Real Time Methods for Face Recognition

Afshin David

A thesis

submitted to the School of Graduate Studies and Research

in partial fulfillment of

the requirement for the degree of

Master of Applied Science

in Electrical Engineering

Ottawa-Carleton Institute of Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

University of Ottawa

Ottawa, Ontario, K1N 6N5

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-19944-4

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Abstract

Identification of individuals on the basis of facial features is the most natural method of distinguishing one individual from another. Automating such a process, based upon quantifiable measures, is of great interest in a variety of applications, such as passport identification and automatic teller machine verification. The most crucial aspect of such applications is their tolerance with respect to variations in facial expressions and the noise introduced by the operating environment.

In this thesis, various face recognition methods are evaluated under conditions of real-time response, varying operating factors, and implementation feasibility. The approaches are based on histogram mapping, wavelet transform, Karhunen and Loeve transform, and optical correlation techniques.

A brief review of the basic concepts in optics is first presented. This is followed by a detailed review of optical methods in pattern recognition. A comprehensive background of algorithmic approaches for face recognition is described. A detailed analysis of the *photobook* system, which is based on the Karhunen and Loeve transform (KLT), is presented. It is argued that, even though the KLT possesses many useful attributes in image processing applications, the performance of KLT face recognition systems is based entirely upon the initial training set. A method for choosing the proper training set is presented. Novel statistical methods that exploit the stationary behaviour of the operating environment are introduced. It is shown that under the condition that control may be exercised on the operating environment, these methods provide a satisfactory result in real-time. The application of histogram, moment, and 2-D discrete wavelet transforms in statistical methods is described. A novel optical correlation based system is presented. It is shown that such a system tolerates changes in facial expressions and can operate under real time constraints.

Acknowledgments

I would like to thank both my supervisors Dr. S. Panchanathan and Dr. J. Chrostowski for their guidance and support. My thanks also go to the members of the Photonic Systems group at National Research Council of Canada.

I am grateful to Dr. T. Aboulnasr, Dr. M. Lynos, and Mr. M. K. Mandal for addressing many critical issues encountered during this research.

Contents

LIST OF FIGURES

LIST OF TABLE

Chapter 1

Introduction

The ability to uniquely and deterministically identify a pattern, based on incomplete or corrupted sets of observations, is of great interest in a variety of scientific and non-scientific disciplines. Most importantly, it is of interest to distinguish different members or patterns of the same class. Motivation for this work arises from applications, such as automatic teller machines, where the uniqueness of the database entry is mainly based upon facial features. The most notable aspect of such applications is that, unlike traditional pattern matching applications, a relative restriction can be imposed on position, illumination, and background. The main objective of this thesis is to investigate techniques that discriminate among members of the same class, namely faces, in contrast to general image retrieval techniques, where the main objective is to separate different classes.

1.1 Definition of the problem

The organization of this study is as follows: (*i*) identify the problems in face recognition, (*ii*) survey current techniques, and (*iii*) propose novel and improved solutions. The goal of the research is to identify a given face in a relatively controlled environment. Specifically, the identification of an individual from a prerecorded face database, given the frontal view of a person with

relative control over the position of the face, is required. The goal of the research was not to identify a face among other patterns or to track a moving face.

1.2 Current techniques

The ability to extract a minimum amount of relevant data, from a pattern, is the distinguishing measure amongst different face recognition techniques. A minimum amount of data must be extracted to reduce the computational cost. Furthermore, to tolerate changes in facial expressions, only relevant data must be extracted. Unlike an image compression application, perfect or near perfect reconstruction is of no interest, even though the extracted data must have a one-to-one relationship with their respective face. Figure 1.1 presents techniques that are considered in this thesis.

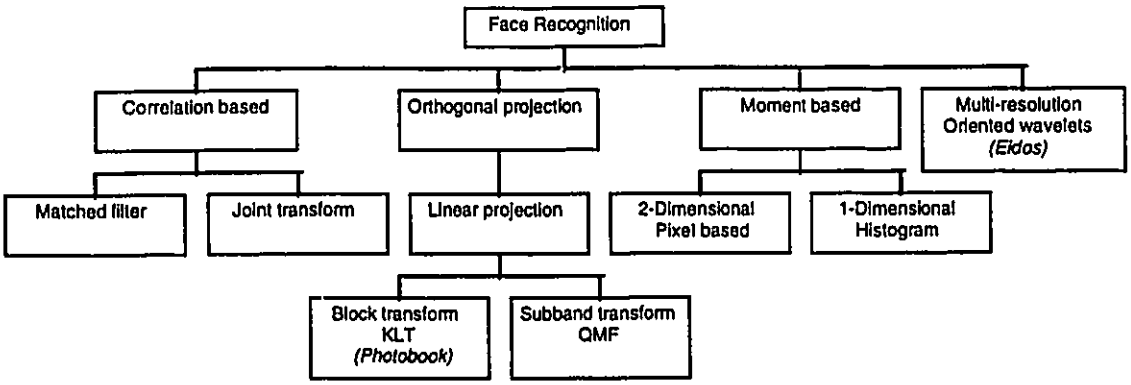


Figure 1.1 Taxonomy of the Considered Approaches

The main premise behind moment based approaches is that the *pdf* of a random variable can be uniquely described by its moments [6]. Depending on the random variable of interest (amplitude and position or just amplitude), either

two dimensional moments of the random variable (*e.g.* pixel) or moments of its *pdf* (*e.g.* histogram) may be defined.

A class of pattern recognition methods that is particularly amenable to optical implementation is the correlation based approach. The objective of such an approach is to maximize the ratio of the expected value of the correlation peak to its variance [5]. There are two classes of optical correlators, distinguished based upon their implementation techniques, matched filter correlators (MFC) and joint transform correlators (JTC).

The class of minimum square error techniques can be broken down into linear and non-linear (not discussed here) approaches. The class of linear mean square decomposition can be further broken down into three major branches (*i*) block transformation, (*ii*) lapped orthogonal transformation (not considered here), and (*iii*) subband transformation [1]. Among block transform methods the most notable is the Karhunen and Loeve transform (KLT). The goal of the KLT is to represent a pattern space with a minimum number of bases [2]. For the purpose of the wavelet decomposition, the class of quadrature mirror filter (QMF) banks, especially those developed by Daubechies [3], is considered in this thesis. The goal of this technique is to decompose a signal to orthogonal bands where each band differs from other bands in its frequency resolution.

Recently, non-orthogonal wavelets have been gaining attention in the field of computer vision as they can model the human visual profile. Recent studies have demonstrated their applicability in the field of face recognition [8].

Clearly, the above taxonomy is not fully comprehensive or unique. For example, different branches of Figure 1.1 may easily overlap. In addition, other approaches such as morphological and artificial intelligence are not considered in this thesis. The main reasons for the selection as illustrated in Figure 1.1 are two fold (*i*) wealth of literature and (*ii*) proven performance.

1.3 Thesis organization

Chapter 2 provides background materials for optical approaches considered in this thesis. Chapter 3 presents a review of electronic methods considered. Chapter 4 presents a detailed description and performance evaluation of the *photobook* system. Chapter 5 describes and evaluates novel methods, histogram-moment and wavelet-histogram, developed in this thesis. Chapter 6 describes a novel optical face recognition system and presents performance evaluation. Finally Chapter 7 draws conclusions and suggests directions for future work.

1.4 Main contributions

The main contributions of this thesis are:

1. Detailed investigation and analysis of a KLT based face recognition system
2. Design and real time implementation of a histogram-moment based face recognition system
3. Design and implementation of a wavelet-histogram based face recognition system

4. Design and real time implementation of an optical database face recognition system

Chapter 2

Review of basic concepts in optics

In this chapter, the basic concepts of optical pattern recognition are presented. A short review of light propagation in a medium and common optical devices in optical image processing is presented. The chapter continues with an introduction to the Fourier optics, and finally optical methods for pattern recognition are discussed.

Among many interesting aspects of optics, as compared to electronics, those that are particularly of interest in image processing are as follows: (*i*) three dimensional span, (*ii*) free space span, and (*iii*) information representation.

Images are complicated data sets both in content and representation. Hence, to achieve a high throughput in image processing applications, it is imperative to perform many operations simultaneously (three dimensional span), not to overwhelm the underlying data network (free space span), and to be able to represent the information in a manner that best suits the application; be it by amplitude, phase, polarization, or frequency modulation.

Another useful aspect of optics is holographic storage where both the magnitude and phase of an image can be recorded and reconstructed. Furthermore, the ability of optics to perform the two dimensional Fourier

transform and parallel multiplication is of great interest in an image processing application.

In the following sections, a quick review of the basic concepts in optics, such as light and its propagation through a medium, is presented. The details can be found in elementary optics book such as [9].

2.1 *Interference*

The phenomenon that describes the enhancement or diminution of two or more waves in an observation plane is referred to as interference. To understand its effect a vectorial treatment of two interfering waves is presented.

Assume two radiation fields E_1 and E_2 are propagating through different paths and coinciding at point p. The electric field at point p will be

$$E_p = E_1 + E_2 \tag{2.1}$$

Defining *irradiance* or radiant power density as the time average of the amplitude square, irradiance at point p is

$$I_p = < E_1^2 + E_2^2 + 2E_1 \cdot E_2 > \tag{2.2}$$

where,
$<x>$ is the time average of x, and
$E_1 \cdot E_2$ is the dot product of the two electric fields.

Under the assumption that the fields are parallel, the above can be simplified as [9]

$$I_p = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Theta \tag{2.3}$$

where,
I_i is the irradiance of E_i equal to $\langle E_i^2 \rangle$, and
Θ is the total phase difference due to path and source differences.

The last term of Equation (2.3) is the interference term. For a given path, changes in Θ are due to source phase variation. If Θ changes randomly, the contribution of the interference term averages to zero and the sources are said to be mutually incoherent. Therefore, some degree of coherence (*i.e.* $\langle \cos\Theta \rangle \neq 0$) is required to observe interference.

Assuming the sources are mutually coherent, the path difference between the two sources will change, as the point of observation changes. Thus spatially separated interference fringes are observed. The observed interference fringes change values between

$$I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \tag{2.4}$$

when $\Theta = 2m\pi$ and

$$I_{\min} = I_1 + I_2 - 2\sqrt{I_1 I_2} \tag{2.5}$$

when $\Theta = (2m+1)\pi$, where m is an integer.

If I_1 is equal to I_2,

$$I_{\max} = 4I_1 \qquad and \qquad I_{\min} = 0 \qquad also \qquad I = 4I_1 \cos^2\left(\frac{\Theta}{2}\right) \tag{2.6}$$

where I is the intensity at the observation point.

Figure 2.1 shows a plot of irradiance versus phase difference.

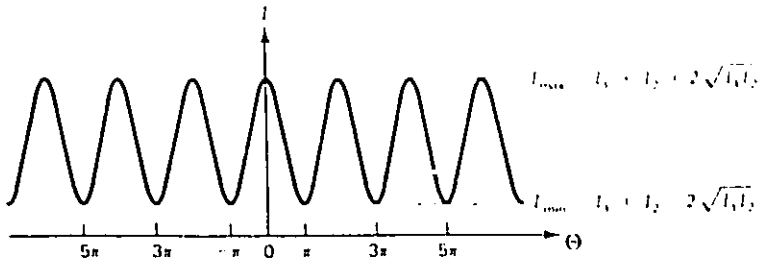


Figure 2.1 Plot of Irradiance Vs Phase Difference

2.2 *Coherence*

As mentioned in the previous section, to observe interference, sources must have a constant phase relationship. A constant phase can be observed either over a given period (*i.e.* temporally coherent), or over a spatial distribution of the source (*i.e.* spatially coherent).

