INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9421755

**A multiresolution approach to computer verification of
handwritten signatures**

Qi, Yingyong, Ph.D.

The University of Arizona, 1993

U·M·I

**300 N. Zeeb Rd.
Ann Arbor, MI 48106**

**A MULTIREOLUTION APPROACH TO COMPUTER
VERIFICATION OF HANDWRITTEN SIGNATURES**

by
Yingyong Qi

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN ELECTRICAL ENGINEERING

In the Graduate College
THE UNIVERSITY OF ARIZONA

1 9 9 3

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Yingyong Qi

entitled A MULTIREOLUTION APPROACH TO COMPUTER VERIFICATION OF

HANDWRITTEN SIGNATURES

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Bobby R. Hunt

11/22/93

Date

Robin N. Strickland

11/22/93

Date

Jeffrey J. Rodriguez

11/22/93

Date

Date

Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Bobby R. Hunt

11/22/93

Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

A handwritten signature in dark ink, appearing to be 'J. P. G.', is written over a horizontal line that serves as a signature line.

ACKNOWLEDGMENTS

I am greatly indebted to my advisor, Dr. Bobby R. Hunt, for his professional guidance, intellectual support, and friendly encouragement throughout my effort to complete a Ph.D. degree in Electrical Engineering. His dedication to science and commitment to excellence will be a model for my professional life. My appreciation and thanks are extended to Drs. Robin N. Strickland and Jeffrey J. Rodriguez for their comment and appraisal of this work. I also thank Drs. Thomas Shipp and Linda Swisher for providing many helpful editorial assistances. Special thanks are given to Dr. H. Hsu of the Department of Electrical Engineering of Ohio State University who helped me to start this pursuit. Finally, my deep appreciation goes to my wife and son for permitting me to divert the time that rightfully belongs to them to this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES	7
LIST OF TABLES	8
ABSTRACT	9
1 INTRODUCTION	10
1.1 Background	10
1.2 Objectives	14
1.3 Summary of Contributions	14
1.4 Overview	15
2 MULTI-RESOLUTION DECOMPOSITION OF IMAGE USING WAVELET TRANSFORMATION	17
2.1 Introduction	17
2.2 Basic Concepts of Wavelet	19
2.3 Multiresolution Decomposition of Signals	26
2.4 Implementation	31
3 SIGNATURE DATABASE	37
3.1 Subjects and Data Collection	37
3.2 Preprocessing	40
4 ALGORITHMS FOR BUILDING MULTIREOLUTION REPRESENTATIONS OF SIGNATURE IMAGE	42
4.1 Global Geometric Features	42
4.2 Wavelet Transformed Features	48
5 SYSTEM DEVELOPMENT AND EVALUATION	52
5.1 Introduction	52
5.2 Scale Selection	53
5.3 Classification Systems	56
5.3.1 Vector Quantization Classifier	57
5.3.1.1 System Description	57
5.3.1.2 System Performance	62
5.3.1.3 System Evaluation	66
5.3.2 Neural Network Classifier	71

5.3.2.1	System Description	71
5.3.2.2	System Performance	74
5.3.2.3	System Evaluation	76
6	CLASSIFICATION SYSTEM USING GRID FEATURES	79
6.1	Introduction	79
6.2	Grid-Feature Representation	80
6.2.1	Composition of Feature Vector	80
6.2.2	Alignment of Feature Vector	82
6.2.2.1	Fundamentals of Dynamic Programming	82
6.2.2.2	Implementation	85
6.3	Grid Classifier	86
6.3.1	System Description	86
6.3.2	System Performance	88
6.3.3	System Evaluation	91
7	SUMMARY AND CONCLUSIONS	94
	REFERENCES	97

LIST OF FIGURES

2.1	A simple wavelet and its corresponding scaling function	21
2.2	The impulse response and frequency response of the conjugate filter H obtained using cubic splines	27
2.3	The impulse response and frequency response of the quadrature mirror filter G with the frequency response of H overlaid	32
2.4	The multi-resolution decomposition process	33
2.5	Schematic illustration of one stage in decomposing a signal from res- olution $j + 1$ to j	34
2.6	Schematic illustration of one stage in decomposing an image from resolution $j + 1$ to j	35
2.7	A sample image and its multi-resolution decomposition	36
3.1	Samples of original signatures	38
3.2	Samples of free-handed forgeries	39
3.3	Samples of simulated forgeries	39
3.4	Samples of bi-level signatures	41
4.1	The spatially delayed algorithm for blank-space removal	45
4.2	Flow chart of slant-angle computation	47
4.3	Multi-resolution representation of a bi-level normalized image	49
4.4	Multi-resolution representation of a vertical projection	50
4.5	Multi-resolution representation of a horizontal projection	51
5.1	Inter-to-intra personal variation ratio as a function of resolution index	55
5.2	Verification rate as a function of resolution index for projections (top plot) and images (bottom plot)	64
5.3	Evaluation results of multi-resolution VQ classifier	69
5.4	The multi-resolution network	74
6.1	A schematic illustration of dynamic programming	84
6.2	Flow chart of the dynamic programming algorithm	87
6.3	An illustration of grid size, grid-step size, and grid overlapping	89
6.4	Total error rate of the grid classifier	90

LIST OF TABLES

5.1	Error rate of the VQ classifier	65
5.2	Statistics of VQ distance at each single resolution	68
5.3	Statistics of VQ distance for multi-resolution systems	70
5.4	Separate error rates of multi-resolution systems	70
5.5	Error rate of network classifier	75
5.6	Statistics of network output	77
5.7	Separate error rates of network classifier	77
6.1	Statistics of the grid classifier	93
6.2	Error rates of the grid classifier	93

ABSTRACT

This dissertation presents procedures and results of works on computer signature verification. Two methods were developed and evaluated. First, verification was made using multi-resolution feature representation. This multi-resolution feature representation included global geometric characteristics and wavelet transformations of a signature image. A number of algorithms were developed to extract the global geometric features. A vector quantization classifier and a neural-network classifier were designed to use the multi-resolution representation for verification. Second, verification was made using a grid approach. In this approach, a signature image was divided by a grid and verification was made based on grid features that approximate local structures of a signature image. The grid feature comparison was made using dynamic programming procedures. Results indicated that both systems could detect free-handed forgeries accurately and could also monitor simulated forgeries with reasonable accuracy.

CHAPTER 1

INTRODUCTION

1.1 Background

Handwritten signatures, such as those appearing on bank checks and credit receipts, often are used as evidence of personal consent or authorization. Signatures, however, are seldom verified unless their authenticity is challenged. Because of the large amount of information in daily operations of relevant organizations, it is desirable to have a computerized system for signature verification.

There are two basic types of forgeries: free-handed and simulated. A free-handed forgery refers to a false signature made by a forger who only knows the spelling of the genuine signature. A simulated forgery refers to a false signature made by a forger who has examined samples of genuine signatures and committed the forgery by simulating (not tracing) the genuine signature. Obviously, it is much more difficult to identify a simulated forgery than a free-handed forgery [1].

The development of a signature verification system requires the solution of five broad types of problems: data acquisition, preprocessing, feature extraction, classification, and performance evaluation [2]. Depending on the method of data acquisition, a system is referred to as *on-line* or *dynamic* when signatures are obtained using a

digitizer or instrumented pen, and as *off-line* or *static* when signatures are obtained as images. Signature verification generally is more difficult in an off-line system than in an on-line system, in part, because timing and other kinematic information is unavailable in an off-line system. For example, it is virtually impossible to detect a traced or extremely skilled forgery using information from a signature image alone. However, an off-line system is more convenient to use and, thus, has a wider range of potential applications than an on-line system. Relatively few studies of limited scope have been undertaken for the development of off-line systems [3, 4, 5].

Signature images as results of human actions are subject to intra-personal variations. A computerized system for signature verification is feasible only if the computer representation of signature images is more sensitive to variations occurring in forgeries than to intra-personal variations. As a result, one crucial step in developing a computerized signature verification system is to search for the computer representation of signature that can maximize the *distance* between signatures of different individuals with the constraint that the *distance* between signatures of the same person is kept constant or minimized once a measure of *distance* is properly defined.

Intuitively, the intra- and inter-personal variations of signatures are scale dependent. On a gross scale, all signatures are similar. For example, they all have an overall rectangular shape consisting of pen/pencil strokes. On a detailed scale, every signature is different. No two signatures by the same person are identical. A

computerized signature-verification system could not succeed when the scale of representation is at either of these two extremes. If the scale were too gross, a system would accept forgeries because of the lack of discriminating power using such a scale. Conversely, on a too-detailed scale, a system would reject true signatures because of the weak similarities of signatures by the same person.

A wide range of feature representation scales have been applied during the development of signature verification systems. Nemcek and Lin [6] used the Fast Hadamard Transform to construct features of signature images. The Hadamard transform, much like the Fourier transform, preserves the energy and entropy of the original image and also tends to decorrelate the features [7]. It specifies the signature image on a detailed or complete basis. Ammar et al. [1] proposed a structural description of a signature to construct features for verification. In this method, the signature image was demarcated based on vertical and horizontal projections of the image [8], and a number of global and local features, such as the geometric size and slant characteristics, were derived from this segmentation. More recently, Ammar proposed reference pattern-based features, which basically were horizontal and vertical projections of a signature image [9].

Verification results using these representations demonstrated the scale-dependent nature of the signature-verification problem. The verification rate using the Hadamard transform was reported to be relatively low, although the representation was approximately complete (an image can be reconstructed within the accuracy of computation).

Horizontal and vertical projections are a subset of information contained in an image, yet a projection-based system seems to have out-performed the transformation-based system [9]. A further information reduction, such as the global and local features used in Ammar et al.'s study, reversed the increasing trend of verification accuracy [1]. There was evidently no monotonic relationship between the detail level of feature representation and the verification performance of a system. Instead, features of each scale appear to have contributed to the verification process in a unique way. It would be desirable if feature representation of different scales could be used simultaneously and constructively in building a signature-verification system.

The advantage of using a multi-scale or multi-resolution representation for signature verification is that a confirmation mechanism can be built into the decision-making process. Signatures of the same person are likely to differ on one, but not on several scales. Likewise, a forgery might be similar to the genuine signature on one, but not on many scales. Thus, it is likely that a system using a multi-resolution representation would perform better than that using a single resolution. The disadvantages of a multi-resolution approach include: (1) a scale selection or evaluation procedure has yet to be developed because feature representations at some scales may contribute negligibly or negatively to the decision-making process, and (2) a relatively complex decision rule must be synthesized.

1.2 Objectives

The specific objectives of this dissertation are:

- To develop algorithms for building effective feature representation of signature images.
- To develop and evaluate systems that use single- or multi-resolution features of signature images for signature verification.
- To develop verification systems that can identify simulated forgeries with a significantly greater-than-chance accuracy, and identify free-handed forgeries with a near-to-perfect accuracy.

1.3 Summary of Contributions

The main contributions of this work include:

- A number of feature-extraction algorithms have been developed in building the multi-resolution representation of signature images.
- Comprehensive evaluations of the multi-resolution approach have been obtained.
- A novel approach to signature verification using a grid feature representation has been developed.

These contributions provide method and guidance to the future development of a signature-verification system.

1.4 Overview

Because the multi-resolution representation of signature image is obtained primarily using the recently developed wavelet transformation, the basic theories of the wavelet transformation are briefly summarized in Chapter 2. Attention is focused on the concepts and definitions that frequently are applied to this work.

Chapter 3 describes the database of this work including the size and data collection procedure. The database is described in detail because it is closely related to the results and limitations of this work.

Chapter 4 presents algorithms developed for building multi-resolution representation of signature images. These algorithms were used primarily for the extraction of global geometric information of a signature image. The computational procedure for obtaining wavelet-transformed features is also included in this chapter.

Chapter 5 deals with issues in system development and evaluation using the multi-resolution approach. It includes a review of mathematical definitions of pattern classification and a description of scale-selection procedures. It also details the classification systems and system evaluations.

Chapter 6 introduces the novel, grid approach to signature verification. It contains the rationale, feature representation, and classification results using this approach.

Chapter 7 contains the summary and conclusions.

