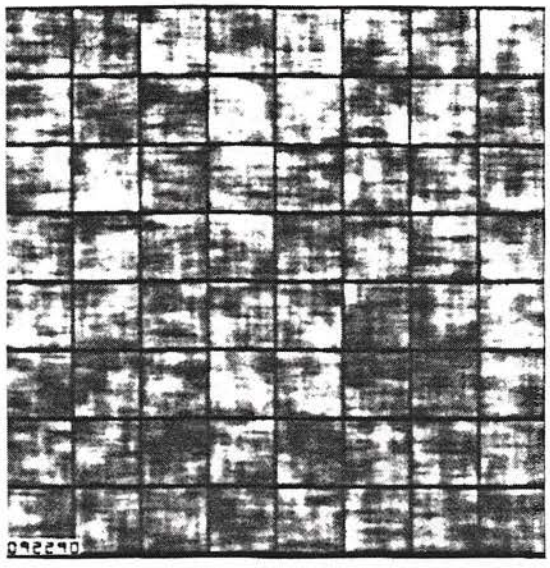
(*)In Table 8-2 and all other tables showing classification accuracy, the number 100% is obtained with the ultimate precision of the DEC PDP10 machine in use.



(a) "0.6" samples



(b) "0.70" samples



(c) "0.9" samples

Figure 8-4. Mean SV Prototype Vectors for Artificial Textures.

Table 8-3 shows the B-distances for combinations of four features taken three at a time hence there are $\binom{4}{3} = \frac{4!}{3!1!} = 4$ possibilities. Table 8-4 tabulates the classification accuracies for Table 8-3.

Taken Two at a Time

Table 8-5 shows the B-distances and Table 8-6 shows the classification accuracies for four features taken two at a time. There are $\binom{4}{2} = \frac{4!}{2!2!} = 6$ possibilities.

Taken One at a Time

Table 8-7 shows the B-distances and Table 8-8 shows their classification accuracy for four features taken one at a time. There are $\binom{4}{1} = \frac{4!}{1!3!} = 4$ possibilities.

Analysis of the Results

Based on the results of Tables 8-2 through 8-8, the following observation can be added to the previous ones:

Observation 4: B-distance is the highest when a maximum number of features is used. Therefore, for four features, the B-distance is higher than that of any combination of three, two, or one feature. In each case, the B-distance is invariant to permutation of elements of the feature group.

Observation 5: Lower combinations of a particular set of features produce a smaller B-distance. For example, features (M_3, M_4) produce a smaller B-distance than (M_2, M_3, and M_4). This is logical because M_2 provides additional information to M_3 and M_4.

Table 8-3. B-Distances for Various Possibilities of Four Features M_1, M_2, M_3, M_4 Taken Three at a Time for Artificial Texture.

ARTIFICIAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		(M_1, M_2, M_3)	(M_1, M_2, M_4)	(M_1, M_3, M_4)	(M_2, M_3, M_4)
"0.5"	"0.6"	2.29	2.23	2.30	2.22
"0.5"	"0.7"	10.44	10.05	10.49	10.47
"0.5"	"0.9"	43.10	47.87	56.50	72.01
"0.6"	"0.7"	3.07	2.97	3.02	3.05
"0.6"	"0.9"	29.41	31.14	33.97	45.08
"0.7"	"0.9"	16.85	16.96	16.75	22.43

Table 8-4. Classification Accuracy in % for Artificial Textures.

ARTIFICIAL TEXTURE PAIR		CLASSIFICATION ACCURACY - 64 SAMPLES							
		(M_1, M_2, M_3)		(M_1, M_2, M_4)		(M_1, M_3, M_4)		(M_2, M_3, M_4)	
		LOW BD.	UP. BD.	LOW BD.	UP. BD.	LOW BD.	UP. BD.	LOW BD.	UP. BD.
"0.5"	"0.6	95.0	99.7	94.7	99.7	95.0	99.8	94.6	99.7
"0.5"	"0.7"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.6"	"0.7"	97.7	99.9	97.5	99.9	97.6	99.9	97.6	99.9
"0.6"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.7"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 8-5. B-Distances for M_1, M_2, M_3, M_4 Taken Two at a Time.

ARITFICIAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES					
		M_1, M_2	M_1, M_3	M_1, M_4	M_2, M_3	M_2, M_4	M_3, M_4
"0.5"	"0.6"	2.12	2.28	2.23	2.05	1.65	1.85
"0.5"	"0.7"	9.52	10.34	10.05	9.49	7.64	8.63
"0.5"	"0.9"	42.48	41.64	45.27	34.38	29.12	46.49
"0.6"	"0.7"	2.71	2.99	2.96	2.98	2.50	2.57
"0.6"	"0.9"	29.18	27.37	28.40	25.86	22.41	28.77
"0.7"	"0.9"	16.49	14.94	14.88	16.05	14.47	14.90

Table 8-6. Classification Accuracy in %.

ARTIFICIAL TEXTURE PAIR		CLASSIFICATION ACCURACY - 64 SAMPLES											
		M_1, M_2		M_1, M_3		M_1, M_4		M_2, M_3		M_2, M_4		M_3, M_4	
		LW.BD.	UP.BD.	LW.BD.	UP.BD.	LW.BD.	UP.BD.	LW.BD.	UP.BD.	LW.BD.	UP.BD.	LW.BD.	UP.BD.
"0.5"	"0.6"	94.0	99.6	94.6	99.7	94.6	99.7	93.6	99.6	90.4	99.1	92.1	99.4
"0.5"	"0.7"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.6"	"0.7"	96.7	99.9	97.5	99.9	97.4	99.9	97.2	99.9	95.9	99.8	96.2	99.9
"0.6"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.7"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 8-7. B-distances for M_1, M_2, M_3, M_4 Taken One at a Time.

ARTIFICIAL TEXTURE PAIR		B-DISTANCE - 64 SAMPLES			
		M_1	M_2	M_3	M_4
"0.5"	"0.6"	2.11	0.52	1.82	1.56
"0.5"	"0.7"	9.52	2.92	8.46	7.14
"0.5"	"0.9"	41.08	27.81	26.53	16.94
"0.6"	"0.7"	2.70	0.98	2.54	2.34
"0.6"	"0.9"	27.18	21.03	19.88	14.10
"0.7"	"0.9"	14.50	13.19	12.47	10.10

Table 8-8. Classification Accuracy in %

ARTIFICIAL TEXTURE PAIR		CLASSIFICATION ACCURACY - 64 SAMPLES							
		M_1		M_2		M_3		M_4	
		LOW BD.	UP. BD.	LOW BD.	UP. BD.	LOW BD.	UP. BD.	LOW BD.	UP.BD.
"0.5"	"0.6"	94.0	99.6	70.2	90.1	91.9	99.3	89.5	98.9
"0.5"	"0.7"	100.0	100.0	97.3	99.9	100.0	100.0	100.0	100.0
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.6"	"0.7"	96.6	99.9	81.3	96.4	96.1	99.8	95.2	99.8
"0.6"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.7"	"0 9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Observation 6: For different feature sets, some two-feature combinations and some one feature combinations are better than some higher number feature forms. For examples combination (M_1,M_3) produces a higher B-distance than (M_1,M_2,M_4); M_1 alone is a more effective feature than (M_2,M_3) or (M_2,M_4).

Observation 7: Looking at Tables 8-7 and 8-8, it is easily verified that the most effective single feature is M_1, followed by M_3, M_4, and M_2, respectively.

Remark: The dimensionality of $z_6 = (M_1,M_2,M_3,M_4)^T$ can further be reduced by a linear transformation that does not significantly alter the performance [40, pp. 273-275].

Overall Results

Single-feature combinations incur minimum cost. The best of them for classification purposes is M_1 (mean) followed by M_3 (skewness), M_4 (kurtosis), and M_2 (deviation) taken from z_2. Classification accuracy can be enhanced by using more than one feature, reaching up to four in which case the accuracy is at an excellent rate of 99.5%, with its worst case at 95.2. The best double-feature is (M_1, M_2), and that of triple-feature being (M_2,M_3,M_4).

8.1.2 Scalar Features

On artificial texture fields "0.5," "0.6," "0.7," and "0.9" of Figure 8-1, a similar set of experiments such as the one in Section

8.1.1 are performed except that instead of M_1, M_2, M_3, M_4, four single features z_7, z_{10}, z_{11}, and z_{15} are used where z_7 = entropy of all singular values, z_{10} = the largest singular value, z_{11} = normalized largest singular value and z_{15} = entropy content of the largest singular value. (See Section 7.4 for a complete description of the scalar features.)

Combination of z_7, z_{10}, z_{11}, and z_{15} Taken Four at a Time

Table 8-9 shows the B-distance and classification accuracies.

Combination of z_7, z_{10}, z_{11}, and z_{15} Taken Three at a Time

Tables 8-10 and 8-11 shows the B-distance and classification accuracies.

Combination of z_7, z_{10}, z_{11}, and z_{15} Taken Two at a Time

Tables 8-12 and 8-13 show the B-distance and classification accuracies.

Combination of z_7, z_{10}, z_{11}, and z_{15} Taken One at a Time

Tables 8-14 and 8-15 tabulate the B-distance and classification accuracies.

Observation 8: Taking all four features results in a near perfect accuracy. For artificial textures, this set of four features are more effective than the moments M_1, M_2, M_3, M_4.

Observation 9: One of the features in this set namely $z_{10} = s_1$ is related only to one singular value while the other three are related to all singular values out of which two are functions of the first

Table 8-9. B-Distance and Classification Accuracies for $z_7, z_{10}, z_{11}, z_{15}$) Taken Four at a Time.

ARTIFICIAL TEXTURE PAIR		64 SAMPLES - ($z_7, z_{10}, z_{11}, z_{15}$)		
		B-DISTANCE	CLASSIFICATION ACCURACY %	
			LOWER BOUND	UPPER BOUND
"0.5"	"0.6"	4.32	99.3	100.0
"0.5"	"0.7"	22.59	100.0	100.0
"0.5"	"0.9"	174.29	100.0	100.0
"0.6"	"0.7"	8.04	100.0	100.0
"0,6"	"0.9"	138.34	100.0	100.0
"0.7"	"0.9"	94.32	100.0	100.0

Table 8-10. B-Distances for z_7, z_{10}, z_{11}, z_{15} Taken Three at a Time.

ARTIFICIAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		z_7, z_{10}, z_{11}	z_7, z_{10}, z_{15}	z_7, z_{11}, z_{15}	z_{10}, z_{11}, z_{15}
"0.5	"0.6"	3.60	3.69	2.82	2.15
"0.5"	"0.7"	19.70	21.33	11.45	8.89
"0.5"	"0 9"	135.20	111.10	49.98	58.09
"0.6"	"0.7"	7.20	7.18	3.73	3.14
"0.6"	"0.9"	110.80	77.73	30.19	41.62
"0.7"	"0.9"	76.64	47.28	15.96	26.76

Table 8-11. Classification Accuracies in %.

ARTIFICIAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		z_7, z_{10}, z_{11}		z_7, z_{10}, z_{15}		z_7, z_{11}, z_{15}		z_{10}, z_{11}, z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP. BD.
"0.5"	"0.6"	98.6	100.0	98.8	100.0	97.0	99.9	94.2	99.7
"0.5"	"0.7"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.6"	"0.7"	100.0	100.0	100.0	100.0	98.8	100.0	97.8	100.0
"0.6"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.7"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 8-12. B-Distances for $z_7, z_{10}, z_{11}, z_{15}$ Taken Two at a Time for Artificial Textures.

ARTIFICIAL TEXTURE PAIRS		BHATTACHARYYA DISTANCE - 64 SAMPLES					
		z_7, z_{10}	z_7, z_{11}	z_7, z_{15}	z_{10}, z_{11}	z_{10}, z_{15}	z_{11}, z_{15}
"0.5"	"0.6"	2.71	2.15	2.11	1.46	1.52	1.19
"0.5"	"0.7"	13.33	9.33	8.73	6.45	7.59	3.85
"0.5"	"0.9"	59.16	22.68	24.78	29.67	37.77	30.42
"0.6"	"0.7"	4.52	2.95	2.84	2.35	2.30	1.34
"0.6"	"0.9"	47.19	17.37	15.53	25.94	19.65	15.70
"0.7"	"0.9"	30.63	10.85	8.56	19.89	8.28	6.91

Table 8-13. Classification Accuracy in %.

ARTIFICIAL TEXTURE PAIRS		% CLASSIFICATION ACCURACY - 64 SAMPLES											
		z_7, z_{10}		z_7, z_{11}		z_7, z_{15}		z_{10}, z_{11}		z_{10}, z_{15}		z_{11}, z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
"0.5"	"0.6"	96.7	99.9	94.1	99.7	93.9	99.6	88.4	98.6	89.1	98.8	84.7	97.6
"0.5"	"0.7"	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	98.9	100.0
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.6"	"0.7"	99.5	100.0	97.4	99.9	97.1	99.9	95.2	99.8	95.0	99.7	86.9	98.2
"0.6"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.7"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 8-14. B-Distances for $z_7, z_{10}, z_{11}, z_{15}$ Taken One at a Time for Artificial Texture.

ARTIFICIAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		z_7	z_{10}	z_{11}	z_{15}
"0.5"	"0.6"	1.98	0.17	0.53	0.53
"0.5"	"0.7"	7.89	0.57	1.82	2.08
"0.5"	"0.9"	14.93	1.30	4.54	23.09
"0.6"	"0.7"	2.43	0.16	0.57	0.57
"0.6"	"0.9"	11.69	0.94	3.66	12.59
"0.7"	"0.9"	7.63	0.55	2.50	5.18

Table 8-15. Classification Accuracy in %.

ARTIFICIAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		z_7		z_{10}		z_{11}		z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP. BD.
"0.5"	"0.6"	93.1	99.5	57.8	76.8	70.5	90.3	70.5	90.3
"0.5"	"0.7"	100.0	100.0	71.8	91.3	91.9	99.3	93.7	99.6
"0.5"	"0.9"	100.0	100.0	86.3	98.1	99.5	100.0	100.0	100.0
"0.6"	"0.7"	95.6	99.8	57.4	76.2	71.7	91.2	71.7	91.2
"0.6"	"0.9"	100.0	100.0	80.5	96.0	98.7	100.0	100.0	100.0
"0.7"	"0.9"	100.0	100.0	71.2	90.9	95.9	99.8	99.7	100.0

normalized singular value i.e. $z_{11} = s_1 / \sum_{i=1}^{k} s_i$ and $z_{15}=z_{11} \log z_{11}$. z_7 utilizes the information regarding all SV's that is why it is more accurate than the others.

Observation 10: From Tables 8-14 and 8-15 it can be deduced that the most effective single feature is z_7 = entropy followed by z_{15}, z_{11}, and z_{10}. Such a result is quite logical since z_{10} utilizes the least amount of information. Interestingly, the double feature (z_7,z_{10}) is more effective than (z_7,z_{15}) or (z_7,z_{11}). This is because most of the information in z_{15} or z_{11} is already in z_7, and we are not adding as much to z_7 by using z_{15} or z_{11} alongside it as we do when we use z_{10}.

Overall Results

One feature z_7 is sufficient to give an average of 98% classification accuracy on the textures under study, and therefore is the best single-feature in these experiments. The best double-feature is (z_7,z_{10}), while the best triple-feature is (z_7,z_{10},z_{11}). Figure 8-5 shows the mean singular value prototype vectors and an examination of it verifies all of the results in Tables 8-2 through 8-15. The curves which are more apart agree with the texture pairs which produce higher B-distances.

8.2 Classification of a Texture Against Textural Background

This section is devoted to the general task of detecting a texture within another one. It has been discussed and shown in Chapter 7 that SVD is not a good measure of the structural properties of a picture. SVD, on the other hand, is a good and inexpensive

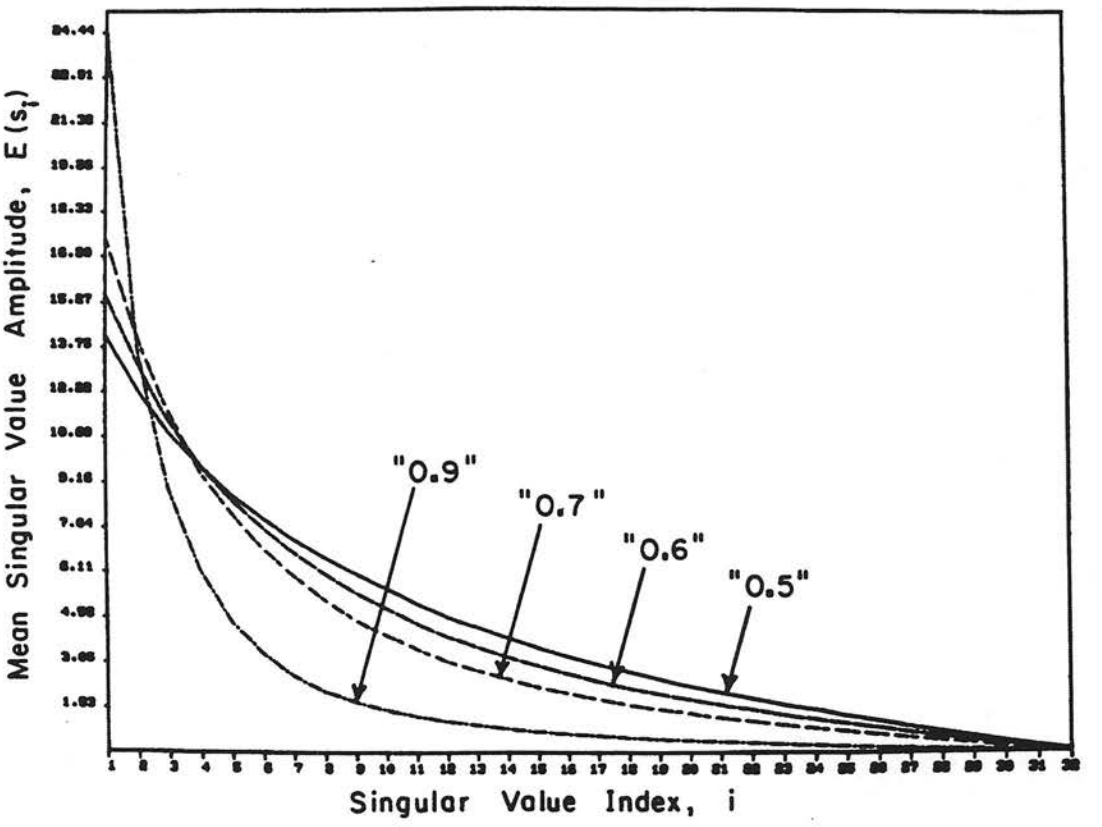


Figure 8-5. Mean SV Prototype Vectors for Artificial Textures.

measure of correlation and stochastic behavior of images. In detection of an object against a background, an edge detection or thresholding technique can be applied in a rather efficient manner. Such techniques, however, do not perform well for texture mixtures. Particularly, thresholding performs poorly for segmentation of a mixture of similar textural patterns. In such a case where there is a texture against textural background, SVD seems to be useful.

A set of experiments are performed on artificial textures with controlled correlation contents. After generating the texture fields with various ρ's, 16 32x32 sample windows of texture with $\rho = 0.5$ is imbedded onto texture fields with $\rho = 0.6$, and $\rho = 0.9$. The same embedments are performed in a reverse manner.

Notational Convention: "0.5" of Figure 8-1 represents the texture associated with $\rho = 0.5$. "0.5/0.9" means the texture associated with $\rho = 0.9$ has been imbedded onto the textural background "0.5." Figure 8-6 shows a schematic depiction of "0.5/0.9."

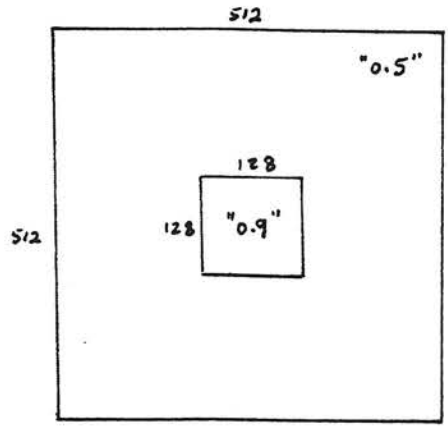


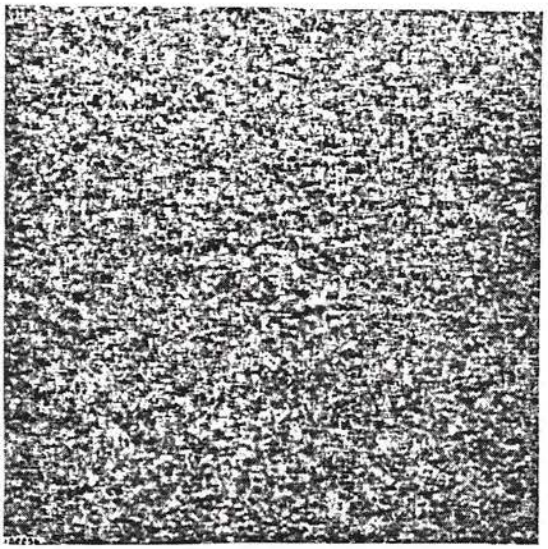
Figure 8-6. Sketch of "0.5"/"0.9" Texture Mixture.

For classification of texture mixtures, **moment features** M_1, (M_1,M_2), (M_2,M_3,M_4), and (M_1,M_2,M_3,M_4) and **scalars features** z_7, (z_7,z_{10}), (z_7,z_{10},z_{11}), and $(z_7,z_{10},z_{11},z_{15})$ are used. Figure 8-7 presents the imbedded forms "0.5/0.6" and "0.6/0.5" for which B-distances between various combinatorial pairs have been tabulated in Tables 8-16 and 8-17 using **moment features**; and Tables 8-18 and 8-19 show the same results applying **scalar features.** Tables 8-20 through 8-23 show the results for "0.5," "0.9," "0.5/0.9" and "0.9/0.5" textures of Figure 8-8. From each of the texture fields in Figure 8-7 and 8-8, 64 32x32 sample windows are randomly selected making sure that the 16 imbedded ones are included. This, in essence, means that for example "0.5/0.9" provides 75% "0.5" + 25% "0.9." The method of evaluation is as given by block diagram of Figure 8-2.