2.2.1 Temporal coherence

Temporal coherence is a measure of the spectral purity of a light emitting source. All sources emit light that can be described by a train of harmonic sequences. Figure 2.2 is a sample of such a train.

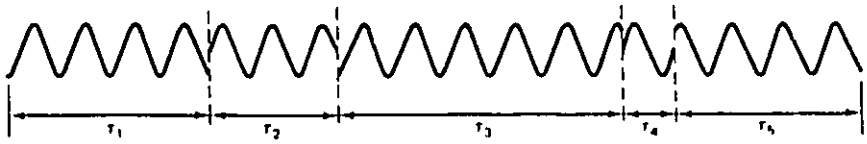


Figure 2.2 Typical Non-continuous Train of Light Pulse Sequences

The average time for which a source is considered to be coherent is called the *coherence time*. The distance traveled during this time is called the *coherence length*. The coherence time is inversely proportional to the spectral width of a source. It then follows [9] that

$$
\begin{aligned}
l_t &= c\,\tau_0 \\
l_t &= \frac{c}{\Delta f} \\
f &= \frac{c}{\lambda} \\
l_t &= \frac{\lambda^2}{\Delta\lambda}
\end{aligned}
\tag{2.7}
$$

where,
l_t is the coherence length,
c is the speed of light in the medium,
τ_0 is the coherence time,
λ is the average wave length,
Δf and $\Delta\lambda$ are spectral bandwidths.

For example, white light, with center frequency around 550 nm and extending from 400 to 700 nm, has a coherence length of 1000 nm; whereas the coherence length of CO_2 lasers is around 11 km [9].

2.2.2 Spatial coherence

Unlike temporal coherence where the main interest is the evolution of the phase correlation in time, spatial coherence measures the spatial distribution of the phase correlation for a given radiation source. That is, it measures changes in the phase of the radiated field due to extension of the source (*i.e.* phase difference of distinct points).

The main advantage of coherent sources (temporal and spatial) is that, when the radiated fields from such sources interact, their complex amplitudes add up; whereas, when incoherent fields interact, their irradiances add up (*i.e.* the phase information is lost) [9][10][11].

2.3 *Diffraction*

When a wave front is altered by an aperture, the observed shadow displays fine details that are related to the shape of the aperture. This effect is called diffraction. To gain an appreciation of diffraction and its role in optical processing, two classical examples are presented, diffraction due to a single slit and diffraction due to multiple slits. No attempt is made to provide a rigorous treatment of the derivations, rather the results are presented [9][11]. The following treatment is based on the far-field assumption (Fraunhofer diffraction)

$$L \gg \frac{A}{\lambda} \tag{2.8}$$

where,
L is the observation distance,
A is the area of the aperture, and
λ is the wavelength.

2.3.1 Diffraction from a single slit

Let us assume the observation point p is at infinity (this can be easily satisfied by the application of a positive lens) and that a plane wave is obstructed by a single slit, as shown in Figure 2.3.

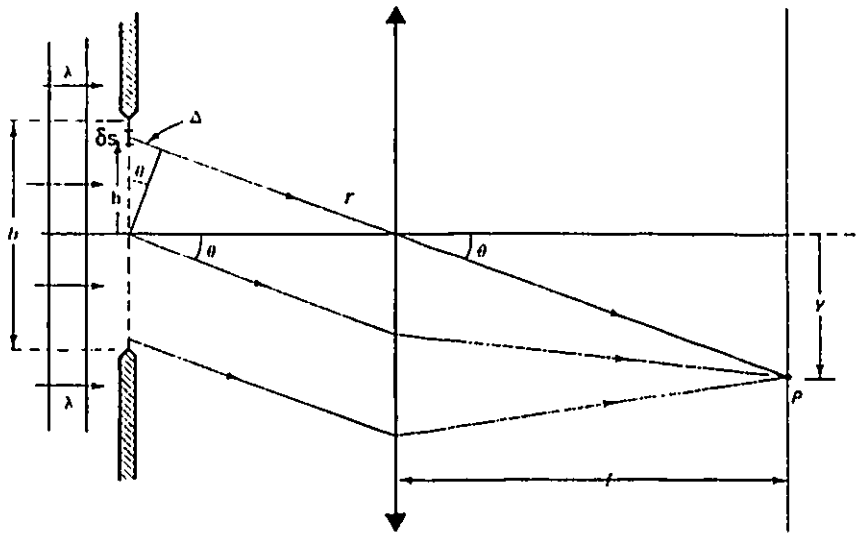


Figure 2.3 Single Slit Diffraction Setup[9]

According to the Huygens-Fresnel principle [9], each point on the incoming wave can be considered as a secondary spherical source (*i.e.* wavelets). The contribution of a wavelet at point p is

$$\delta E_p = \left(\frac{E_L \delta s}{r} \right) e^{j[kr - wt]} \tag{2.9}$$

where,
E_L is the amplitude per unit width of the slit,
δ_s is the differential area of the wave front,
k is the wave number, and
r is the distance of the wavelet to point p.

For a small slit, the effect of r on the amplitude of Equation (2.9) is insignificant; however, due to the exponential nature of phase, such an assertion cannot be made regarding the phase at point p. Thus Equation (2.9) can be simplified as follows:

$$r = r_0 + \Delta$$
$$\Delta = h \sin\theta \tag{2.10}$$
$$\delta E_p = \left(\frac{E_L \delta s}{r_0} \right) e^{j[kr_0 + ks\sin\theta - wt]}$$

where r_0 is the center to center distance between the slit and observation plane.

Integrating over the opening, the field at point p is

$$E_p = \frac{E_L h}{r_0} e^{j(kr_0 - wt)} \text{sinc}(\beta) \tag{2.11}$$

where,

$\beta = \frac{1}{2} kb \sin\theta$, and

b is the width of the slit.

Apart from the constant phase factor and ratio factor, E_P is the Fourier transform of a square pulse function. The irradiance observed at point p is

$$I = I_0 \text{sinc}(\beta)^2 \tag{2.12}$$

where $I_0 \propto \left(\frac{E_L h}{r_0} \right)^2$.

2.3.2 Diffraction from many slits

The procedure described above can be generalized to many slits[9].

$$E_p = \frac{E_L h}{r_0} e^{j(kr_0 - wt)} \text{sinc}(\beta) \frac{\sin N\alpha}{\sin\alpha} \tag{2.13}$$

where,

$$\alpha = \frac{1}{2} ka \sin\theta,$$

a is the center to center distance of the slits, and
N is the number of slits.

The irradiance observed at point p is

$$I = I_0 \left(\frac{\sin N\alpha}{\sin \alpha} \right)^2 \operatorname{sinc}(\beta)^2 \tag{2.14}$$

where $I_0 \propto \left(\frac{E_L h}{r_0} \right)^2$.

Comparing Equations (2.14) and (2.12), the diffraction pattern observed at point p is proportional to the magnitude of the Fourier transform of an aperture sampled by a function f, where

$$f = \left(\frac{\sin N\alpha}{\sin \alpha} \right)^2 \tag{2.15}$$

Figure 2.4 depicts the case for N=8 and a=3b.

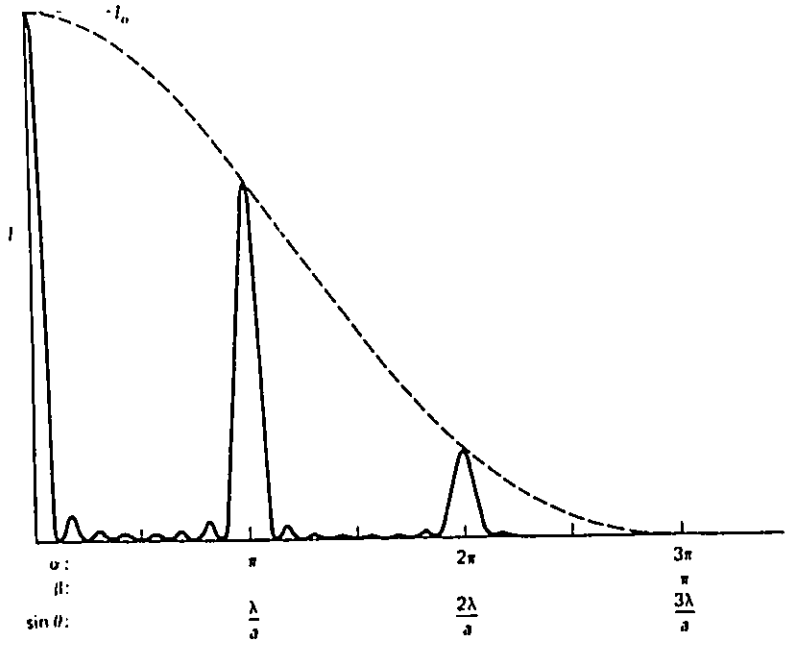


Figure 2.4 Diffraction Pattern of 8 Slits, Spaced Three Times Their Width

As N increases, the secondary peaks move away from the center of the diffraction pattern (go to a null), while the magnitude of the central peak increases by N^2. For a large N, Equation (2.15) resembles a delta sampling train.

2.4 *Fourier optics*

The two general approaches in treating this subject, are linear system [12] and diffraction based [11] analyses. The former is based on linear shift invariant modeling of the light propagation, and the latter on scalar diffraction theory. The former is described in detail.

2.4.1 Linear shift invariant modeling

The material presented in this section is based on [11] and [12]. Again no attempt is made to provide a rigorous derivation, but rather just the results pertinent to this discussion.

The propagation of light through a medium can be characterized as a linear spatial filter $h(x,y,z)$ such that

$$u(x,y,z) = h(x,y,z) * u(x,y,0) \tag{2.16}$$

where,
u is the radiation field,
h is the impulse response of the medium, and
z is the distance between source and observation plane.

It is shown in [12] that the transfer function of $h(x,y,z)$ is

$$H(f_x, f_y) = \begin{cases} e^{j2\pi \frac{\sqrt{1-(f_x\lambda)^2+(f_y\lambda)^2}}{\lambda}} & \text{for} \quad f_x^2 + f_y^2 < \frac{1}{\lambda^2} \\ 0 & \text{Otherwise} \end{cases}$$

$$f_x = \frac{x_f}{z\lambda} \tag{2.17}$$

$$f_y = \frac{y_f}{z\lambda}$$

where,
f_x and f_y are the spatial frequencies,
x_f and y_f are the respective spatial variables, and
z is the distance traveled by light.

With respect to Equation (2.17), it is seen that a large phase distortion is introduced for the high frequency components. However, the magnitude response remains unity within the circular region of radius λ^{-1}.

2.4.2 Fresnel approximation

Given the typical size of an optical setup relative to the wave length, a number of simplifications can be made to Equation (2.17). The main goal in Fresnel approximation is to estimate the spherical surface of a wave front with a quadratic function. With this in mind, Equation (2.17) can be simplified as

$$H(f_x, f_y) = e^{jkz} e^{-jz\lambda\pi(f_x^2, f_y^2)} \tag{2.18}$$

The following binomial approximation has been used.

$$\begin{aligned} &\sqrt{1+b} \approx 1 + \frac{1}{2}b \\ &b = -\left[(f_y\lambda)^2 + (f_x\lambda)^2\right] \end{aligned} \tag{2.19}$$

To obtain the spatial representation of $h(x,y,z)$, after the above simplification, the inverse Fourier transform of Equation (2.18) is taken.

$$\begin{aligned} &h(\Delta x, \Delta y, z) = \frac{e^{jkz}}{jz\lambda} e^{\frac{jk(\Delta x^2+\Delta y^2)}{2z}} \\ &\Delta x = x_d - x_s \\ &\Delta y = y_d - y_s \end{aligned} \tag{2.20}$$

where Δx and Δy are the horizontal and vertical distances traveled by the light (d and s subscripts represent the observation and source planes, respectively).