CHAPTER 2

MULTI-RESOLUTION DECOMPOSITION OF IMAGE USING WAVELET TRANSFORMATION

2.1 Introduction

In image processing, mathematical transformations are often used to project the image into a space whose basis function exhibits certain desired mathematical properties. The wavelet transformation, in essence, is a multi-resolution decomposition of an image, thereby making it possible to discover and to use scale-independent features in image analysis and recognition.

To simplify the discussion on the multi-resolution decomposition of an image, the mathematical notations used throughout this Chapter are as follows:

- $\mathbf{Z}$ and $\mathbf{R}$ denote the set of integers and real numbers, respectively.
- $\mathbf{Z} \times \mathbf{Z}$ and $\mathbf{R} \times \mathbf{R}$ denote two-dimensional integers and real numbers, respectively.
- $\mathbf{L}^2(\mathbf{R})$ denotes the space of measurable, square-integrable, one-dimensional functions $f(x)$,

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < +\infty \quad \forall f(x) \in \mathbf{L}^2(\mathbf{R}). \quad (2.1)$$

- The inner product of two functions $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$ is written as $\langle f(x), g(x) \rangle$ and

$$\langle f(x), g(x) \rangle = \int_{-\infty}^{+\infty} f(u)g(u)du. \quad (2.2)$$

- Two functions, $f(x)$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$, are orthogonal when

$$\langle f(x), g(x) \rangle = \begin{cases} 1 & \text{if and only if } f(x) \equiv g(x) \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

and the notation, $f(x) \perp g(x)$, is used to denote the orthogonality.

- The norm of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is written as $\|f(x)\|$ and

$$\|f(x)\| = (\langle f(x), f(x) \rangle)^{1/2}. \quad (2.4)$$

- The Fourier transform of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ is written as $\hat{f}(\omega)$ and

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)e^{-j\omega x}dx. \quad (2.5)$$

The inverse transform is

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{j\omega x}d\omega. \quad (2.6)$$

- The discrete Fourier transform of a sequence h_k ($k \in \mathbf{Z}$ and $\sum_k |h_k|^2 < +\infty$) is written as $H(\omega)$ and

$$H(\omega) = \sum_{k=-\infty}^{+\infty} h_k e^{-j\omega k}. \quad (2.7)$$

The inverse transform is

$$h_k = \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(\omega)e^{j\omega k}d\omega. \quad (2.8)$$

2.2 Basic Concepts of Wavelet

A *wavelet* is a function $\psi(x)$ whose translations and dilations can be used for expansions of $\mathbf{L}^2(\mathbf{R})$ [10]. The translation and dilation operations of a function, $\psi(x)$, are defined as:

$$\begin{aligned} \text{Translation : } \mathcal{T}[\psi(x)] &= \psi(x - t) \quad t \in \mathbf{R} \\ \text{Dilation : } \mathcal{D}[\psi(x)] &= \sqrt{s}\psi(sx) \quad s \in \mathbf{R}^+ \end{aligned} \tag{2.9}$$

where $\mathbf{R}^+$ denotes the positive half of the real axis. The constant factor $\sqrt{s}$ in the dilation operation is to ensure that the norm of the function, $\psi(x)$, remains unchanged by the operation. A brief notation of all the translations and dilations of $\psi(x)$ is

$$\sqrt{s}\psi(sx - t)_{(t,s) \in \mathbf{R} \times \mathbf{R}^+} \tag{2.10}$$

There actually exists $\psi(x)$ such that translations and dilations of $\psi(x)$ at $t = k$ and $s = 2^j$, $(k, j) \in \mathbf{Z} \times \mathbf{Z}$, form an orthonormal basis of $\mathbf{L}^2(\mathbf{R})$ [11]. This basis can be briefly written as

$$\sqrt{2^j}\psi(2^j x - k)_{(j,k) \in \mathbf{Z} \times \mathbf{Z}} \tag{2.11}$$

It has the property that, at each dilation index (resolution) j , $(\sqrt{2^j}\psi_j(2^j x - k))_{k \in \mathbf{Z}}$, is an orthonormal basis of a subspace $O_j \subset \mathbf{L}^2(\mathbf{R})$. These subspaces, $(O_j)_{j \in \mathbf{Z}}$, are related to each other and to $\mathbf{L}^2(\mathbf{R})$ as follows

$$O_j \perp O_l \quad j \neq l \tag{2.12}$$

$$\mathbf{L}^2(\mathbf{R}) = \cdots \oplus O_{-1} \oplus O_0 \oplus O_1 \oplus \cdots \tag{2.13}$$

where $\oplus$ indicates an orthonormal sum (the summation elements, $(O_j)_{j \in \mathbf{Z}}$, are mutually orthogonal). The decomposition of $\mathbf{L}^2(\mathbf{R})$ by the orthonormal subspaces $(O_j)_{j \in \mathbf{Z}}$, as shown in equation 2.13, is referred to as an *orthogonal decomposition* [12].

$\mathbf{L}^2(\mathbf{R})$ can also be decomposed using a *multi-resolution decomposition*. For a multi-resolution decomposition of $\mathbf{L}^2(\mathbf{R})$, the subspaces $(V_j)_{j \in \mathbf{Z}} \subset \mathbf{L}^2(\mathbf{R})$ are required to have the following properties [13]:

1. $V_j \subset V_{j+1} \quad \forall j \in \mathbf{Z}$;
2. $\cup_{j=-\infty}^{j=+\infty} V_j$ is dense in $\mathbf{L}^2(\mathbf{R})$ and $\cap_{j=-\infty}^{j=+\infty} V_j = \{0\}$;
3. $V_{j+1} = V_j + O_j \quad \forall j \in \mathbf{Z}$;
4. $f(x) \in V_j \leftrightarrow f(2x) \in V_{j+1} \quad \forall j \in \mathbf{Z}$.

The meanings of some of these properties are discussed in detail in the next section. It has been shown that there exists a unique function $\phi(x) \in \mathbf{L}^2(\mathbf{R})$ such that for any given $j \in \mathbf{Z}$, $(\sqrt{2^j} \phi_j(2^j x - k))_{k \in \mathbf{Z}}$ is an orthonormal basis of V_j [13]. The function, $\phi(x)$, is called a *scaling function*.

In summary, a wavelet, $\psi(x)$, has the property that its translations at each dilation index, j , form the orthonormal basis of a subspace $O_j \subset \mathbf{L}^2(\mathbf{R})$. These subspaces, $(O_j)_{j \in \mathbf{Z}}$, can be used to obtain an orthogonal decomposition of $\mathbf{L}^2(\mathbf{R})$. A scaling function, $\phi(x)$, has the property that its translations at each dilation index, j , form the orthonormal basis of another subspace $V_j \subset \mathbf{L}^2(\mathbf{R})$. These subspaces, $(V_j)_{j \in \mathbf{Z}}$,

Figure 2.1: A simple wavelet and its corresponding scaling function

can be used to obtain a multi-resolution decomposition of $\mathbf{L}^2(\mathbf{R})$. The subspaces, O_j and V_j , are related by $V_{j+1} = V_j + O_j \forall j \in \mathbf{Z}$.

To visualize some of these concepts, let us look at a simple example. A complete, orthonormal basis for functions that are constant on each quarter interval is shown in Figure 2.1, where $\psi(x)$ is the orthogonal wavelet and $\phi(x)$ is the scaling function. This basis is complete because any quarter-interval constant function is a linear combination of these four functions. It is orthonormal because the inner product of each basis functions with itself is unity, and the inner product between any two different basis functions is zero. Note that the two essential operations: translation and dilation are already shown in this example. The step from $\psi(2x)$ to $\psi(2x - 1)$ is translation. The step from $\psi(x)$ to $\psi(2x)$ is dilation.

The next level of expanding the basis functions contain $\psi(4x), \psi(4x - 1), \psi(4x - 2), \psi(4x - 3)$, each supported on an interval of length $\frac{1}{4}$. There can be an infinite

family of functions:

$$\psi_{jk}(x) = \psi(2^j x - k) \quad \text{for } j \geq 0 \text{ and } 0 \leq k < 2^j \quad (2.14)$$

or equivalently,

$$\psi_{jk}(x) = \sqrt{2^{-j}} \psi(x - 2^{-j} k) \quad \text{for } j \geq 0 \text{ and } 0 \leq k < 2^j \quad (2.15)$$

when x goes from $-\infty$ to ∞ . It has been proven that these functions, together with a scaling function $\phi(x)$, form a complete, orthonormal basis for any piecewise constant function [14].

The formal construction of a wavelet proceeds with two steps: the solution of the *dilation equation* for the scaling function, $\phi(x)$, and the construction of the wavelet, $\psi(x)$. The first step is serious and the second step is simple [14].

The dilation equation results from the properties of the multi-resolution decomposition of $L^2(\mathbf{R})$ presented earlier. Because $V_j \subset V_{j+1}$ and $(\sqrt{2^j} \phi_j(2^j x - k))_{k \in \mathbf{Z}}$ is an orthonormal basis of V_j , the scaling function, $\phi(x) = \sqrt{2^j} \phi_j(2^j x - k)|_{j=0, k=0}$, can be written as a linear combination of $\phi(2x - k) = \sqrt{2^j} \phi_j(2^j x - k)|_{j=1, k \in \mathbf{Z}}$, i.e.,

$$\phi(x) = \sum_k h_k \phi(2x - k) \quad (2.16)$$

where h'_k 's are summation constants and $\sum_{k=-\infty}^{+\infty} |h_k|^2 < +\infty$. This equation is the dilation equation and its solution is the scaling function $\phi(x)$.

With the solution of the dilation equation, $\phi(x)$, the constants h'_k 's can be obtained as

$$h_k = \langle \phi(x), \phi(2x - k) \rangle \quad \forall k \in \mathbf{Z} \quad (2.17)$$

because of the orthonormal property of the scaling function. When the coefficients, h'_k 's, are arbitrarily chosen, however, a solution of the dilation equation is difficult to obtain. The constants, h'_k 's, need to be specially designed so that a solution of the dilation equation exists and satisfies the orthonormal properties of a scaling function. The choice of h'_k 's actually dominates all that follows, in part, because of the connection, $V_{j+1} = V_j + O_j \forall j \in \mathbf{Z}$, between the orthonormal and multi-resolution decompositions of $\mathbf{L}^2(\mathbf{R})$. It has been shown that the wavelet, $\psi(x)$, can be computed as

$$\psi(x) = \sum_k (-1)^k h_{1-k} \phi(2x - k), \quad (2.18)$$

once the scaling function, $\phi(x)$, is obtained (see [14] for details).

For example, the choice of $h_0 = h_1 = 1$ and $h_k = 0 \forall k \neq \{0, 1\}$ produces a dilation equation

$$\phi(x) = \phi(2x) + \phi(2x - 1) \quad (2.19)$$

The solution of this dilation equation is the box scaling function in the example shown in Figure 2.1. The corresponding wavelet is computed as

$$\psi(x) = \phi(2x) - \phi(2x - 1) \quad x \in [0, 1] \quad (2.20)$$

This wavelet is also called the Haar's wavelet. Except for this simple example, a direct solution of the dilation equation 2.16 is difficult.

The solution of the dilation equation, however, can be derived relatively easily using the Fourier transform. Taking the Fourier transform of the dilation equation 2.16,

we have

$$\int_{-\infty}^{+\infty} \phi(x) e^{-j\omega x} dx = \sum_k h_k \int_{-\infty}^{+\infty} \phi(2x - k) e^{-j\omega x} dx \quad (2.21)$$

Let $\xi = 2x - k$, then

$$\int_{-\infty}^{+\infty} \phi(x) e^{-j\omega x} dx = \frac{1}{2} \sum_k h_k e^{-j\omega k/2} \int_{-\infty}^{+\infty} \phi(\xi) e^{-j\omega \xi/2} d\xi \quad (2.22)$$

or

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right). \quad (2.23)$$

Using successive substitution,

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right) = H\left(\frac{\omega}{2}\right) H\left(\frac{\omega}{4}\right) \hat{\phi}\left(\frac{\omega}{4}\right) = \cdots = \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right) \quad (2.24)$$

The inverse Fourier transform of $\hat{\phi}(\omega)$ produces the scaling function $\phi(x)$.