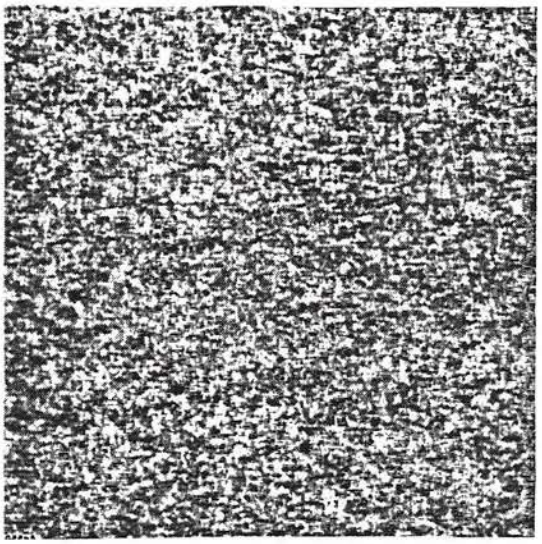
8.2.1 Analysis

Figure 8-9 shows the mean singular value prototype vector for texture mixtures of Figures 8-7 and 8-8.

The B-distance between "0.5" and "0.5" is obviously zero. Table 8-22 shows that using all four scalars, the B-distance between "0.5" and 100% "0.9" is 205.03 which is considered to be enormous. B-distance between "0.5" and "0.9/0.5" (i.e. 75% "0.9" + 25% "0.5") is 4.629. The drop from 205.03 to 4.62 is clearly an indication that a great change in B-distance happens when 75% "0.9" is used instead of 100%. The drop is 200.41 and is so great that the probability of error in separating them is almost zero. Likewise, when "0.5" is taken with 100% "0.5," the B-distance is zero; but when "0.5" is taken with 75% "0.5" + 25% "0.9," B-distance increases to 3.41. The



(a) "0.5/0.6"



(b) "0.6/0.5"

Figure 8-7. Embedments "0.5/0.6" and "0.6/0.5".

Table 8-16. B-Distances Between Various Combinations of "0.5," "0.6," "0.5/0.6," and "0.6/0.5" Using Moment Features.

ARTIFICIAL TEXTURE PAIR		B-DISTANCE - 64 SAMPLES			
		(M_1)	(M_1,M_2)	(M_2,M_3,M_4)	(M_1,M_2,M_3,M_4)
"0.5"	"0.6"	2.34	2.40	2.36	2.60
"0.5"	"0.5/0.6"	0.20	0.20	0.19	0.21
"0.5"	"0.6/0.5"	0.64	0.65	0.66	0.70
"0.6"	"0.5/0.6"	0.59	0.60	0.61	0.65
"0.6"	"0.6/0.5"	0.17	0.17	0.17	0.18
"0.5/0.6"	"0.6/0.5"	0.14	0.14	0.15	0.16

Table 8-17. Classification Accuracies in %.

ARTIFICIAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		M_1		(M_1,M_2)		(M_2,M_3,M_4)		(M_1,M_2,M_3,M_4)	
		LOW BD.	UP.BD.	LOW BD	UP.BD	LOW BD.	UP.BD.	LOW BD.	UP.BD.
"0.5"	"0.6"	95.2	99.8	95.5	99.8	95.2	99.7	96.3	99.9
"0.5"	"0.5/0.6"	58.9	78.4	59.0	78.7	59.0	78.6	59.7	79.6
"0.5"	"0.6/0.5"	73.6	92.5	73.9	92.6	74.0	92.8	75.2	93.4
"0.6"	"0.5/0.6"	72.3	91.6	72.5	91.8	72.9	92.0	73.9	92.6
"0.6"	"0.6/0.5"	57.7	76.7	57.8	76.7	57.9	77.1	58.5	77.9
"0.5/0.6"	"0.6/0.5"	56.3	74.3	56.5	74.7	57.2	75.8	57.8	76.8

Table 8-18. B-Distances Between Various Combinations of "0.5," "0.6," "0.5/0.6," and "0.6/0.5" Using Scalar Features.

ARTIFICIAL TEXTURE PAIR		B-DISTANCE - 64 SAMPLES			
		(z_7)	(z_7,z_{11})	(z_7,z_{10},z_{11})	$(z_7,z_{10},z_{11},z_{15})$
"0.5"	"0.6"	2.18	2.27	3.82	4.69
"0.5"	"0.5/0.6"	0.21	0.21	0.28	0.61
"0.5"	"0.6/0.5"	0.65	0.67	0.80	1.33
"0.6"	"0.5/0.6"	0.54	0.57	0.77	1.04
"0.6"	"0.6/0.5"	0.14	0.15	0.25	0.37
"0.5/0.6"	"0.6/0.5"	0.13	0.15	0.17	0.25

Table 8-19. Classification Accuracies in %.

ARTIFICIAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(z_7)		(z_7,z_{11})		(z_7,z_{10},z_{11})		$(z_7,z_{10},z_{11},z_{15})$	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
"0.5"	"0.6"	94.3	99.7	94.8	99.7	98.9	100.0	99.5	100.0
"0.5"	"0.5/0.6"	59.5	79.3	59.5	79.4	62.2	82.8	72.8	92.0
"0.5"	"0.6/0.5"	73.9	92.6	74.4	92.9	77.7	94.7	86.9	98.3
"0.6"	"0.5/0.6"	70.9	90.7	71.8	91.3	77.0	94.4	82.4	96.8
"0.6"	"0.6/0.5"	56.6	74.7	57.0	75.6	61.2	81.5	65.5	86.2
"0.5/0.6"	"0.6/0.5"	56.2	74.1	56.8	75.1	58.1	77.3	61.2	81.5

Table 8-20. B-Distances Between Various Combinations of "0.5," "0.9," "0.5/0.9," and "0.9/0.5" Using Moment Features.

ARTIFICIAL TEXTURE PAIR		B-DISTANCE = 64 SAMPLES			
		(M_1)	(M_1,M_2)	(M_1,M_2,M_3)	(M_1,M_2,M_3,M_4)
"0.5"	"0.9"	43.02	46.13	71.85	71.96
"0.5"	"0.5/0.9"	0.94	0.94	1.44	1.48
"0.5"	"0.9/0.5"	1.61	1.62	2.27	2.34
"0.9"	"0.5/0.9"	1.37	0.42	1.83	1.87
"0.9"	"0.9/0.5"	0.69	0.74	1.04	1.04
"0.5/0.9"	"0.9/0.5"	0.18	0.18	0.26	0.29

Table 8-21. Classification Accuracies in %.

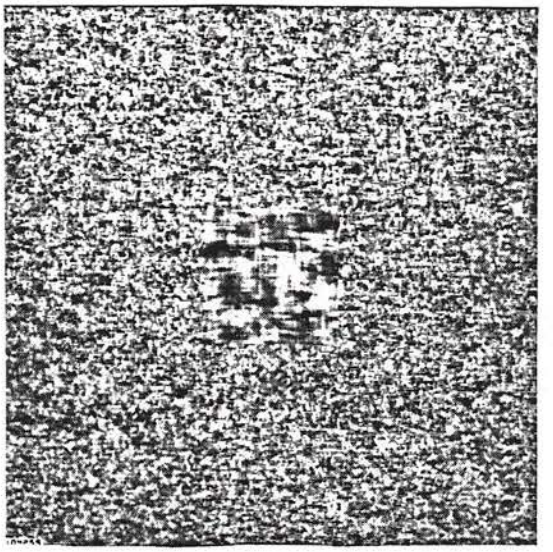
ARTIFICIAL TEXTURE PAIR		(M_1)		(M_1,M_2)		(M_1,M_2,M_3)		(M_1, M_2, M_3, M_4)	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.5"	"0.5/0.9"	80.5	96.0	80.5	96.0	88.2	98.6	88.7	98.7
"0.5"	"0.9/0.5"	90.0	99.0	90.1	99.0	94.9	99.7	95.2	99.8
"0.9"	"0.5/0.9"	87.3	98.4	87.9	98.5	92.0	99.4	92.3	99.4
"0.9"	"0.9/0.5"	74.9	93.2	76.1	93.9	82.4	96.8	82.4	96.8
"0.5/0.9"	"0.9/0.5"	58.1	77.3	58.4	77.7	61.5	81.9	62.9	83.5

Table 8-22. B-Distances Between Various Combinations of "0.5," "0.9," "0.5/0.9," and "0.9/0.5" Using Scalar Features.

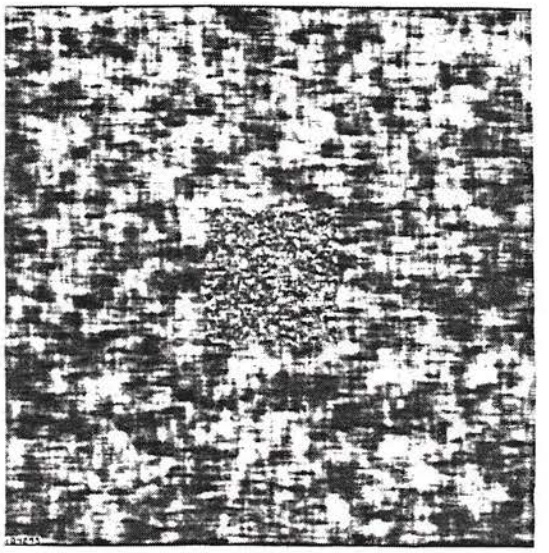
ARTIFICIAL TEXTURE PAIR		B-DISTANCE - 64 SAMPLES			
		(z_7)	(z_7,z_{11})	(z_7,z_{10},z_{11})	$(z_7,z_{10},z_{11},z_{15})$
"0.5"	"0.9"	15.02	26.39	154.71	205.03
"0.5"	"0.5/0.9"	1.18	1.40	1.74	3.41
"0.5"	"0.9/0.5"	1.85	2.34	2.84	4.62
"0.9"	"0.5/0.9"	1.04	1.33	1.99	2.16
"0.9"	"0.9/0.5"	0.38	0.52	1.06	1.14
"0.5/0.9"	"0.9/0.5"	0.18	0.27	0.37	0.41

Table 8-23. Classification Accuracies in %.

ARTIFICIAL TEXTURE PAIR		% CLASSIFICATION ACCURACY							
		(z_7)		(z_7,z_{11})		(z_7,z_{10},z_{11})		$(z_7,z_{10},z_{11},z_{15})$	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
"0.5"	"0.9"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
"0.5"	"0.5/0.9"	84.7	97.6	87.7	98.5	91.3	99.2	98.4	100.0
"0.5"	"0.9/0.5"	92.1	99.4	95.2	99.8	97.1	99.9	99.5	100.0
"0.9"	"0.5/0.9"	82.4	96.8	86.8	98.2	93.2	99.5	94.2	99.7
"0.9"	"0.9/0.5"	65.9	86.6	70.3	90.3	82.7	96.9	84.0	97.4
"0.5/0.9"	"0.9/0.5"	58.4	77.7	62.0	82.5	65.6	86.3	66.9	87.5

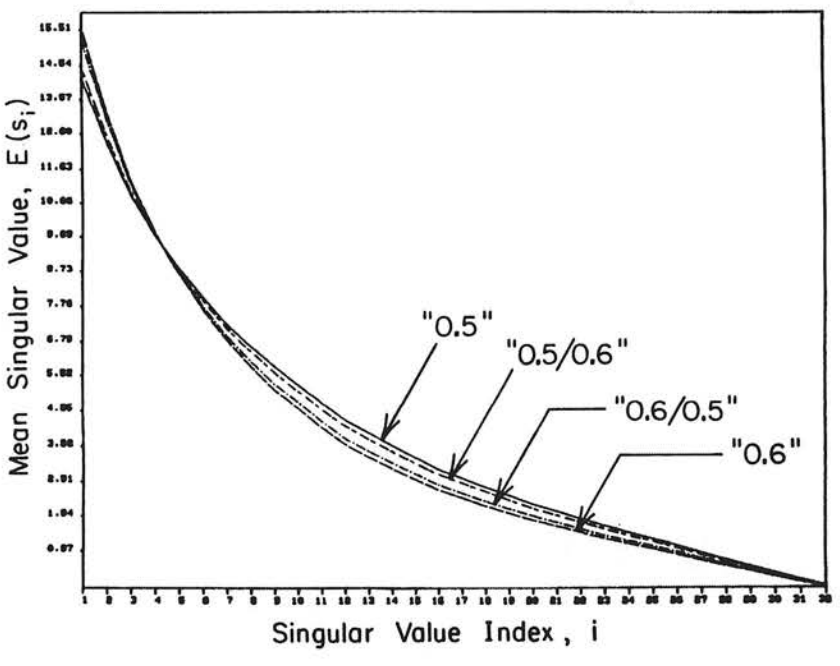


(a) "0.5/0.9"

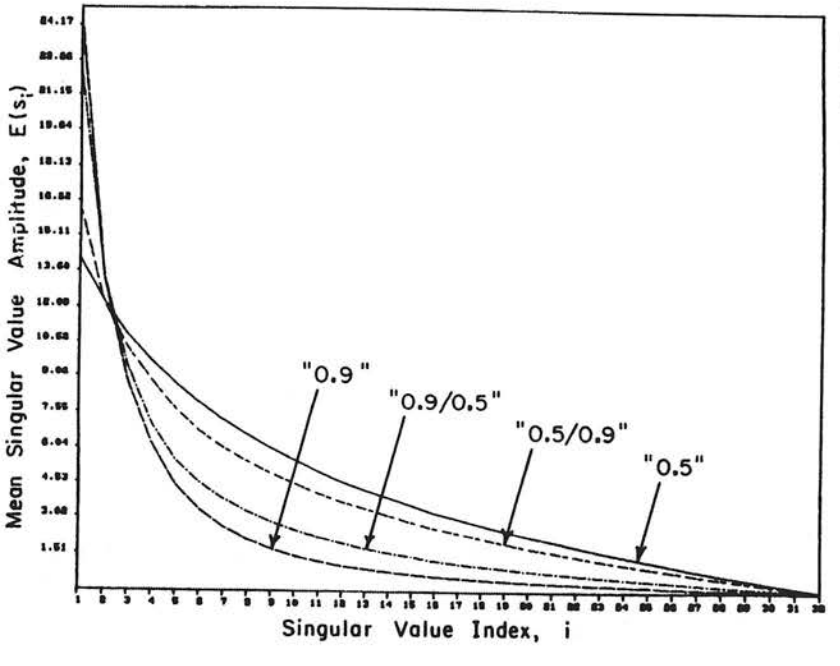


(b) "0.9/0.5"

Figure 8-8. Embedments "0.5/0.9" and "0.9/0.5."



(a) embedment of "0.5" and "0.6"



(b) embedment of "0.5" and "0.9"

Figure 8-9. Mean SV Prototype Vector for Embedded Artificial Textures.

classification accuracy for the latter is, in its minimum, 98.4%. This means that the two textures which used to be the same and could not be told apart can now be discriminated and classified with a good probability of at least 98.4%. This is due to existence of a foreign texture in one of them. The same analysis holds for moment features given in Table 8-20.

The more interesting result on artificial texture is for textures of Figure 8-7 in which the embedments can hardly be distinguished by human eye. But, using SVD scalar features (Table 8-18 and 8-19), the foreign texture "0.5" in "0.6/0.5" can be found with an accuracy of between 86.9% to 98.3%. The average being 92.6% gives an indication of SVD features usefulness in detection and classification of a texture within another one.

8.3 Significance of the Largest Singular Value Feature

Figures related to the curves of mean-SV-prototpye-vectors show that their steepness systematically increases with texture correlation. The most sensitive element of this change is that of S_1. For example, in Figure 8-5, the mean is higher for ρ =0.9 correlation and drastically drops from 24.44 to 17.43 for a 0.2 change in ρ.

8.4 Summary

This chapter has shown the usefulness of SVD features for classifying textural image data. Artificial textures have been employed because it is possible to control their correlation so that the behavior of singular values can be observed in a controlled environment. These results will aid us in real-world experiments of

Chapter 9 where natural texures will be under experiment. It is established here, that both scalar and moment features perform very well in a controlled textural environment, while scalar ones have a slight advantage over moments.

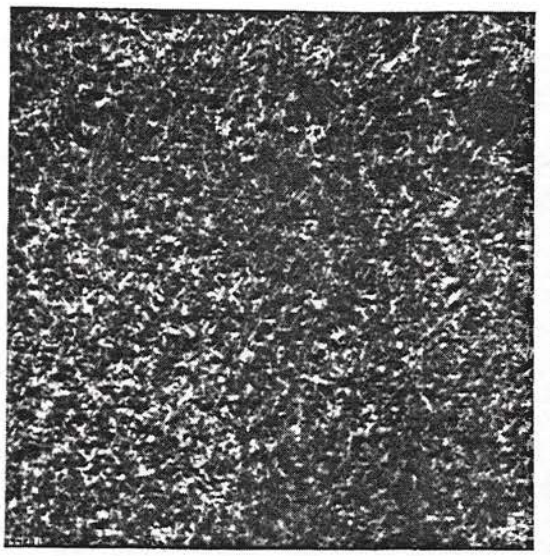
CHAPTER 9

CLASSIFICATION OF NATURAL TEXTURE BY SVD

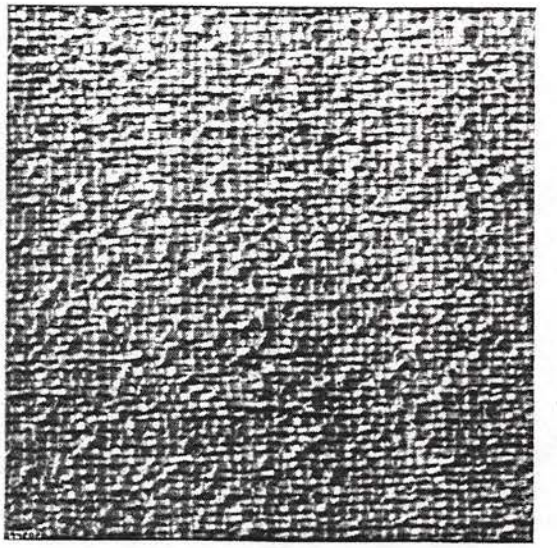
Natural textures, as defined in Chapter 2, are referred to image patterns whose subjects have been created either by nature (such as grass) or exist in every day life without being altered or generated by machines prior to being photographed (such as patterns of a wool cloth).

9.1 Natural Textures

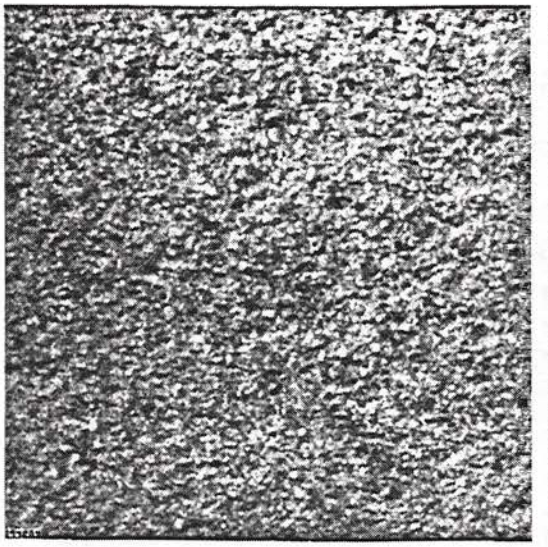
Four natural texture fields of grass, raffia, sand, and wool have been employed for experimentation. Figure 9-1 shows the four textures. They are digitized and have 512x512 pixels quantized to 256 gray levels. The principle of evaluation is as in Chapter 8. However, the natural textures have been standardized so that measurements on them can be made in an unbiased fashion. By standardization, it is meant to create textures that have the same first-order statistics and equal variance. For such a purpose, we Gaussianize the four textures' histograms having zero mean and unit variance (see Appendix G). The 32x32 windows at the boundaries of natural textures are avoided. Such a practice provides 196 possible non-overlapping 32x32 sample windows from each texture. An experiment is performed using moments varying the number of samples. For such a purpose, N = 32, 64, 128, and 196 is used, and the results are tabulated in Table 9-1. Figure 9-2 shows semi-log scale of the B-distances of Table 9-1. The fluctuation in the values is stabilized after 64 samples. Sixty-four non-overlapping 32x32



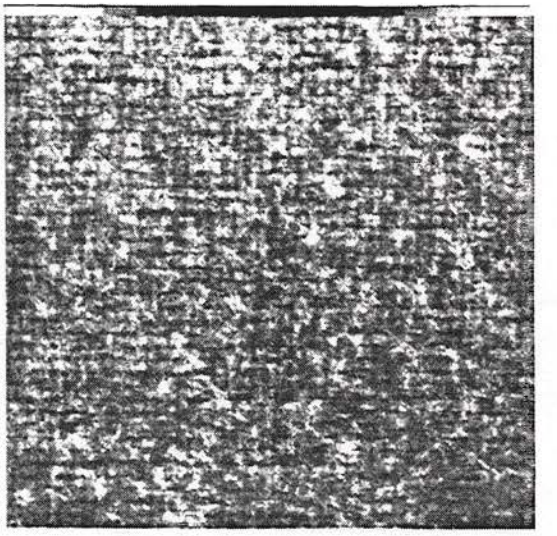
(a) grass



(b) raffia



(c) sand



(d) wool

Figure 9-1. Natural Texture Fields (512x512).

Table 9-1. B-Distances for Various Numbers of 32x32 Sample Windows Extracted From Natural Textures.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - (M_1, M_2, M_3, M_4)			
		32 SAMP	64 SAMP.	128 SAMP	196 SAMP.
G	R	2.47	2.47	2.36	2.42
G	S	1.62	1.42	1.25	1.25
G	W	4.11	3.7	3.45	3.31
R	S	10.91	7.20	6.64	6.33
R	W	13.81	11.20	10.69	9.24
S	W	4.25	3.91	2.94	2.56

Table 9-2. B-Distances and Classification Accuracies for M_1, M_2, M_3, M_4 Taken Four at a Time.