Thus Equation (2.16) can be described as

$$u(x_d, y_d, z) = h(\Delta x, \Delta y, z) * u(x_s, y_s, 0)$$

$$u(x_d, y_d, z) = \iint_{\infty} u(x_s, y_s, 0) e^{jk\frac{\Delta r^2 + \Delta y^2}{2z}} dx_s dy_s$$

$$u(x_d, y_d, z) = \frac{e^{jkz}}{jz\lambda} e^{(x_d^2 + y_d^2)} \frac{jk}{2z} \iint_{\infty} u(x_s, y_s, 0) e^{jk\frac{x_s^2 + y_s^2}{2z}} e^{\frac{-jk}{z}(x_s x_d + y_s y_d)} dx_s dy_s \tag{2.21}$$

which, aside from the constant amplitude and phase factors, is the Fourier transform of

$$u(x_s, y_s, 0) e^{\frac{jk}{2z}(x_s^2 + y_s^2)} \tag{2.22}$$

where the transform must be evaluated at spatial frequencies f_x and f_y.

2.4.3 Transfer function of a converging lens

If the quadratic phase term

$$e^{\frac{jk}{2z}(x_s^2 + y_s^2)} \tag{2.23}$$

of Equation (2.22) is eliminated, the desired Fourier transform of an aperture described by $u(x_s, y_s, 0)$ can be obtained. Thus an optical device with transfer function

$$H_d(x, y) \propto e^{\frac{-jk}{2d}(x^2 + y^2)} \tag{2.24}$$

can be incorporated in the optical path of an image where the phase terms diminish for appropriate values of z and d.

The simplest device that would satisfy the above criterion is a converging lens that can be characterized as follows [12] :

$$H_l(x,y) = e^{jkn\Delta} e^{\frac{-jk}{2f}(x^2+y^2)} \tag{2.25}$$

where,
n is the index of refraction,
Δ is the mid-section thickness,
f is the focal length, and
k is the wave number.

In qualitative terms, given the convex shape of any converging lens, the portion of the wave going through the mid-section is more delayed than the edges of the wave front. This delay can be regarded as the phase lag of the converging lens required to compensate for the phase lead in Equation (2.22). It can then be shown [12], for an image located d_0 away from the converging lens, that at the back focal plane of the lens the radiation field is described as

$$u_f(x_f, y_f) = \frac{1}{jf\lambda} e^{\frac{jk}{2f}(1-\frac{d_0}{f})(x_f^2+y_f^2)} U_0(f_x, f_y) \tag{2.26}$$

where,
$U_0(f_x, f_y)$ is the Fourier transform of $u(x,y)$, the image, and
f is the focal length of the converging lens.

In Equation (2.26), the spatial frequencies are evaluated at a distance z equal to the focal length of the lens. For the special case where d_0 is equal to f, the following is obtained,

$$\begin{aligned} u_f(x_f, y_f) &= AU_0(f_x, f_y) \\ A &= \frac{1}{jf\lambda} \end{aligned} \tag{2.27}$$

That is, the amplitude scaled Fourier transform of the image is obtained at the back focal plane. A typical optical Fourier transform is depicted in Figure 2.5.

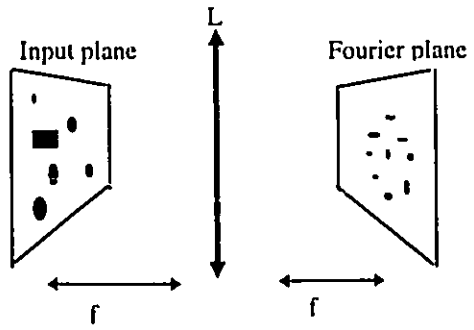


Figure 2.5 Typical 2f Optical Fourier Transformer

With respect to Equation (2.26), it is clear that if the phase information is not important, then there is no restriction on the location of the input plane. As seen in Equations (2.13) and (2.22), the Fresnel approximation at infinity is the same as the Fraunhofer approximation, or as mentioned previously, the focal point of a converging lens is equivalent to an observation point at infinity.

2.5 *Spatial light modulators*

One of the most important reasons that has brought optics to the forefront of parallel processing is its ability to dynamically manipulate the characteristics of the incident light. Spatial light modulators (SLM) are devices that change the amplitude, phase or polarization of the incident light.

Methods by which the above are accomplished are complicated and involved. In this thesis, only a subset of these methods, as they pertain to this work, is considered. In general, two major methods are used for manipulating

the incoming wave; by scattering or by changing the polarization of the incoming light [13]. In what follows, only the latter is considered in detail.

2.5.1 Polarization based light modulators

This class of light modulators can be divided into two types depending on how the input image or control signal is being presented. If the image is presented as an incoherent optical sequence, the displays are referred to as liquid crystal light valves (LCLV). On the other hand, if the control signals are provided electrically, the SLM is referred to as liquid crystal television (LCTV) [5]. LCTV is the type of SLM used in this research, and its operational basis is considered in the next section.

2.5.1.1 Liquid crystal TV

The liquid crystal TV is a 90 degree twisted nematic liquid crystal [15], sandwiched between two parallel polarizers. When no electric field is present, the plane of polarization for linearly polarized light is rotated through 90 degree by twisted liquid crystal molecules, as shown in Figure 2.6*a)*, and no light is transmitted through the second polarizer. Under an applied electric field, the twist and the tilt of the liquid crystal molecules are altered, resulting in a greater fraction of the light retaining the initial polarization direction and thus passing through the second polarizer [16], as shown in Figure 2.6*b)*.

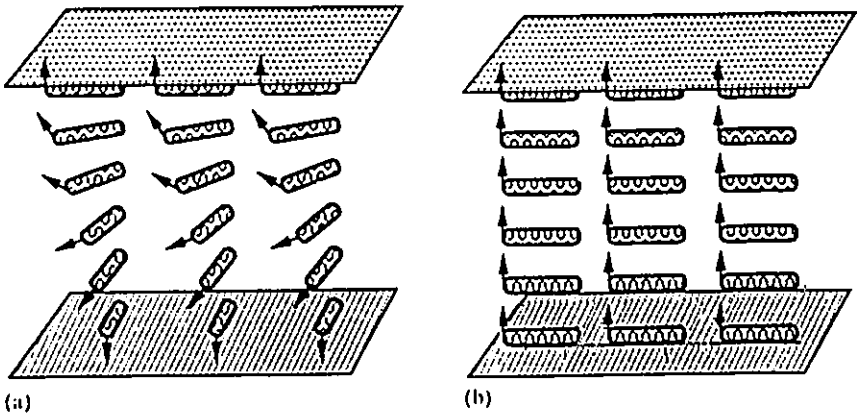


Figure 2.6 Operation of the Twisted Nematic Liquid Crystal
a) In the Absence of External Electric Field
b) In the Presence of External Electric Field
Arrows Point to the Polarization Direction

Regardless of the implementation technique used in SLMs, they can be modeled as two adjacent perpendicular gratings. Thus with respect to Equation (2.15) for a large number of pixels, an SLM becomes a simple sampling grid. It should be noted that this sampling grid can be quite different from those encountered in traditional digital signal processing, since a random bias will be introduced in the frequency plane, which depends upon the relative intensity of the neighboring pixels (see Section 2.1).

2.6 *Quantum well devices*

Since their introduction in the late 60's by Esaki and Tsu [17], quantum well devices have proven their feasibility in a variety of electrical and optical applications. Cutoff frequencies in the order of tens of GHz are common in such devices [18]. To fully appreciate their capabilities and implications in optics and electronics is beyond the scope of this thesis. In what follows, a brief qualitative

implication of quantum wells on non-linear optics will be described namely, the quantum confined stark effect (QCSE).

2.6.1 Multiple quantum well devices

A multiple quantum well (MQW) device is a multi-layer structure where a series of narrow band gap materials (wells) is sandwiched between layers of larger band gap materials (barriers), with the condition that the width of the wells is smaller than the radius of host material excitons (bounded hole-electron pairs). Under such conditions, the quantum size effect (QSE) dominates the electrical and optical properties of these structures [19]. Figure 2.7 depicts a MQW structure.

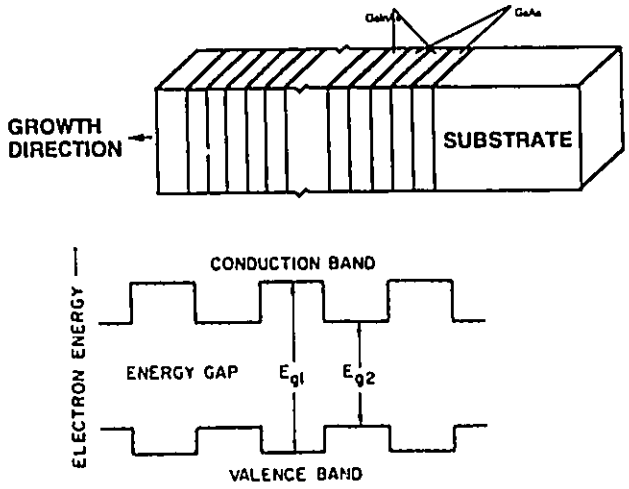


Figure 2.7 Typical MQW Structure

2.6.2 Quantum confined stark effect

Assuming that an electric field is applied perpendicularly to the layers of an MQW structure, excitons trapped in the wells are pulled apart by the applied field. However, unlike bulk materials, excitons are not separated. This is due to

the barriers' walls. Nonetheless, the wave function associated with excitons will be modified due to the compression exerted by the barriers. This in turn will cause a shift in the absorption spectrum of the structure, as shown in Figure 2.8 [14][20][21].

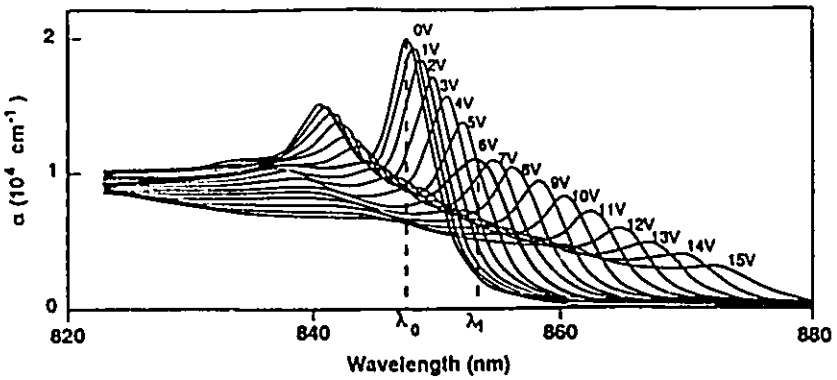


Figure2.8 The Absorption Spectrum Shifts Due to Applied External Field [22].

2.6.3 Semi-insulating quantum wells

First introduced and incorporated in optical correlators by Glass *et al.* at AT&T Bell Laboratories [23][24], these devices exploit the QCSE to modulate the incident read-beam on an MQW structure by the interference pattern produced through a write-beam. Figure 2.9 depicts a typical setup for such a structure.