Although the Fourier transform of the scaling function has a simple analytical form, a numerical computation of $\hat{\phi}(\omega)$ is still difficult. The following theorem gives a practical characterization of the Fourier transform of the scaling function

Theorem 1 *Let $\phi(x)$ be a scaling function, and let $H(\omega)$ be the discrete Fourier transform of a digital filter whose impulse response $h_k = \langle \phi(x), \phi(2x - k) \rangle$. $H(\omega)$ satisfies the following two properties:*

$$|H(0)| = 1 \quad (2.25)$$

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1 \quad (2.26)$$

Conversely, if $H(\omega)$ satisfies the conditions 2.25 and 2.26 and

$$|H(\omega)| \neq 0 \quad \omega \in [0, 2\pi], \quad (2.27)$$

then the function defined by

$$\hat{\phi}(\omega) = \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right) \quad (2.28)$$

is the Fourier transform of a scaling function.

Equations 2.25 and 2.26 indicate that $H(\omega)$ has a unit gain at zero frequency, and the summation of amplitude squares of $H(\omega)$ and $H(\omega + \pi)$ is also unit. A digital filter that satisfies these equations is referred to as a *conjugate filter* in digital signal processing [15].

A numerical approximation of $\prod_{j=1}^{\infty} H(\frac{\omega}{2^j})$ in equation 2.24 can be made using polynomial splines of order $2p + 1$ ($p \in \mathbf{Z}$). Let V_1 denote the vector space of all functions of $\mathbf{L}^2(\mathbf{R})$ which are p times continuously differentiable and are equal to a polynomial of order $2p + 1$ on each interval $[k, k + 1]_{(k \in \mathbf{Z})}$. Let V_j denote the vector spaces that are derived from V_1 based on properties of a multi-resolution decomposition of $\mathbf{L}^2(\mathbf{R})$. It has been shown [16] that the Fourier transform of the scaling function associated with such a multi-resolution approximation is

$$\hat{\phi}(\omega) = \frac{1}{\omega^n \sqrt{\Sigma_{2n}(\omega)}} \quad n = 2 + 2p \quad (2.29)$$

where

$$\Sigma_n(\omega) = \sum_{k=-\infty}^{+\infty} \frac{1}{(\omega + 2k\pi)^n}. \quad (2.30)$$

A closed form of $\Sigma_n(\omega)$ can be obtained by calculating the derivative of order $n - 2$ of the equation

$$\Sigma_2(\omega) = \frac{1}{4 \sin^2(\omega/2)}. \quad (2.31)$$

Because of the relationship between $\hat{\phi}(\omega)$ and $\hat{\phi}(\frac{\omega}{2})$ defined by equation 2.24,

$$\hat{\phi}(2\omega) = H(\omega)\hat{\phi}(\omega), \quad (2.32)$$

and $H(\omega)$ is

$$H(\omega) = \sqrt{\frac{\Sigma_{2n}(\omega)}{2^{2n}\Sigma_{2n}(2\omega)}} \quad (2.33)$$

The amplitude transfer function of $H(\omega)$ and its impulse response are shown in Figure 2.2 for $p = 1$ and $n = 4$. It corresponds to a multi-resolution approximation built from cubic splines. The function $\Sigma_8(\omega)$ is given by

$$\Sigma_8(\omega) = \frac{N_1(\omega) + N_2(\omega)}{105(\sin \frac{\omega}{2})^8} \quad (2.34)$$

where

$$N_1(\omega) = 5 + 30(\cos \frac{\omega}{2})^2 + 30(\sin \frac{\omega}{2})^2(\cos \frac{\omega}{2})^2 \quad (2.35)$$

$$N_2(\omega) = 2(\sin \frac{\omega}{2})^4(\cos \frac{\omega}{2})^2 + 70(\cos \frac{\omega}{2})^4 + \frac{2}{3}(\sin \frac{\omega}{2})^6. \quad (2.36)$$

This $H(\omega)$ is a low-pass filter and was used to construct the wavelet function for the multi-resolution decomposition of signature images as described in the following sections.

2.3 Multiresolution Decomposition of Signals

The multi-resolution decomposition of a signal is made by projecting the signal into wavelet basis functions once the wavelet basis is obtained. For simplicity, only

Figure 2.2: The impulse response and frequency response of the conjugate filter H obtained using cubic splines

the multi-resolution decomposition of one-dimensional signals is discussed here. The basic concepts can be readily extended to two-dimensional signals.

Supposing that A_j is an operator that projects a function, $f(x) \in \mathbf{L}^2(\mathbf{R})$, into a space, $V_j \subset \mathbf{L}^2(\mathbf{R})$, that contains all possible approximations of $f(x)$ at the resolution j . A_j is expected to have certain qualitative properties. Some of them are:

1. A_j is a self-consistent operator, i.e, if $A_j f(x)$ is an approximation of the function $f(x)$ at resolution j , the approximation would not change when the operator is applied again. This means

$$A_j \cdot A_j = A_j \quad (2.37)$$

2. A_j is an orthogonal projector, i.e., among all possible approximations at resolution j , $A_j f(x)$ is the most similar to $f(x)$. This means

$$\|g(x) - f(x)\| \geq \|A_j f(x) - f(x)\| \quad \forall g(x) \in V_j \quad (2.38)$$

3. A_j is a causal operator, i.e., the approximation of a signal at resolution $j + 1$ contains all the information at resolution j . This means

$$V_j \subset V_{j+1} \quad \forall j \in \mathbf{Z} \quad (2.39)$$

Note that the term causality here means reasonably or mutually consistent. It is used differently than that in signal processing [17].

These properties mirror the properties presented in the previous section for a multi-resolution decomposition of $\mathbf{L}^2(\mathbf{R})$. In fact, the following theorem [13] says that the

orthonormal basis of the subspace $(V_j)_{j \in \mathbf{Z}}$ is defined by a unique scaling function $\phi(x)$.

Theorem 2 *Let $(V_j)_{j \in \mathbf{Z}}$ be a multi-resolution approximation of $\mathbf{L}^2(\mathbf{R})$. There exists a unique function $\phi(x) \in \mathbf{L}^2(\mathbf{R})$ such that if $\phi_j(x) = 2^j \phi(2^j x)$ for $j \in \mathbf{Z}$, then $\phi_{jk} = (\sqrt{2^{-j}} \phi_j(x - 2^{-j}k))_{k \in \mathbf{Z}}$ is an orthonormal basis of V_j .*

With the basis functions defined for $(V_j)_{j \in \mathbf{Z}}$, the multi-resolution decomposition of a function, $f(x) \in \mathbf{L}^2(\mathbf{R})$, can be obtained simply by computing the inner product between $f(x)$ and the basis functions, $\phi_{jk}(x)$, at resolution j . Such a decomposition, however, is redundant because the information at a lower resolution is included as part of the information at a higher resolution. This redundancy results from the causality property 2.38 of the projector A_j . A non-redundant approach is to include the complement, orthogonal space O_j in the multi-resolution decomposition process, where O_j satisfies

1. $O_j \perp V_j \quad \forall j \in \mathbf{Z}$
2. $O_j \oplus V_j = V_{j+1} \quad \forall j \in \mathbf{Z}$

The basis functions of O_j are specified by the following theorem [13]:

Theorem 3 *Let $(V_j)_{j \in \mathbf{Z}}$ be a multi-resolution vector space, $\phi(x)$, the scaling function, and $H(\omega)$ the corresponding conjugate filter. The translation and dilation of $\phi(x)$ at resolution j , $(\psi_j(x - 2^{-j}k))_{k \in \mathbf{Z}}$ is a orthonormal basis of O_j . In addition, the*

Fourier transform $\psi(x)$ is related to the Fourier transform of $\phi(x)$ by

$$\hat{\psi}(\omega) = G\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \quad (2.40)$$

where

$$G(\omega) = e^{-i\omega} \overline{H(\omega + \pi)} \quad (2.41)$$

This theorem unequivocally describes the role of the wavelet, $\psi(x)$, in the multi-resolution decomposition of functions of $L^2(\mathbf{R})$. It expands the complementary subspace of $(V_j)_{j \in \mathbf{Z}}$. Thus, a multi-resolution decomposition of a signal can be made by projecting the signal into subspaces $V_m, O_m, O_{m+1}, \dots, O_n$, where m and n indicate the lowest and the highest resolution indices, respectively, in the multi-resolution approximation.

Equation 2.41 indicates that the amplitude of the transfer function, $G(\omega)$, is equal to the amplitude of the transfer function of, $H(\omega)$, when $H(\omega)$ is shifted by π and flipped in frequency. This operation is equivalent to make a mirror image of $H(\omega)$ with reference to $\pi/2$, a quadrature dividing location in the complex plane. Thus, $G(\omega)$ and $H(\omega)$ are called *quadrature mirror filters* [15, 18, 19].

The relationship between impulse responses of the quadrature mirror filters can be derived directly from Equation 2.41. Let g_k denote the inverse Fourier transform of $G(\omega)$ and h_k denote the inverse Fourier transform of $H(\omega)$. The inverse Fourier transform of $e^{-i\omega} \overline{H(\omega + \pi)}$ is $(-1)^{(1-k)} h_{1-k}$. Thus, the impulse responses of filter G

and H are related by

$$g_k = (-1)^{(1-k)} h_{1-k} \quad (2.42)$$

and g_k is the digital representation of the wavelet $\psi(x)$.

The impulse response and the amplitude transfer function of the quadrature mirror filter, G , of the conjugate filter, H , shown in Figure 2.2 are illustrated in Figure 2.3. As described in the next section, the multi-resolution decomposition of signals are obtained using these filters.

2.4 Implementation

In practice, the physical device typically measures a signal at one particular resolution. Assuming this resolution is 0, the causality principle (equation 2.38) implies that approximations at resolutions j ($j < 0$) can all be computed from the measured signal by a decomposition process illustrated in Figure 2.4.

It can be shown [20] that the numerical equivalent of this decomposition process is the successive, discrete convolution of the signal with the digital filters $\tilde{H}$ (low-pass) and $\tilde{G}$ (high-pass) followed by subsampling at every other data point. Here, $\tilde{H}$ and $\tilde{G}$ are impulse-response reflected version of the quadrature mirror filters, H and G , i.e.,

$$\tilde{h}_k = h_{-k} \quad (2.43)$$

$$\tilde{g}_k = g_{-k} \quad (2.44)$$

where h , g , $\tilde{h}$, and $\tilde{g}$ are the impulse responses of H , G , $\tilde{H}$, and $\tilde{G}$, respectively.

Figure 2.3: The impulse response and frequency response of the quadrature mirror filter G with the frequency response of H overlaid

Figure 2.4: The multi-resolution decomposition process

Let us look at the projection of a signal at resolution j using the scaling function as the basis. Because $V_j \subset V_{j+1}$ and $(\sqrt{2^{-j-1}}\phi_{j+1}(x - 2^{-j-1}k))_{k \in \mathbf{Z}}$ is an orthonormal basis of V_{j+1} , for any $n \in \mathbf{Z}$, $\phi_j(x - 2^{-j}n) \in V_j$ can be written as

$$\phi_j(x - 2^{-j}n) = 2^{-j-1} \sum_{k=-\infty}^{+\infty} \langle \phi_j(u - 2^{-j}n), \phi_{j+1}(u - 2^{-j-1}k) \rangle \cdot \phi_{j+1}(u - 2^{-j-1}k) \quad (2.45)$$

By changing variables in the inner product integral, it can be shown that

$$2^{-j-1} \langle \phi_j(u - 2^{-j}n), \phi_{j+1}(u - 2^{-j-1}k) \rangle = \langle \phi_{-1}(u), \phi_0(u - (k - 2n)) \rangle \quad (2.46)$$

When computing the inner product of a signal, $f(x)$, with both sides of equation 2.45, we have

$$\begin{aligned} \langle f(u), \phi_j(u - 2^{-j}n) \rangle &= \sum_{k=-\infty}^{+\infty} \langle \phi_{-1}(u), \phi_0(u - (k - 2n)) \rangle \cdot \\ &\quad \langle f(u), \phi_{j+1}(u - 2^{-j-1}k) \rangle \end{aligned} \quad (2.47)$$

Let H denote the discrete filter whose impulse response is

$$h_n = \langle \phi_{-1}(u), \phi_0(u - n) \rangle \quad \forall n \in \mathbf{Z} \quad (2.48)$$

Figure 2.5: Schematic illustration of one stage in decomposing a signal from resolution $j + 1$ to j

Let $\tilde{H}$ denote a mirror filter of H with the impulse response $\tilde{h}_n = h_{-n}$. By inserting equation 2.48 in the previous equation, we obtain

$$\langle f(u), \phi_j(u - 2^{-j}n) \rangle = \sum_{k=-\infty}^{+\infty} \tilde{h}(2n - k) \cdot \langle f(u), \phi_{j+1}(u - 2^{-j-1}k) \rangle \quad (2.49)$$

Thus, the projection of $f(x)$ at resolution j using the scaling function as the basis can be computed as the convolution of $f(x)$ at resolution $j + 1$ with $\tilde{h}$ and keep the other sample of the output.