NATURAL TEXTURE PAIR		64 SAMPLES - (M_1, M_2, M_3, M_4)		
		B-DISTANCE	CLASSIFICATION ACCURACY %	
			LOWER BOUND	UPPER BOUND
G	R	2.47	95.8	99.8
G	S	1.42	88.0	98.5
G	W	3.71	98.8	100.0
R	S	7.20	100.0	100.0
R	W	11.20	100.0	100.0
S	W	3.91	99.0	100.0

G = grass, R = raffia, S = sand, W = wool

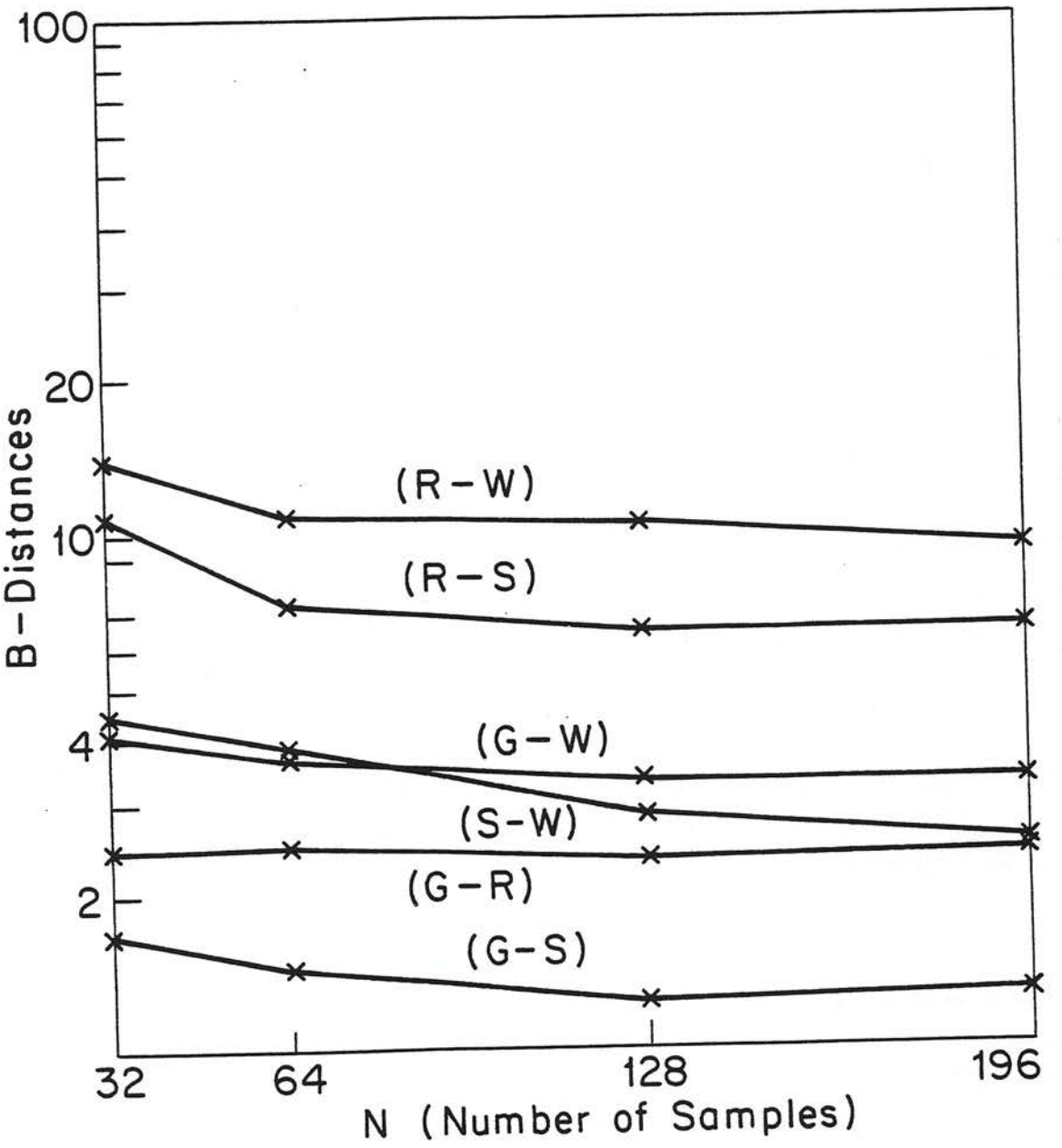


Figure 9-2. Semi-Log Scale of the B-Distances vs Number of Samples for Natural Textures.

sample windows for four textures are displayed in Figure 9-3. Five sets of experiments are performed with M_1, M_2, M_3, M_4 taken from vector features z_1, z_2, z_3, z_4, and z_5; and two sets on scalar features z_7, z_8, z_{10}, z_{11}, z_{12}, z_{13}, z_{14}, z_{15}.

9.1.1 Vector Features and Their Comparison

Using natural textures grass, raffia, sand, and wool, various combinations of M_1, M_2, M_3, M_4 are tested.

z_2, The Normalized Singular Value Histogram

M_1, M_2, M_3, M_4 Taken Four at at Time

Table 9-2 shows the B-distance and classification accuracy.

Taken Three at a Time

There are four possibilities. Tables 9-3 and 9-4 show the results.

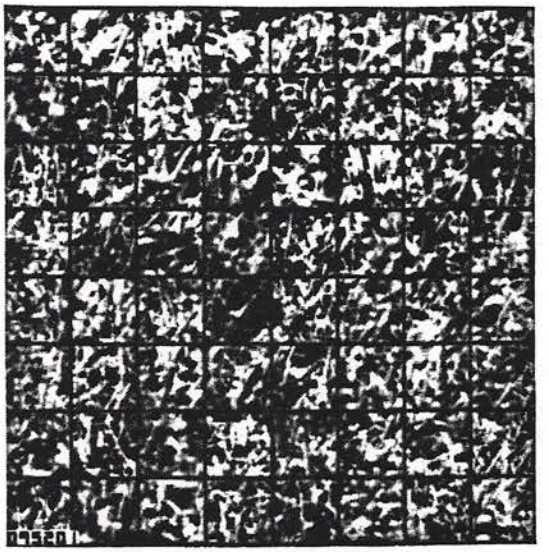
Taken Two at a Time

There are six possibilities. Tables 9-5 and 9-6 show the results.

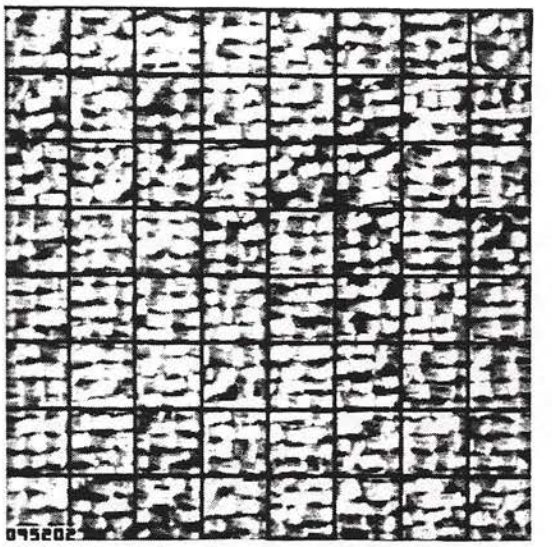
Taken One at a Time

There are four possibilities. Tables 9-7 and 9-8 tabulate the results.

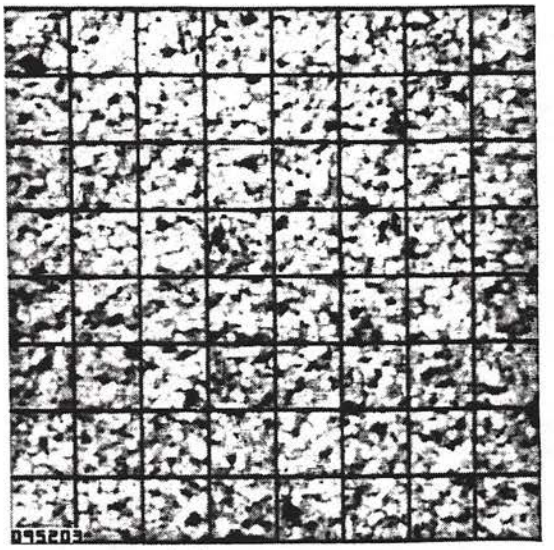
Observation 1: The same basic observations of Section 8.1.1 hold here too. For single feature case, in most texture pairs, M_1 gives the highest accuracy, followed by M_3, M_4, and M_2 respectively. This result agrees with those of artificial textures. The best double-



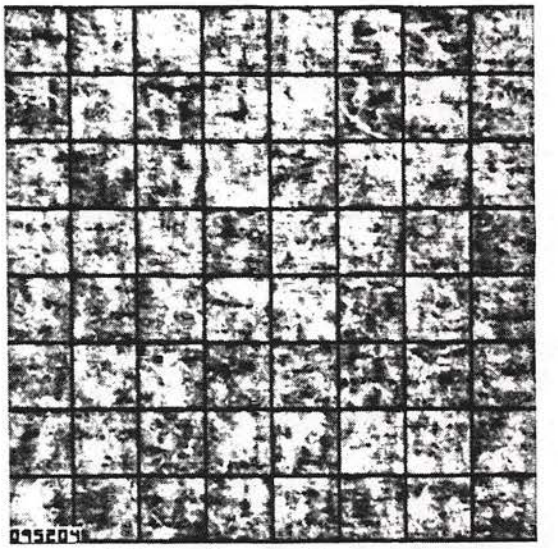
(a) grass samples



(b) raffia samples



(c) sand samples



(d) wool samples

Figure 9-3. 64 Non-overlapping 32x32 Samples Randomly Extracted.

Table 9-3. B-Distances for Various Combinatorial Possibilities of M_1, M_2, M_3, M_4 Taken Three at a Time for Natural Textures.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		M_1, M_2, M_3	M_1, M_2, M_4	M_1, M_3, M_4	M_2, M_3, M_4
G	R	2.03	1.86	2.36	2.42
G	S	1.14	1.19	1.28	1.37
G	W	2.20	2.48	1.62	2.78
R	S	5.52	6.07	5.85	7.00
R	W	8.14	8.85	7.24	10.71
S	W	3.83	3.86	3.07	3.46

Table 9-4. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		M_1, M_2, M_3		M_1, M_2, M_4		M_1, M_3, M_4		M_2, M_3, M_4	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
G	R	93.5	99.6	92.3	99.4	95.3	99.8	95.6	99.8
G	S	84.0	97.4	84.8	97.6	86.2	98.1	87.3	98.4
G	W	94.5	99.7	95.8	99.8	90.1	99.0	96.9	99.9
R	S	99.8	100.0	99.9	100.0	99.9	100.0	100.0	100.0
R	W	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	W	98.9	100.0	99.0	100.0	97.7	99.9	98.4	100.0

Table 9-5. B-Distances for M_1, M_2, M_3, M_4 Taken Two at a Time for Natural Textures.

NATURAL TEXTURE PAIRS		BHATTACHARYYA DISTANCE - 64 SAMPLES					
		M_1, M_2	M_1, M_3	M_1, M_4	M_2, M_3	M_2, M_4	M_3, M_4
G	R	1.38	1.21	0.84	1.69	1.12	2.03
G	S	0.99	0.67	0.66	1.12	1.16	1.03
G	W	2.01	1.08	1.38	1.92	2.24	1.47
R	S	5.18	5.27	5.61	3.75	3.65	3.86
R	W	7.79	6.75	6.96	7.28	7.49	6.48
S	W	2.93	1.68	2.69	2.92	3.07	2.98

Table 9-6. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES											
		M_1, M_2		M_1, M_3		M_1, M_4		M_2, M_3		M_2, M_4		M_3, M_4	
		LOW BD.	UP.BD	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
G	R	87.5	98.2	85.1	97.7	78.5	95.1	90.8	99.1	83.8	97.3	93.4	99.6
G	S	81.5	96.4	74.3	92.9	74.2	92.8	83.8	97.3	84.3	97.5	82.2	96.7
G	W	93.3	99.6	83.1	97.0	87.5	98.4	92.7	99.5	94.7	99.7	88.5	98.7
R	S	99.7	100.0	99.7	100.0	99.8	100.0	98.8	100.0	98.7	100.0	98.9	100.0
R	W	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
S	W	97.3	99.9	90.7	99.1	96.6	99.9	97.3	99.9	97.7	99.9	97.5	99.9

Table 9-7. B-Distances for M_1, M_2, M_3, M_4 Taken One at a Time for Natural Textures.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		M_1	M_2	M_3	M_4
G	R	0.65	0.20	0.89	0.52
G	S	0.48	0.89	0.32	0.58
G	W	0.91	1.82	0.95	1.18
R	S	4.96	3.45	3.02	2.56
R	W	6.22	7.05	5.40	4.33
S	W	0.81	2.83	1.40	2.17

Table 9-8. Classification Accuracy in %.

NATURAL TEXTURE PAIR		CLASSIFICATION ACCURACY - 64 SAMPLES							
		M_1		M_2		M_3		M_4	
		LOW BD.	UP.BD.	LOW BD.	UP. BD.	LOW BD.	UP.BD.	LOW BD.	UP. BD.
G	R	73.9	92.7	59.0	78.6	79.5	95.6	70.4	90.3
G	S	69.0	89.2	79.6	95.6	63.9	84.6	72.1	91.5
G	W	79.8	95.7	91.9	99.3	80.6	96.1	84.6	97.6
R	S	99.6	100.0	98.4	100.0	97.6	99.9	96.1	99.8
R	W	99.9	100.0	100.0	100.0	99.8	100.0	99.3	100.0
S	W	77.7	94.7	97.0	99.9	87.7	98.5	94.3	99.7

G = grass, R = raffia, S = sand, W = wool

feature is (M_1,M_2). The best triple-feature is (M_2,M_3,M_4) as in the artificial case.

$\underline{z}_3$ and $\underline{z}_5$, Normalized Length and Energy

$\underline{z}_3$ is the normalized length vector and $\underline{z}_5$ is the normalized energy vector. The same experiments as in $\underline{z}_2$ are performed on these vectors. The best results for single-feature combination are M_2 and M_1, the best double-feature is (M_1,M_2), the best triple-feature is (M_2,M_3,M_4), and the four-feature combination is the best of all in terms of the B-distance. Tables 9-9 through 9-12 show the results.

Overall Results

Vector feature $\underline{z}_3$ is better than $\underline{z}_5$ while both are less effective than $\underline{z}_2$ in terms of the B-distances and classification accuracies obtained.

$\underline{z}_1$ and $\underline{z}_4$

In these vector features, the kurtosis, M_4, is the best single feature for $\underline{z}_1$ and deviation, M_2, is the best one for $\underline{z}_4$. (M_3,M_4) is the best double-feature for $\underline{z}_1$ and (M_1,M_2) is the best one for $\underline{z}_4$. The best triple feature for both appear to be (M_2,M_3,M_4). Tables 9-13, 9-14, 9-13a, and 9-14b show the results.

Comparison

$\underline{z}_1$ is better than $\underline{z}_4$. In comparing the vector features, the order of effectivity among them is $\underline{z}_2$, $\underline{z}_1$, $\underline{z}_3$, $\underline{z}_5$, and $\underline{z}_4$. **Such a result shows that the normalized vectors perform better than the unnormalized ones. The best moment combinations are (M_1), (M_1,M_2),**

Table 9-9. B-Distances for $\underline{z}_3$ Using the Best Moment Combinations.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		(M_2)	(M_1, M_2)	(M_2, M_3, M_4)	(M_1, M_2, M_3, M_4)
G	R	0.38	1.24	2.51	2.70
G	S	0.75	0.99	1.46	1.50
G	W	1.32	2.01	2.29	3.93
R	S	4.86	5.30	7.46	8.26
R	W	7.05	7.57	8.30	12.65
S	W	1.42	2.71	3.25	3.65

Table 9-10. Classification Accuracies in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(M_2)		(M_1, M_2)		(M_2, M_3, M_4)		(M_1, M_2, M_3, M_4)	
		LOW.BD.	UP. BD.	LOW. BD.	UP. BD	LOW. BD.	UP. BD.	LOW. BD.	UP. BD.
G	R	65.8	86.4	85.5	97.9	95.9	99.8	96.7	99.9
G	S	76.4	94.0	81.4	96.4	88.4	98.6	89.5	98.9
G	W	86.7	98.1	93.3	99.5	94.9	99.7	99.0	100.0
R	S	99.6	100.0	99.8	100.0	100.0	100.0	100.0	100.0
R	W	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	W	87.9	98.5	96.7	99.9	98.1	100.0	98.7	100.0

Table 9-11. B- Distances for Z_5 Using the Best Moment Combinations

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		(M_1)	(M_1,M_2)	$(M_2,M_3,M_4$	(M_1,M_2,M_3,M_4)
G	R	0.81	0.97	2.24	2.34
G	S	0.07	0.48	1.09	1.13
G	W	0.01	2.42	2.47	2.75
R	S	1.87	5.06	5.90	5.98
R	W	0.94	8.08	8.16	9.02
S	W	0.07	3.01	3.81	3.84

Table 9-12. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(M_1)		(M_1,M_2)		(M_2,M_3,M_4)		(M_1,M_2,M_3,M_4)	
		LOW BD	UP BD	LOW BD	UP BD	LOW BD	UP BD	LOW BD	UP BD
G	R	77.8	94.8	81.0	96.3	94.7	99.7	95.2	99.8
G	S	53.4	68.1	69.1	89.3	83.2	97.1	83.8	97.3
G	W	50.5	57.0	95.6	99.8	95.8	99.8	96.8	99.9
R	S	92.3	99.4	99.7	100.0	99.9	100.0	99.9	100.0
R	W	80.5	96.0	100.0	100.0	100.0	100.0	100.0	100.0
S	W	53.4	68.1	97.5	99.9	98.9	100.0	98.9	100.0

Table 9-13. B-Distances for $\underline{z}_1$ Using the Best Moment Combinations.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		(M_4)	(M_3,M_4)	(M_2,M_3,M_4)	(M_1,M_2,M_3,M_4)
G	R	0.52	1.31	2.54	2.66
G	S	0.83	0.96	1.46	1.55
G	W	1.08	1.20	1.55	4.15
R	S	3.45	6.63	7.89	8.26
R	W	4.92	8.62	10.61	13.16
S	W	2.16	2.72	3.26	3.83

Table 9-14. Classification Accuracies in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(M_4)		(M_3,M_4)		(M_2,M_3,M_4)		(M_1,M_2,M_3,M_4)	
		LOW.BD.	UP.BD.	LOW.BD.	UP.BD.	LOW.BD.	UP.BD.	LOW.BD.	UP.BD
G	R	70.2	90.2	86.5	98.1	96.1	99.8	96.5	99.9
G	S	78.2	94.9	80.9	96.2	88.3	98.6	89.4	98.9
G	W	83.0	97.0	85.0	97.7	89.3	98.9	99.2	100.0
R	S	98.4	99.9	99.9	100.0	100.0	100.0	100.0	100.0
R	W	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	W	94.2	99.6	96.7	99.9	98.1	100.0	98.9	100.0

Table 9-13a. B-Distances for $\underline{z}_4$ Using the Best Moment Combinations.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		$(M_2$	(M_1, M_2)	(M_2, M_3, M_4)	(M_1, M_2, M_3, M_4)
G	R	0.55	0.65	1.95	2.29
G	S	0.47	0.60	0.91	0.98
G	W	0.68	2.86	2.83	3.30
R	S	4.97	6.19	5.81	7.63
R	W	6.83	10.07	8.42	11.24
S	W	0.22	2.91	3.56	3.84

Table 9-14a. Classification Accuracies in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(M_2)		(M_1, M_2)		(M_2, M_3, M_4)		(M_1, M_2, M_3, M_4)	
		LOW.BD.	UP.BD.	LOW.BD.	UP.BD.	LOW.BD.	UP.BD.	LOW.BD.	UP.BD.
G	R	71.2	90.8	73.9	92.6	92.9	99.5	94.9	99.7
G	S	68.7	89.0	72.7	91.9	79.9	95.8	81.2	96.3
G	W	74.6	93.1	97.1	99.9	97.1	99.9	98.2	100.0
R	S	99.7	100.0	99.9	100.0	99.9	100.0	100.0	100.0
R	W	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	W	59.8	79.8	97.3	99.9	98.6	100.0	98.9	100.0

(M_2,M_3,M_4) and (M_1,M_2,M_3,M_4) for each feature group. The four-feature gives the highest classification accuracy.

9.1.2 Scalar Features and Their Comparison

On the same textural images of grass, raffia, sand and wool, two sets of experiments are performed on the scalar features defined in Chapter 7. One, on z_7, z_{10},z_{11}, and z_{15}, which include entropies and largest singular value features. The second, on z_8, z_{12}, z_{13}, and z_{14}, which contains the energy features.

z_7,z_{10},z_{11},z_{15}

z_7 is the entropy of all singular values and is a measure of correlation in texture. z_{10} is the largest singular value. z_{11} is the largest component of the normalized singular value histogram and is less than one. z_{15} is the entropy of the largest SV.

Tables 9-15 through 9-21 tabulate the results of this set of experiments. The best single features for natural textures are determined by Tables 9-20 and 9-21 to be z_7 followed by z_{15}, z_{11}, and z_{10}. With a few exceptions, the best double-feature artificial case, is (z_7,z_{10}) and the best triple-feature is (z_7, z_{10}, z_{11}). All results agree with the artificial case.

z_8,z_{12},z_{13},z_{14}

z_8 is derived from z_2 and is given by equation (7-26). z_8 is the total energy in the normalized singular values histogram and is always less than one. z_{12} is the largest component of the normalized SV length and it, too, is always less than one. z_{13} is the energy in the largest normalized singular value, while z_{14} is the largest

Table 9-15. B-Distances and Accuracies Using $z_7, z_{10}, z_{11}, z_{15}$ Taken Four at a Time for Natural Textures

NATURAL TEXTURE PAIR		64 SAMPLES - $z_7, z_{10}, z_{11}, z_{15}$		
		B-DISTANCE	CLASSIFICATION ACCURACY	
			LOWER BOUND	UPPER BOUND
G	R	1.18	84.6	97.6
G	S	1.09	83.4	97.1
G	W	2.05	93.6	99.6
R	S	7.73	100.0	100.0
R	W	9.65	100.0	100.0
S	W	2.68	96.6	99.9

Table 9-16. B-Distances Using $z_7, z_{10}, z_{11}, z_{15}$ Taken Three at a Time for Natural Textures.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		z_7, z_{10}, z_{11}	z_7, z_{10}, z_{15}	z_7, z_{11}, z_{15}	z_{10}, z_{11}, z_{15}
G	R	0.83	0.81	1.08	0.92
G	S	0.82	0.79	0.88	0.65
G	W	2.02	2.00	1.67	0.30
R	S	6.69	6.84	5.32	3.02
R	W	9.32	9.24	7.06	1.30
S	W	2.32	2.14	2.67	0.78

Table 9-17. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		z_7, z_{10}, z_{11}		z_7, z_{10}, z_{15}		z_7, z_{11}, z_{15}		z_{10}, z_{11}, z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
G	R	78.4	95.1	77.9	94.9	83.2	97.1	80.2	95.9
G	S	78.1	95.0	77.5	94.7	79.3	95.5	74.1	92.8
G	W	93.4	99.6	93.3	99.5	90.7	99.1	63.0	83.7
R	S	99.9	100.0	99.9	100.0	99.8	100.0	97.6	99.9
R	W	100.0	100.0	100.0	100.0	100.0	100.0	86.4	98.1
S	W	95.1	99.8	94.2	99.7	96.6	99.9	77.3	94.5

Table 9-18. B-Distances Using $z_7, z_{10}, z_{11}, z_{15}$ Taken Two at a Time.