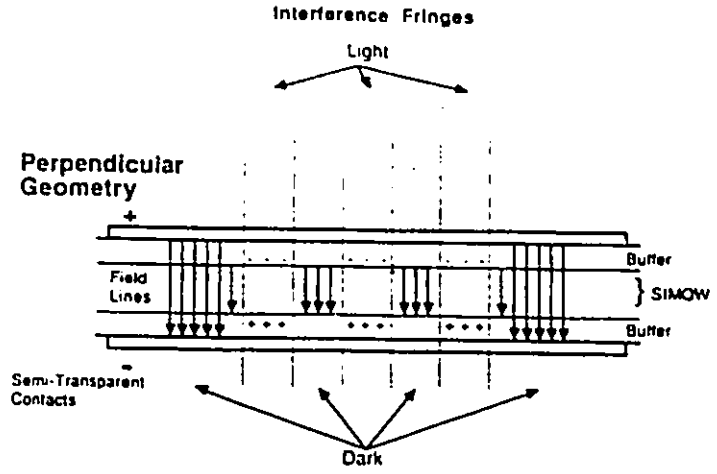


Figure 2.9 Semi-insulating MQW Structure with Perpendicular External Electric Field[23]

A slowly time-varying perpendicular electric field is applied to the structure. The application of this field will cause an absorption spectrum shift throughout the structure due to the QCS effect. An incident light pattern (write-beam) will photo-generate electron-hole pairs that will drift in a direction such as to screen the external electric field. At points on the device where the electric field has been screened, the QCSE is reduced. Thus the intensity variation on the device is recorded as a variation of the absorption spectrum [23]. A second beam of light (read-beam) incident on the device will be modulated according to the absorption variation on the device [24]. This means that, in real time (about $2\mu s$), a read-beam has been modulated according to the interference pattern of the write-beam.

To prevent the diffusion of carriers due to accumulation at the MQW structure and the buffer interface, a slowly time-varying electric field is applied [24]. The device is made semi-insulating by proton or ion implantation, so that

the photo-generated carriers can be trapped in these defects. With this approach the need to pixelate the device is alleviated [25].

2.7 *Review of optical methods in face recognition*

The correlation based pattern recognition method is most commonly used in optics due to the ease of implementation of the Fourier transform. In other words, the main objective of an optical system is to implement the following:

$$y(x,y) = \iint_{\infty} f(u,v)h(u-x,v-y)dudv \tag{2.28}$$

where,
$h(x,y)$ is the reference signal, and
$f(x,y)$ is the scenery (*i.e.* the desired signal and the background noise).

The operation of Equation (2.28) can be described as the downward movement of the filter $h(x,y)$ on the scenery $f(x,y)$, performing multiplication at each point. Assuming a wide sense stationary (WSS) background [30], the results of the multiplication are insignificant, except for the segments where the filter $h(x,y)$ overlaps with its image on the scenery $f(x,y)$. At such points, a peak centered at the location of the reference is observed. Figure 2.10 demonstrates this point in the discrete sense.

12	17	14	4	0
6	1	-7	-1	-8
1	18	12	13	-8
4	3	-6	-13	9
-7	9	6	10	-1

$\otimes$

12	-13
-6	-13

a)

	-0.3	-0.57	-0.55	-0.26	-0.05	0.00
	-0.45	-0.24	0.21	0.33	0.31	-0.09
518*	-0.18	-0.35	-0.31	0.05	0.53	-0.09
	-0.13	-0.55	0.23	*1.00*	-0.18	-0.29
	0.08	-0.13	-0.03	-0.13	-0.62	0.22
	0.18	-0.39	0.06	-0.11	0.26	-0.02

b)

Figure 2.10 Correlation Based Pattern Matching

a) Correlating Scenery and Reference Function

b) Highest Correlation Point is Where a Matched Is Made

It is noted that the highest value in Figure 2.10*b)* is the location of the target in the scenery.

Two optical techniques are considered for the implementation of Equation (2.28), the matched filter correlator (MFC) and joint transform correlator (JTC).

2.7.1 Matched filter correlator

To implement Equation (2.28) optically, consider its Fourier transform

$$Y(f_x, f_y) = F(f_x, f_y)H^*(f_x, f_y) \tag{2.29}$$

where,
$Y(f_x, f_y)$ is the Fourier transform of Equation (2.28),
$F(f_x, f_y)$ is the Fourier transform of the scene, and
$H^*(f_x, f_y)$ is the conjugate transform of the reference signal.

All are evaluated with respect to spatial frequencies f_x and f_y. That is,

$$y(x, y) = IFT\left[F(f_x, f_y)H^*(f_x, f_y)\right] \tag{2.30}$$

where *IFT* is the inverse Fourier transform.

Optically, Equation (2.28) may be constructed as shown in Figure 2.11.

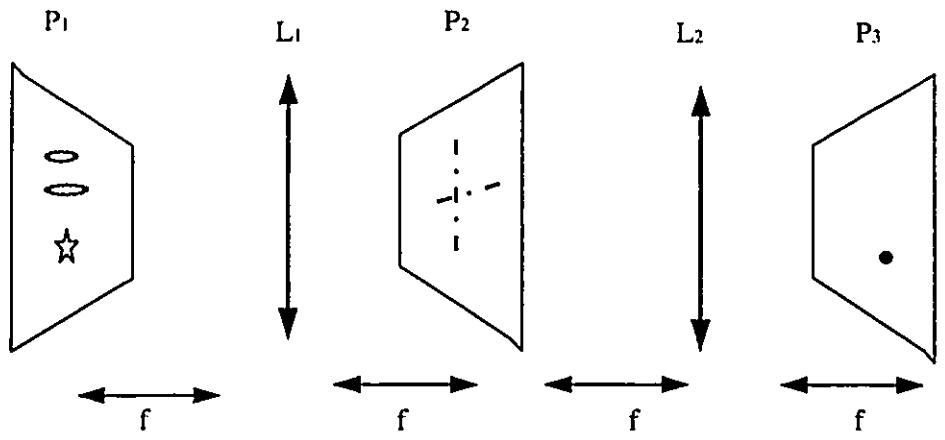


Figure 2.11 Typical MFC Setup

Let P_1 contain the scenery $f(x,y)$ and P_2 the conjugate transform of $h(x,y)$. The light entering the plane P_2 will be the Fourier transform of the scene, and the light leaving P_2 is proportional to the following:

$$F(f_x, f_y)H^*(f_x, f_y) \tag{2.31}$$

Noting that

$$FT\left[FT[x(d)]\right] = x(-d) \tag{2.32}$$

where $FT[x]$ is the Fourier transform of x, the light incident on the plane P_3 is proportional to Equation (2.28).

2.7.1.1 Holographic storage

Since the transfer function of the filter at P_2 has a complex value, it cannot be recorded on photographic media, which are intensity devices. To be able to record the conjugate transform of the reference (*i.e.* $H^*(f_x, f_y)$) on such materials, holographic recording methods must be used.

The most commonly used method for holographic recording is achieved by placing an image on plane P_1 and interfering its Fourier transform on plane P_2 with an off-axis reference beam, which is coherent with respect to the plane wave illuminating plane P_1. This setup is shown in Figure 2.12.

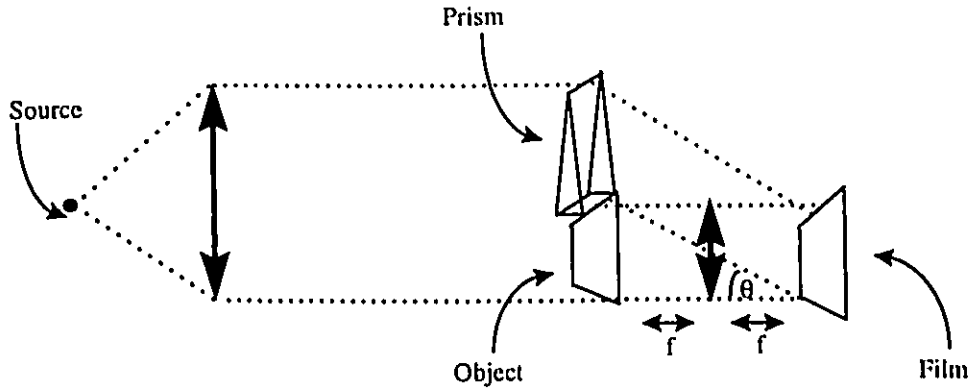


Figure 2.12. A Method to Record a Hologram[12]

The intensity on the film plane is

$$I(f_x, f_y) = k\left|H(f_x, f_y) + Ae^{-jf_x\alpha}\right|^2$$
$$= k\left|H(f_x, f_y)\right|^2 + kA^2 + kAH(f_x, f_y)e^{jf_x\alpha} + kAH^*(f_x, f_y)e^{-jf_x\alpha} \quad (2.34)$$

where,
$\alpha = \frac{\sin\theta}{\lambda}$,
k is the normalizing constant,
θ is the angle between plane P_2 and the reference beam, and
A is the amplitude of the reference beam.

When this mask is placed on plane P_2 of Figure 2.11, the light leaving plane P_2 has three distinct components. The first component is

$$kF(f_x, f_y)\left[A^2 + \left|H(f_x, f_y)\right|^2\right] \tag{2.35}$$

and its inverse Fourier transform appears on the center of plane P_3 (*i.e.* the DC term). The second term is

$$kAF(f_x, f_y)H(f_x, f_y)e^{j\alpha f_x} \tag{2.36}$$

and its inverse Fourier transform corresponds to the 2-D convolution of the reference signal and the scenery, at distance α away from the DC term.

The last term is

$$kAF(f_x, f_y)H^*(f_x, f_y)e^{-j\alpha f_x} \tag{2.37}$$

and its inverse Fourier transform corresponds to the 2-D correlation of the reference signal and the scenery, at distance $-\alpha$ away from the DC term (mirror image of the convolution signal).

2.7.2 Joint transform correlator

Another technique used to perform correlation optically is the JTC. First proposed by Goodman and Weaver [31], this method eliminates the need for holographic storage media.

Figure 2.13 illustrates the operation of an optical JTC.