A similar derivation would indicate that the projection of $f(x)$ at resolution j using the wavelet as the basis can be computed as the convolution of $f(x)$ at resolution $j + 1$ with $\tilde{g}$ and keep the other sample of the output. A schematic illustration of one stage of the filtering process is shown in Figure 2.5.

This multi-resolution decomposition process can easily be generalized to images. The two-dimensional wavelet transformation can be treated as one-dimension transformation along x and y axes. Each step of the decomposition now comprises of the convolution of the rows and then the columns of the image with the one-dimensional

Figure 2.6: Schematic illustration of one stage in decomposing an image from resolution $j + 1$ to j

quadrature mirror filters. One stage of the decomposition process for images is illustrated in Figure 2.6. This implementation scheme is used in the multi-resolution decomposition of signature images as described in Chapter 4. As an example, a sample image and its multi-resolution decomposition are shown in Figure 2.7.

Figure 2.7: A sample image and its multi-resolution decomposition

CHAPTER 3

SIGNATURE DATABASE

3.1 Subjects and Data Collection

The database comprised 450 signatures that were collected from 25 people. Among the 25 people, fifteen were asked to provide signatures for building the authentic signature database. Five were asked to produce a set of free-handed forgeries, and the other five were asked to produce a set of simulated forgeries of the 15 subjects.¹

The 15 people for building the authentic signature database were selected randomly. They represented a wide variety of signature styles, from completely incomprehensible line strokes to clear and tidy handwriting. They were asked to sign non-overlapping signatures on a white sheet of paper. A total of 20 signatures were collected from each person over a course of one month to sample intra-personal variations. Fifteen of these signatures were used for system training and the remaining five for system testing.

The five, free-handed forgers were also selected randomly. They were asked to produce one forgery of each of the 15 subjects' signature when only the spellings of

¹Authentic signature is the true signature of a person. Free-handed and simulated forgeries are defined in the introduction.

Figure 3.1: Samples of original signatures

the original signatures were known. The five people producing simulated forgeries were skilled forgers. One was an art student experienced in simulating signatures for paintings. Three were cartographers who are capable of making hand-drawn maps. The last was a secretary skilled in producing substitution signatures. They were asked to produce one forgery on each of the 15 subjects when samples of genuine signatures were available for examination. They were allowed to practice as many times as they wished before writing forgeries. A sample set of original signatures and their free-handed and simulated forgeries are shown in Figures 3.1 to 3.3.

Figure 3.2: Samples of free-handed forgeries

Figure 3.3: Samples of simulated forgeries

3.2 Preprocessing

These signatures were scanned into the computer at a 8-bit, 300 dot-per-inch resolution (Howtek, Scan Master 3). The scanned signatures were cut out from the scan page in its original orientation and size using an image editor. These signatures were converted into bi-level images by thresholding [8]. The thresholding involved:

1. linearly normalizing the gray-level of each image into the range 0 – 255,
2. computing the histogram of the grey-level normalized image,
3. setting a threshold at the level above which there were 65% of the total pixels,
and
4. assigning a pixel to 1 (black) if its grey-level was greater than the threshold
and to 0 (white) otherwise.

Sample bi-level signatures derived from the signatures shown in Figure 3.1 are shown in Figure 3.4. These bi-level images were the raw database for system development and evaluation. The size of this database is comparable to those reported in the literature [5].

Nancy S. Crandall

John R. Johnson

John R. Johnson

Figure 3.4: Samples of bi-level signatures

CHAPTER 4

ALGORITHMS FOR BUILDING MULTIREOLUTION REPRESENTATIONS OF SIGNATURE IMAGE

4.1 Global Geometric Features

The top level of the multi-resolution representation of a signature image comprised the global geometric features. These features (described later) included:

- height and width of a signature image,
- width of the signature image with blank spaces between horizontal elements removed,
- slant angle of the signature,
- vertical center-of-gravity of black pixels,
- maximum horizontal projection,
- area of black pixels, and
- baseline shift of signature image.

The height and width of a signature were directly measured as the size of the image after blank edges of the image were removed. The blank-edge removal was

accomplished by sequentially eliminating rows (or columns) on the edge of the image if the total number of pixels in the row (or column) was less than 2 (an arbitrary threshold). This simple edge-removal method was effective, in part, because of the high signal-to-noise ratio of the collected images.

The width of a signature is claimed to have large intra-personal variation when blank spaces exist between horizontal elements [4]. Thus, the signature width was also measured when blank spaces between horizontal elements were removed. The blank space removal was made based on the vertical projection of the image. The vertical and horizontal projections of an image are defined as

$$P_v[j] = \sum_i I_{ij} \quad \text{for } j = 0, 1, \dots, N-1 \quad (4.1)$$

$$P_h[i] = \sum_j I_{ij} \quad \text{for } i = 0, 1, \dots, M-1 \quad (4.2)$$

where P_v and P_h denote the vertical and horizontal projections, respectively. $I_{ij} \in [0,1]$ indicates the pixel level at the i th row and j th column. N and M denote the width and height of the image.

A spatially delayed, thresholding algorithm was used to identify the blank spaces.

This algorithm has the following steps:

1. Initialize both over-threshold and sub-threshold counters to zero.
2. Start the over-threshold counter when the vertical projection exceeds a given amplitude threshold.

3. Start the sub-threshold counter once the vertical projection goes below the amplitude threshold. The over-threshold counter, however, will not stop until the sub-threshold counter exceeds a length threshold (spatial delay).
4. Compute the difference between over-threshold and sub-threshold counters, and cumulate this difference to the width counter each time the sub-threshold counter is stopped.
5. Go back to the beginning until the end of the projection is reached.

The spatial delay was introduced to ensure that a horizontal segment was removed only when the width of a sub-threshold segment in the vertical projection exceeds a given limit. This algorithm is schematically illustrated in Figure 4.1

Although slant angle represents a prominent, idiosyncratic feature of a signature, it is mainly a psycho-physical concept. A well-established algorithm for measuring slant angle is not available. In my view, a useful measure of slant angle should reflect both local and global slant characteristics of a signature. Here, the local slant refers to the angle of long or dominant strokes, and the global slant refers to the overall direction of line strokes of an image. Thus, the direction that produced the maximum projection is used as a measure of the local slant, and the direction that produced the maximum number of connected short lines is used as a measure of global slant. In most cases, the local and global slant angles would coincide with each other. There were cases, however, that differed significantly. For example, when an A-shaped

Figure 4.1: The spatially delayed algorithm for blank-space removal

letter existed in the signature, the maximum projection might occur at either sides of the letter and result in approximately a 30-degree measurement error in the local slant. To ensure a reliable estimation of the slant angle, we used the direction that provided the least compromise between the local and global slant measures as the final slant angle. Image rotations were made using Alan Paeth's three-shear method [21]. A flow chart of the algorithm for the slant angle measurement is shown in Figure 4.2. A visual inspection indicated that this algorithm correctly captured the slant characteristics of all signature images.

The vertical center of gravity, C_v , and the maximum horizontal projection, P_m , were used to indicate the location and strength of the signature baseline, and were obtained from the horizontal project P_h as

$$C_v = \frac{\sum_i i \cdot P_h[i]}{\sum_i P_h[i]} \quad (4.3)$$

$$P_m = \max_i \{P_h[i]\}. \quad (4.4)$$

The area of black pixels, A_b , computed as the summation of vertical and horizontal projections, was used to reflect the proportion of strokes in a signature,

$$A_b = \sum_i P_v[i] + \sum_j P_h[j] \quad (4.5)$$

The baseline shift, measured as the difference in vertical center-of-gravities (see equation 4.3) between the left and right halves of the image, was included to compensate for the slant-angle variation due to the overall orientation of a signature. These geometric features were scaled into a comparable range (0—1) so that the norm of

Figure 4.2: Flow chart of slant-angle computation

the feature vector would not be dictated by one element. Most of the scalings were made relative to another parameter in the feature vector although some were made relative to a constant. For example, the height was scaled relative to the width, which resulted in the aspect ratio, but the slant angle was scaled by 75 degree, which was the upper limit of image rotation.

4.2 Wavelet Transformed Features

The wavelet transformation was used to build the multi-resolution representation for both signature images and their projections. The images were size normalized from their blank-edge-removed size into a 128x256 matrix using linear interpolations prior to the wavelet transformation. The grey-level images, as a result of size-normalization, were transformed back into bi-level images using a thresholding procedure similar to that described in Chapter 3. The scaling function of the quadrature mirror filters was obtained using cubic splines, as discussed in detail in Chapter 2 [20].

The multi-resolution representations of vertical and horizontal projections were made by convolving the projections with the impulse response of one-dimensional quadrature mirror filters recursively until the length of the horizontal projection was 2 and the length of vertical projection was 4. Similarly, the multi-resolution representations of a signature image were made by first convolving rows and then the columns of an image with the one-dimensional quadrature mirror filters recursively

Figure 4.3: Multi-resolution representation of a bi-level normalized image until the height of the image was 2 and the width of the image was 4. An example of multi-resolution representation of image is shown in Figure 4.3, and the multi-resolution representations of its projections are shown in Figures 4.4 and 4.5.

Figure 4.4: Multi-resolution representation of a vertical projection

Figure 4.5: Multi-resolution representation of a horizontal projection

CHAPTER 5

SYSTEM DEVELOPMENT AND EVALUATION

5.1 Introduction

Given the feature representations described in the previous chapter, the signature-verification problem can be translated directly into a conventional pattern-classification problem where each randomly sampled signature x belongs to one of two groups. The solution to the problem is a classification procedure or function δ that projects a sample signature from its single- or multi-resolution feature space into the discrete set $\{0, 1\}$, where 0 and 1 denote the genuine and forgery groups, respectively. The classification function, δ , can be constructed based on a set of randomly sampled signatures (the training set) whose categorical identities are known. The performance of the classification function can be evaluated by the overall misclassification rate r , the probability that δ misclassifies a randomly sampled signature. This rate can be estimated by computing the proportion of misclassification that δ made on a second set of randomly sampled signatures (the test set) when these signatures are not used to construct δ .

It is relatively simple to construct the classification function using single-resolution, feature representation. For example, when templates have been derived from the

training set, the minimum distance, d_m , between a test sample x and the templates projects x from its feature space onto the real axis, $\mathbf{R}$. This projection would be binary if $\mathbf{R}$ is partitioned by a threshold $c \in \mathbf{R}$ and

$$\delta(x; c) = \begin{cases} 1 & \text{if } d_m(x) \geq c \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

It is more complex to construct the classification function using multi-resolution feature representation. Additional decisions need to be made on how to select scales that are likely to have positive contributions to the classification problem and how to use these scales constructively. These decisions need to be made *a priori*, i.e, based on the training set only, although they can be evaluated *posteriori*.

In this work, we attempted to accomplish the scale selection and combination processes based on training samples only. We then provided a comprehensive evaluation of each feature representation and classification scheme.

5.2 Scale Selection

In practice, it is difficult to make an optimal selection of scales *a priori*, in part, because forgeries are unavailable. An effective scale representation, however, should have a large ratio between inter- and intra-personal variations at least for the training samples. It is, thus, assumed that the inter-to-intra personal variation ratio of training samples could be used to predict the efficacy of a feature representation for classification.