NATURAL TEXTURES PAIRS		BHATTACHARYYA DISTANCE - 64 SAMPLES					
		z_7, z_{10}	z_7, z_{11}	z_7, z_{15}	z_{10}, z_{11}	z_{10}, z_{15}	z_{11}, z_{15}
G	R	0.74	0.75	0.72	0.50	0.55	0.85
G	S	0.74	0.60	0.59	0.40	0.35	0.34
G	W	1.22	1.65	1.65	0.29	0.26	0.04
R	S	5.72	4.50	4.43	1.86	2.14	1.82
R	W	8.97	6.59	6.72	0.92	0.91	0.56
S	W	1.80	2.31	2.11	0.45	0.34	0.54

Table 9-19. Classification Accuracy in %.

NATURAL TEXTURES PAIRS		% CLASSIFICATION ACCURACY - 64 SAMPLES											
		z_7, z_{10}		z_7, z_{11}		z_7, z_{15}		z_{10}, z_{11}		z_{10}, z_{15}		z_{11}, z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
G	R	76.2	93.9	76.3	94.0	75.7	93.7	69.6	89.7	71.2	90.9	78.7	95.2
G	S	76.2	94.0	72.4	91.7	72.2	91.6	66.4	87.0	64.8	85.5	64.5	85.2
G	W	85.2	97.8	90.4	99.1	90.4	99.1	62.6	83.2	61.3	81.7	52.1	64.3
R	S	99.8	100.0	99.4	100.0	99.4	100.0	92.2	99.4	94.1	99.7	91.9	99.3
R	W	100.0	100.0	99.9	100.0	99.9	100.0	80.0	95.8	79.9	95.8	71.5	91.1
S	W	91.7	99.3	95.0	99.8	93.9	99.6	68.0	88.4	64.5	85.2	70.7	90.5

Table 9-20. B-Distances Using $z_7, z_{10}, z_{11}, z_{15}$ Taken One at a Time for Natural Textures.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		z_7	z_{10}	z_{11}	z_{15}
G	R	0.66	0.11	0.43	0.47
G	S	0.48	0.03	0.07	0.05
G	W	0.65	0.07	0.03	0.03
R	S	4.32	0.06	0.91	0.94
R	W	4.86	0.02	0.20	0.22
S	W	0.42	0.07	0.19	0.15

Table 9-21. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		z_7		z_{10}		z_{11}		z_{15}	
		LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
G	R	74.3	92.9	55.4	72.6	67.5	88.0	68.6	88.9
G	S	69.1	89.3	51.4	61.9	53.3	67.9	52.3	65.0
G	W	73.8	92.6	53.5	68.4	51.7	62.9	51.6	62.4
R	S	99.3	100.0	53.0	67.0	79.8	95.8	80.5	96.1
R	W	99.6	100.0	51.0	60.1	59.1	78.7	60.0	80.0
S	W	67.3	87.8	53.3	67.9	58.8	78.4	57.2	75.8

component of the energy ratio of SV's. z_{13} and z_{14} are also less than one. Being less than unity, this set of features assure algorithmically good behavior. Only the best combination of features will be shown in this part. From our experiment, the best single-feature is found to be z_8. The best double-feature is (z_8,z_{14}). The best triple feature is (z_8,z_{13},z_{14}) and the four feature produces the highest B-distance as expected. Tables 9-22 and 9-23 tabulate the best combinations.

9.1.3 Analysis of the Results

Figure 9-4 shows the mean-SV-prototype vectors computed for 64 sample windows extracted from raffia, grass, sand, and wool texture fields. From Figure 9-4, raffia and wool curves seem to be farthest apart suggesting that, infra structurally, these two textures are the most dissimilar--although visually they may not look that way. Raffia-wool highest dissimilarity is systematically verified in all of our experiments in this chapter. This pair always produce the highest B-distance comparing to other possible pairs.

The summary of results for feature selection from the SVD family of texture features are:

i) Entropy scalar feature set generally performs better than energy scalar feature set. Thus various combinations of z_7, z_{10}, z_{11}, z_{15} are usually better than those of z_8, z_{12}, z_{13}, z_{14}.

ii) Moment features perform better on natural textures than scalar ones.

iii) Moment features are computed from vector features among

Table 9-22. B-Distances for Z_8, Z_{12}, Z_{13}, Z_{14} Using the Best Scalar Combinations.

NATURAL TEXTURE PAIR		BHATTACHARYYA DISTANCE - 64 SAMPLES			
		(Z_8)	(Z_8,Z_{14})	(Z_8,Z_{13},Z_{14})	$(Z_8,Z_{12},Z_{13},Z_{14})$
G	R	0.85	0.89	1.35	1.46
G	S	0.27	0.38	0.77	0.95
G	W	0.14	1.20	1.36	1.98
R	S	2.90	3.54	4.74	5.24
R	W	2.13	5.04	5.76	6.46
S	W	0.09	1.31	1.86	2.77

Table 9-23. Classification Accuracy in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY - 64 SAMPLES							
		(Z_8)		(Z_8,Z_{14})		$(Z_8,Z_{13},Z_{14}$		$(Z_8,Z_{12},Z_{13},Z_{14})$	
		LOW BD	UP BD	LOW BD	UP BD	LOW BD	UP BD	LOW BD	UP BD
G	R	78.6	95.2	79.5	95.6	87.0	98.3	88.4	98.6
G	S	61.8	82.3	65.8	86.5	76.8	94.3	80.7	96.1
G	W	56.5	74.7	84.9	97.7	87.2	98.3	93.1	99.5
R	S	97.2	99.9	98.5	100.0	99.6	100.0	99.7	100.0
R	W	94.1	99.6	99.7	100.0	99.8	100.0	99.9	100.0
S	W	54.3	70.3	86.5	98.1	92.2	99.4	96.9	99.9

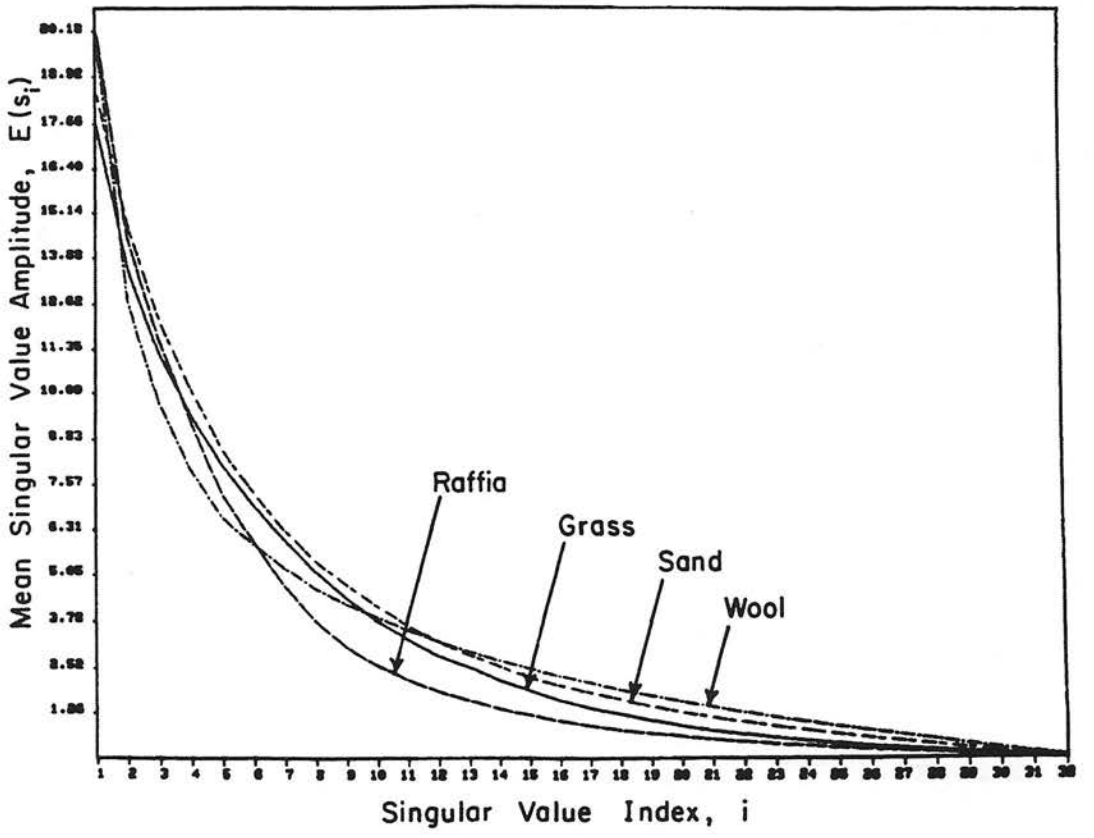


Figure 9-4. Mean Singular Value Prototype Vectors for Natural Textures.

which z_2 has the best performance.

iv) Subcombination of moments giving the best results are (M_1), (M_1,M_2), (M_2,M_3,M_4) and (M_1,M_2,M_3,M_4) with the last set being the most effective.

9.2 Classification of Natural Texture Against a Textural Background

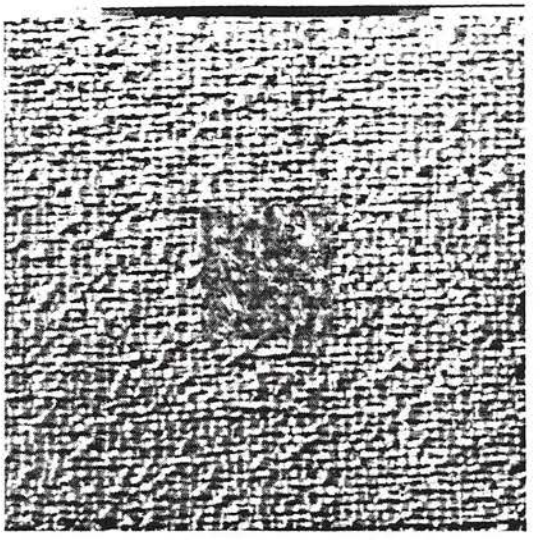
In this set of experiments, with the same scheme of Figure 8-6, 16 homogeneous sample windows of standardized and histogram Gaussianized wool are imbedded onto the raffia image and vice versa. Then, the B-distances are computed as in the artificial case. Figure 9-5 displays raffia and wool imbedded forms.

Notational Convention: Rf means raffia and Wl means wool. Rf/Wl or Raffia/Wool means 16 samples of wool are imbedded onto raffia background. Upon extracting sample windows, this embedment is equal to 75% raffia + 25% wool.

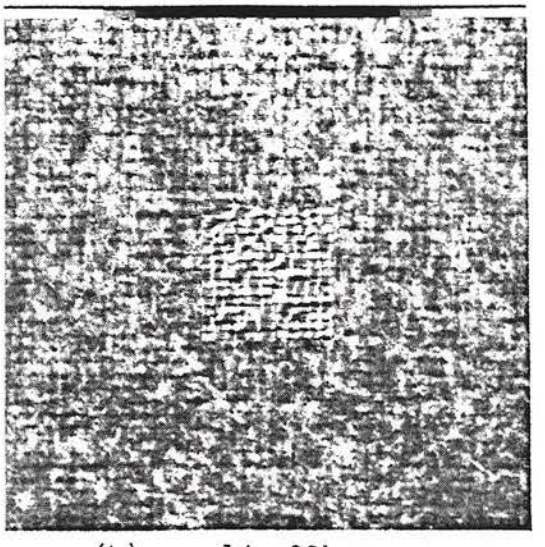
Tables 9-24 and 9-25 present B-distances and classification accuracies for various possible pairs of Rf, Wl, Rf/Wl, and Wl/Rf.

Remark 1: Each texture field is a real picture with zero mean, unit variance, and Gaussian histogram.

Remark 2: The B-distance values for (Rf,Wl) pairs are slightly different from those of Section 9.1. This is due to usage of a set of different sample windows. In fact, in this chapter the classifier is forced to select the 16 center windows and the other 48 are randomly chosen where in Section 9.1 all windows are randomly selected.



(a) raffia/wool



(b) wool/raffia

Figure 9-5. Embedded Versions Raffia/Wool and Wool/Raffia.

Table 9-24. B-Distances Between Various Combinations of Rf, Wl, Rf/Wl, and Wl/Rf Using Moment Features.

NATURAL TEXTURE PAIR		B-DISTANCE - 64 SAMPLES			
		(M_1)	(M_1, M_2)	(M_2, M_3, M_4)	(M_1, M_2, M_3, M_4)
Rf	Wl	6.65	8.34	11.380	11.8747
Rf	Rf/Wl	0.43	0.44	0.695	0.7128
Rf	Wl/Rf	0.98	1.01	1.433	1.6211
Wl	Rf/Wl	0.82	1.07	1.694	1.8350
Wl	Wl/Rf	0.33	0.50	0.951	0.9625
Rf/Wl	Wl/Rf	0.14	0.16	0.322	0.5033

Table 9-25. Classification Accuracies in %.

NATURAL TEXTURE PAIR		% CLASSIFICATION ACCURACY							
		(M_1)		(M_1, M_2)		(M_2, M_3, M_4)		(M_1, M_2, M_3, M_4)	
		LOW BD.	UP. BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.	LOW BD.	UP.BD.
Rf	Wl	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Rf	Rf/Wl	67.6	88.1	67.8	88.3	75.1	93.3	75.5	93.6
Rf	Wl/Rf	81.3	96.4	81.8	96.6	88.1	98.6	90.1	99.0
Wl	Rf/Wl	77.9	94.8	82.8	96.9	90.8	99.1	92.0	99.4
Wl	Wl/Rf	64.0	84.7	69.7	89.7	80.7	96.1	80.9	96.2
Rf/Wl	Wl/Rf	56.4	74.4	57.6	76.5	63.8	84.4	69.8	89.8

Analysis

Figure 9-6 shows the mean-singular value-prototype-vectors for raffia-wool textures. For natural textures, using four moment features, Table 9-24 displays that B-distance between raffia and 100% wool is 11.8747. The distance drops sharply to 1.6211 when 75% wool + 25% raffia is employed instead of 100% wool. The foreign texture (i.e. raffia) introduced onto wool causes such a drop in the B-distance. This is also clearly an indication that singular value decomposition can be used as a means of classification of texture mixtures.

9.3 A Comparative Study

From all experiments performed in this chapter, one important table can be drawn: The four-feature combinations as well as the best single, double, and triple feature combinations for M_1, M_2, M_3, M_4 and z_7, z_{10}, z_{11}, z_{15} features on natural texture are shown in Table 9-26.

From B-distance analysis, it has been determined that the most effective single features for moments are M_1, M_3, M_4 and M_2, respectively. Hence M_1 is the best single-feature. The best double-features have been determined to be (M_1,M_2), while the best triple-feature is (M_2,M_3,M_4). For scalars, z_7, z_{15}, z_{11}, and z_{10} are respectively effective among single features. Hence, the best single-feature is z_7. The best double-feature is experimentally shown to be (z_7,z_{10}) and the best triple-feature is (z_7,z_{10},z_{11}).

In Table 9-26, there are two columns for evaluation of the features in each group. One being the worst case and the other being

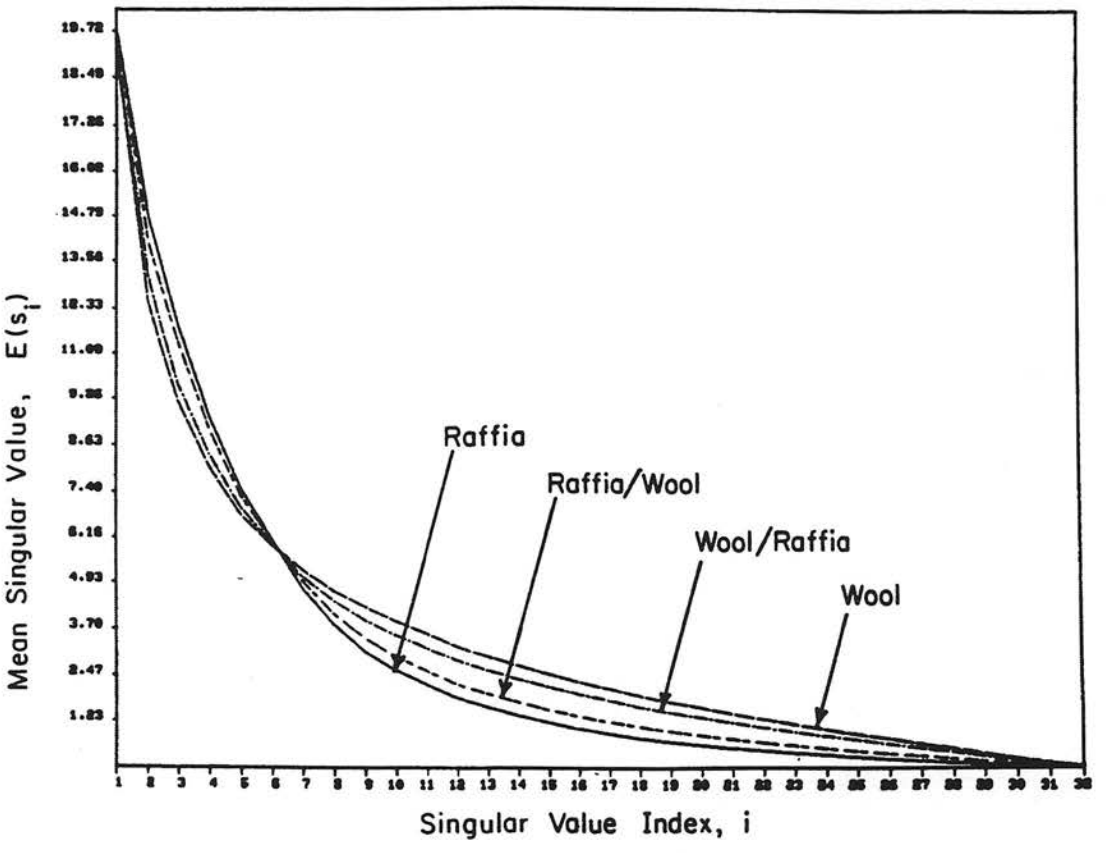


Figure 9-6. Mean SV Prorotype Vector for Natural Embedded Textures.

Table 9-26. Comparison of the Best Features Applied on Natural Texture Images.

FEATURE	MOMENTS			SCALARS		
	BEST FEATURE	% ACCURACY NATURAL TEXTURE		BEST FEATURE	% ACCURACY NATURAL TEXTURE	
		WORST CASE	AVERAGE		WORST CASE	AVERAGE
Single	M_1	69.0	89.4	z_7	69.1	87.2
Double	M_1, M_2	81.5	96.2	z_7, z_{10}	76.2	92.8
Triple	M_2, M_3, M_4	87.3	98.0	z_7, z_{10}, z_{11}	78.1	94.5
Quadruple	M_1, M_2, M_3, M_4	88.0	98.3	$z_7, z_{10}, z_{11}, z_{12}$	83.4	96.0

the average in classification accuracy. The "worst case," in our experiments, is taken as the lower bound to the probability of classification accuracy for the most similar textures among all natural textures under experiments. The "worst case" does not provide a fair judgement on the strength of features. So we have defined a value for an overall comparison, and that is the "average" to the probability of classification accuracy. Such averge values are determined by considering the mid-values between lower and upper bound to probability of classification accuracies for all pairs and multiplying them by the weight 1/6 (for 6 combinatorial possibilities among four textures) then summing them. Hence, average classification accuracy = $\sum_{i=1}^{6} \frac{1}{6} (\text{mid-value})_i$ for i being an index for each possible pair. The comparative classification accuracies are presented in Table 9-26. As an example, in Table 9-26, the numbers 88.0% and 98.3% are obtained from Table 9-2 in the manner which was defined above.

Using Bayesian probability of error, it is desired to compare SVD family of features with other methods in texture processing in order to evaluate its performance in terms of accuracy and cost.

Haralick Texture Measures

One of the well established texture feature algorithms has been developed by Haralick [51-52]. The reader can refer to Chapter 2, Section 2.3, for a brief review of this technique, which is based on gray-level co-occurrrance matrices of images in various angular directions. Haralick, et al. performed essentially three sets of experiments on textures [51]: (a) on photomicrographs of sandstones,

which produced an average classification accuracy of 89%, (b) on aerial photograph pictures, which produced an average of 82.3% classification accuracy, and (c) on multispectral scanner data derived from satellite imagery, which produced an average of 83.5% of classifier accuracy, which was a substantial improvement on 74-77% accuracy using, a now obsolete, spectral texture classification algoritm. Other experiments with Haralick measures have been performed [48], and have produced relatively similar classification results. In terms of computationality, this form of classification is expensive because, typically, it uses eleven features, each computed for four directions, thus giving rise to a 44-dimensional feature vector to be used in classification. Computing time for each of the 44 components of every feature vector by itself can be enormous. SVD features, on the other hand, are simple and require much less computation. To be quantitative, using a 4-dimensional vector of moment features, an average classification accuracy of 98.3% is obtained for an input of four natural textures grass, raffia, sand, wool using 18.62 minutes of CPU on a DEC system computer

It is imperative to mention that although the computing systems in both Haralick's and our experiments are DEC and the input data are four images at a time on both cases, relying on above experiments, it is difficult to objectively compare the two techniques. This is because the input textural images are not identical.

Laws Segmentation Algorithm

Contrary to the previous case, it is possible to, more

objectively, evaluate SVD features performance in comparison with a recent texture segmentation procedure suggested by Laws [146]. Four of the natural textures used in [146] are grass, raffia, sand, and wool, and are the same as, in fact identical to, the ones used in our SVD experiments.

Using the predicted class confusion percent values among the four natural textures [Ibid, page 148], and subtracting each from 100.0, we obtain a predicted pairwise classification accuracy. Noting that in Laws' class confusion table, two textures have different class confusion values if their pairwise order are changed (e.g., grass-raffia, and raffia-grass), an average value of the two is computed. All values obtained are compared against the SVD four moments taken all at the same time given by Table 9-2 in this document. In Table 9-2, the lower bounds to the probability of classification accuracy for all texture pair are slightly less than Laws' average predicted value, while the upper bound is slightly higher. Since the true classification accuracy lies somewhere between the lower and upper bounds, an average is computed between them and the results are compared in Table 9-27.

For most texture pairs, SVD has higher values. The overall results are close, while an average of 97.78% predicted accuracy are recorded for Laws, SVD gives an average of 98.33%, which is slightly over half a percent more than Laws. Computation time difference is more notable. For eight natural textures, Laws uses 48.05 minutes of CPU to achieve classification while SVD uses approximately 4 sec per 32x32 sample windows to extract and compute 32 singular values; and

Table 9-27. Comparison of Average SVD and Average of Laws Classification Values and Computation Time.