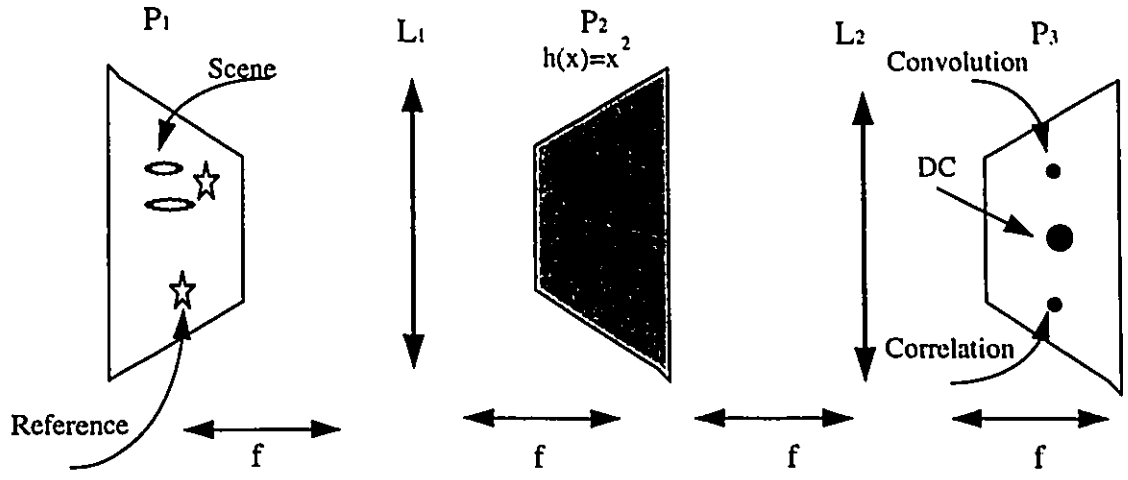


Figure 2.13. Typical JTC Setup

Assuming both the reference $h(x,y)$ and scenery $f(x,y)$ reside on the plane P_1, that is,

$$i(x,y) = f(x-x_0,y) + h(x+x_0,y) \tag{2.38}$$

the amplitude of the field just before plane P_2 is described by

$$I(f_x,f_y) = F(f_x,f_y)e^{-j2\pi f_x x_0} + H(f_x,f_y)e^{j2\pi f_x x_0} \tag{2.39}$$

Since plane P_2 is a square law device, the field immediately after plane P_2 is the intensity of the incident field. That is

$$\begin{aligned}\left|I(f_x,f_y)\right|^2 &= I(f_x,f_y)I^*(f_x,f_y)\\ &= \left|F(f_x,f_y)\right|^2 + \left|H(f_x,f_y)\right|^2 + F(f_x,f_y)H^*(f_x,f_y)e^{j4\pi f_x x_0}\\ &\quad + F^*(f_x,f_y)H(f_x,f_y)e^{-j4\pi f_x x_0}\end{aligned} \tag{2.40}$$

Taking the inverse Fourier transform of Equation (2.40) via lens L_2, at plane P_3 the amplitude of the incident field is

$$
\begin{aligned}
& f(x,y) \otimes f(x,y) + h(x,y) \otimes h(x,y) \\
& + f(x,y) \otimes h(x - 2x_0, y) \\
& + f(x,y) \otimes h(x + 2x_0, y)
\end{aligned}
\tag{2.41}
$$

where $\otimes$ is the correlation operation.

The DC term is the sum of the auto-correlation signals. The cross-correlation signals reside a distance $2x_0$ from the DC term, as depicted in Figure 2.11.

Comparing the JTC with MFC, the tradeoff is the elimination of a holographic storage device at the cost of resolution reduction on the plane P_1, the input plane.

2.7.3 Sampling considerations

From Section 2.4.1, the spatial frequency is related to spatial displacement by

$$
f_x = \frac{x_f}{d\lambda} \tag{2.42}
$$

where,
x_f is the input plane coordinate,
λ is the wavelength, and
d is the distance between the spatial and frequency planes.

It then follows that

$$
f_{x_{\max}} = \frac{x_{\max}}{2d\lambda} \tag{2.43}
$$

where x_{max} is the width of the frequency plane.

Due to the transform symmetry of real signals, only half of the frequency plane can be spanned by the frequency content of a signal. Thus the maximum frequency span is $\frac{x_{max}}{2}$.

At the input plane where the image is represented by a pixelated device, the maximum frequency resolution is

$$f_{max} = \frac{1}{2d_1} \tag{2.44}$$

where d_1 is the center to center distance between pixels.

With respect to Equations (2.44) and (2.43), the required focal length for a given frequency plane width is

$$d = f = \frac{d_1 x_{max}}{\lambda} \tag{2.45}$$

where f is the focal length of the converging lens .

Assuming the frequency plane is a pixelated device (CCD or SLM), then

$$f = \frac{d_1 d_2 N}{\lambda} \tag{2.46}$$

where,
$d_2N=x_{max}$ is the width of the device,
d_2 is the center to center distance between the pixels in the frequency plane, and
N is the number of pixels in each row at the frequency plane .

A number of techniques is used to reduce the effective size of the required focal length, while satisfying Equation (2.46). However, most of these techniques either corrupt the phase of the incident field or require very sensitive setups [32] [33].

2.7.4 Measures of performance for optical correlator based systems

To evaluate the performance of optical correlators, a number of common metrics is used. They are signal to noise ratio (SNR), peak to correlation energy (PCE), and Horner efficiency [34].

SNR is defined as the ratio of the squared average value of correlation peak to its variance. That is,

$$SNR = \frac{|E(c)|^2}{Var(c)} \tag{2.47}$$

where,
c is the peak output of the correlator,
$E(x)$ is the expected value of x, and
$Var(x)$ is the variance of x.

It can be shown that Equation (2.47) is maximized under the following condition[5][34]:

$$H(f_x, f_y) = \alpha \frac{S^*(f_x, f_y)}{P_n(f_x, f_y)} \tag{2.48}$$

where,
$H(f_x, f_y)$ is the frequency response of the filter (*e.g.* plane P_2 in Figure 2.11).
$S^*(f_x, f_y)$ is the conjugate transform of the reference signal,
$P_n(f_x, f_y)$ is the frequency spectrum of the background noise, and
α is the equality constant.

Assuming a white noise background, Equation (2.48) can be simplified as

$$H(f_x, f_y) = S^* \beta(f_x, f_y) \tag{2.49}$$

where,

$\beta = \alpha N_0$, and

N_n is the power spectrum density of the background white noise.

For β=1, Equation (2.49) is identical to the filter used in Equation (2.30). This means that correlator based systems, such as MFC, are optimal under the SNR metric, as defined by Equation (2.47).

2.7.4.1 Implications

Certain implicit and explicit assumptions have been made in the derivation of Equation (2.49). It is assumed that the background noise and the reference signal are overlapping [30]. This is a direct consequence of the definition of SNR used in Equation (2.47). It is shown in [30] that Equation (2.49) is a sub-optimal solution, which causes broad and noisy correlation peaks, under the conditions that the noise and reference are not overlapping. Figure 2.14 shows an example of such a situation.

Reference [30] covers, in detail, approaches to overcome this problem (at least theoretically).

Since, in image processing applications, the reference signals are predominately low frequency signals, filters derived by Equation (2.49) are low pass. Thus they block much of the incident light on the filter plane (reduction in Horner efficiency).

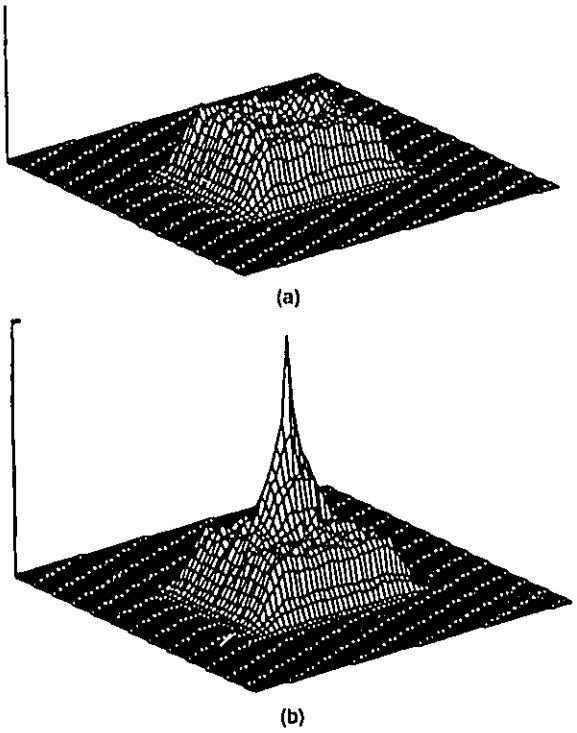


Figure 2.14. Correlation Peak Broadening
a) Non-overlapping Background Noise
b) Overlapping Background Noise [30]

In the derivation of Equation (2.49), it is assumed that the background noise is white noise. Although true for temporal signals, it is rarely the case for spatial signals [30], as such Equation (2.49) is a sub-optimal solution.

Even though correlation based filters are translation invariant, they are not scale, or rotation invariant. Their extreme sensitivity to scale, rotation, and illumination changes is one of their major drawbacks in identification systems. 20 dB drop in the SNR, for a 2% change in scale or 2° rotation variation, is not uncommon in correlator based systems [34].

Obviously, once the problems are known, solutions can be devised to overcome them. Each of the aforementioned problems has a solution, be it

redefining the reference filter based upon a new definition of the SNR, use of phase only filters (POF) to increase Horner efficiency, adaptive approaches for colored noise, and other post or pre transformations to achieve three point invariance. Nonetheless, the main problem lies in the fact that all the above objectives cannot be satisfied at the same time, either because of conflicting performance metrics, or practical problems in realizing these solutions.

Chapter 3

Review of algorithms for face recognition

In this chapter we present a review of popular methods for face recognition. The following sections present a review of the histogram, 2-D moment, wavelet transform, and KLT based methods for face recognition.

3.1 Histogram and moment based methods

In applications, such as face recognition, where the operating environment is stationary, a more compact representation of face images would be obtained if common environmental factors were quantified and eliminated. Such factors may include average background illumination, most probable location of the face in a scene, or the orientation of faces. These factors may be quantified based on the statistical properties of the scene.

Unlike orthogonal methods where the bases of interest rarely make any physical sense, statistical properties such as moments and joint probability density have a physical interpretation. This may be in terms of expected value (average), contrast (variance) or skew (third moment). However, these moments are not optimized in the MSE sense, information content, or noise immunity.

In the following sections, two classes of statistical methods are considered. The distinguishing factor between these classes is the definition of the random variable of interest.

If a face is to be considered as a three dimensional entity of the form

$$z = f(x, y) \qquad x, y \in R \tag{3.1}$$

where,
f(x,y) describes a face in Cartesian coordinates,
z is the intensity of the face, and
R is the region of interest.

then the random variable of interest is a two dimensional signal (*i.e.* pixel).

However, under the assumption that all the relative positions, on average, are the same (head position, relative position of facial features), it is presumed that the distinguishing characteristics are contained in the intensity variation of a face. Thus the random variable of interest becomes the intensity variation of the image. That is,

$$h(z) = w_z \tag{3.2}$$

where,
z is the intensity level,
h(z) is the histogram of an image, and
w_z is the frequency of intensity level z.

A detailed description of statistical methods is presented next.

3.1.1 Histogram-based face recognition

In this section, the random variable of interest is the intensity variation of a face image. The justification for this assumption is based on the observation that, due to geometrical similarities between all face images in a face recognition application (*e.g.* face position, relative feature position), such common factors may be eliminated. Thus the major distinguishing factor is the intensity variation

of each face. Although, theoretically, there may exist two different faces with the same histograms, the likelihood of such an event is negligible. To emphasize this point, histograms of different faces are shown in Figure 3.1.