Let us denote each feature vector at resolution k of subject i as $\mathbf{f}_i^k$, and the intra-personal and overall means of the feature vector as $\bar{\mathbf{f}}_i^k$ and $\bar{\mathbf{f}}^k$, respectively. The inter-to-intra personal variation ratio was defined as

$$J^k = \text{tr}(\mathbf{W}^{k-1} \mathbf{B}^k) \quad (5.2)$$

where $\mathbf{W}^k$ was the average intra-personal covariance matrix $\mathbf{W}_i^k$,

$$\mathbf{W}_i^k = \langle (\mathbf{f}_i^k - \bar{\mathbf{f}}_i^k)(\mathbf{f}_i^k - \bar{\mathbf{f}}_i^k)^T \rangle \quad \text{for } i = 1, \dots, 15 \quad (5.3)$$

and $\mathbf{B}^k$ was the inter-personal covariance matrix

$$\mathbf{B}^k = \langle (\bar{\mathbf{f}}_i^k - \bar{\mathbf{f}}^k)(\bar{\mathbf{f}}_i^k - \bar{\mathbf{f}}^k)^T \rangle \quad (5.4)$$

This ratio provides an invariant measure under ordinary (nonsingular) scaling and rotation transformations of the feature vector[22]. The invariant property of J^k can be demonstrated based on the fact that the trace of a matrix is invariant under orthogonal linear transformations [23].

Let $\mathbf{T}$ denote an orthogonal linear transformation, where $\mathbf{T}$ may be either a scaling operation or a rotation operation or both. When the transformation is applied to each feature vector, the ratio J^k becomes

$$J^k = \text{tr}((\mathbf{T}\mathbf{W}^k\mathbf{T}^T)^{-1}(\mathbf{T}\mathbf{B}^k\mathbf{T}^T)) \quad (5.5)$$

where $\mathbf{T}^T$ is the transposition of $\mathbf{T}$. Because $\mathbf{T}$ is orthogonal, $\mathbf{T}^{-1} = \mathbf{T}^T$ and, thus, J^k remains the same with or without the transformation.

Figure 5.1: Inter-to-intra personal variation ratio as a function of resolution index

The ratio, J^k , was computed for each scale in the multi-resolution, wavelet representation except when $\mathbf{W}^k$ was close to singular. The J^k per degree-of-freedom as a function of resolution index k is illustrated in Figure 5.1 for both projections and images (note that the ratio was not computed when $k = 0$ for projections and $k = 0, -1, -2, -3$ for images because of the singularity of $\mathbf{W}^k$). This ratio was also computed for the global geometric features.

The computed ratios show that J^k is not a monotonic function of the resolution index k for the projections. This inter-to-intra personal variation ratio tended to be high when the scale was less detailed. Similar results were obtained for the multi-resolution representations of images, although several ratios of detailed scales were not

computable. These ratios provided partial information on the discrimination power of each scale for signature verification, and were used later as weights in building the multi-resolution classification function (see next section for details).

5.3 Classification Systems

Two systems were developed for signature verification using multi-resolution feature representation: a vector quantization (VQ) classifier [24] and an artificial neural network [25, 26]. The VQ classifier allows the use of a consistent procedure in processing feature vectors of different length or resolution, and is easy to implement because its training and classification procedures are relatively simple. The major disadvantage of the VQ classifier is that it can only partition the feature space using hyperspheres, and is incapable of drawing complicated, non-linear class boundaries.

As is well known, the artificial neural network is capable of delineating arbitrarily complicated class boundaries [26]; but the performance of a neural network classifier heavily depends on the network architecture and training method [27, 28, 29]. An extensive search of an optimal network for signature verification using multi-resolution representation is beyond the scope of this work. Thus, only a simple multi-layer network was used for the verification problem. The emphasis here was to compare the performance of the VQ classifier and a basic network classifier when single- and multi-resolution feature representations were used. Such a comparison was expected

to provide useful information for the future development of signature verification systems.

5.3.1 Vector Quantization Classifier

5.3.1.1 System Description

A VQ classifier is a collection of code vectors (codebook) that *optimally* represent patterns in a given feature space. The codebook can be constructed using a statistical clustering algorithm and its optimality can be measured by quantization errors that are computed as the minimal differences between the code vectors and randomly sampled patterns in a given feature space. Obviously, a large number of training patterns are needed for a rigorous construction of VQ codebook. In this work, the optimization of the codebook is not pursued because of the limited size of the database. The number of code vectors were arbitrarily set at 3 to approximate the 15 training signatures from each person, i.e., there would be 3 templates (cluster centroids) for each person at each single resolution. Thus, upon completion of training, a total of 45 centroids were obtained as the verification templates at each single resolution.

The training of a VQ classifier has supervised learning and unsupervised learning components. The supervised learning involves the pre-defined association of training vectors with each person. The unsupervised learning involves the clustering of each

person's training vectors into a number of clusters, and computing the centroids of each cluster. The clustering algorithm used here is described as follows:

1. Compute the centroid, $\mathbf{M}$, of the first cluster as the mean of all training vectors.
2. For any n number of clusters, computing the centroid of the cluster of maximum variance, $\mathbf{m}_m$, and add a perturbation, $\mathbf{p}$, to create the $n + 1$ th centroid, $\mathbf{m}_{n+1}$,

$$\mathbf{m}_{n+1} = \mathbf{m}_m + \mathbf{p} \quad (5.6)$$

$$\mathbf{p} = \underbrace{(-\epsilon, \dots, -\epsilon)}_{N/3}, \underbrace{(\epsilon, \dots, \epsilon)}_{N/3}, \underbrace{(-\epsilon/2, \dots, -\epsilon/2)}_{N/3} \quad (5.7)$$

where ϵ is a small number and N is the length of the feature vector.

3. Recluster all training samples based on the $n + 1$ centroids, $\mathbf{m}_i$ ($i = 1, \dots, n + 1$).

The reclustering process has the following steps:

- Re-assign all training samples to the $n + 1$ clusters based on the minimum distances between each training sample and the $n + 1$ centroids.
- Re-compute the $n + 1$ centroids as the centroids of the $n + 1$ clusters.
- Go back to the re-assignment step until the relative reduction of total error (dE) is less than a given threshold, where dE is defined as

$$dE = \frac{E_{k-1} - E_k}{E_k} \quad (5.8)$$

E_k is the total error at the k th reclustering loop and is defined as

$$E_k = \sum_{i=1}^P \min_j \|\mathbf{f}_i - \mathbf{m}_j\| \quad (5.9)$$

where P is the total number of training samples.

4. Going back to step 2 until the number of cluster is equal to the given number of code vectors.

To use multi-resolution features effectively for classification, three issues need to be resolved. First, there are two possible ways to use the multi-resolution information. The multi-resolution information could be used accumulatively, i.e., the information at a lower resolution is included to build the next level of representation. This accumulative process would be similar to the human examination of a signature when first examining it at a distance and then bringing it closer at each step. Note that the accumulation of information only applies to the same type of feature representation (either projection or image).

The multi-resolution information also can be used independently, i.e., the information at a higher resolution is analyzed independently without including the information of a lower resolution. This process would be similar to the human examination of a signature by first looking at the gross features and then analyzing only the next level of details at each step of the verification process. Because of the orthonormal nature of the wavelet transformation, significant differences between these two approaches were not expected.

Second, there are two alternatives to integrate multi-resolution information for signature verification. One alternative is to compute the minimum distance between

a sample signature and the templates at each single resolution and use the average minimum distance for making the final acceptance/rejection classification. In this method, each single resolution contributes to the classification process indiscriminately and no other prior knowledge is necessary to construct the classification function. The second alternative is to use a weighted summation of the minimum distances of each single resolution for the final classification so that each scale has a pre-specified role in the classification process. Here, the weight vector would need to be pre-determined and should reflect the discriminating power of each scale. The first alternative is referred to as the uniform approach and the second as the weighted approach.

In mathematical terms, let $\mathbf{d} = (d_1, d_2, \dots, d_n)^T$ denote the minimum distance vector between a sample signature x and the templates at n different scales. Let $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$ denote a vector of weights where $w_i \geq 0 \forall i$ and $\|\mathbf{w}\| = 1$. A multi-resolution classification function, δ , can be constructed as

$$\delta(x; \mathbf{w}, c) = \begin{cases} 1 & \text{if } \sum_i w_i d_i \geq c \\ 0 & \text{otherwise} \end{cases} \quad (5.10)$$

where c is a given threshold. This classification function corresponds to the uniform approach when $w_i = \text{const} \forall i$, and corresponds to the weighted approach when $w_i \neq \text{const} \forall i$. As discussed later in this Chapter, an optimal weight vector that linearly separates between genuine and forgery signatures can be obtained once the identities of all test samples are known.

Third, all classification thresholds should be predetermined. One way to obtain these thresholds is to use the variance of each cluster as an estimation of the threshold. This approach is feasible only if the size of the training set is sufficiently large. It is also possible to set each threshold in a position so that the total error in both false rejection and false acceptance is minimal. Apparently, the determination of threshold by minimizing the total classification error tends to underestimate the misclassification rate r because the threshold estimated from a small set of training samples is unlikely to be optimal and, thus, the actual misclassification rate r is unlikely to reach the minimum. These threshold settings, however, are legitimate for the performance comparisons and evaluations of different classification systems with the understanding that comparisons and evaluations are made based on the upper performance limits of the systems.

In this work, attempts to predetermine optimal settings of the acceptance/rejection thresholds were not made because the main purpose here was to evaluate the use of multi-resolution representation for signature verification. Thresholds were set to minimize the total error rate of a classification function using the following procedures:

1. Compute the minimum distance between each test sample and the templates.
2. Compute the range of minimum distances of test genuine signatures.
3. Let the threshold vary from the minimum to the maximum of the computed range using a small incremental step (1% of the range).

4. Count the total number of misclassified samples at each threshold value.
5. Choose the threshold that minimizes the total number of misclassified samples as the final classification threshold.

The final total error rate, r , was defined as the ratio between the total number of misclassified samples at the selected threshold and the total number of test samples.

Each classification function was evaluated for the accumulative and independent multi-resolution feature representations, and for the uniform and weighted methods of integrating multi-resolution information. In the uniform approach, the weight vector was simply set to $\frac{1}{N}$, where N is the number of scales in the multi-resolution system. In the weighted approach, the weight vector was the inter-to-intra personal variation ratios derived from training samples as described in the previous section. These ratios reflected the relative discriminating power of each scale for signature verification.

5.3.1.2 System Performance

Verification rates, $R = 1 - r$, are shown in Figure 5.2 for each scale of the multi-resolution representation when the inter-to-intra personal variation ratio is relatively large (> 1). For comparison, a few verification rates are also included when the inter-to-intra personal variation ratio is relatively low ($k = 0$ for projection and $k = -3$ for image).

Verification error rates are tabulated in Table 5.1 for all single- and multi-resolution classification systems. In the Table, the false-rejection rate is defined as the ratio between the number of mis-rejected genuine signatures and the number of all test genuine signatures. The false-acceptance rate for simulated forgeries is defined as the ratio between the number of simulated forgeries that were accepted as true signatures and the total number of simulated forgeries. Similarly, the false acceptance rate for free-handed forgeries is defined as ratio between the number of free-handed forgeries that were accepted as true signatures and the total number of free-handed forgeries. The total error rate is defined as the ratio between the number of all misclassified signatures (genuine, free-handed, and simulated) and the total number of test signatures.

These results indicate that the verification rate correlated well with the inter-to-intra personal variation ratio, J^k , which lends support to the claim that the inter-to-intra personal variation ratio of training samples can be used in evaluating the efficacy of a scale. Relatively low error rates were obtained when the computed inter-to-intra personal variation ratio was relatively large. For each single resolution system, the error rate of the accumulative approach was consistently lower than that of the independent approach. An exception occurred for the lowest resolution when the feature representations of the two approaches were the same.