NATURAL TEXTURE PAIR		SVD AVERAGE CLASSIFICATION %		LAWS AVERAGE PREDICTED CLASSIFICATION %
Grass	Raffia	97.8		99.4
Grass	Sand	93.3		92.8
Grass	Wool	99.4		99.7
Raffia	Sand	100.0		98.1
Raffia	Wool	100.0		100.0
Sand	Wool	99.5		96.7
OVERALL RESULT		98.33		97.78
TOTAL COMPUTATION TIME (CPU MINUTES)		EIGHT SEPARATE	ONE COMPOSITE	ONE COMPOSITE
		34.68	17.61	48.05

subsequently producing the four moments M_1, M_2, M_3, M_4. 1.54 minutes are used for all B-distance computations. We have used 64 sample windows taken from four textures in our experiments with 18.62 minutes of CPU time to calculate all combinations of features. Using 8 natural textures and extracting 64 samples from each, and calculating only the necessary features will take 34.68 minutes of CPU. The same procedure can be applied on one composite 512x512 picture in which 128x128 texture windows are placed beside each other at the top half and reducing the texture windows to 32x32 and 16x16 at the bottom half. In such a case, the SVD sample windows can be reduced to 16x16. There are 1024 16x16 non-overlapping windows. Each one will take about one second of CPU for singular value computations to which the B-distnace computation is added resulting in a total of 17.61 minutes of CPU. In both cases, as Table 9-27 shows, computation times are substantially less than Laws in order to give the same order of classification accuracy. SVD features have advantages of flexibility in both the number of samples and number of features. For example, only one feature, M_1, can be used, but average classification accuracy will decrease about 10%. Both Laws and SVD techniques have determined that among grass, raffia, sand and wool, raffia-wool are the easiest to classify, and grass-sand are the most difficult. Also, grass-wool are more easily separated from each other than grass-raffia while the same holds for raffia-sand and sand-wool, respectively. Although such agreements on results hold in most cases between Laws algorithm and

sand-wool, where Laws' predicted values indicate that the former must classify easier than the latter, while ours show the opposite.

9.4 Summary

The objective of this chapter has been to determine the best combination of SVD features for natural texture classification and segmentation. It is established that moment features derived from vector $\underline{z_2}$ generally perform better on natural textures than scalar features. On the other hand, the opposite is generally true for artificial textures. Adding scalar features, such as z_7 (entropy) to z_6 to make it a 5-dimensional vector improves its sensitivity, but the improvement is so slight that it does not justify the cost.

The best single, double and triple combination of feature as well as all four together are compared in Table 9-26. Section 9.3 presents a comparison of SVD features with those of Haralick's and Laws and the results are promising.

Section 9.2 introdues a technique for detection of foreign texture in a textured image and classification of texture mixtures. An application can be named in medicine where unhealthy tissues demonstrate different textural properties than healthy ones. In food processing, an unwanted substance with a different texture can be found using the SVD technique.

CHAPTER 10

CONCLUSIONS

Often times the engineer/scientist of today has to deal with multi-parameter and large-scale systems and data. In many instances, the data is stochastic in nature, which makes it even more toilsome to analyze. In this document, such a type of problem has been encountered and analyzed, and the solution to part of it has been presented.

This dissertation has studied the problem of stochastic singular value decomposition for texture processing. The material has been divided into four parts. Part one has surveyed various studies, techniques and approaches to texture analysis. Part two has presented the theoretical developments on the topic. Part three has included the experiments and applications, and part four contains the glossary, appendices and references.

Efforts have been made to systematize the theoretical derivations. Thus, they can be shown to be members of one family of derivations, and that they lead into each other. Most of the probability functions are given in terms of transcendental functions of zonal harmonics of covariance matrices.

It would be possible to have textures with similar appearances but different 3rd or higher order statistics. It was plausible to ask whether this were true for appearance of texture and its singular value distribution. The answer to such a question was not evident at the beginning. We are at a position now to assert that similar looking textures have similar singular value distributions. The

appearance of a non-structural stochastic image texture is closely related to its 2nd order statistic. This fact distinguishes the set of singular values as a powerful measurement of correlation.

In summary, stochastic SVD provides an answer to complex behavior of a lattice of correlated random variables.

10.1 Applications

The application of this work can be discussed in two contexts: mathematical and engineering. For mathematical applications, this work has investigated the problem of stochastic perturbation, and, as a result, the p.d.f. of a random matrix and the joint p.d.f. of its singular values has been derived. Probability distribution function of the largest singular values as well as various applications in solving multivariate statistical problems have been explored. Any problem which involves numerous correlated parameters in lattice form can be handled or approached using the present technique. For example, the problem of round-off noise in matrix computations where many interdependent parameters are involved can be solved through the derivations of this work.

For engineering applications, the basic utilization is the SVD's ability to reduce the degrees of freedom in image feature extraction from k^2 to k, and, as shown in the application part, in many cases, k degrees of freedom can further be reduced, reaching in some instances to one. Such enormous dimensionality reduction is important in terms of speed and cost of classifying images. Stochastic SVD also has the ability to substantially save bandwidth in image transmission. Oftentimes, standard edge detection and thresholding do not perform

well on non-structural patterns. In such cases, identifying and locating a textural pattern within another texture is much more difficult than locating a well bounded object in a background. The technique of sections 8.2 and 9.2 can provide a solution to the problem of texture detection within another texture. In food processing, for example, an unwanted substance having different textural properties can be readily detected using such techniques. SVD texture feature extraction can be applied on infrared multispectral scanner data in remote sensing and ultrasonic textural patterns as well. In another application, a bank of singular value features can be stored on memory for the textural patterns which are obtained through identical conditions. Then, the SV features of an unknown texture obtained under similar conditions can be correlation matched with the data in the bank, and a hint on the texture's identification can be designated. In medical applications, unhealthy tissues which usually exhibit different textural properties than healthy ones can be distinguished and recognized through the application of the SVD feature extraction technique.

SVD technique can be implemented on microprocessors and a real-time image feature extraction can be established. The only restriction on engineering application of SVD feature extraction technique is that it performs accurately on textural, but not on structural patterns. Therefore, applying it on pictures of man-made objects shall be avoided.

10.2 Further Research

Research can be continued on this topic on two grounds: theoretical and practical.

On the theoretical ground, one can extend the derivations of this dissertation to various combinations of non-central, complex, time variant, infinite dimensional with finite rank, and non-normal cases. The same analysis can be extended to discrete probability functional cases. The analytical calculation of moments of individual singular values and the singular value covariance matrix can be researched further in the context of applied mathematics. Speeding up the convergence of zonal polynomials needs further investigation.

On the practical ground, a high correlation texture field produces a steep singular value curve. At the same instance, the field possesses a smooth autocorrelation function and low power at the high frequencies of its power spectrum. On the other hand, a low correlation texture field has flat singular value curves, while exhibiting an abrupt autocorrelation function and high power in the high frequencies of its power spectrum. Hence, there must be a relation between singular values of an image field and its power spectrum behavior. Further research on SVD, in such contexts, is worth performing.

PART FOUR: REFERENCE MATTER

This part provides the necessary supplements, support and compliments for the disseration. It includes the glossary of symbols and abbreviations, seven appendices A-G, and the bibliography.

GLOSSARY OF SYMBOLS AND ABBREVIATIONS

Symbol	Description
SVD	Acronym for Singular Value Decomposition
(*)	Symbol of footnoting in this document
$\underline{m}$	White Random Process m
$\underline{M}$	White Random Matrix M
$\underline{f}$	Vector f (Lexicographic vector of $\underline{F}$)
$\underline{F}$	Matrix F (Separable Correlated Zero Mean Texture Field)
$\underline{F} \rightarrow \underline{f}$	Lexicographic Transformation of Matrix $\underline{F}$ to vector $\underline{f}$
$\underline{d}(\underline{F})$	Matrix of Differential Element of Matrix
dF	Differential Value Element on $\underline{F}$.
$\underline{A}^T$	Transpose of Matrix $\underline{A}$
$\sum_{i=1}^{n} a_i$	$a_1+a_2+\ldots+a_n$ (Summation)
$\prod_{i=1}^{n} a_i$	$a_1 \cdot a_2 \cdots a_n$ (Product)
π	Constant = 3.141592653589793
g(.)	Probability Density Function of (.)
G(.)	Distribution Function of (.)
$\lvert . \rvert$	Determinant of Square Matrix [•]
$\triangleq$	Definition Symbol
$\equiv$	Identity Symbol
$\otimes$	Direct or Kronecker Product
$\underline{K}_C$	Covariance Matrix Along the Columns
$\underline{K}_R$	Covariance Matrix Along the Rows

$\underline{\mu}_f$	Mean of Random Process $\underline{f}$
$\underline{K}_f$	Covariance of Random Process $\underline{f}$
Var(.)	Variance of the Random Variable (.)
Cov(.)	Covariance Matrix of Variables (.)
$\perp\!\!\!\perp$	Statistical Independence Symbol
abs(.)	Absolute Value of (.)
$[a_{ij}]$	Matrix Whose $(ij)^{th}$ element is a_{ij}
C_{eu}	Euler Constant = 0.5772156649
$diag(a_1,\ldots,a_k)$	Diagonal Matrix of $(a_1,\ldots,a_k)$
s_i	i^{th} Singular Value
λ_i	i^{th} Eigenvalue
$\underline{S}$	Diagonal Matrix of Singular Values
$\underline{\Lambda}$	Diagonal Matrix of Eigenvalues
[A|B]	Partitioned Matrix
$J_{\underline{b}}(\underline{a})$	Jacobian of Transformation $\underline{b} = T(\underline{a})$
$\Gamma(x)$	Gamma Function of Real Number x
n!	Factorial of n = 1x2x...xn
log	Logarithm in Base 10
Ln	Logarithm in Base e (Neperian)
∞	Infinity Symbol
O(k)	Orthogonal Group of Dimention k
$V_{k,n}$	Stiefel Manifold of k-frame in Euclidean Space R^n
$\sim$	Symbol of Being Unnormalized
ε	Symbol of Membership in a Group or Set
$\approx$	Approximately Equal

Ibid	Ibidem (a Latin word) "in the same place," used for Referencing
lim	Limit
p.d.f.	Pobability Density Function
l.h.s.	Left Hand Side
r.h.s.	Right Hand Side
w.r.t.	With Respect To

APPENDIX A

MATRIX DIFFERENTIAL FORMS

The results of the following six derivations have been used in the theoretical part of this dissertation. Derivation A.1 is the essence for Derivations A.2, A.3 and A.3-1 and has been proved in [68] by the use of exterior products and Weierstrass Theorem of determinant. The anticommutative property of exterior products has been established in [69, Chapter 3]. In this appendix, with a different method, formal proofs will be given for all six. Derivation A.1 will be proved by use of the standard Jacobian concept.

Derivation A.1:Considering $\underline{z}$ and $\underline{y}$, two column vectors of dimension k, and let $\underline{z} = \underline{A}\,\underline{y}$. Therefore, if $\underline{dz}$ and $\underline{dy}$ be two corresponding column vectors of k differential forms, we have

$$\underline{d}(\underline{z}) = \underline{A}\,\underline{d}(\underline{y}). \qquad \text{(A-1)}$$

Then, the product of the components of $\underline{dz}$ will be equal to the product of the components of $\underline{dy}$ multiplied by (absolute value of) determinant of $\underline{A}$, i.e.

$$\prod_{i=1}^{k} dz_i = \text{abs}(|\underline{A}|) \prod_{i=1}^{k} dy_i \qquad \text{(A-2)}$$

Proof: It is known from calculus that the volume of the differential hyper-parallelepiped of dimension k is equal to the volume of differential hyper-parallelepiped of its transformation multiplied by the absolute value of Jacobian of the transformation.

The volume of the differential hyper-parallelepiped of $\underline{z}$ is equal to $C_z \prod_{i=1}^{k} dz_i$ and for $\underline{y}$, it is $C_y \prod_{i=1}^{k} dy_i$, where C_z and C_y are Jacobians of the representation of z and y w.r.t. a set of orthonormal basis vectors. Thus,

$$\prod_{i=1}^{k} dz_i = \left(J_{\underline{z}}(\underline{y})\right) \prod_{i=1}^{k} dy_i \qquad \text{(A-3)}$$

Hence, for the transformation $\underline{z} = \underline{A}\,\underline{y}$, the differential volume elements are related by a Jacobian in the form of (A-3). Considering the linear relation $\underline{z} = \underline{A}\,\underline{y}$, it can be seen that the i^{th} component of $d\underline{z}$ is a linear combination of all components of $\underline{y}$ as

$$z_i = a_{i1}y_1 + a_{i2}y_2 + \ldots + a_{ik}y_k \qquad \text{(A-4)}$$

Using the definition of Jacobian, and considering the r.h.s. of (A-4) to be equitable to a function $f_i(y_1, y_2, \ldots, y_k)$, the Jacobian of the transformation (A-1) will be equal to

$$J_{\underline{z}}(\underline{y}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix} \qquad \text{(A-5)}$$

Using (A-5) and (A-3), we obtain (A-2) and hence the proof. **Q.E.D.**

Remark: Using definition (3-25), relation (A-2) can be written as

$$dz = \text{abs}(|\underline{A}|)dy \qquad \text{(A-6)}$$

Summary:

If $\underline{z} = \underline{A}\underline{y}$

Then $\underline{d}(\underline{z}) = \underline{A}\,\underline{d}(\underline{y})$

and, $dz = |A|dy$

Derivation A.2: Let $\underline{Z}$ be a kxn, $\underline{A}$ a kxk, and $\underline{Y}$ a kxn matrix such that $\underline{Z} = \underline{A}\underline{Y}$. Differentiating and using definition (3-28), we will have

$$\underline{d}(\underline{Z}) = \underline{A}\,\underline{d}(\underline{Y}) \qquad \text{(A-7)}$$

Then,

$$dZ = \text{abs}(|\underline{A}|^n)dY \qquad \text{(A-8)}$$

when dZ and dY are defined in (3-25).

Proof: The differential volume element of each column of $\underline{Z}$ is related to the corresponding column of $\underline{Y}$ by equation (A-2) proved in Derivation A.1. There are n columns in $\underline{Z}$ and $\underline{Y}$; thus, the total differential volume element of $\underline{d}(\underline{Z})$ will be equal to $\prod_{i,j}^{k,n} dZ_{ij}$ which itself is equal to $|\underline{A}|^n$ multiplied to the total volume element $dY = \prod_{i,j}^{k,n} dY_{ij}$ and hence (A-8). **Q.E.D.**

Derivation A.3: Let $\underline{W}$ be a kxn matrix, $\underline{Z}$ a kxn matrix and $\underline{B}$ a nxn matrix such that $\underline{W} = \underline{Z}\,\underline{B}$. Upon differentiation of matrices $\underline{W}$ and

$\underline{Z}$ we will have

$$\underline{d(W)} = \underline{d(Z)B} \qquad \text{(A-9)}$$

Then,

$$dW = |\underline{B}|^{k} dZ \qquad \text{(A-10)}$$

where, $dW = \prod_{i,j}^{k,n} d\underline{W}_{ij}$ and $dZ = \prod_{i,j}^{k,n} dZ_{ij}$ as defined in (3-25).

Proof: Matrix $\underline{B}$ is postmultiplied to $\underline{d(Z)}$,, it therefore operates on the rows of $d(\underline{Z})$ to produce $\underline{d(W)}$. The volume element for each row of $d(\underline{W})$ is equal to $|B|$ times the volume elements of each row of $d(\underline{Z})$. There are k rows in $\underline{d(W)}$ and $\underline{d(Z)}$; thus, the proof for (A-10).

Q.E.D.

Derivation A.3-1: Using $\underline{Y}$, $\underline{Z}$, and $\underline{W}$ of Derivations A.2 and A.3, we have

$$\underline{W} = \underline{A}\,\underline{Y}\,\underline{B}$$

upon differentiation, we will have

$$\underline{d(W)} = \underline{A}\,\underline{d(Y)B} \qquad \text{(A-11)}$$

Then,

$$dW = \text{abs}(|A|^{n}|B|^{k})dY \qquad \text{(A-12)}$$

where $\underline{W}$ and $\underline{Y}$ are $\underline{kxn}$ matrices, $\underline{A}$ is a $\underline{kxk}$ and $\underline{B}$ is a $\underline{nxn}$ matrix. dW and dY are according to definition (3-25), i.e., $dW = \prod_{i,j}^{k,n} dW_{ij}$ and $dY = \prod_{i,j}^{k,n} dY_{ij}$.

Remark: dW and dY are the differential volume element of $\underline{W}$ and $\underline{Y}$.

Proof: Using relation (A-10) which was proved in Derivation A.3, we have $dW = abs(|\underline{B}|^k)dZ$. But, from (A-8) which was proved in Derivation A.2, $dZ = abs(|A|^n)dY$. Subsequently, using dZ of the latter in the former, relation (A-12) will be obtained. **Q.E.D.**

Derivation A.4: If two column vectors $\underline{a}$ and $\underline{b}$ are orthogonal, i.e.

$$\underline{a}^T\underline{b} = 0 \qquad \text{(A-13)}$$

Then, their linear differential form is anti-commutative, i.e.

$$\underline{a}^T\underline{d}(\underline{b}) = -\underline{b}^T\underline{d}(\underline{a}) \qquad \text{(A-14)}$$

Elaboration: Assuming $\underline{a}$ and $\underline{b}$ are of dimension k and using definition (3-27),

$$\underline{d}(\underline{b}) = \begin{bmatrix} db_1 \\ db_2 \\ \vdots \\ db_k \end{bmatrix}, \quad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}, \text{ and } \underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

Then, (A-14) implies that

$$a_1db_1 + a_2db_2 +\ldots+ a_kdb_k = -(b_1da_1 + b_2da_2 +\ldots+ b_kda_k) \quad \text{(A-15)}$$

Proof:

Since

$$\underline{a}^T\underline{b} = 0$$

then,

$$d(\underline{a}^T\underline{b}) = 0 \quad \text{(A-16)}$$

but,

$$d(\underline{a}^T\underline{b}) = \underline{d}(\underline{a}^T)\underline{b} + \underline{a}^T\underline{d}(\underline{b}) \quad \text{(A-17)}$$

and,

$$\underline{d}(\underline{a}^T)\underline{b} = \underline{b}^T\underline{d}(\underline{a}) \quad \text{(A-18)}$$

Remark: (A-18) is true because both sides of it are equivalent to $b_1da_1 + b_2da_2 +\ldots+ b_kda_k$. Therefore, using r.h.s. of (A-18) in (A-17) and equating the result with zero, one will obtain (A-14).

Q.E.D.

Derivation A.5: If a vector $\underline{a}$ has the constant magnitude, then its linear differential form is zero, i.e.

$$\underline{a}^T\underline{d}(\underline{a}) = 0 \quad \text{(A-19)}$$

Elaboration: assuming the same definition for $\underline{a}$ as in Derivation A.4, (A-19) means,

$$a_1 da_1 + a_2 da_2 + \ldots + a_k da_k = 0 \qquad \text{(A-20)}$$

Proof:

Let

$$\underline{a}^T \underline{a} = C \qquad \text{(A-21)}$$

Then,

$$d(\underline{a}^T \underline{a}) = 0 \qquad \text{(A-22)}$$

But

$$d(\underline{a}^T \underline{a}) = \underline{d}(a^T)\underline{a} + a^T \underline{d}(\underline{a}) \qquad \text{(A-23)}$$

Knowing that

$$\underline{d}(a^T)\underline{a} = \underline{a}^T \underline{d}(\underline{a}) \qquad \text{(A-24)}$$

Using (A-24) in (A-23) and equating it with zero,

$$\underline{a}^T \underline{d}(\underline{a}) + a^T \underline{d}(\underline{a}) = 2\underline{a}^T \underline{d}(\underline{a}) \qquad \text{(A-25)}$$

$$= 0$$

Hence the proof. **Q.E.D.**

Derivation A.6: Each $\underline{U}^T d(\underline{U})$ and $\underline{V}^T d(\underline{V})$ is a skew symmetric matrix, where $\underline{U}$ and $\underline{V}$ are the real version of the matrices defined in (3-1). ($\underline{U}$ and $\underline{V}$ are orthogonal matrices rather than unitary.)

Proof: For $\underline{U}^T \underline{d}(\underline{U})$, we proceed with the proof as

$$\underline{U}^T \underline{U} = \underline{I}_k \qquad \text{(A-26)}$$

Applying differential operations on (A-26),

$$\underline{d}(\underline{U}^T)\underline{U} + \underline{U}^T\underline{d}(\underline{U}) = 0 \qquad \text{(A-27)}$$

Note that $\underline{U}$ is the variable and $\underline{I}_k$ at the r.h.s. of (A-26) is constant. Therefore, from (A-27), we have

$$\begin{aligned} \underline{d}(\underline{U}^T)\underline{U} &= -\underline{U}^T\underline{d}(\underline{U}) \\ &= -[\underline{d}(\underline{U}^T)U]^T \end{aligned} \qquad \text{(A-28)}$$

and hence the proof. The proof for $\underline{V}^T d(\underline{V})$ will be the same. **Q.E.D.**

Remark: Derivation A.6 can easily be extended to unitary matrices over the complex fields.

Selected Reference: [68].

APPENDIX B

THE MULTIVARIATE GAMMA FUNCTION

The Single-Variate Gamma Function:

Single-variate gamma function was developed as a generalization of the factorial function of the natural numbers, that is, n! for positive integers n. The gamma function is, in essence, the extension of the above problem to x! for arbitrary real numbers x.

For n! where n is a positive integer, we come upon the following imporper integral discovered by Euler:

$$\int_0^\infty e^{-t} t^{n-1} dt = (n-1)! \qquad \text{(B-1)}$$

Obviously, replacing n by (n+1) on both sides of (B-1) will give us the relation for n!.

The equation (B-1) is a very interesting one! Variable t on the l.h.s. of (B-1) is continuous and $e^{-t}t^{n-1}$ is therefore, also, continuous. But, when integrated, it is mapped to a set of discrete integer numbers depending on the value of n.

The gamma function is subsequently defined as follows

$$\Gamma(x) \triangleq \int_0^\infty e^{-t} t^{x-1} dt \qquad \text{(B-2)}$$

where the only difference between (B-1) and (B-2) is that positive integer n of the former has been replaced by an arbitrary real number x. It can be proved that r.h.s. of (B-2) converges for all positive real x.