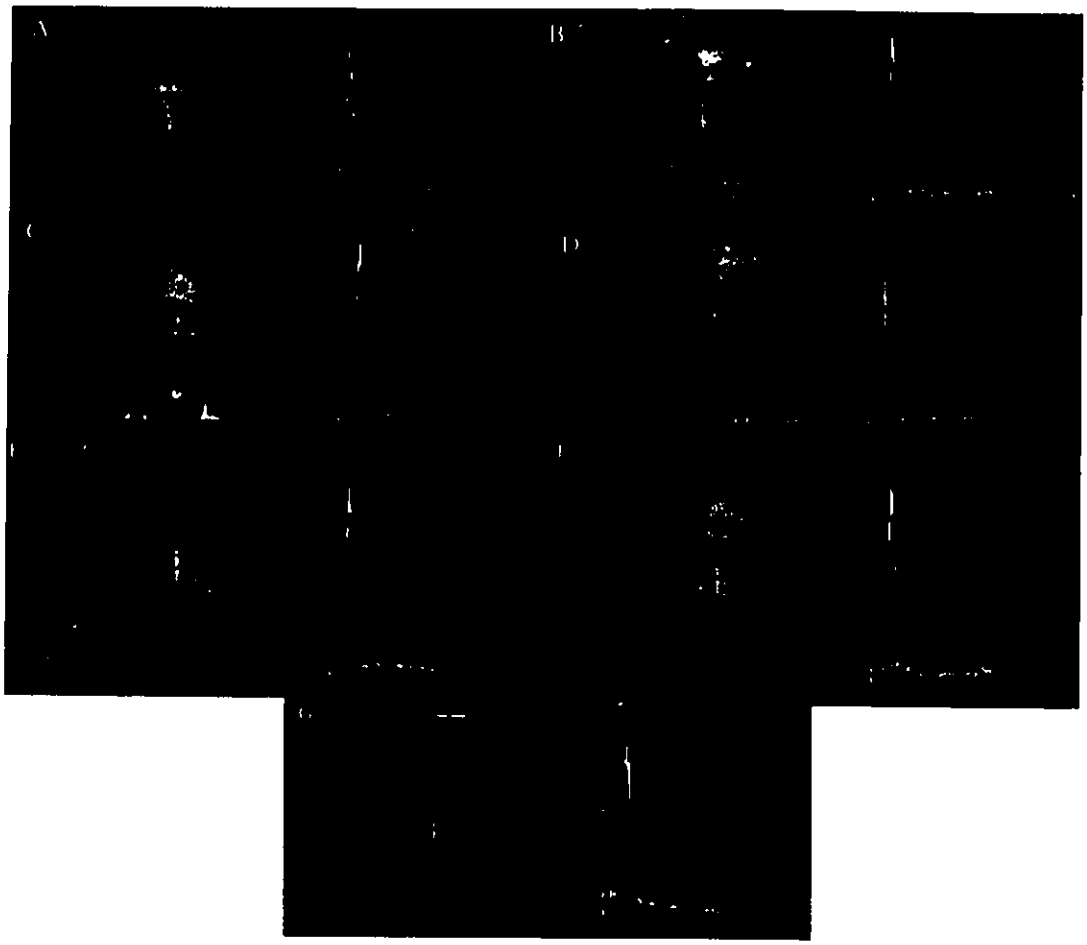


Figure 3.1 Histograms of Different Faces

This particular sequence of images was chosen to show the uniqueness of the histogram mapping under race, age, and facial variation.

A number of immediate advantages is realized under histogram mapping. The histogram of a face image is invariant with respect to translation, rotation, and scale (with a constant scale factor). Perhaps the more interesting aspect of the

histogram mapping is its invariance towards facial expressions, as illustrated in Figure 3.2.

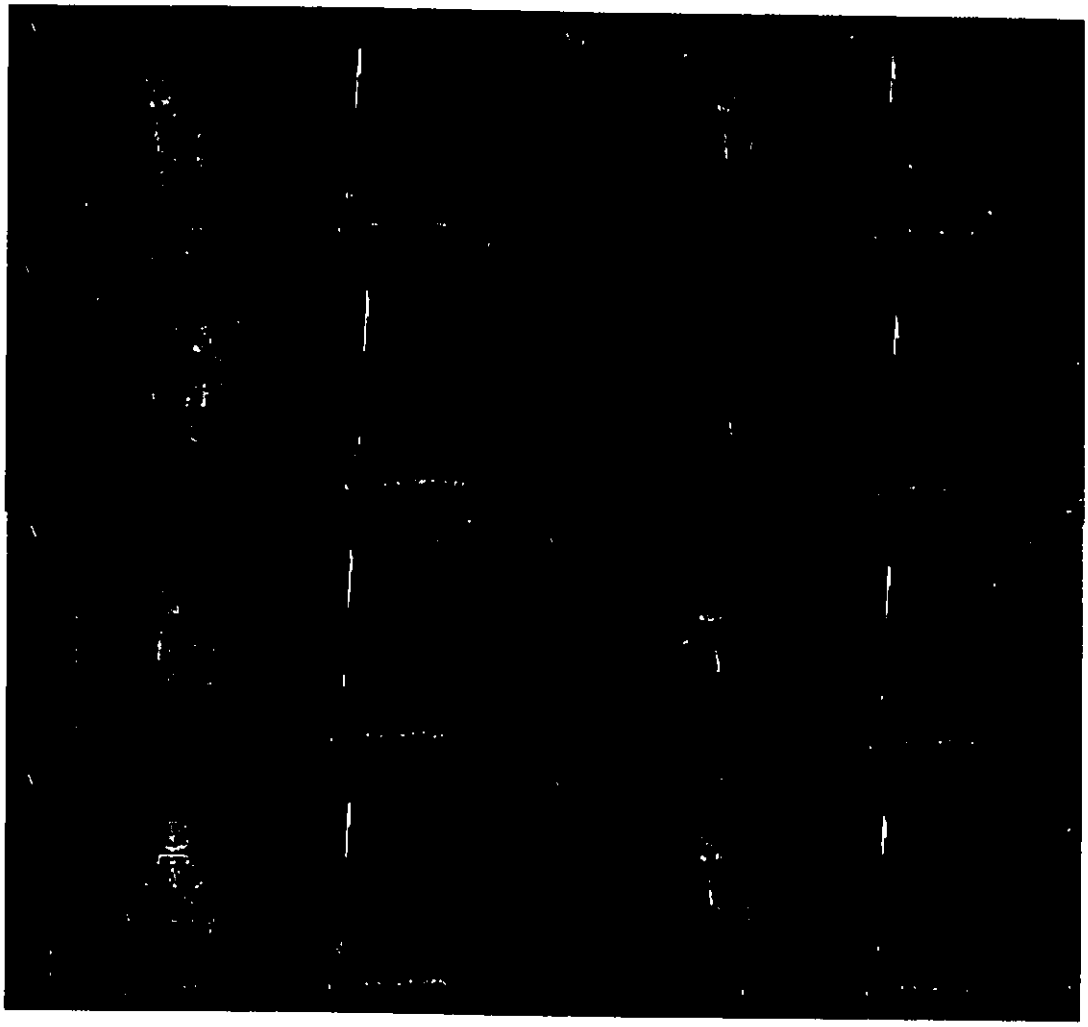


Figure 3.2 Tolerance of Histogram Mapping Towards Facial Expressions Changes, In-plane, and Off-plane Rotation

Also of definite advantage in a face recognition application is the invariance of the histogram mapping with respect to off-plane rotations, as shown in the above figures.

The invariance property of the histogram mapping toward off-plane rotation, and facial expressions is to be expected, since the intensity distribution

of a face image is essentially a stationary distribution. Such a distribution can be roughly approximated by a Gaussian function, as evident from the above histograms. This can also be verified by examining the power spectrum of the face images as shown in Figure 3.3.

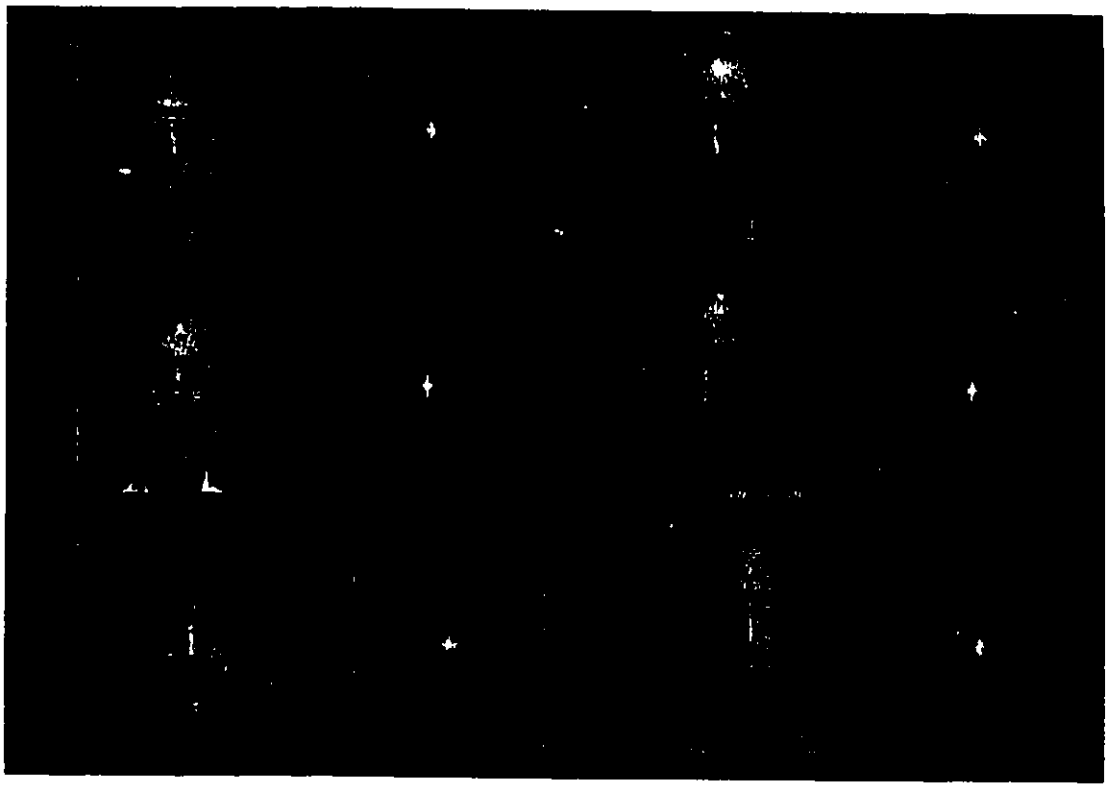


Figure 3.3 A Face-signal Is Predominantly a Low Frequency Signal

The Original Images Are on the Left Hand Side, and The Respective Power

Spectrum on the Right Hand Side.

As shown, a face signal is predominantly a low frequency signal. That is, there are few sudden changes in intensity distribution of a face image.

3.1.1.1 Histogram comparison

The main concern in histogram comparison is to identify a measure that is resilient to minor changes in environmental factors, facial expression, and most importantly, is feasible for real time implementations. In the next section common methods used for this purpose are introduced.

3.1.1.1.1 Direct comparison

Most commonly used methods in histogram comparison are the Hamming distance and the Maximum likelihood, which are defined as follows:

$$H_d = \frac{\sum_{i=a}^{b} |h_1(i) - h_2(i)|}{\sum_{i=a}^{b} h_1(i)}$$
$$S_d = \frac{\sum_{i=a}^{b} \max(h_1(i), h_2(i))}{\sum_{i=a}^{b} h_1(i)} \quad (3.3)$$

where,
H_d is the Hamming distance,
S_d is the Maximum-likelihood distance,
$h_x()$ is the histogram of image x,
a is the starting histogram bin, and
b is the last histogram bin.

With respect to the Hamming distance, histograms are said to be identical if H_d is zero; whereas, in the case of the Maximum likelihood method, histograms are said to be identical if S_d is equal to one. However, as expected, these measures provide no resilience towards differential noise, that is noise induced by a small

shift in the histograms. Furthermore, this class of noise is always present in the environment (change of illumination, variations of the same face, and so on). To partially overcome this problem, the histogram may be segmented, and the average value of each segment may be considered as the proper value of the histogram. Alternatively, as suggested in [45], the *CDF* of the histogram should be considered as the proper measure of similarity. In both cases, the effect of the differential noise will be averaged out through the width of the segments. Even though some modest improvement is achieved by the use of windowing, it performs poorly, under changes in illumination and head movement.

In the next section we consider the random variable of 2-D coordinates

3.1.2 Pixel based face recognition

The most common way to characterize a random variable (*i.e.* pixel) is through its moments. Two classes of moments are considered, namely non-orthogonal and orthogonal moments.