Figure 5.2: Verification rate as a function of resolution index for projections (top plot) and images (bottom plot)

Wavelet Rep.	Feature Type	Res. Index	False Rej. (%)	False Acc. Sim. (%)	False Acc. Free. (%)	Total (%)
	Geom.	0	11.7	10.3	4.0	8.7
Acc.	Image	-6	21.3	14.7	4.0	13.3
		-5	6.7	20.0	1.3	9.3
		-4	12.0	20.0	4.0	12.0
		-3	25.3	28.0	5.3	19.6
	Proj.	-6	16.0	18.7	5.3	13.3
		-5	12.0	16.0	2.7	10.2
		-4	14.7	21.3	0.0	12.0
		-3	16.0	21.3	1.3	12.9
		-2	14.7	26.7	2.7	14.7
		-1	14.7	28.0	5.3	16.0
		0	14.7	26.7	5.3	15.6
	Unif. Wt.	Multi	8.0	17.3	1.3	8.9
		Multi	6.7	13.3	0.0	6.7
Ind.	Image	-6	21.3	14.7	4.0	13.3
		-5	14.7	20.0	0.0	11.6
		-4	14.7	22.7	10.7	16.0
		-3	33.3	45.3	40.0	39.6
	Proj.	-6	16.0	18.7	5.3	13.3
		-5	14.7	22.7	2.7	13.3
		-4	20.0	28.0	10.7	19.6
		-3	32.0	37.3	17.3	28.9
		-2	33.3	46.7	18.7	32.9
		-1	34.7	41.3	17.3	31.1
	Unif. Wt.	Multi	16.0	16.0	4.0	12.0
		Multi	10.7	10.7	0.0	7.1

Table 5.1: Error rate of the VQ classifier

For the multi-resolution systems, the accumulative approach performed slightly better than the independent approach. In either accumulative or independent approach, the error rate was lower for the weighted summation of minimum distances than for the uniform summation of minimum distances. The multi-resolution system could have a higher error rate than that of a single-resolution system when all scales were included uniformly or indiscriminately. This observation merely confirms that some scales do contribute negatively to the classification process if these scales carry the same weight as others. Both scale selection and evaluation are necessary to build an effective multi-resolution system as demonstrated by the classification results using the weighted approach. The best classification function was the accumulative, weighted multi-resolution system.

5.3.1.3 System Evaluation

The error rates presented above do not provide information on the reliability of the classification functions using single- and multi-resolution feature representations. The reliability of each classification function was evaluated here using procedures of statistical, linear discrimination analysis. This evaluation was accomplished primarily by computing the statistical separability ratio (F-ratio) between genuine and forgery groups [30]. For easy comparison, all projections from the feature space to the real axis $\mathbf{R}$ were normalized to $[0, 100]$ based on the range of projections of all signatures from the three groups.

Let d_{ik} denote the minimum distance between the feature vector, $\mathbf{f}_i^k$, of the i th sample from the k th group and the classification templates, $\mathbf{t}_j$,

$$d_{ik} = \min_j (\|\mathbf{f}_i^k - \mathbf{t}_j\|) \quad (5.11)$$

where $k = 0, 1, 2$ denoting genuine, simulated forgery, and free-handed forgery groups, respectively. Let m_k and s_k^2 denote the mean and variance of d_{ik} . The F-ratio between genuine and forgery signatures is defined as

$$F_k = \frac{m_0 - m_k}{s_0^2 + s_k^2} \quad k = 1, 2 \quad (5.12)$$

This F-ratio was also computed for the multi-resolution classification functions. Let $\mathbf{m}_k$ denote the mean of $\mathbf{d}_{ik}$, the minimum, multi-resolution distance vector of the i th sample from the k th group ($k = 0, 1, 2$). Let S_k denote the covariance matrix of $\mathbf{d}_{ik}$. The F-ratio for the multi-resolution classification function is a function of the weight vector $\mathbf{w}$,

$$F_k(\mathbf{w}) = \frac{\mathbf{w}^T(\mathbf{m}_0 - \mathbf{m}_k)(\mathbf{m}_0 - \mathbf{m}_k)^T \mathbf{w}}{\mathbf{w}^T(S_0 + S_k)\mathbf{w}} \quad (5.13)$$

Note that the inter-to-intra personal variation ratios were used earlier as a weight vector because these ratios were the only readily available information about the discriminating power of the scales. Now, the identities of all test samples are known and it is possible to choose a weight vector that maximizes the linear separability between genuine and forgery groups. It can be shown that the weight vector that maximizes the F-ratio is [31]

$$\hat{\mathbf{w}}_k = (S_0 + S_k)^{-1}(\mathbf{m}_0 - \mathbf{m}_k) \quad (5.14)$$

			Genuine		Simulated			Free Handed		
Rep.	Type	Res.	Mean	SD	Mean	SD	F_1	Mean	SD	F_2
	Geom.	0	7.37	6.28	25.24	14.01	1.35	48.22	26.56	2.24
Acc.	Image	-6	9.09	7.71	20.56	14.93	0.47	41.95	27.83	1.29
		-5	7.34	7.08	20.78	13.08	0.82	45.39	23.08	2.48
		-4	9.41	8.18	20.25	11.80	0.57	39.37	19.60	1.99
	Proj.	-6	7.19	6.18	19.08	15.15	0.53	38.83	30.68	1.02
		-5	7.35	7.36	20.41	12.21	0.84	42.75	27.05	1.60
		-4	7.98	7.38	19.54	11.27	0.73	41.24	24.38	1.70
		-3	8.62	8.49	17.89	10.27	0.48	37.87	22.49	1.48
		-2	7.63	7.99	14.66	9.38	0.33	34.01	21.08	1.37
		-1	7.73	7.78	13.88	8.86	0.27	33.58	20.33	1.41
Ind.	Image	-6	9.09	7.71	20.56	14.93	0.47	41.95	27.83	1.29
		-5	8.90	8.54	21.84	13.84	0.63	47.59	22.21	2.64
		-4	11.26	9.80	20.85	12.73	0.36	32.75	16.56	1.25
	Proj.	-6	7.19	6.18	19.08	15.15	0.53	38.83	30.68	1.02
		-5	10.68	9.07	21.11	15.68	0.33	44.40	26.65	1.43
		-4	15.03	9.94	23.78	15.35	0.23	43.24	20.77	1.50
		-3	20.01	11.87	21.29	15.84	0.00	32.96	18.02	0.36
		-2	18.36	10.97	16.56	12.64	0.01	33.33	16.79	0.56
		-1	16.81	9.72	14.92	10.34	0.02	33.83	18.42	0.67

Table 5.2: Statistics of VQ distance at each single resolution

Therefore, the performance of the current multi-resolution classification functions can be compared to this optimal linear discrimination function. The known group identities of test samples also made it possible to evaluate separately the multi-resolution classification functions for simulated and free-handed forgery classifications.

The F-ratio, together with the mean and standard deviation of d_{ik} , for each single-resolution classification function are shown in Table 5.2. These statistics for the multi-resolution classification functions and separate error rates are illustrated in Figure 5.3, and are tabulated in Tables 5.3 and 5.4.

Figure 5.3: Evaluation results of multi-resolution VQ classifier

		Genuine		Simulated			Free Handed		
Rep.	Type	Mean	SD	mean	SD	F_1	Mean	SD	F_2
Acc.	Unif.	5.83	5.58	17.96	10.85	0.99	39.97	23.79	1.95
	Wt.	5.32	5.21	19.44	11.14	1.32	42.81	23.56	2.42
	Opt.	7.46	7.15	28.24	13.53	2.02	58.90	22.17	4.88
Ind.	Unif.	6.10	5.82	17.38	10.66	0.86	39.77	21.11	2.37
	Wt.	5.47	5.68	19.40	11.03	1.26	43.32	21.55	2.89
	Opt.	5.40	5.41	29.11	16.01	1.72	48.37	20.19	4.23

Table 5.3: Statistics of VQ distance for multi-resolution systems

		Simulated			Free Handed		
Rep.	Type	False Rej.	False Acc.	Total	False Rej.	False Acc.	Total
Acc.	Unif.	2.7	22.7	12.7	0.0	8.0	4.0
	Wt.	2.7	17.3	10.0	0.0	2.7	1.3
	Opt.	2.7	9.3	6.0	0.0	2.7	1.3
Ind.	Unif.	16.0	16.0	16.0	2.7	6.7	4.7
	Wt.	5.3	16.0	10.7	1.3	4.0	2.7
	Opt.	1.3	17.3	9.3	1.3	2.7	2.0

Table 5.4: Separate error rates of multi-resolution systems

These results indicate that, for single-resolution systems, the F-ratio, in general, was < 1 for simulated forgery; but > 1 for free-handed forgeries. Reliable, single-resolution representations for detecting free-handed forgery included the geometric features, wavelet transformation of image at resolution -5 , and wavelet transformation of projection at resolution -4 .

For multi-resolution systems, the F-ratio is relatively high, again, when the multi-resolution information is used selectively. The F-ratio of the weighted multi-resolution classification function could be doubled if additional information were available to optimize its performance. The multi-resolution classification function is in the 90% range for discriminating simulated forgeries and can reach above 98% for discriminating free-handed forgeries. Thus the multi-resolution classification system designed here could have practical values for monitoring free-handed forgeries. An informal inspection of verification errors suggest that most errors were signatures produced by individuals who had either clear and tidy handwriting or inconsistent (significantly variable) signatures.

5.3.2 Neural Network Classifier

5.3.2.1 System Description

An artificial neural network generally is used for pattern recognition/classification, in which features of an unknown pattern are the input and the categorical identity of the pattern is the output [32, 33, 29]. Typically, training samples of each category

are used collectively to train the recognition/classification network. A verification network would differ slightly from a recognition network in that (1) the category (or the claimed identity) of a genuine or forgery signature is known and the network is used to determine if the signature belongs to a particular pattern class, and that (2) forgery samples generally are not available for training.

It is possible to use the same procedure for building a network recognizer/classifier as for building the network verifier. A signature would be accepted if it were recognized as one of the target categories and the output level at the corresponding output node were relatively high. The signature would be rejected if none of the output nodes produced a high level output. There are many effective methods to build such a network. This network recognizer/classifier, however, could not be expanded easily. It would require modifications of the overall network architecture and network weights using all training samples when additional pattern categories have to be incorporated in the system.

An alternative would be to build a network for each pattern class (or individual). Each network would have only one output node and the acceptance/rejection would be made based on the level of the output. The same network architecture could be used for all individuals. The advantages of this approach are (1) the system could be easily expanded to include additional persons, and (2) the acceptance/rejection threshold could be tuned individually to count for personal differences. As described below, the system also could process multi-resolution features directly.

We took the second approach. The network for each individual was a three-layer, feedforward network. The number of input nodes was equal to the length of the feature vector. The number of hidden nodes was equal to the number of input nodes. There was only one output node for each network. The activation function of each node was the conventional sigmoid function. This network was referred to as the *elementary network*.

For multi-resolution representation, an elementary network was built for each scale of the multi-resolution representation. The output from each elementary network of the same individual was fed into one output node which was the output layer of the multi-resolution network. We only used scales where the computed inter-to-intra personal variation ratio, J^k , was relatively large (> 1). This scale selection was necessary for reducing complexity of network structure and for shortening the training time. A schematic illustration of the multi-resolution network is shown in Figure 5.4.

The training set for each elementary network consisted of the training samples of all 15 subjects. The training algorithm was the classical Back-Propagation method with a learning rate $\alpha = 0.9$ [25]. A momentum term was added in updating the weights ($\beta = 0.6$). The training loop would not terminate until the total error was less than 10^{-5} and the error difference between consecutive training iterations was less than 1×10^{-8} or a total of 2×10^4 training iterations had been exhausted. Inspections of the training error as a function of training iterations indicated that the

Figure 5.4: The multi-resolution network

network converged relatively quickly in most cases and the average number of training iterations used was about 2000. This fast convergence might be due to the relatively easy partitioning of the feature space when forgeries were not used in training.

5.3.2.2 System Performance

Verification error rates for each single-resolution network and the multi-resolution network are shown in Table 5.5. These results can be summarized as follows:

1. Overall, the performance of the multi-resolution network classifier was comparable to that of the multi-resolution VQ classifier.
2. The multi-resolution network yielded the lowest verification error rate when independent features were used.