From the last 2 relations,

$$\Gamma(n+1) = n! \tag{B-3}$$

Properties of the Gamma Function:

$\Gamma(x)$ is defined for all real numbers, with the exception of 0 and the negative integers (B-4)

$\Gamma(x)$ is continuous and differentiable where it is defined (B-5)

$$\Gamma(1) \triangleq 1 \tag{B-6}$$

$$\Gamma(x+1) = x\Gamma(x) \tag{B-7}$$

$\Gamma(x+n) = (x+n-1)(x+n-2)\ldots(x+1)x\Gamma(x)$ for every positive integer n (B-8)

$$\frac{\Gamma(x+n)}{\Gamma(x)} = (x)(x+1)(x+2)\ldots(x+n-1) \triangleq (x)_n \tag{B-9}$$

Note that in (B-9) $(x)_0 \triangleq 1$

$$2^{(2x-1)}\Gamma(x)\Gamma(x+\tfrac{1}{2}) = \sqrt{\pi}\Gamma(2x) \tag{B-10}$$

where, (B-10) is called the gamma duplication formula.

Remark: (B-9) can be utilized to give the value of gamma function for negative real arguments.

$$\Gamma(x) > 0 \quad \text{for} \quad x > 0 \tag{B-11}$$

$\Gamma(x)$ has the sign $(-1)^n$ for $-n < x < -n+1$ where n is a positive integer (B-12)

$\Gamma(x) = \pm\infty$ at $x = 0$ or $x =$ negative integers (B-13)

$\Gamma(x)$ is log convex for $x > 0$ (B-14)

$\Gamma(x)$ has derivatives of arbitrary high order (B-15)

$\Gamma(x)\Gamma''(x) > 0$ (B-16)

Stirling's formulas:

$$\Gamma(x) = \sqrt{2\pi}\, x^{x-\frac{1}{2}}\, e^{-x+\mu(x)} \tag{B-17}$$

where,

$$\mu(x) = \sum_{n=0}^{\infty} (x+n+\tfrac{1}{2})\log\left(1 + \frac{1}{x+n}\right) - 1 = \frac{\theta}{12x}\,,\ 0 < \theta < 1 \tag{B-18}$$

$$n! = \sqrt{2n}\, n^{n+\frac{1}{2}}\, e^{-n+\frac{\theta}{12n}} \tag{B-19}$$

The last 3 formulas are approximations of the gamma function for large values of x. The relative accuracy for $x \geq 10$ is quite high. Formula (B-19) provides a simple and highly accurate approximation to n! for $n \geq 10$. Using (B-10),

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \tag{B-20}$$

and using (B-9)

$$\Gamma(-x) = \frac{\Gamma(1-x)}{-x} \tag{B-21}$$

Trigonometric Relations:

$$\sin \pi x = \frac{\pi}{-x\Gamma(x)\Gamma(-x)}$$

$$= \frac{\pi}{\Gamma(x)\Gamma(1-x)} \tag{B-22}$$

$$\int_0^{\pi/2} (\sin\theta)^{2x-1}(\cos\theta)^{2y-1} d\theta = \frac{1}{2}\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \qquad \text{(B-23)}$$

Sketch:

Using (B-11), (B-12), and (B-13), a schematic plot of the gamma function can be obtained as in Figure B-1.

The Multivariate Gamma Function:

The multivariate case can be obtained through the generalization of the defining single-variate integral, (B-2)

$$\Gamma_k(a) \triangleq \int_{\underline{B}>0} e^{-\text{tr } \underline{B}} |\underline{B}|^{a-\frac{1}{2}(k+1)} d(\underline{B}) \qquad \text{(B-24)}$$

(B-24) can be represented in terms of product of single-variate gamma functions.

$$\Gamma_k(a) = \pi^{\frac{1}{4}k(k-1)} \prod_{i=1}^{k} \Gamma\left(a - \frac{1}{2}(i-1)\right) \qquad \text{(B-25)}$$

Another useful relation is

$$\int_{\underline{B}>0} \exp\left\{-\frac{1}{2}(\text{tr } \underline{B})\right\} d\underline{B} = 2^k \prod_{i=1}^{k} \Gamma(k+1-i) \qquad \text{(B-26)}$$

Selected References: [71-80].

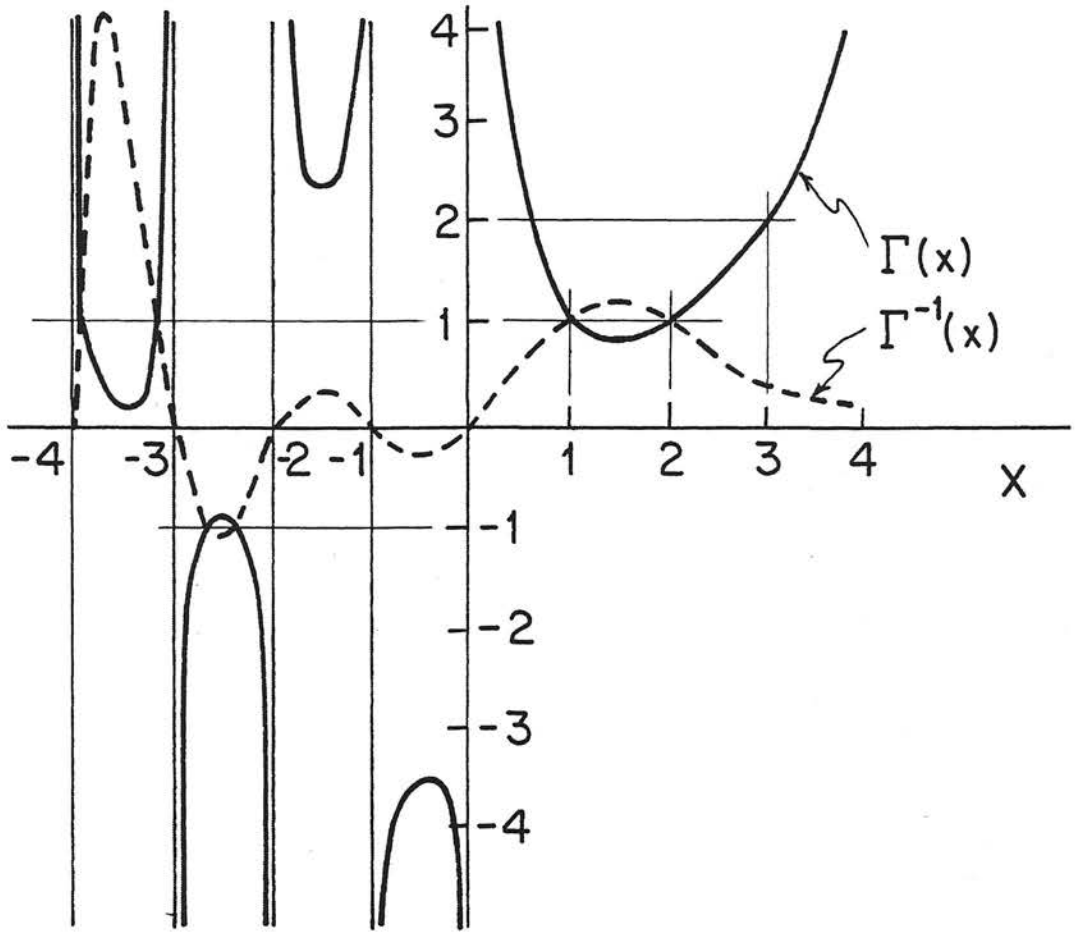


Figure B-1. Sketch of $\Gamma(x)$ and $\Gamma^{-1}(x)$ vs x.

APPENDIX C

STIEFEL MANIFOLD AND ORTHOGONAL GROUP

Definition C.1: Isomorphism

If R and R' are 2 real vector spaces, then R and R' are isomorphic if there exists a one-to-one correspondence $\alpha \leftrightarrow \alpha'$ between the elements α of R and α' of R' such that:

i) $\alpha_1 + \alpha_2 \leftrightarrow \alpha_1' + \alpha_2'$,

ii) $r\alpha_1 \leftrightarrow r\alpha_1'$ for $\alpha_1, \alpha_2 \in V$, and $\alpha_1', \alpha_2' \in V'$ and $r \in$ Real line [56].

Definition C.2: Homomorphism

An isomorphism which is not one-to-one or onto.

Definition C.3: Manifold (*)

A topological space such that every point has a neighbor which is homomorphic to the interior of a sphere in Euclidean space of the same number of dimensions.

Consider the SVD definition, $\underline{F} = \underline{U}\ \underline{S}\ \underline{V}^T$. From definition (3-25) of Chapter 3,

$$dF = \prod_{i,j}^{k,n} dF_{ij} \qquad (C\text{-}1)$$

Since $\underline{F}$ is $\underline{kxn}$, $\prod_{i,j}^{k,n} dF_{ij}$ is the volume in the Euclidean space R^{kn}. The Euclidean space R^{kn} is homomorphic to the product of the

---

(*) A variation of the word many-fold meaning one object which has several outlets.

orthogonal group $V_{k,k}$ which contains $\underline{U}$, the vector space R^k which contains $\underline{S}$ and the Stiefel monifold O(k) over which $\underline{V}$ ranges. Hence, r.h.s. of (C-1) will be represented in terms of the differential forms of the manifolds O(k), R^k, and $V_{k,n}$.

Definition of the Stiefel Manifold

Considering k n-dimensional orthonormal vectors ($k \leq n$) as columns of a matrix $\underline{V}$. The $\underline{nxk}$ matrix $\underline{V}$ is a point in the Stiefel manifold $V_{k,n}$

$$\underline{V} \in V_{k,n} \tag{C-2}$$

$\underline{V}$ is alternatively called a k-frame in space R^n. Obviously,

$$\underline{V}^T\underline{V} = \underline{I}_k \tag{C-3}$$

and $\underline{V}$ has $nk - \frac{1}{2}k(k+1)$ degrees of freedom on its nk elements. Therefore, $V_{k,n}$ is a $\frac{1}{2}k(k+1)$ algebraic variety on the nk-dimensional Euclidean space.

Definition of the Orthogonal Group:

Considering k k-dimensional orthonormal vectors as column of a matrix U. The $\underline{kxk}$ matrix $\underline{U}$ is called a point in the orthogonal group O(k).

$$\underline{U} \in O(k) \tag{C-4}$$

Obviously,

$$\underline{U}^T\underline{U} = \underline{I}_k \tag{C-5}$$

There are $\frac{1}{2}$ k(k-1) degrees of freedom for $\underline{U}$. Therefore, $\underline{O}(k)$ is a $k^2 - \frac{1}{2}$ k(k-1) $= \frac{1}{2}$ k(k+1) dimensional algebraic variety in a k^2-dimensional Euclidean space. Alternatively, one can say that since

$$\underline{U} = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k] \tag{C-6}$$

and

$$u_i^T u_i = 1 \text{ for } i = 1,2,\ldots,k \tag{C-7}$$

Then,

$$\sum_{i=1}^{k} u_i^T u_i = k \tag{C-8}$$

(C-8) means the summation of square of all elements of $\underline{U}$ is equal to constant k. The latter implies that $\underline{U}$, as a point of O(k), is on the surface of a k^2-dimensional hypersphere with radius $\sqrt{k}$. O(k) is the group for all the points on the surface of such hypersphere.

Invariance of the Orthogonal Group:

Consider

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

$$\underline{x}^T \underline{x} = x_1^2 + \ldots + x_k^2 \tag{C-9}$$

where, (C-9) represents a quadratic form. Upon a linear transformation on $\underline{x}$ as

$$\underline{Y} = \underline{U}\,\underline{X} \tag{C-10}$$

one can say that

$$\underline{Y}^T\underline{Y} = \underline{X}^T\underline{U}^T\underline{U}\,\underline{X} = \underline{X}^T\underline{X} \qquad \text{(C-11)}$$

(C-11) implies

$$y_1^2 + \ldots + y_k^2 = x_1^2 + \ldots + x_k^2$$

which is equivalent to saying that $\underline{U}\varepsilon O(k)$ is a matrix of linear transformation that leaves the quadratic form $x_1^2+\ldots+x_k^2$ invariant.

Invariant Measure on the Orthogonal Group:

It is desired to determine a differential form for $\underline{U}\varepsilon O(k)$ such that the properties of $O(k)$ are preserved. The differential form for an orthogonal matrix is the product of elements of the matrix of its linear differentials in all possible algebraic variety on the Euclidean space. Noting that the main property of $\underline{U}$ is

$$\underline{U}^T\underline{U} = \underline{I}_k \qquad \text{(C-12)}$$

differentiating (C-12), $\underline{U}^T\underline{d}(\underline{U})$ will be obtained and is skew symmetric according to Derivation A.6. For a skew symmetric matrix, the algebraic variety is $\frac{1}{2}k(k-1)$ dimensional. Thus, the product of all differential forms without duplication will be the product of the upper diagonal elements of $\underline{U}$. Considering (C-6), the product will be $\prod_{i<j}^{k} \underline{u}_i^T\underline{d}(\underline{u}_j)$. Defining the measure as $\tilde{d}(u)$

$$\tilde{d}(u) \triangleq \text{differential form for the measure on orthogonal group} \qquad \text{(C-13)}$$

Thus,

$$\tilde{d}(u) = \prod_{i<j}^{k} \underline{u}_i^T \underline{d}(\underline{u}_j) \qquad \text{(C-14)}$$

Explanation:

Another way of looking at the above argument is that according to the Derivation A.5 of Appendix A

$$\underline{u}_i^T \underline{d}(\underline{u}_i) = 0 \qquad \text{(C-15)}$$

and because of Derivation A.4 of the same appendix

$$\underline{u}_i^T \underline{d}(\underline{u}_j) = -\underline{u}_j^T \underline{d}(\underline{u}_i) \qquad \text{(C-16)}$$

Thus, only the product of elements on the lower or upper diagonal of $\underline{U}$ will suffice to provide the differential form $\tilde{d}(u)$ without duplication of any element in the product.

Invariant Measure on the Stiefel Manifold:

One wishes to obtain a differential form for the invariant measure on $\underline{V} \epsilon V_{k,n}$ such that the properties of $V_{k,n}$ are preserved. Noting the main property of $V_{k,n}$ is that its members $\underline{V}$ exhibit

$$\underline{V}^T \underline{V} = \underline{I}_k \qquad \text{(C-17)}$$

It is noted that $\underline{V}$ is a nxk matrix which contains k n-dimensional orthonormal column vectors.

$$\underline{V} = [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k] \tag{C-18}$$

Considering (n-k) n-dimensional orthonormal column vectors $q_1, q_2, \ldots, q_{n-k}$ such that they form a matrix

$$\underline{Q} = [q_1, q_2, \ldots, q_{n-k}] \tag{C-19}$$

The partition $[\underline{V}|\underline{Q}]$ is a member of orthogonal group O(n). Where

$$[\underline{V}|\underline{Q}]^T[\underline{V}|\underline{Q}] = \underline{I}_n \tag{C-20}$$

(C-20) means

$$\left[\begin{array}{c} \underline{V}^T \\ \hline \underline{Q}^T \end{array}\right] [\underline{V}|\underline{Q}] = \begin{bmatrix} \underline{I}_k & \underline{0} \\ \underline{0} & \underline{I}_{n-k} \end{bmatrix} \tag{C-21}$$

Upon evaluation of the product of differential forms for $[\underline{V}|\underline{Q}]$; and noting that there are $\frac{1}{2}k(k+1)$ dimensional algebraic variety for $[\underline{V}]$, and considering the following definition for the differential form

$$\tilde{d}(V) \triangleq \text{differential form for the measure on Stiefel manifold} \tag{C-22}$$

Thus,

$$\tilde{d}(V) = \prod_{i<j}^{k} v_j^T \underline{d}(\underline{v}_i) \prod_{i=1}^{k} \prod_{j=1}^{n-k} q_j^T \underline{d}(\underline{v}_i) \tag{C-23}$$

Integral of the Measure on Stiefel Manifold:

The value of the integral $\tilde{d}(V)$ is obtained by integrating the measure (C-23) over the whole space. To derive such integral, it is necessary to complete the following mathematical derivation.

If m n-dimensional orthonormal vectors $\underline{q}_1, \underline{q}_2, \ldots, \underline{q}_m$ are orthonormal to the vector $\underline{x}$, all of the involved vectors span a Euclidean space R^{m+1} provided that n = m+1. Through a geometrical approach and change of coordinates to a polar one, it is easily proved [68] that

$$\int_{\text{space}} \prod_{j=1}^{m} \underline{q}_j^T \underline{d}(\underline{x}) = A(m+1) \qquad \text{(C-24)}$$

where,

$$\underline{d}(\underline{x}) = [dx_1, dx_2, \ldots, dx_n]^T \qquad \text{(C-25)}$$

and

$$A(\ell) = \frac{2\pi^{\frac{1}{2}\ell}}{\Gamma(\frac{1}{2}\ell)} \qquad \text{(C-26)}$$

where, $A(\ell)$ is the area of the unit hypersphere in R^{ℓ}. $\Gamma(\frac{1}{2}\ell)$ of (C-26) is the single-variate gamma function introduced in Appendix B. It is easy to try the proof of (C-24) for m=1 which is the case of a unit circle in R^2 whose surface area means its perimeter and will be equal to 2π.

Remark: The unit sphere in R^1, consists of a line between 2 points ± 1. For (C-26) its area will be derived correctly as 2.

The integral of (C-23) will therefore be derived as follows:

$$\int_{V_{k,n}} \tilde{d}(V) = \prod_{i=1}^{k} \int \prod_{i<j}^{k} v_j^T \underline{d}(\underline{v}_i) \prod_{j=1}^{n-k} q_j^T \underline{d}(\underline{v}_i) \qquad \text{(C-27)}$$

According to (C-18), (C-19) and (C-20), the vectors $\underline{q}_j$ are orthonormal to $\underline{v}_i$. Furthermore, vectors $\underline{v}_j$ are also orthonormal to $\underline{v}_i$ for $i \neq j$. In the case of (C-27), for i=1, one will have i<j and hence the orthogonal vector to v_1 will be $v_2, v_3, \ldots, v_k$ and $q_1, q_2, \ldots, q_{n-k}$. This provides (n-1) space to be orthogonal to v_1. The integral, therefore, will be that of (C-24) for m = n-1. For i=2 m=n-2, and for i=k m=n-k. Thus, one will have

$$\int_{V_{k,n}} \tilde{d}(V) = A(n-1+1)A(n-2+1)\ldots A(n-k+1) \qquad \text{(C-28)}$$

which means

$$\int_{V_{k,n}} \tilde{d}(V) = \prod_{i=1}^{k} A(n-i+1) \qquad \text{(C-29)}$$

Using (C-26), (C-29) and (B-25)

$$\int_{V_{k,n}} \tilde{d}(V) = \frac{2^k \pi^{kn/2}}{\Gamma_k(n/2)} \qquad \text{(C-30)}$$

where $\Gamma_k(n/2)$ is the multivariate gamma function introduced in Appendix C.

Integral of the Measure on the Orthogonal Group:

Because orthogonal group is a special case of Stiefel manifold where n=k, the integral of the measure on orthogonal group will be obtained when n=k. Hence,

$$\int_{O(k)} \tilde{d}(U) = \frac{2^k \pi^{k^2/2}}{\Gamma_k(k/2)} \qquad \text{(C-31)}$$

Remark:

(C-31) provides the integral over the whole (improper) orthogonal group.

If the elements of the first row of $\underline{U}$ is restricted to be positive, in order to assure the uniqueness of SVD (Section 3.1.1), (C-24) and (C-26) will be evaluated over half of the space, hence the area of unit hypersphere in R^{ℓ} will be half of that of (C-26) as

$$A(\ell) = \frac{\pi^{\frac{1}{2}\ell}}{\Gamma(\frac{1}{2}\ell)} \text{ (for elements of } \underline{U}\text{'s 1st row } > 0) \qquad \text{(C-32)}$$

(C-32) will cause that value of (C-31) to be

$$\int_{\substack{O(k) \text{ for} \\ \text{elements of} \\ \underline{U}\text{'s 1st Row}}} \tilde{d}(U) = \frac{\pi^{k^2/2}}{\Gamma_k(k/2)} \qquad \text{(C-33)}$$

Normalized Measure on Stiefel Manifold:

$$d(V) \triangleq \text{normalized measure} \qquad \text{(C-34)}$$

In order to normalize $\tilde{d}(V)$ to obtain $d(V)$ it is enough to divide $\tilde{d}(V)$ by the value of its integral (C-30). Hence

$$\tilde{d}(V) = \frac{2^k \pi^{kn/2}}{\Gamma_k(n/2)} d(V) \qquad \text{(C-35)}$$

Thus,

$$\int_{V_{k,n}} d(V) = 1 \qquad \text{(C-36)}$$

Normalized Measure on Orthogonal Group (Haar Measure):

$$d(U) \triangleq \text{normalized measure} \qquad \text{(C-37)}$$

Hence

$$\tilde{d}(U) = \frac{2^k \pi^{k^2/2}}{\Gamma_k(k/2)} d(U) \qquad \text{(C-38)}$$

Thus,

$$\int_{\text{Whole } O(k)} d(U) = 1 \qquad \text{(C-39)}$$

Remark: $d(U)$ is also called Haar measure, and

$$\int_{\substack{O(k) \text{ for the} \\ \text{1st row positive}}} d(U) = 2^{-k} \qquad \text{(C-40)}$$

Therefore,

$$\int_{\substack{O(k) \text{ for 1st} \\ \text{row positive}}} d(U) = 2^{-k} \int_{\text{Whole } O(k)} d(U) \qquad \text{(C-41)}$$

Selected References: [68], [81-89].

APPENDIX D

TENSOR OR KRONECKER PRODUCT MATHEMATICAL RELATIONS

If $\underline{A}$ is $\underline{kxk}$ and $\underline{B}$ is $\underline{nxn}$ then the left Kronecker product of A and B is defined by

$$\underline{A} \otimes \underline{B} \triangleq \begin{bmatrix} b_{11}\underline{A} & b_{12}\underline{A} & \cdots & b_{1n}\underline{A} \\ b_{21}\underline{A} & b_{22}\underline{A} & \cdots & b_{2n}\underline{A} \\ \vdots & \vdots & & \vdots \\ b_{n1}\underline{A} & b_{n2}\underline{A} & \cdots & b_{nn}\underline{A} \end{bmatrix} \tag{D-1}$$

where, $\underline{B} = [b_{ij}]$, $i,j = 1,\ldots,n$. Hence, $\underline{A} \otimes \underline{B}$ is a $\underline{knxkn}$ matrix.