3.1.2.1 Non-orthogonal moments

The $(p+q)_{th}$ order regular moments of a positive-real-bounded function $f(x,y)$ are defined as

$$m_{p,q} = \iint_R f(x,y) x^p y^q dxdy \qquad p,q = 0,1,2,3... \tag{3.4}$$

It can be shown that $f(x,y)$ is uniquely reconstructed based on its moments [6]. That is,

$$f(x,y) = \iint_{\infty} e^{-j2\pi(x\xi_1 + y\xi_2)} [\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} m_{p,q} \frac{(j2\pi)^{p+q}}{p!q!} \xi_1^p \xi_2^q] d\xi_1 d\xi_2 \quad (3.5)$$

Since the order of summation and integration cannot be interchanged (the Fourier transform of ξ_1^p is not bounded), it is not possible to truncate Equation (3.5). Nonetheless, regular moment based approaches have extensively been used in a variety of pattern matching, character recognition, and texture analysis applications [37][42].

Moments, defined by Equation (3.4), are not invariant with respect to translation, scale, and rotation. However, in [6], seven functions of centralized moments (mean adjusted), up to the third order moments, have been defined that are three point invariant.

Although the uniqueness of finite numbers of moments with respect to their original *pdf* may be argued, their applicability in a variety of pattern matching applications has been demonstrated. As an example, Figure 3.4 depicts four variations of a same aerial image 3.4*a*). Variations are with respect to scale 3.4*b*), reflection 3.4*c*), rotation by 2° 3.4*d*), and rotation by 45° 3.4*e*).

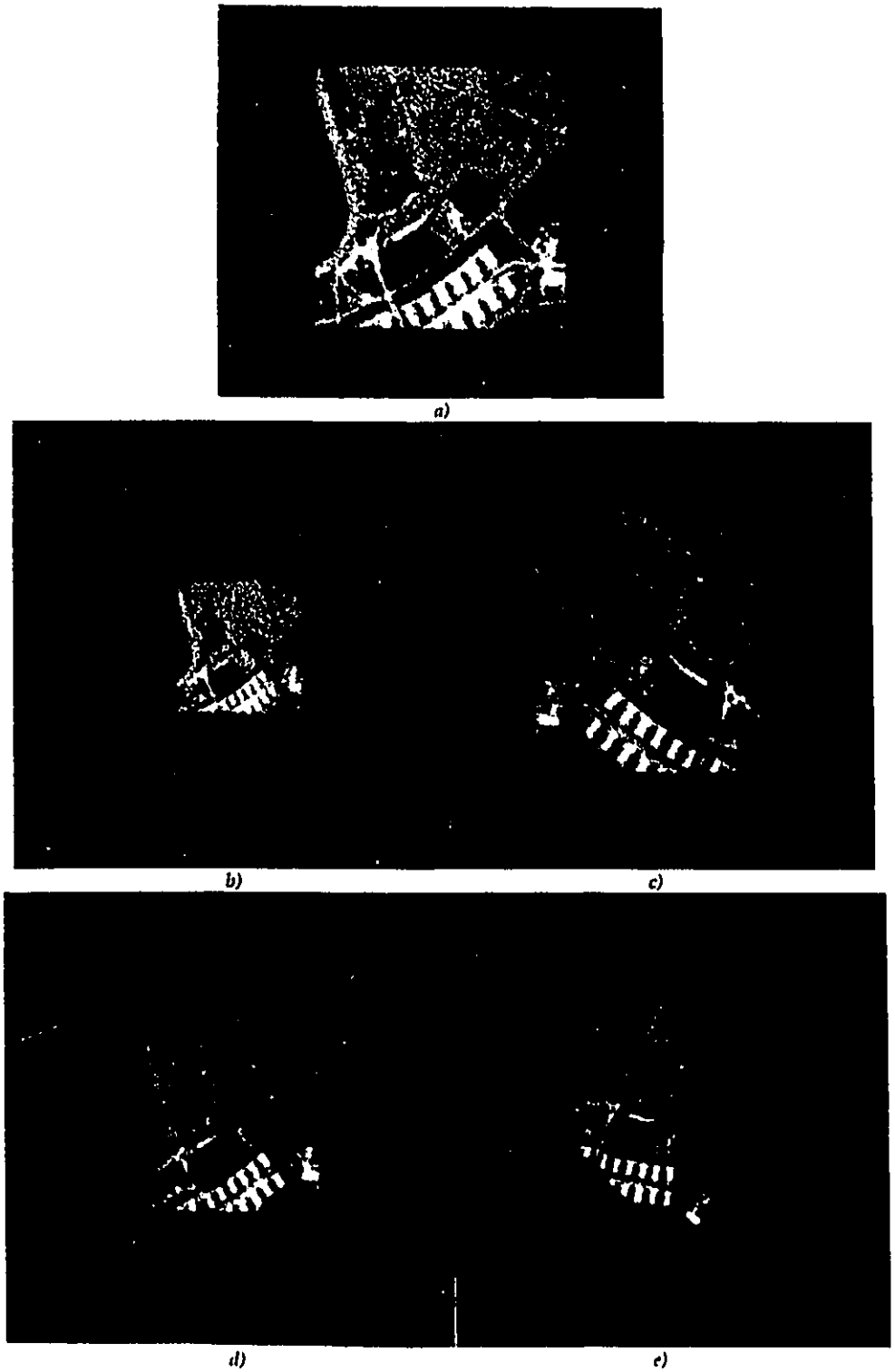


Figure 3.4 Moment Invariant Functions

a) The Original Image, b) Scaled Down by Factor 2

c) Reflected, d) Rotated by 2"

e) Rotated by 45" [42]

Table 3.1 shows the logarithmic values of the seven invariant moment functions. With respect to Table 3.1, it is clear that invariant moment functions remain the same under scale, rotation, and translation variation. Since the moment functions have a large dynamic range, the logarithmic values of these functions are presented in Table 3.1 [42].

Moment functions	Original a)	Half size b)	Reflected c)	Rotated 2° d)	Rotated 45° e)
ϕ_1	6.249	6.226	6.919	6.253	6.318
ϕ_2	17.180	19.954	19.955	17.270	16.803
ϕ_3	22.655	23.531	26.689	22.836	19.724
ϕ_4	22.919	24.236	26.901	23.130	20.437
ϕ_5	45.749	48.349	53.724	46.136	40.525
ϕ_6	31.830	32.916	37.134	32.068	29.315
ϕ_7	45.589	48.343	53.590	46.017	40.470

Table 3.1 Three Point Invariant Moment Functions[42]

The small differences, across the columns in Table 3.1, have been attributed to digital quantization noise [42].

3.1.2.2 Orthogonal moments

The main motivation in defining orthogonal moments, is to gain progressively better approximations to *f(x,y)* in the MSE sense, as the number of the moments increases. The most popular orthogonal moments, which have been successfully applied to a variety of pattern matching applications, are the Legendre, and Zernike moments [43][37][44].

The Legendre moments are defined in the interior of a unit square as

$$\begin{aligned} &P_0(x) = 1 \\ &P_n(x) = \frac{1}{n!2^n}\frac{d^n}{dx^n}(x^2+1)^n, \qquad n = 1,2,... \\ &\int_{-1}^{1} P_n(x)P_m(x)dx = \frac{2}{2n+1}\delta(m-n) \end{aligned} \tag{3.6}$$

The transform of *f(x,y)* then becomes

$$\lambda_{p,q} = \frac{(2p+1)(2q+1)}{4}\int_{-1}^{1}\int f(x,y)P_p(x)P_q(x)dxdy \tag{3.7}$$

It follows that *f(x,y)* can be approximated by *g(x,y)* as [37]

$$\begin{aligned} &g(x,y) = \sum_{p=0}^{N}\sum_{q=0}^{N-p}\lambda_{p,q}P_p(x)P_q(y) \\ &g(x,y) \longrightarrow f(x,y) \qquad as \qquad N \longrightarrow \infty \end{aligned} \tag{3.8}$$

The Zernike moments are not only orthogonal, but their magnitudes are also rotation invariant. The bases of the Zernike moments are defined in the interior of the unit circle as

$$V_{nm}(x,y) = R_{nm}(x,y)e^{jm\tan^{-1}\frac{y}{x}}$$

$$R_{nm}(x,y) = \sum_{s=0}^{n-|m|/2}(-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}(x^2+y^2)^{(n/2)-s} \quad (3.9)$$

$$\iint_{x^2+y^2\le 1} V_{nm}(x,y)^* V_{pq}(x,y)dxdy = \frac{\pi}{n+1}\delta(n-p,m-q)$$

where,
n is a positive integer or zero, and
m is a positive and negative integer subject to constraint $n-|m|$ even, and $|m| \le n$.

The transform of $f(x,y)$ then becomes

$$A_{nm} = \frac{n+1}{\pi}\iint_{x^2+y^2\le 1} f(x,y)V_{nm}(x,y)^* dxdy \quad (3.10)$$

and $f(x,y)$ can be approximated by $g(x,y)$ [44]

$$g(x,y) = \sum_{n=0}^{N}\sum_{m=0}^{N-n} A_{nm}V_{nm}(x,y)$$
$$g(x,y) \longrightarrow f(x,y) \qquad as \qquad N \longrightarrow \infty \quad (3.11)$$

If A_{nm} is the Zernike moment of $f(r,\theta)$, then it can be shown that the Zernike moment of $f(r,\theta-\alpha)$ is $A_{nm}e^{-jm\alpha}$. Thus the magnitude of the Zernike transform is rotation invariant [44].

If the two transforms are compared, the Legendre moments have two unique attributes. Firstly, they are separable as shown by Equation (3.6), which means their computational cost may be decreased. Secondly, they are linearly dependent on regular moments as defined by Equation (3.4) [37].

Even though orthogonal moments have been used in character recognition applications, their applicability in face recognition is hindered by their computational complexity. Although some numerical methods have been proposed to overcome this problem [43], these methods require constraints that are not readily possible in face recognition applications (*e.g.* binary image assumption).

3.2 *Wavelet transform*

In this section, the wavelet transform and its application to image processing is presented. Useful attributes of the wavelet transform are spatial-frequency decomposition, orientation sensitivity, and computational simplicity. The recent attention to the wavelet transform can be attributed to the similarity of visual cortex response to the Gabor wavelets [7] and the advancement in the theory of finite-support wavelet basis [1][3]. In what follows, the development is primarily based on 1-D signals, and time is considered as the independent variable. The generalization to the spatial domain, of the development, is achieved by generating separable 2-D bases.

The motivation for the wavelet analysis arises in applications where different frequency resolutions are required at different time instants. This ensures localization in both frequency and time. Motivated by the earlier work by Gabor in speech analysis (short time Fourier transform), the wavelet transform of a function is defined as:

$$\varphi_f(a,b:t) = \langle \psi(a,b:t), f(t) \rangle \tag{3.12}$$

where,

$\psi(a,b:t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a})$ is one of the wavelet bases,

$\psi(t)$ is the mother wavelet,

$\varphi(a,b:t)$ is the projection of $f(t)$ over a scaled and shifted mother wavelet,

a is the frequency resolution (localization in frequency), and

b is the time shift parameter (localization in time).

Equation (3.12) states that a signal $f(t)$ is correlated with a mother wavelet at different resolutions (determined by a) and different time instants (determined by b). The scale and shift parameters can be either continuous (CWT) or discrete (DWT). In the latter case,

$$a = a_0^m$$
$$b = a_0^m n b_0 \tag{3.13}$$

where n, m are integers and $a_0 > 1$, $b_0 > 0$ [1].