Wavelet Rep.	Feature Type	Res. Index	False Rej. (%)	False Acc. Sim.(%)	False Acc. Free. (%)	Total (%)
	Geom.	0	21.3	8.0	6.7	12.0
Acc.	Image	-6	26.7	10.7	2.7	13.3
		-5	14.7	16.0	0.0	10.2
		-4	5.3	20.0	1.3	8.9
	Proj.	-6	22.7	8.0	4.0	11.6
		-5	28.0	16.0	4.0	16.0
		-4	14.7	20.0	2.7	12.4
		-3	14.7	22.7	2.7	13.3
		-2	14.7	22.7	2.7	13.3
		-1	9.3	22.7	4.0	12.0
	All	Multi	9.3	10.7	1.3	7.1
Ind.	Image	-6	26.7	10.7	2.7	13.3
		-5	13.3	22.7	2.7	12.9
		-4	13.3	16.0	1.3	10.2
	Proj.	-6	22.7	8.0	4.0	11.6
		-5	29.3	8.0	1.3	12.9
		-4	21.3	14.7	6.7	14.2
		-3	29.3	12.0	12.0	17.8
		-2	24.0	26.7	24.0	24.9
	All	Multi	4.0	9.3	1.3	4.9

Table 5.5: Error rate of network classifier

3. For each single scale of the multi-resolution representation, there was no apparent correlation between the detailed level of a scale and the error rate.
4. There was a case (projection at resolution index -1) where the network did not converge within a reasonable amount of time when using the independent feature set.

5.3.2.3 System Evaluation

The F-ratio defined in the previous section was computed here for the output of each network. Classification error rates were also computed separately for simulated and free-handed forgeries. These results are tabulated in Tables 5.6 and 5.7.

These results can be summarized as follows:

1. The F-ratio for the single resolution network classifier was not a monotonic function of the resolution index. The F-ratio for some single resolutions can be significantly higher than others.
2. The F-ratio for the multi-resolution network classifier is comparable to the multi-resolution VQ classifier when processing simulated forgeries.
3. The F-ratio is significantly higher than the VQ classifier (including the optimized linear discrimination function) when processing free-handed forgeries.

These results suggest that the network classifier is advantageous when the class boundary is complicated, but not substantially overlapped (non-separable). The

Rep.	Type	Res.	Genuine		Simulated			Free Handed		
			Mean	SD	mean	SD	J_1	Mean	SD	J_2
	Geom.	0	73.01	24.15	35.19	30.89	0.93	16.45	28.23	2.32
Acc.	Image	-6	71.57	27.11	42.93	33.31	0.44	15.85	28.38	2.02
		-5	76.04	21.06	41.16	31.80	0.84	12.55	20.97	4.56
		-4	76.51	17.51	41.29	24.76	1.35	11.18	13.30	8.83
	Proj.	-6	71.42	32.44	32.69	35.33	0.65	22.29	21.54	1.59
		-5	65.48	31.43	44.39	35.73	0.20	17.85	26.82	1.33
		-4	76.14	20.47	43.56	32.14	0.73	13.35	23.87	3.99
		-3	67.73	26.04	36.44	32.74	0.56	9.99	17.36	3.40
		-2	64.77	24.73	33.90	30.39	0.62	10.49	15.65	3.44
		-1	70.64	23.65	37.99	30.46	0.72	15.13	18.56	3.41
	All	Multi	72.30	17.12	37.51	23.73	1.41	12.98	15.42	6.63
Ind.	Image	-6	71.57	27.11	42.93	33.31	0.44	15.85	28.38	2.02
		-5	78.53	17.99	41.71	33.60	0.93	15.97	24.02	4.34
		-4	68.39	26.79	31.73	29.45	0.85	4.95	7.95	5.15
	Proj.	-6	71.42	32.44	32.69	35.33	0.65	22.29	21.54	1.59
		-5	67.51	34.23	39.18	40.47	0.29	26.85	34.67	0.70
		-4	62.54	34.25	33.11	35.56	0.36	10.90	21.13	1.65
		-3	49.46	36.48	18.41	31.10	0.42	15.40	25.78	0.58
		-2	33.84	32.73	17.81	23.01	0.16	23.39	31.10	0.05
	All	Multi	70.15	16.33	32.72	20.58	2.03	13.99	13.96	6.84

Table 5.6: Statistics of network output

Rep.	Simulated			Free Handed		
	False Rej.	False Acc.	Total	False Rej.	False Acc.	Total
Acc.	5.3	13.3	9.3	2.7	2.7	2.7
Ind.	2.7	10.7	6.7	1.3	2.7	2.0

Table 5.7: Separate error rates of network classifier

network classifier is capable of effectively integrating independent, multi-resolution information for classification even if each single resolution was not effective. One reason that the network classifier does not demonstrate any advantage over the VQ classifier for simulated forgery classification may be that the network could not be properly constrained due to the lack of forgeries during training. This is not reflected in the classification of free-handed forgeries because the constraints provided by the training samples of other individuals may be sufficient for building the classifier.

CHAPTER 6

CLASSIFICATION SYSTEM USING GRID FEATURES

6.1 Introduction

In the previous Chapter, detailed descriptions of classification systems using multi-resolution feature representations are presented. The multi-resolution representation largely comprised features derived from the wavelet transformation. While wavelet transformation is capable of decomposing an image using a wide range of scales, it treats an image as an integrated identity and is not capable of providing structural information of image elements such as the angle of a corner and the curvature of an arc. The structural information, although important to the human visual system, cannot be easily incorporated into a computer system. There are technical difficulties in the representation and comparison of the structural information, in part, because a signature image often comprises many intersecting, irregular curves and conventional methods for specifying simple curves are not applicable. For example, a tree of chain codes would have to be used to represent a signature. Even if the representation were accurate, a useful comparison of tree structures of different size and topology would be practically impossible [34, 35].

An alternative is to use size-normalized, raw images for verification. The size normalization allows comparisons of images of different sizes while preserving the topology of the images. The spatial sparsity of a bi-level signature image, however, would require a 2-dimensional alignment algorithm so that meaningful comparisons of raw images can be made. Such an alignment is computationally expensive and difficult to implement.

In this work, we took a grid approach in an attempt to use local structural information for signature verification. In this approach, the structural information of image elements were partially obtained based on a grid that was superimposed on the size-normalized signature image. Images were compared at the grid level. This procedure can be conceptualized as one of the human methods for image verification — cutting the images into small pieces and comparing them on a piece-by-piece basis when it is difficult to compare complicated images as a whole. In this Chapter, we provide the procedure and results using this grid approach.

6.2 Grid-Feature Representation

6.2.1 Composition of Feature Vector

The grid-feature vector included the *boundary code* of each grid and the total number of pixels inside the grid. The boundary code was a binary vector that was

defined as

$$b_i = \begin{cases} 1 & \text{pixel at } i\text{th position} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

where i was the distance from the upper-left corner of each grid when moving counter-clockwise on the boundary. The length of the boundary code was equal to 4 times the side of each square grid.

This boundary code is an incomplete, linear approximation to the structure within each grid when the size of the grid is relatively small. It is incomplete because there are multiple ways to linearly connect a given set of boundary locations within a grid. This ambiguity could be alleviated to a certain extent by the continuity constraint of line strokes, i.e., line strokes typically move continuously through each grid and this continuity can be used to eliminate some uncertainties in connecting the boundary locations. Two supplemental measures were taken to reduce this ambiguity further:

1. The intersection between a line stroke and the grid was thinned so that each intersection was represented by only one code element b_i . The boundary code was considered circularly continuous in the thinning process [36].
2. The number of pixels within each grid was included in the feature vector.

It was realized that the size normalization would introduce structural distortions to the signature image. Because distortions due to size-normalization tend to be similar to signatures of the same person, it was assumed that these distortions would not substantially degrade the verification process. A significant problem, however, existed

in feature comparison. Because crossings between grid and line strokes were not exact even for signatures from the same person, a rigid comparison of the boundary codes would not be useful. An effective classification system using grid features had to be flexible enough to tolerate expected variations.

6.2.2 Alignment of Feature Vector

There were a number of alternatives to make the grid feature comparison flexible. For example, the boundary code could be shifted circularly for a given number of steps. Hamming distance between boundary code vectors could be computed at each step and the minimum could be used as the final measure of difference. This approach was global and could not count for local differences. On the other hand, if each individual element in the boundary vector was allowed to shift locally, there would be difficulties in resolving matching conflicts of neighboring pixels. An ideal approach was to allow each element in the boundary code to shift locally in a coordinated way so that the global difference between boundary codes could be minimized. This was accomplished using dynamic programming procedures [37, 38].

6.2.2.1 Fundamentals of Dynamic Programming

The alignment of one-dimensional functions using dynamic programming is well documented in the literature [39, 40, 41], and its basic principles are summarized here.

The inputs to this process are two functions or vectors (i.e., boundary code vectors), and the outputs are a warping function and the degree of remaining mismatch. The warping function provides a path for the optimal match between the inputs. The remaining mismatch, referred to as the total cost in dynamic programming terminology, reflects the inherent discrepancies between the two inputs.

Suppose that two vectors to be matched are

$$A = a_1, a_2, \dots, a_i, \dots, a_N \quad (6.2)$$

$$B = b_1, b_2, \dots, b_i, \dots, b_N \quad (6.3)$$

where a_i and b_i are the i th element of vectors A and B , respectively. Computation of the warping function can be viewed as the process of finding a minimum-cost path through the lattice of points depicted in Figure 6.1.

Let C be the warping function

$$C = c_1, c_2, \dots, c_k, \dots, c_N$$

where $c_k = [i_k, j_k]$ is a pair of pointers to the samples being matched. For each c_k , the cost can be calculated as

$$d(c_k) = |a_{i_k} - b_{j_k}| \quad (6.4)$$

The final warping function is determined by minimizing the total-cost function

$$D(C) = \sum_{k=1}^N d(c_k) \quad (6.5)$$

Figure 6.1: A schematic illustration of dynamic programming

subject to certain constraints [39, 40, 41]. Despite the seemingly innumerable possibilities of pair matching, an optimal path is guaranteed to exist [37].

6.2.2.2 Implementation

In this work, the search for the optimal match between two boundary code vectors involved the following steps:

1. For a given searching limit d , the lower limit and the upper limit for searching in the vertical direction were given by $\max(1, i - d)$ and $\min(N, i + d)$, respectively, at each step i in the horizontal direction. These searching boundaries defined a polygon, as shown in Figure 6.1.
2. Within these searching limits, the path for connecting each point (i, j) to previous points in the lattice was determined by minimizing the total cost for reaching the current position. Specifically,
 - (a) Start with $i = 1, j = 1$. Because there were no predecessors, the Hamming distance between the two samples was computed as the starting cost.
 - (b) Loop through all (i, j) . In each loop, the costs from the current point (i, j) to the predecessors $(i - 1, j)$, $(i, j - 1)$, and $(i - 1, j - 1)$ were computed. The connection between point (i, j) and one of its predecessors was made such that the cost for making the connection plus the cost for reaching the

particular predecessor was minimized. This minimum cost was stored as the cost for reaching point (i, j) .

- (c) When $i = j = N$, the search was terminated and a complete path was retrieved from the terminating point.

3. The final total cost was equal to the cost for reaching the destination point (N, N) .

A flow chart of the algorithm is shown in Figure 6.2.

6.3 Grid Classifier

6.3.1 System Description

Because the grid feature vector was largely binary and relatively long, it would not be appropriate to use either a VQ classifier or neural network classifier. Considering that the training set was relatively small (15 for each subject) in our database, all training samples were used as templates for verification.

The Hamming distance rather than the Euclidean distance was computed to measure the difference (or similarity) between grid-boundary codes. This distance was computed using binary operations to reduce computational cost. The Hamming distance, d_h , between boundary codes, $\mathbf{b}_l$ and $\mathbf{b}_k$, was defined as

$$d_h = \frac{1}{N} \sum_{i=1}^N \mathbf{b}_l i \oplus \mathbf{b}_k i \quad (6.6)$$

Figure 6.2: Flow chart of the dynamic programming algorithm

where N is the length of the boundary code and the $\oplus$ denotes the EXCLUSIVE OR logic. This distance plus the normalized difference in pixel count were computed on a grid-by-grid basis. The mean of distances from all grid elements yielded the projection from the grid feature space to $\mathbf{R}$. The classification threshold was again chosen to minimize the total error rate.