Identities:

$$(\underline{A}+\underline{C}) \otimes \underline{B} \equiv \underline{A} \otimes \underline{B} + \underline{C} \otimes \underline{B} \tag{D-2}$$

$$(\underline{A} \otimes \underline{B})^{adj} \equiv \underline{A}^{adj} \otimes \underline{B}^{adj} \tag{D-3}$$

$$(\underline{A} \otimes \underline{B})^{*} \equiv \underline{A}^{*} \otimes \underline{B}^{*} \tag{D-4}$$

$$(\underline{A} \otimes \underline{B})^{T} \equiv \underline{A}^{T} \otimes \underline{B}^{T} \tag{D-5}$$

$$(A \otimes B)^{-1} \equiv A^{-1} \otimes B^{-1} \quad \text{when A and B are non-singular square matrices} \tag{D-6}$$

$$(\underline{A} \otimes \underline{B})(\underline{C} \otimes \underline{D}) \equiv (\underline{A}\,\underline{C} \otimes \underline{B}\,\underline{D}) \tag{D-7}$$

$$tr(\underline{A} \otimes \underline{B}) \equiv tr(\underline{A})tr(\underline{B}) \tag{D-8}$$

$$Rank(A \otimes B) \equiv Rank(\underline{A})Rank(\underline{B}) \tag{D-9}$$

Complex Properties:

If $\underline{A}$ and $\underline{B}$ are Hermitian, then

$$A \otimes B \text{ is Hermitian and vice versa} \qquad \text{(D-10)}$$

Advanced Properties:

$$|\underline{A} \otimes \underline{B}| = |\underline{A}|^n |\underline{B}|^k \quad \text{when } \underline{A} \text{ and } \underline{B} \text{ are square matrices where } k \text{ is the dimension of } \underline{A} \text{ and } n \text{ is that of } \underline{B} \qquad \text{(D-11)}$$

Remark: (D-11) can be proved by spectral factorization of both sides of the equation.

Consider

$$\underline{A} = \text{a } \underline{kxk} \text{ matrix}$$

$$\underline{B} = \text{a } \underline{nxn} \text{ matrix}$$

$$\underline{M} \xrightarrow{\text{lexico trans.}} \underline{m}, \quad \underline{M} \text{ is } \underline{kxn}$$

and,

$$\underline{F} \xrightarrow{\text{lexico trans.}} \underline{f}, \quad \underline{F} \text{ is } \underline{kxn}$$

Then, for

$$\underline{f} = (\underline{A} \otimes \underline{B})\underline{m}$$

one obtains [91]

$$\underline{F} = \underline{A}\,\underline{M}\,\underline{B}^T \qquad \text{(D-12)}$$

and,

$$\mathrm{Tr}(\underline{m}\underline{f}^T) = \mathrm{Tr}(\underline{M}\underline{F}^T) \qquad \text{(D-13)}$$

relation (D-13) is proved in Derivation 5.1.

Selected Publications [93-95] and [91].

APPENDIX E

GENERATION OF A RANDOM PROCESS HAVING AN ARBITRARY MEAN AND CORRELATION

It is desired to obtain a N dimensional random process $\underline{f}$ with mean $\mu_f = E\{\underline{f}\}$ and correlation $R_{\underline{f}} = E\{\underline{f}\ \underline{f}^T\}$.

Consider $\underline{m}$ to be a N dimensional white random process. Then,

$$\underline{\mu}_m = E(\underline{m}) = \underline{0} \tag{E-1}$$

and,

$$\underline{R}_m = E\{\underline{m}\ \underline{m}^T\} = \underline{I}_N \tag{E-2}$$

Also, the covariance $K_{\underline{m}}$ of $\underline{m}$ is

$$\begin{aligned} \underline{K}_m &= E\{(\underline{m}-\mu_m)(\underline{m}-\mu_m)^T\} \\ &= E\{\underline{m}\ \underline{m}^T\} \\ &= \underline{R}_m \\ &= \underline{I}_N \end{aligned} \tag{E-4}$$

The covariance of $\underline{f}$ will be

$$\underline{K}_f = E\{(\underline{f}-\underline{\mu}_f)(f-\underline{\mu}_f)^T\} = \underline{R}_f - \underline{\mu}_f\underline{\mu}_f^T \tag{E-4}$$

Spectral factorization of $\underline{K}_f$ yields:

$$\underline{K}_f = \underline{E}_f\underline{\Lambda}_f\underline{E}_f^T \tag{E-5}$$

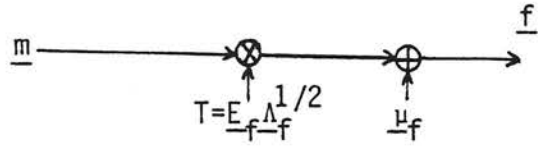
Consider the following operation on $\underline{m}$

$$\underline{f} = T\underline{m} + \underline{\mu}_f \tag{E-6}$$

where,

$$T = \underline{E}_f \underline{\Lambda}_f^{1/2} \tag{E-7}$$

schematically, (E-6) is



$\underline{f}$ of (E-6) = the desired vector with mean $\underline{\mu}_f$ and covariance matrix $\underline{K}_f$ (correlation matrix $\underline{R}_f$)

Conversely,

$$\underline{f} = T\underline{m} + \underline{\mu}_f$$

$$= (\underline{E}_f \underline{\Lambda}_f^{1/2})\underline{m} + \underline{\mu}_f$$

Therefore,

$$\text{Mean} = E\{\underline{f}\} = (\underline{E}_f \underline{\Lambda}_f^{1/2})E\{\underline{m}\} + \underline{\mu}_f$$

$$= \underline{\mu}_f \tag{E-8}$$

and,

$$
\begin{aligned}
\text{Covariance} &= E\{(\underline{f}-\underline{\mu}_f)(\underline{f}-\underline{\mu}_f)^T\} \\
&= E\{\underline{E}_f \underline{\Lambda}_f^{1/2} \underline{m}\ \underline{m}^T \underline{\Lambda}_f^{1/2} \underline{E}_f^T\} \\
&= \underline{E}_f \underline{\Lambda}_f^{1/2} E\{\underline{m}\ \underline{m}^T\} \underline{\Lambda}_f^{1/2} \underline{E}_f \\
&= \underline{E}_f \underline{\Lambda}_f^{1/2} \underline{I}_N \underline{\Lambda}_f^{1/2} \underline{E}_f^T = \underline{E}_f \underline{\Lambda}_f \underline{E}_f^T = \underline{K}_f
\end{aligned}
\qquad \text{(E-9)}
$$

The whiteness for process $\underline{m}$ is an essential requirement. If $\underline{m}$ is normally distributed, then $\underline{f}$ will be a normally distributed, correlated random process.

APPENDIX F

ZONAL POLYNOMIALS

The theory of zonal harmonics of a positive definite symmetric matrix has been introduced by Cartan [97]. The development and study of this concept has been presented by Hua [98-99] and independently by James [100].

Consider a kxk positive definite symmetrix matrix $\underline{A}$. Let V_P be the vector space of homogeneous polynomials of degree P in the distinct elements of $\underline{A}$. Thrall [101] has proved that V_P decomposes into a direct sum of irreducible invariant subspaces V_ρ corresponding to each partition ρ of P where

$$\rho = (P_1, P_2, \ldots, P_k) \qquad \text{(F-1)}$$

ρ is called a partition of P into not more than k parts such that

$$P_1 \geq P_2 \geq \ldots \geq P_k \geq 0 \qquad \text{(F-2)}$$

and

$$P_1 + P_2 + \ldots + P_k = P \qquad \text{(F-3)}$$

For every degree P, there are a number of possible partitions.

For example, if P=5 and k=3, the possible partitions for ρ would be: (5), (4,1), (3,2), (3,1,1), (2,2,1). Therefore, following (F-2) and (F-3), there are five possible partitions of 5 into 3 parts. If k=5, then to the above 5 partitions, one can add (2,1,1,1) and (1,1,1,1,1).

The polynomial $(\mathrm{Tr}\ \underline{A})^P \in V_P$ has a unique decomposition into polynomials $C_\rho(\underline{A}) \in V_\rho$ shown as

$$(\mathrm{Tr}\ \underline{A})^P = \sum_{\rho} C_{\rho}(\underline{A}) \qquad \text{(F-4)}$$

where Tr(A) means trace of $\underline{A}$, $C_{\rho}(\underline{A})$ is the zonal polynomial, $\underline{A}$ is a $\underline{k}\underline{x}\underline{k}$ positive definite symmetric matrix and $\sum_{\rho}$ means summation over all possible partitions ρ of P into not more than k parts. If k=1, the zonal polynomial $C_{\rho}(\underline{A})$ is analogous to the P power of a single variable, i.e., for a real number r,

$$C_{\rho}(r) = r^P \qquad \text{(F-5)}$$

For dimensionality k higher than 1, zonal polynomial is analogous to cos kθ and sin kθ in ordinary Fourier series [102].

The principal properties of the zonal polynomials are as follows:

i) Zonal polynomials are invariant to orthogonal transformation

$$C_{\rho}(\underline{A}) = C_{\rho}(\underline{H}^T\ \underline{A}\ \underline{H}) \qquad \text{(F-6)}$$

where, $\underline{H} \in O(k)$ (Appendix C elaborates on the orthogonal group O(k)). This property can be proved simply by noting that (Tr $\underline{A}$) is invariant under orthgonal transformation.

ii) Invariance under circular transformation

$$C_{\rho}(\underline{A}\ \underline{B}\ \underline{C}) = C_{\rho}(\underline{C}\ \underline{A}\ \underline{B}) = C_{\rho}(\underline{B}\ \underline{C}\ \underline{A}) \qquad \text{(F-7)}$$

This property can be proved by noting that Tr($\underline{A}\ \underline{B}\ \underline{C}$) is invariant to

the circular transformation. It is noteworthy that although $\underline{A}\ \underline{B}\ \underline{C}$ and $\underline{C}\ \underline{B}\ \underline{A}$ may be square matrices of different dimensions, their traces are equal.

iii) Decomposability (Z.P. Fundamental Property)

$$\int_{O(k)} C_p(\underline{H}^T\ \underline{A}\ \underline{H}\ \underline{B})dH = \frac{C_p(\underline{A})C_p(\underline{B})}{C_p(\underline{I}_k)} \qquad \text{(F-8)}$$

where $\underline{A}$, $\underline{B}$, and $\underline{H}$ are $\underline{kxk}$ matrices and dH is normalized invariant (Haar) measure on the orthogonal group O(k) such that the measure of the whole group is unity $\int_{O(k)} d\underline{H} = 1$ (Appendix C elaborates on the Haar measure). The proof of this property is given in [15].

Note: If $\underline{H}$ is $\underline{nxk}$ where $k < n$ and $\underline{H} \epsilon V_{k,n}$ where $V_{k,n}$ is the Stiefel Manifold of k-frame in Euclidean space R^n (Appendix C defines $V_{k,n}$), and the integral is taken over $V_{k,n}$, then the same decomposability holds except that dimension of $\underline{A}$ which in this case is n will be the order of the identity matrix in the denominator of (F-8), i.e. $\underline{I}_k$ of (F-8) shall be replaced by $\underline{I}_n$. If $\underline{H}$ is $\underline{kxn}$, where $k < n$, (F-8) as it is shown will hold.

It is useful to note that

$$C_0(\underline{A}) = 1 \qquad \text{(F-9)}$$

$$C_p(b\underline{A}) = b^P C_p(\underline{A}) \qquad \text{(F-10)}$$

where b is a scalar constant.

There is a recursive method (see Remark 4) for calculating the zonal polynomials up to a desired degree of accuracy for any power P

and any dimensionality k. Such a recursive technique can be used for programming the calculations of the zonal polynomials on the computer. Table F-1 gives the coefficients of the zonal polynomials of a positive definite symmetric matrix up to order 5.

Example on Homogeneous Monomial Symmetric Functions

As an example, let the degree be P=3, then there are a total of three possible partitions for 3 as $\rho=(3)$, $\rho=(2,1)$, and $\rho=(1,1,1)$.

Let A be a 3x3 positive definite matrix hence k=3 and there will be three real, positive eigenvalues λ_1, λ_2, and λ_3. The monomial symmetric function for partition $\rho=(3)$ will be

$$M_{(3)} \triangleq \lambda_1^3 + \lambda_2^3 + \lambda_3^3 \qquad \text{(F-11)}$$

for partition $\rho=(2,1)$

$$M_{(2,1)} \triangleq \lambda_1^2\lambda_2 + \lambda_2^2\lambda_1 + \lambda_1^2\lambda_3 + \lambda_3^2\lambda_1 + \lambda_2^2\lambda_3 + \lambda_3^2\lambda_2 \qquad \text{(F-12)}$$

and for partition $\rho=(1,1,1)$

$$M_{(1,1,1)} \triangleq \lambda_1\lambda_2\lambda_3 \qquad \text{(F-13)}$$

Each M_ρ is symmetric w.r.t. λ_1, λ_2, and λ_3, where the degree of each monomial term in M_ρ is 3 and their form of powers of λ_i's corresponds to their respective partition.

In terms of M_ρ's, the zonal polynomial $C_{(2,1)}$ is given by Table F-1 as

$$C_{(2,1)} = 4M_{(2,1)} + 6M_{(1,1,1)} \qquad \text{(F-14)}$$

where, M_ρ's are given above.

Table F-1. Zonal Polynomials Up to Degree 5 in Terms of Homogeneous Monomial Symmetric Functions.

DEGREE P	PARTITION ρ	ZONAL POLYNOMIAL C_ρ in NORMALIZED FORM
0	(0)	1
1	(1)	$M_{(1)}$
2	(2)	$\frac{1}{3}(3M_{(2)}+2M_{(1,1)})$
	(1,1)	$\frac{2}{3}(2M_{(1,1)})$
3	(3)	$\frac{1}{15}(15M_{(3)}+9M_{(2,1)}+6M_{(1,1,1)})$
	(2,1)	$\frac{9}{15}(4M_{(2,1)}+6M_{(1,1,1)})$
	(1,1,1)	$\frac{5}{15}(6M_{(1,1,1)})$
4	(4)	$\frac{1}{105}(105M_{(4)}+60M_{(3,1)}+54M_{(2,2)}+36M_{(2,1,1)}+24M_{(1,1,1,1)})$
	(3,1)	$\frac{20}{105}(18M_{(3,1)}+12M_{(2,2)}+22M_{(2,1,1)}+24M_{(1,1,1,1)})$
	(2,2)	$\frac{14}{105}(24M_{(2,2)}+16M_{(2,1,1)}+24M_{(1,1,1,1)})$
	(2,1,1)	$\frac{56}{105}(10M_{(2,1,1)}+24M_{(1,1,1,1)})$

	(1,1,1,1)	$\frac{14}{105}$ ($^{24M}(1,1,1,1)$)
5		$a_1M_{(5)}+a_2M_{(4,1)}+a_3M_{(3,2)}+a_4M_{(3,1,1)}+a_5M_{(2,2,1)}+a_6M_{(2,1,1,1)}+a_7M_{(1,1,1,1,1)}$

		a_1	a_2	a_3	a_4	a_5	a_6	a_7
	(5)	$\frac{1}{945}$ (945	525	450	300	270	180	120)
	(4,1)	$\frac{35}{945}$ (	120	72	138	108	126	120)
	(3,2)	$\frac{90}{945}$ (		72	48	88	96	120)
	(3,1,1)	$\frac{225}{945}$ (			42	28	78	120)
	(2,2,1)	$\frac{252}{945}$ (				40	60	120)
	(2,1,1,1)	$\frac{300}{945}$ (					36	120)
	(1,1,1,1,1)	$\frac{42}{945}$ (						120)

General Monomial Symmetric Function

Let $\underline{A}$ be an arbitrary positive definite symmetric matrix of order k. $\underline{A}$ has k real positive eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_k$. Let $\rho = (P_1,P_2,\ldots,P_k)$ be a partition of P into k parts, then the homogeneous monomial symmetric functions of $\lambda_1,\lambda_2,\ldots,\lambda_k$ for ρ will be shown by M_ρ as

$$M_\rho \triangleq \lambda_1^{P_1},\lambda_2^{P_2},\ldots,\lambda_k^{P_k} + \text{permutation of } \lambda_1,\lambda_2,\ldots,\lambda_k$$

$$\text{under the powers } P_1,P_2,\ldots,P_k \qquad \text{(F-15)}$$

In this appendix, p of 1 through 5 will be presented. The zonal polynomials are represented by the symbol $C_\rho(\underline{A})$ in their normalized form where, $\underline{A}$ is a $\underline{kxk}$ positive definite symmetric matrix.

Remark 1: In the Table F-1, the coefficients outside of the parenthesis serve as the normalizing factors which can be found by

$$\text{normalizing parameter} = \frac{2^P P!}{(2P)!}\,x(2P) \qquad \text{(F-16)}$$

where $\frac{2^P P!}{(2P)!}$ is equal to $\frac{1}{1\times3\times5\times\ldots(2P-1)}$ and determines the denominator of the parameter. (2P) determines the numerator and is the dimension of the representation 2p of the symmetric group on 2k symbols [108].

Remark 2: Zonal polynomials are positive on the domain of positive definite matrices.

Remark 3: Zonal polynomials are invariant under interchanging the eignevalues. In other words, interchanging λ_i and λ_j does not alter the zonal polynomial. This can be seen from the symmetry of z.p. representation by monomial functions and also from the definition of the z.p.

Remark 4: To calculate z.p. coefficients of a given matrix recursively, the Laplace-Beltrami operator [103] can be used which gives the coefficients in terms of the monomial symmetric functions of eigenvalues as presented in Table F-1.

Remark 5: For an identify matrix of dimension k, $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 1$ which makes every monomial term unity. The total number of monomial terms can easily be calculated, therefore $C_p(\underline{I}_k)$ can be determined.

Remark 6: If the dimension of matrix, i.e., k, is less than the degree P, the zonal polynomials for higher partitions will be zero. For example if P=3, and k=2, then $C_{1,1,1}(\underline{A}) = 0$.

Remark 7: The normalizing coefficients in Table F-1, for each degree and all possible partitions of that degree, add up to unity.

Remark 8: Other than the method discussed, the zonal polynomials can also be represented in terms of elementary symmetric functions of eigenvalues of $\underline{A}$ [102]; and also in terms of sums of powers of eigenvalues $\underline{A}$ [Ibid.].

Selected Publication: [103].

APPENDIX G

HISTOGRAM GAUSSIANIZATION

The objective of this appendix is to outline a procedure to standardize first order statistics of natural image fields in experiments of Chapters 8 and 9. Standardization of several images simply means making them equal mean and variance, having the same first-order histograms. Such a tactic makes a texture's SVD features totally related to its internal correlation, rather than to other factors such as photographic contrast and biases. One of the most practical methods is histogram Gaussianization. In our experiments, we modify texture field histograms to become Gaussian with zero mean and unit variance. The procedure is as follows:

Let $P_f(f)$ and $p_g(g)$ represent the actual and modified probability density functions, respectively, where f and g represent the actual and modified pixel values. Obviously for mean zero and unit variance, $p_g(g)$ is desired to be

$$p_g(g) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-g^2}{2}\right) \qquad \text{(G-1)}$$

The histogram transformation can be obtained as

$$\int_{g_{min}}^{g} p_g(g)dg = \int_{f_{min}}^{f} P_f(f)df = \mathbb{P}_f(f) \qquad \text{(G-2)}$$

where $P_f(f)$ is the cumulative distribution function of $\underline{F}$. Hence,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g} \exp\left(\frac{-g^2}{2}\right) dg = \mathbb{P}_f(f) \qquad \text{(G-3)}$$

so

$$\text{erf}(g) = \mathbb{P}_f(f) - \frac{1}{2} \tag{G-4}$$

The integral on the l.h.s. of (G-3) is referred to as Gauss(g). For discrete cases $\mathbb{P}_f(f)$ is $H_F(m)$ and $\mathbb{P}_f(f)$ is $\sum_{m=0}^{j} H_F(m)$ where j is the maximum grey level. It is obviously impossible to have $-\infty$ or numbers which are too small for computer. Hence, the ideal case shall be changed to a practical one:

$$\text{Gauss}(g)-\text{Gauss}(g_{min}) = \sum_{m=0}^{j} H_F(m) \tag{G-5}$$

or

$$g = \text{Gauss}^{-1}\left[\sum_{m=0}^{j} H_F(m)+\text{Gauss}(g_{min})\right] \tag{G-6}$$

g_{min} must be selected in such a way that $\text{Gauss}(g_{min})$ is close to zero but not exactly zero otherwise g_{min} would be $-\infty$ resulting in a computationally erroneous value. The value 0.0013 was chosen for $\text{Gauss}(g_{min})$ which is 0.13% of the total area under p.d.f. of g

$$g = \text{Gauss}^{-1}\left[\sum_{m=0}^{j} H_F(m) + 0.0013\right] \tag{G-7}$$

where $g_{min} = \text{Gauss}^{-1}(0.0013)$ another possible method would be

$$g = \text{Gauss}^{-1}\left[\sum_{n=0}^{j} H_F(m)\right] \tag{G-8}$$

in which every g is tested and clipped in such a way that if

$$|g| > |g_{min}| \qquad \text{(G-9)}$$

Then, $|g|$ shall be replaced by $|g_{min}|$.

Reference: [21].

BIBLIOGRAPHY

[1] G. H. Golub and C. Reinsch, "Singular Value Decomposition and Least Square Solutions," Numerish Matematic, vol. 14, 1970, pp. 403-420.

[2] A. Albert, Regression and Moore-Penrose Pseudo-Inverse, Academic Press, New York, 1972.

[3] H. C. Andrews and C. L. Patterson, "Outer Product Expansions and their Uses in Digital Image Processing," American Math Monthly, Vol. 82, No. 1, January 1974, pp. 1-13.

[4] T. S. Huang and P. M. Narendra, "Image Restoration by Singular Value Decomposition," Appl. Opt., Vol. 14, No. 9, September 1975, pp. 2213-2216.

[5] W. K. Pratt, "Singular Value Decomposition Image Feature Extraction," Semi-Annual Technical Report, USCIPI Report 770, Image Processing Institute, Univ. of So. Calif., Los Angeles, CA, 30 September 1977, pp. 20-30.

[6] B. Ashjari and W. K. Pratt, "Singular Value Decomposition Image Feature Extraction," Semiannual Technical Report, USCIPI Report 800, Image Processing Institute, USC, Los Angeles, CA, 31 March 1978, pp. 72-89.

[7] 'Student', "Probable Error of a Mean," Biometrika, vol. VI, 1908, pp. 1-25.

[8] R. A. Fisher, "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," Biometrika, vol. X, 1915, pp. 507-521.

[9] J. Wishart, "The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population," Biometrika, vol. XX-A, 1928, pp. 32-52.

[10] R. A. Fisher, "The Sampling Distributions of Some Statistics Obtained from Non-linear Equations," Annals of Eugenics, vol. 9, 1939, pp.238-249.

[11] M. A. Girshick, "On the Sampling Theory of the Roots of Determinantal Equations," Annals of Math. Stat., vol. 10, 1939, pp. 203-224.