A number of grid size and overlapping strategies were used to evaluate this grid approach. The initial grid size was set to 4 and was doubled at each step until it was 32. The initial grid step size also was 4 and doubled, if possible, until it was 32. The grid step size was limited to be less or equal to the grid size. The relationship among grid size, grid step size, and grid overlapping are illustrated in Figure 6.3. It was expected that error rates would be low when using large overlapping ratios (area of overlapping/grid size) because grid overlapping provided continuity in signature comparison. It was also expected that the added flexibility in feature comparison using dynamic programming would reduce verification error rates significantly.

6.3.2 System Performance

The results using this grid approach are shown in Figure 6.4a. These include the error rates of all grid division schemes when verification was made using

1. original (not thinned) boundary code,
2. thinned boundary,

Figure 6.3: An illustration of grid size, grid-step size, and grid overlapping

3. thinned boundary code and dynamic programming matching procedure (searching limit = 2)
4. thinned boundary code and dynamic programming matching procedure (searching limit = 4)

As can be seen, high error rates occurred when using the original (not thinned) boundary code as the feature representation. The error rates were reduced by 5-6% when boundary codes were thinned and by an additional 5-6% when dynamic programming procedure was used for code matching. These error reductions were most prominent when the grid size was relatively large and grid step size was relatively small. The lowest error rate under each experimental condition consistently occurred when grid overlapping ratio (area of overlapping / area of grid) was around the maximum.

(a)

(b)

Figure 6.4: Total error rate of the grid classifier

The error rates as a function of the search limit in dynamic programming, when grid size was 32 and grid step size was 4, are shown in Figure 6.4b. We see that when the overlapping ratio was maximum, the searching limit of dynamic programming produced a relatively limited effect on verification accuracy. As expected, the system performed the best when the searching limit was neither too small nor too large.

6.3.3 System Evaluation

In this evaluation, the reliability of the grid classifier as an independent system was determined using the F-ratio defined in the previous Chapter. Verification error rates of the grid classifier were computed separately for the classification of simulated and free-handed forgeries.

The potential contribution of the grid classifier as part of a multi-resolution system was also estimated. Because of the unique characteristics of the grid feature representation, it was difficult to integrate the grid classifier into the multi-resolution systems described in the previous Chapter. For example, the grid feature vector could not be easily included into the multi-resolution, network systems because a subnet using the grid feature representation was difficult to built. The length, sparsity, and variability of the grid feature vector made it practically impossible to train a convergent network. These characteristics of the grid feature vector also prevented the development of a basic VQ classifier using conventional clustering and template building methods.

It is possible, however, to include the minimum distance between each test sample and the templates (all training samples) in the grid feature space as an extra element to the minimum distance vector, $\mathbf{d}$, of a multi-resolution VQ system because this minimum distance is an equivalent measure to those in the minimum distance vector. Using this augmented minimum distance vector, the potential contribution of the grid feature vector as part of a multi-resolution VQ system can be estimated.

In the evaluation here, the minimum distance vector of the independent, multi-resolution VQ system was augmented by the minimum distance in the grid feature space. The optimal weight vector, $\mathbf{w}$, for the augmented multi-resolution classification function (equation 5.10) was obtained using equation 5.14. F-ratios of this augmented system were then computed. Similar computations were not made for the accumulative, multi-resolution VQ system because the grid feature vector was an independent component in this estimation process.

Results of these evaluations are shown in Tables 6.1 and 6.2. These results indicate that the grid classifier is less effective than the multi-resolution systems for the classification of simulated forgeries. It is, however, comparable to the multi-resolution systems for the classification of free-handed forgeries (see Table 6.1). The F-ratios of the augmented multi-resolution VQ system conform to this observation. The inclusion of the grid feature in the independent, multi-resolution VQ system slightly decreased the F-ratio for the classification of simulated forgeries (from 1.72 to 1.62);

Type	Genuine		Simulated			Free Handed		
	Mean	SD	Mean	SD	F_1	Mean	SD	F_2
Indep.	7.02	7.32	18.49	12.61	0.62	40.76	20.81	2.34
Multi.	8.90	7.75	27.12	12.01	1.62	59.42	19.66	5.71

Table 6.1: Statistics of the grid classifier

Simulated			Free Handed		
False Rej.	False Accp.	Total	False Rej.	False Accp.	Total
8.0	16.0	12.0	2.7	1.3	2.0

Table 6.2: Error rates of the grid classifier

but increased the F-ratio for the classification of free-handed forgeries (from 4.23 to 5.71).

An informal inspection of forgeries seems to indicate that skilled forgers tend to be very careful about simulating the details of a signature. A free-handed forgery, however, is unlikely to be close to a genuine signature at detailed level. Thus, the the grid approach can be used either as an alternative or as a special part of a multi-resolution system for the classification of free-handed forgeries. It is simple to implement and can be equally effective.

CHAPTER 7

SUMMARY AND CONCLUSIONS

The initial objective of this work was to develop algorithms for the computer verification of handwritten signatures. The goal was to build a system that could detect free-handed forgeries accurately and could monitor simulated forgeries as well. To a large extent, this goal has been accomplished. We took two different routes in reaching the destination. First, we applied the recently-developed wavelet transformation and developed a multi-resolution system. Second, we invented a grid method that uses approximation of local structure of a signature for verification. Both systems exhibited high accuracy and reasonable reliability for the classification of free-handed and simulated forgeries.

One crucial step in computer signature verification is to extract features of a signature image that are effective in reflecting personal idiosyncrasy such as the slant angle and the overall orientation of a signature image. A number of algorithms for effectively characterizing a signature image are developed and presented in detail. Some of these algorithms may have wider applications in other areas of image processing and computer vision.

The development and evaluation of the multi-resolution system was tested using standard statistical procedures. We predicted that an initial estimation of inter-to-intra personal variations of signatures based on training samples provides valuable information for the development of a multi-resolution system when forgeries were not available. We expected that a multi-resolution system, if designed properly, is superior to a single-resolution system, in part, because the human visual system processes multi-resolution information constantly, and is the most comprehensive system for image classification and recognition. Experimental evidences in this work are supportive to our prediction and expectation.

We then presented a grid approach to the signature-verification problem. This approach was motivated by the premise that the local structural information of an image is important to the human visual system and should be included in the verification system. Again, effective methods for extracting local structure information of a signature were not available. We invented a simple method, the grid approach, for approximating local structural information. This method makes intuitive sense and has been shown to be effective when flexibility is introduced in the feature comparison process.

Obviously, the algorithms developed here need to be more extensively evaluated using a significantly larger database before practical applications are attempted. We regret that a standard database is not available and our database is substantially limited by physical resources. We believe, however, that this work has added a few

solid stones to the future development of computer signature verification systems and to the field of computer vision in general.

REFERENCES

- [1] M. Ammar, Y. Yoshida, and T. Fukumura, "Structural description and classification of signature images," *Pattern Recognition*, vol. 23, pp. 697–710, 1990.
- [2] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification — the state of the art," *Pattern Recognition*, vol. 22, pp. 107–131, 1989.
- [3] M. Ammar, Y. Yoshida, and T. Fukumura, "Off-line preprocessing and verification of signatures," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, pp. 589–902, 1988.
- [4] M. Ammar, Y. Yoshida, and T. Fukumura, "Feature extraction and selection for simulated signature verification," in *Computer Recognition and Human Production of Handwriting* (R. Plamondon, C. Suen, and L. Simner, eds.), pp. 61–67, Singapore: World Scientific Publ. Co., 1989.
- [5] M. Parizeau and R. Plamondon, "A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 710–717, 1990.
- [6] W. Nemcek and W. Lin, "Experimental investigation of automatic signature verification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 4, pp. 121–126, 1974.
- [7] H. Andrews, "Walsh functions in image processing, feature selection, and pattern recognition," in *Proc. 1971 Applications of Walsh Functions Symp.*, pp. 26–32, 1971.
- [8] R. Nagel and A. Rosenfeld, "Computer detection of freehand forgeries," *IEEE Transactions on Computers*, vol. C26, pp. 895–905, 1977.
- [9] M. Ammar, "Performance of parametric and reference pattern based features in static signature verification: A comparative study," in *Proc. of IEEE Conference on Pattern Recognition and Artificial Intelligence*, pp. 646–648, 1990.

- [10] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math.*, vol. 15, pp. 723–736, 1984.
- [11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure and Appl. Math.*, vol. 41, pp. 909–996, 1988.
- [12] C. Chui, *An Introduction to Wavelets*. New York: Academic Press Inc., 1989.
- [13] S. G. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $l^2(r)$," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.
- [14] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Review*, vol. 31, pp. 614–627, 1989.
- [15] D. Esteban and C. Galand, "Application of quadrature mirror filters to split band voice coding schemes," in *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, pp. 191–195, May 1977.
- [16] G. Battle, "A block spin construction of ondelettes, part i: Lemarie functions," *Commun. Math. Phys.*, vol. 110, pp. 601–615, 1987.
- [17] A. Oppenheim and R. Schaffer, *Discrete -Time Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1989.
- [18] G. Pirani and V. Zingarelli, "An analytical formula for the design of quadrature mirror filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, pp. 645–648, June 1984.
- [19] P. C. Millar, "Recursive quadrature mirror filters — criteria specification and design method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 413–420, April 1985.
- [20] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.
- [21] A. Paeth, "A fast algorithm for general raster rotation," *Graphics Interface 86*, pp. 77–81, 1986.

- [22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic Press, second ed., 1972.
- [23] G. Strang, *Linear Algebra and Its Applications*. New York: Harcourt Brace Jovanovich, 3rd ed., 1988.
- [24] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.
- [25] D. Rumelhart, G. Hinton, and R. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (D. Rumelhart and J. McClelland, eds.), vol. 1, pp. 318–362, MIT Press, 1986.
- [26] R. Lippman, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 1, pp. 4–22, 1987.
- [27] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," *Proc. of IEEE*, vol. 78, pp. 1415–1442, Sept. 1990.
- [28] B. Hunt, Y. Qi, and D. Dekruger, "Fuzzy classification using set membership functions in the back propagation algorithm," *Heuristics, Journal of Knowledge Engineering*, vol. 5, pp. 62–74, 1992.
- [29] Y. Qi, B. Hunt, and N. Bi, "The use of fuzzy membership in network training for isolated word recognition," *Proc. of IEEE International Conference on Neural Networks*, vol. 2, pp. 1823–1828, March 1993.
- [30] M. DeGroot, *Probability and Statistics*. Menlo Park, California: Addison-Wesley, 1986.
- [31] I. K. Sethi and A. K. Jain, *Artificial Neural Networks and Statistical Pattern Recognition*. New York: Elsevier Science Pub. Co., 1991.
- [32] Y. H. Pao, *Adaptive Pattern Recognition and Neural Networks*. New York: Addison-Wesley, 1989.

- [33] Y. Qi and B. Hunt, "Voiced-Unvoiced-Silence classifications of speech using hybrid features and a network classifier," *IEEE Transaction on Speech and Audio Processing*, vol. 1, pp. 250–255, April 1993.
- [34] A. Jian, *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.
- [35] B. Pan, S. Wu, and G. Yan, "A method fo recognizing handprinted characters," in *Computer Recognition and Human Production of Handwriting* (R. Plamondon, C. Suen, and L. Simner, eds.), pp. 61–67, Singapore: World Scientific Publ. Co., 1989.
- [36] R. Stefanelli and A. Rosenfeld, "Some parallel thinning algorithms for digital pictures," *J. ACM*, vol. 18, pp. 255–264, 1971.
- [37] G. Nemhauser, *Introduction to Dynamic Programming*. New York: Wiley, 1966.
- [38] Y. Qi, "Time normalization in voice analysis," *J. Acoust. Soc. Amer.*, vol. 92, pp. 2569–2576, 1992.
- [39] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–23, pp. 67–72, Feb. 1965.
- [40] G. White and R. Neely, "Speech recognition experiments with prediction band-pass filtering, and dynamic programming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–24, pp. 183–188, 1976.
- [41] T. Parsons, *Voice and Speech Processing*. New York: McGraw-Hill, 1987.