[12] P. L. Hsu, "On the Distribution of Roots of Certain Determinantal Equations," Annals of Eugenics, vol. 9, 1939, pp. 250-258.

[13] A. M. Mood, "On the Distribution of the Characteristic Roots of Normal Second-Moment Matrices," Ann. Math.Stat., vol. 22, pp. 266-273.

[14] S. N. Roy, "P-Statistics, or Some Generalizations on the Analysis of Variance Appropriate to Multi Variate Problems," Sankhya, vol. 3, 1939, pp. 341-396.

[15] A. T. James, "The Distribution of the Latest Roots of the Covariance Matrix," Ann. Math. Stat., vo. 31, 1960, pp. 151-158.

[16] D. K. Faddeev, "On the Conditionality of Matrices," Math.Inst. of Steklov, 53, 1959, pp. 387-391 (Russian).

[17] S. Kung, "A New Identification and Model Reduction Algorithm via Singular Value Decompositions," Proc. 12th Asilomar Conf on Circuits, Systems and Computers, Pacific Grove, CA, November 1978, pp. 705-714; also "A New Low-Order Approximation Algorithm Via Singular Value Decompositions," submitted to IEEE Transactions on Automatic Control.

[18] T. W. Anderson, S. O. Gupta, G. P. H. Styan, A Bibliography of Multivariate Statistical Analysis, John Wiley and Sons, N.Y., 1972.

[19] R. M. Pickett, "Visual Analysis of Texture in the Detection and Recognition of Objects," in Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, 1970, pp. 289-308.

[20] J. K. Hawkins, "Textural Properties for Pattern Recognition," in Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, 1970, pp. 347-370.

[21] W. K. Pratt, Digital Image Processing, Wiley-Interscience, New York, 1978.

[22] J. L. Muerle, "Some Thought on Texture Discrimination," in Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, 1970, pp. 371-379.

[23] P. Brodatz, Texture: A Photographic Album for Artists and Designers, Dover, New York, 1956.

[24] R. Schmidt, The USC-Image Processing Institute Data Base, Revised Edition, USCIPI Report 780, Image Processing Institute, USC, Los Angeles, California, Oct. 1977.

[25] W. Metzger, Gesetze des Sehens, Waldemar Kramer, Frankfort am Main, 1953.

[26] J. J. Gibson, The Senses Considered as Perceptual System, Houghton Mifflin, Boston, 1966.

[27] A. Rosenfeld and B. S. Lipkin, "Texture Synthesis," in Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, 1970,pp. 309-345.

[28] R. M. Pickett, "Perceiving Visual Texture: A Literature Survey," Rept. AMRL-TR-68-12, Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, March 1968.

[29] B. Julesz, "Visual Pattern Discrimination," IRE Trans. on Inf. Theory, vol.IT-8, Feb. 1962, pp. 84-92.

[30] B. Julesz, Foundations of Cyclopean Perception, Univ. of Chicago Press, 1970.

[31] B. Julesz, et al., "Inability of Humans to Discriminate Between Visual Textures that Agree in Second-Order Statistics Revisited," Perception, vol. 2, 1973, pp. 391-405.

[32] B. Julesz, "Experiments in the Visual Perception of Texture," Scientific American, vol. 232, 1975, pp. 34-43.

[33] I. Pollack, Perceptual Psychophysics, vol. 13, 1973, pp. 276-280.

[34] S. R. Purks, and W. Richards, "Visual Texture Discrimination Using Random Dot Patterns," Journal Optical Society of America, vol. 67, June 1977, pp. 765-771.

[35] W. K. Pratt, O. D. Faugeras, and A. Gagalowicz, "Visual Discrimination of Stochastic Texture Fields," IEEE Trans. on Systems, Man, and Cybernetics, November 1978.

[36] R. S. Ledley, "Introduction to the Special Issue on Feature Extraction," Pattern Recognition, Pergamon Press, vol. 3, 1971, pp. 1-2.

[37] S. K. Abdali, "Feature Extraction Algorithms," Pattern Recognition, Pergamon Press, vol. 3, 1971, pp. 3-21.

[38] I. B. Muchnik, "Local Characteristic Formation Algorithm for Visual Patterns," Automn. Remote Control, no. 27, 1966.

[39] I. B. Muchnik, "Simulation of Process of Forming the Language for Description and Analysis of the Forms of Images," Pattern Recognition, Pergamon Press, vol. 4, 1972, pp. 101-140.

[40] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.

[41] J. T. Tou, and R. P. Heydron, "Some Approaches to Optimum Feature Extraction," in Computer and Information Science, vol. 2, J. T. Tou, Ed., Academic Press, New York, 1967.

[42] S. S. Wilks, "Multidimensional Statistical Scatter," in Contributions to Probability and Statistics, I. Olkin, Ed., vol. I, Stanford Publ., California, 1960.

[43] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, Wiley-Interscience, New York, 1973.

[44] H. Kobayashi, and J. B. Thomas, "Distance Measures and Related Criteria," Proc. 5th Annual Allerton Conf. on Circuit and System THeory, 1967, pp. 491-500.

[45] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Trans. Communication Technology, vol. COM-15, Feb. 1967, pp. 52-60.

[46] L. Garlucci, "A Formal System for Texture Languages," Pattern Recognition, Pergamon Press, vol. 4, 1972, pp. 53-72.

[47] S. Tsuji, and F. Tomita, "A Structural Analyzer for a Class of Textures," Computer Graphics and Image Processing, vol. 2, Dec. 1973, pp.216-231.

[48] J. S. Weska, C. R. Dyer and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification," IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-6, April 1976, pp. 269-285.

[49] A. Rosenfeld, and E. Troy, "Visual Texture Analysis," Technical Report 70-116, Univ. of Maryland, College Park, Maryland, June 1970. Also in Conference Record for Symposium on Feature Extraction and Selection in Pattern Recognition (IEEE Publication 70c-51c), Argonne, Illinois, Oct. 1970, pp. 115-124.

[50] A. Rosenfeld, and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," IEEE Trans. on Computers, vol. C-20, no. 5, May 1971, pp. 562-569.

[51] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-3, November 1973, pp. 610-621.

[52] R. M. Haralick and K. Shanmugam, "Computer Classification of Reservoir Sandstones," IEEE Trans. Geosci. Electronics, vol. GE-11, October 1973, pp. 171-177.

[53] O. D. Faugeras, W. K. Pratt, "Stochastic Based Feature Extraction," Prepublication Manuscript, Jan. 1979.

[54] L. S. Davis, "Structural-Statistical Approaches to Analysis of Macrotextures," Proposal to the National Science Foundation, No. 79-04851, Univ. of Texas, Austin, Texas,October 30, 1978.

[55] P. Lancaster, Theory of Matrices, Academic Press, New York, 1969.

[56] J. T. Moore, Elements of Linear Algebra and Matrix Theory, McGraw-Hill, New York, 1968.

[57] E. G. Kogbetliantz, "Solution of Linear Equations by Diagonalization of Coefficients Matrix," Quart. Applied Math., 13, 1955, pp. 123-132.

[58] M. R. Hestener, "Inversion of Matrices by Biorthogonalization and Related Results," J. Soc. Indust. Applied Math., 6, 1958, pp. 51-90.

[59] G. E. Forsythe and P. Henric, "The Cyclic Jacobi Method for Computing the Principal Values of a Complex Matrix," Proc. Am. Math. Soc., 94, 1960, pp. 1-23.

[60] V. N. Kublanovskaja, "Some Algorithms for Solution of the Complete Problem of Eigenvalues," V. Vycisl. Mat. i. Mat. FIZ., 1, 1961, pp. 555-570.

[61] C. L. Lawson and R. J. Hanson, Solving Least Square Problems, Prentice-Hall, Englewood Cliffs, N.J., 1974, Appendix C.

[62] C. B. Moler, "Three Research Problems in Numerical Linear Algebra," Proceedings of Symposium in Applied Math., volume 22, 1978.

[63] V. C. Klema and A. T. Laub, "The Singular Value Decomposition: Its Computations and Some Application," IEEE Transactions on Automatic Control, Vol. AC-25, No. 2, April 1980, pp. 164-176.

[64] C. B. Moler and G. W. Stewart, "An Efficient Matrix Factorization for Digital Image Processing," Informal Report LA-7637-MS, Los Alamos Scientific Laboratory, Los Alamos, New Mexico, Jan. 1979.

[65] S. C. Sahasrabudhe and P. M. Vaidya, "Estimation of Singular Values of an Image Matrix," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSSP-27, no. 4, August 1979.

[66] W. L. Deemer and I. Olkin, "The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis," based on lectures of P. L. Hsu at the University of North Carolina, Biometrika, Vol. 38, 1951, p. 345.

[67] I. Olkin, "The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis," Biometrika, Vol. 40, 1953, pp. 43-46.

[68] A. T. James, "Normal Multivariate Analysis and the Orthogonal Group," Ann. Math. Stat., 25, 1954, pp. 40-75.

[69] E. Goursat, lecons sur le Probleme de Pfaff, Hermann, Paris, 1922.

[70] W. Kaplan, Advanced Calculus, Addison-Wesley, Massachusetts, 1959.

[71] E. Artin, The Gamma Function, English Translation by M. Bulter, Hold, Reinhart, and Winston, 1964.

[72] E. Artin, Einfuhrung in Die Theorie der Gammafunktion, Hamburger Mathematische Einzelschrifen, published by Verlag B. G. Teubner, Leipzig, Heft/1931.

[73] B. C. Carlson, Special Functions of Applied Mathematics, Academic Press, New York, 1977.

[74] A. Erdelyi, Operational Calculus and Generalized Functions, Holt, Reinhart and Winston, New York, 1962.

[75] A. Erdelyi, Tables of Integral Transform, Vol. 1 and 2, based, in part, on notes left by H. Bateman, McGraw-Hill, N.Y., 1954.

[76] A. Erdelyi, (Editor), Higher Transcendental Functions, Vols. I, II and III, based, in part, on notes left by H. Bateman, McGraw-Hill, N.Y., 1953-55.

[77] A. E. Ingham, "An Integral Which Occurs in Statistics," Proc. Cambridge Philos. Soc. 29, 1933, pp. 271-276.

[78] C. S. Herz, "Bessel Functions of Matrix Argument," Ann. of Math., 61, 1, pp. 474-523.

[79] W. Grobner and N. Hofreiter, Integraltafel, Ersterteil, Unbestimmte Integrale, Springer-Verlag, Wien, N.Y., 1965-66.

[80] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover Publications, New York, 1972.

[81] T. W. Anderson, An Introduction to Multivariant Statistical Analysis, John Wiley and Sons, N.Y., 1958.

[82] E. Artin, Modern Higher Algebra: Galois Theory, New York University, Institute for Mathematics and Mechanics, New York, 1956.

[83] E. Artin, Algebraic Numbers and Algebraic Functions, Lecture notes of a course given at Princeton University, 1950-51, which was a revised version of lectures given at New York University during the preceding summer, Gordon and Breach, New York, 1967.

[84] L. de Branges, Hilbert Spaces of Entire Functions, Prentice-Hall, Englewood Cliffs, N.J., 1968.

[85] Y. L. Luke, The Special Functions and Their Approximations, Academic Press, N.Y. 1969.

[86] N. J. Vilenkin, "Special Functions and the Theory of Group Representations," Amer. Math. Soc., Providence, R.I., Vol. 22, 1968.

[87] H. Hotelling, "Relations Between Two Sets of Variates," Biometrika, Vol. 28, 1936, pp. 321-377.

[88] G. Szego, Orthogonal Polynomials, American Math. Soc. Providence, 4th ed., 1975.

[89] A. Weil, "L'Integration dans les Groupes Topologiques et ses Applications," Actualites Scientifique et Industrielles, Hermann, Paris, 1953.

[90] G. G. Hall, Matrices and Tensors, Vol. 4, editor H. Jones, The International Encyclopedia of Physical Chemistry and Chemical Physics, Editor in Chief, E. A. Guggenheim, et al., Pergamon Press, Oxford, 1963.

[91] W. K. Pratt, "Vector Space Formulation of Two-Dimensional Signal Processing Operations," Computer Graphics and Image Processing, 4, 1975, pp. 1-24.

[92] H. Grenader and G. Szego, Toeplitz Forms and Their Applications, University of California Press, Berkeley, 1958.

[93] M. Marcus, Basic Theorems in Matrix Theory, National Bureau of Standards, Appl. Math.Ser., 57, Jan. 22, 1960.

[94] J. Todd, "The Condition of Certain Matrices," Arch. Math., 5, 1954.

[95] C. C. MacDuffee, The Theory of Matrices, Chelsea, New York, 1946.

[96] A. Papoulils, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, N. Y., 1965.

[97] E. Cartan,"Les Fonctions Zonales d'un espace symetrique irreductible," in Oeuvre Completes, Partie 1, Vol. 2, 1952, pp. 1064-1070.

[98] L. K. Hua, "On the Theory of Functions of Several Complex Variables," Acta Math Sinica, I - Vol. 2, 1953, pp. 288-323; II- Vol. 5, 1955, pp. 1-25; III - Vol. 5, 1955, pp. 205-242 (three papers in Chinese).

[99] L. K. Hua, "On the Theory of Functions of Several Complex Variables," American Math. Soc. Translation, Ser. 2, Vol. 32, 1963, English translation of [98], I - pp. 163-194; II - pp. 195-220; III - pp. 221-263.

[100] A. T. James, "Zonal Polynomials of the Real Positive Definite Symmetric Matrices," Ann. Math. Ser. 2, 74, 1961, pp. 456-469.

[101] R. M. Thrall, "On Symmetrized Kronecker Powers and Structure of the Free Lie Ring," Am. J. of Math., Vol. 64, 1942, pp. 372-388.

[102] A. T. James, "Distributions of Matrix Variates and Latent Roots Derived From Normal Samples," Ann. Math. Stat., 35, 1964, pp. 475-501.

[103] A. T. James, "Calculation of Zonal Polymomial Coefficients by Use of the Laplace-Beltrami Operator," Ann. Math. Stat., Vol. 39, No. 5, 1968, pp. 1711-1718.

[104] Dec System 10/Dec System-20 Hardware Reference Manual, Vol. 1, Central Processor, EK-10/20-Hr-001.

[105] J. H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice-Hall, Englewood Cliffs, NJ, 1963.

[106] W. S. Dorn and D. D. McCracken, Numerical Methods with Fortran II Case Studies, John Wiley and Sons, New York, 1972.

[107] E. K. Blum, Numerical Analysis and Computation: Theory and Practice, Addison-Wesley, Menlo Park, CA, 1972.

[108] F. D. Murnaghan, The Theory of Group Representation, Dover Publications, New York, 1963.

[109] R. Bellman, Introduction to Matrix Analysis, McGraw-Hill Book Co., New York, 1960.

[110] S. Karlin, and H. M. Taylor, A First Course in Stochastic Processes, Academic Press, New York, 1975.

[111] K. C. S. Pillai,, "On the Distribution of the Largest Characteristic Root of a Matrix in Multivariate Analysis," Biometrika, 52 (3 and 4), 1965, p. 405.

[112] T. Sugiyama, "The Distribution of the Largest Latent Root and Corresponding Latent Vector for Principal Component Analysis," Ann. Math. Stat., Vol. 37, 1966, pp. 995-1001.

[113] T. Sugiyama,,"On the Distribution of the Largest Latent Root of the Covariance Matrix," Ann. Math. Stat., Vol. 38, 1967, pp. 1148-1151.

[114] A. G. Constantine, "Some Noncentral Distribution Problems in Multivariate Analysis," Ann. Math. Stat., Vol. 34, 1963, pp. 1270-1285.

[115] W. K. Pratt, "Generalized Wiener Filtering Computation Techniques," IEEE Trans. Computers, Vol. C-21, No. 7, 1972, pp. 636-641.

[116] W. K. Pratt and F. Davarian, "Fast Computational Techniques for Pseudoinverse and Wiener Image Restoration," IEEE Trans. Computers, C-26, June 6, 1977, pp. 571-580.

[117] S. S. Wilks, Mathematical Statistics, John Wiley and Sons, New York, 1962, Chapter 18.

[118] C. G. Khatri, "Series Representations of Distributions of Quadratic Forms in the Normal Vectors and Generalized Variance," Journal of Multivariate Analysis, Vol. 1, 1971,pp. 199-214.

[119] C. G. Khatri, "On Certain Distribution Problems Based on Positive Definite Quadratic Functions in Normal Vectors," Ann. Math. Stat., 37, 1966, pp. 468-479.

[120] B. K. Shah, "Distribution Theory of Positive Quadratic Forms in Multivariate Normal Samples," Ann. Math. Stat., 39, No. 2, 1968, pp. 1095.

[121] B. K. Shah, "Distribution of Noncentral Positive Definite Quadratic Functions from a Multivariate Normal Distribution," Ann. Math. Stat., 39, No. 18, 1968, pp. 1090.

[122] T. Hayakawa, "On the Distribution of the Latent Roots of a Positive Definite Random Symmetric Matrix I," Ann. Inst. Statistical Mathematics, Vol. 21, 1969, pp. 1-12.

[123] T. Hayakawa, "On the Distribution of a Quadratic Form in a Multivariate Normal Sample," Ann. Inst. Stat. Math., Vol. 18, 1966, pp. 191-200.

[124] A. T. James, "The Non-Central Wishart Distribution," Proc. Roy. Soc. London, Ser. A, 229, 1955, pp. 364-366.

[125] S. N. Roy, Some Aspects of Multivariate Analysis, John Wiley and Sons, Inc., New York, 1957.

[126] H. Nyquist, S. O. Rice, and J. Riordan, "The Distribution of Random Determinants," Quart. Appl. Math., Vol. 12, 1954, pp. 97-104.

[127] R. Bellman, "A Note on the Mean Value of Random Determinants," Quart. Appl. Math., Vol. 13, 1955, pp. 322-324.

[128] R. Englman, "The Eigenvalues of a Randomly Distributed Matrix," Nuovo Cimento, Vol. 10,. 1958, pp. 615-621.

[129] I. C. Gohberg and M. G. Krein, Introduction to the Theory of Linear Nonself-Adjacent Operator, American Mathematical Society, 1969.

[130] H. C. Andrews and C. L. Patterson, "Outer Product Expansions and Their Use in Digital Image Processing," IEEE Trans. Computers, Vol. C-25, No. 2, February, 1976, pp. 140-148.

[131] H. C. Andrews and C. L. Patterson, "Singular Value Decomposition and Digital Image Processing," IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-24, No. 1, February 1976, pp. 26-53.

[132] H. C. Andrews and C. L. Patterson, "Singular Value Decomposition (SVD) Image Coding," IEEE Trans. Comm., Vol. COM-26, No. 4, April 1976, pp. 425-432.

[133] H. C. Andrews and B. R. Hunt, Digital Image Restoration, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[134] G. H. Golub and W. Kahan, "Calculating the Singular Value of Pseudo-Inverse of a Matrix," Journal of SIAM Anal., Ser. B, Vol. 2, No. 2, 1965, p. 250.

[135] A. Wald, Sequential Analysis, John Wiley and Sons, New York, 1947.

[136] T. Marill,"On the Effectiveness of Receptors in Recognition Systems," IEEE Trans. Info. Theory, Vol. IT-9, 1963, pp. 11-27.

[137] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," Ann. Math. Stat., Vol. 23, 1962, pp. 493-507.

[138] M. Rosenblatt and D. Slepian, "Nth Order Markov Chains with any Set of N Variables Independent," J. Soc. Ind. Appl. Math., Vol. 10, No. 3, September 1962, pp. 537-549.

[139] J. Woods, "Two-Dimensional Discrete Markov Random Fields," IEEE Trans. Inform. Theory, Vol. IT-18, No. 2, March 1972, pp. 232-240.

[140] R. M. Gray, "Toeplitz and Circulant Matrices," Technical Report No. 6502-1, Info. Sys. Lab., Stanford Electronics Lab., Stanford Univ., Stanford, CA, June 1971.

[141] W. K. Pratt, O. D. Faugeras, A. Gagalowicz, "Application of Stochastic Texture Field Models to Image Processing," IEEE Proceedings, Vol. 69, No. 5, May 1981, pp. 542-551.

[142] A Chellapa, "Stochastic Models for Image Analysis and Processing," Ph.D. Dissertation, School of Electrical Engineering, Purdue University, W. Lafayette, Indiana, 1981.

[143] D. Garber, "Computational Models for Texture Analysis and Texture Synthesis," Ph.D. Dissertation, School of Electrical Engineering, Image Processing Institute, University of Southern California, Los Angeles, California, May 1981

[144] A. Rosenfeld and A. C. Kak, Digital Picture Processing, Academic Press, New York, 1976.

[145] B. Lashgari, L. Silverman, and J. F. Abramatic, "Approximation of 2-D Separable in Denominator Filter," submitted to IEEE Transactions in Circuit and Systems, 1981.

[146] K. I. Laws, "Textural Image Segmentation," Ph.D. Dissertation, School of Electrical Engineering, Image Processing Institute, University of Southern California, Los Angeles, California, January 1980.

[147] G. Dahlquist and A. Bjorck, Numerical Methods, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974.

[148] H. Wechsler and M. Kidode, "A Random Walk Procedure for Texture Discrimination," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, No. 3, July 1979, pp. 272-280.

[149] L. S. Davis, S. A. Johns, and J. K. Aggarwal, "Texture Analysis Using Generalized Co-occurrence Matrices," IEEE Trans. Pattern and Machine Intelligence, Vol. PAMI-1, No. 3, July 1979, pp. 251-259.

[150] S. Wang, F. R. Kias Velasco, A. Y. Wu, and A. Rosenfeld, "Relative Effectiveness of Selected Texture Primitive Statistics for Texture Discrimination," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-11, No. 5, May 1981, pp. 360-370.

[151] L. S. Davis, "Computing the Spatial Structure of Cellular Textures," CGIP-11, 1979, pp. 111-122.

[152] F. Vilnrotter, R. Nevatia, and K. Price, "Automatic Grid Relation Extraction and Texture Reconstruction for Homogeneous Regular Textures," Semi-Annual Technical Report, USCIPI Report 1010, 31 March 1981, pp. 27-42.

[153] R. W. Conners and C. A. Harlow, "A Theoretical Comparison of Four Texture Algorithms," Image Analysis Lab., Technical Report IAL-4-79, University of Missouri, Columbia, August 1976.

[154] R. T. Masumoto, "The Design of a 16x16 Multiplier," LAMBDA, The Magazine of VLSI Design, First Quarter 1980, pp. 15.

[155] W. K. Pratt (private communication).

[156] S. U. Lee, "Design of SVD/SGK Convolution Filters for Image Processing," Ph.D. Dissertation, School of Electrical Engineering, Image Processing Institute, University of Southern California, Los Angeles, California, January 1980.