Figure 3.5 illustrates the time-frequency behaviour of the DWT.

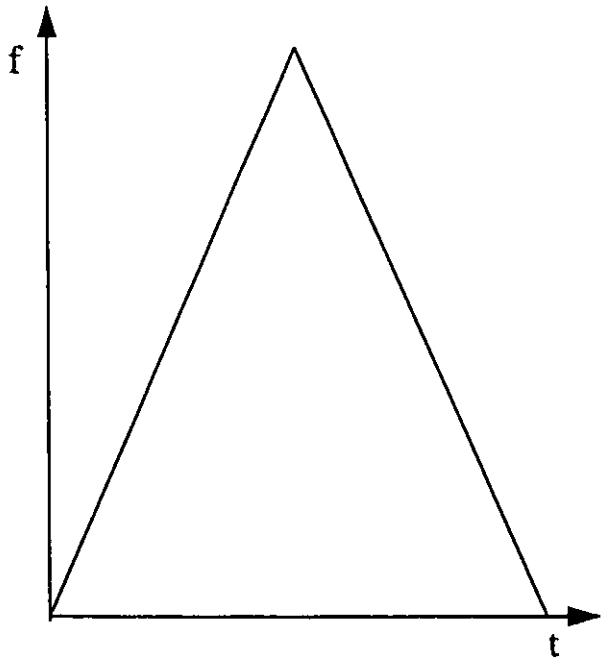


Figure 3.5 Time- frequency Window Mapping

For high frequency signals, a narrow time window (filter) is used, and for the low frequency signals a wide time window is used. Under relatively modest conditions, filter $\psi(t)$ can generate a family of orthonormal bases, simply by scaling and shifting $\psi(t)$, the mother wavelet [47].

Both the theoretical and practical implications of the wavelet transform are best described under the multi-resolution frame work. A detailed treatment of this subject can be found in [48].

If $f(t)$ has a finite energy, $f(t) \in L^2$, and

$$\phi_b = \phi(t-b) \tag{3.14}$$

furthermore, assuming the scale function spans the subspace V_0 of L^2, then

$$\phi_{a,b}(t) = \sqrt{a}\phi(at-b) \tag{3.15}$$

spans the V_1 subspace for a=2, such that

$$V_0 \subset V_1 \tag{3.16}$$

In general, it can be shown that for

$$\phi_{j,b}(t) = 2^{\frac{j}{2}}\phi(2^j t-b) \tag{3.17}$$

the following holds:

$$....\subset V_0 \subset V_1 \subset V_2 \subset V_3 \subset ...\subset V_j ...\subset L^2 \tag{3.18}$$

The implementation of such orthogonal decomposition may be achieved by a simple dyadic filter structure [1]. It can be shown that

$$\phi(2^j a_0 - b) = \sum_n h(n-2k)\phi(2^{j+1}a_0 - n) \tag{3.19}$$

where $h(n)$ is the scaling coefficient or scaling filter.

Even though any L^2 function can be described by subspaces V_j, important features of a signal are best described by a different set of functions, which spans

the difference subspaces between resolution j to $j+1$. These are the wavelet functions. The *difference* subspaces are defined as

$$\begin{aligned} V_j \cup W_j &= V_{j+1} \\ V_j \cap W_j &= \varnothing \end{aligned} \tag{3.20}$$

It can be shown that

$$L^2 = V_0 \cup \left\{ \bigcup_1^\infty W_j \right\} \tag{3.21}$$

Alternatively,

$$L^2 = \bigcup_{-\infty}^{\infty} W_j \tag{3.22}$$

A similar relationship as Equation (3.19) exists between wavelet functions. That is

$$\psi(2^j a_0 - b) = \sum_n g(n-2k)\psi(2^{j+1} a_0 - n) \tag{3.23}$$

With respect to Equations (3.21) and (3.22), subspaces W_j capture the details of a signal at resolution j.

With regards to the requirement of both the wavelet functions and the scale functions, two observations can be made. Firstly, in practical applications, rarely are infinite details of a signal of interest. It is more often the case that a crude presentation of a signal is given (scale), and then the details are added as required (wavelets). Secondly, addition of the scale function allows the wavelet transform to span spaces other than L^2. For example, a pure wavelet basis cannot

describe the function $f(t)=1$ since it is not a L^2 function; whereas, such a function can easily be described by the scale function.

To avoid the cost of continuous projection of a function over wavelet bases, subband filter structures based on $h(n)$, and $g(n)$ are often used in digital applications. Analogous relationships to Equations (3.19) and (3.23) exist between the wavelet transform coefficients. It can be shown that

$$\begin{aligned} c_j(k) &= [c_{j+1}(k) * h(-k)]2n \\ d_j(k) &= [d_{j+1}(k) * g(-k)]2n \\ n &\in Z \end{aligned} \tag{3.24}$$

The transform coefficients of c_j and d_j, of subspace V_j and W_j respectively, are obtained by convolving $h(-k)$ and $g(-k)$ (correlating by $h(k)$ and $g(k)$) with the transform coefficients c_{j+1} of subspace V_{j+1} and down sampling the results by two [47].

Figure 3.6 illustrates this operation.

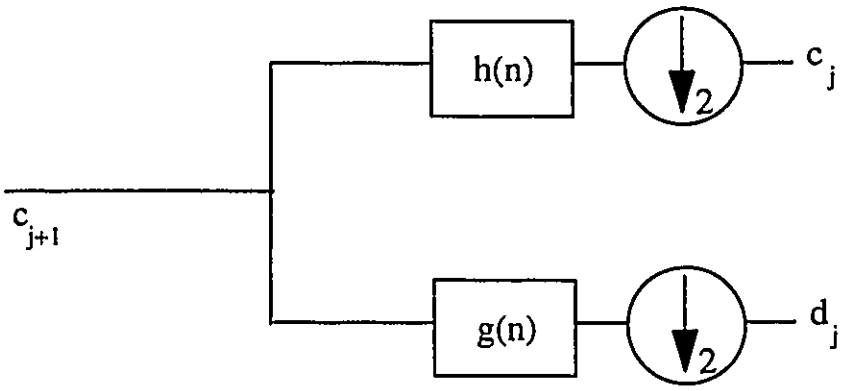


Figure 3.6 Analysis via Dyadic Subband Structure

Furthermore, it can be shown that $g(n)$, and $h(n)$ are related as follows:

$$g(n) = (-1)^n h(N-1-n) \tag{3.25}$$

where N is the size of the filter.

Similarly, to reconstruct $f(t)$ at resolution $j+1$, the following operation is applied:

$$\left[c_j(k)2^{-1}n\right]*h(n)+\left[d_j(k)2^{-1}n\right]*g(n) \qquad n\in Z \tag{3.26}$$

Upsampled coefficients at resolution j are convolved with the scale and wavelet filters, and the results are added to obtain the reconstruction at resolution $j+1$. Figure 3.7 illustrates this operation.

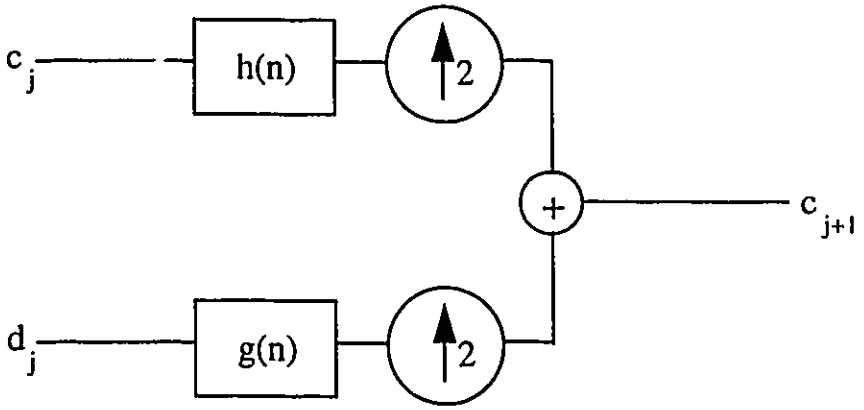


Figure 3.7 Synthesis via Dyadic Subband Structure

It is noted that c_{j+1} is an approximation of the function $f(t)$ at resolution $j+1$. The highest resolution approximation of $f(t)$ (*i.e.* c_{-}) can be obtained from the sampled function $f(n)$. At rates above the Nyquist-frequency, the scale function of Equation (3.19) behaves like a delta train. Figure 3.8 depicts a dyadic subband structure for the DWT decomposition.

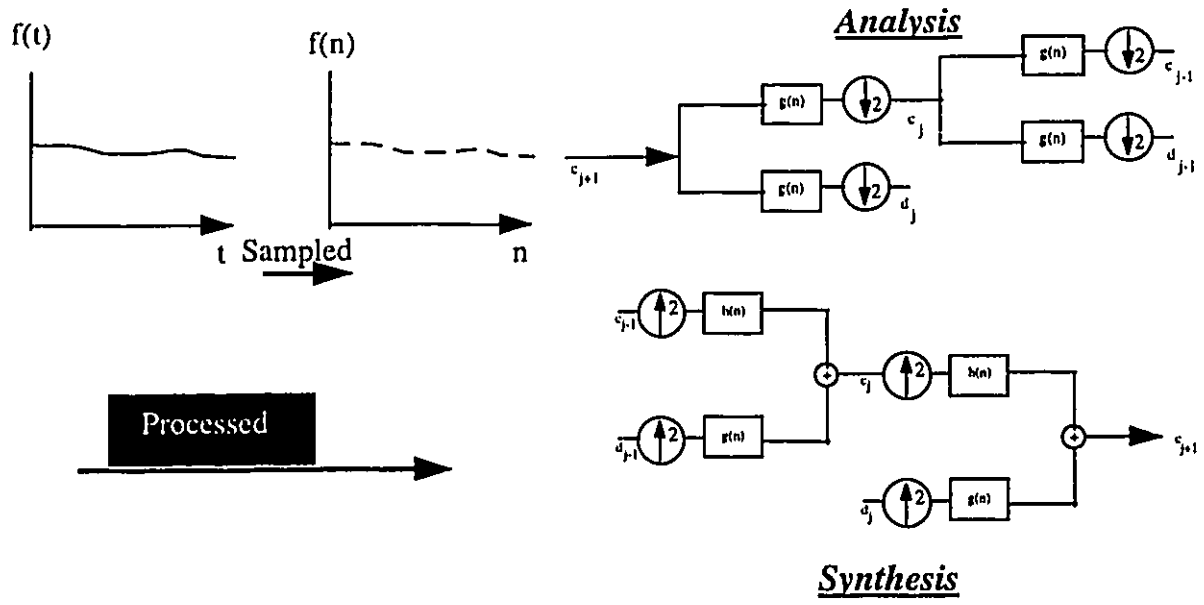


Figure 3.8 Two Stages, Three Subbands Analysis and Synthesis Operation

3.2.1 Properties

The wavelet transform contains information from both the input (temporal/spatial) domain, and the frequency domain. 2-D wavelets may be constructed from standard families of 1-D wavelets, such that they are orientation sensitive [46]. The computational complexity of the DWT for an *NxN* image is of order N^2 and, when compared with $(N^2 log_2(N^2))$ of the FFT, the computational gain is appreciable for standard size images.

3.2.2 Wavelet transform in image processing

Equation (3.12) can be extended to decompose 2-D signals. This means that

$$\varphi_{wf;i}(a,b{:}\,x,y) = \langle \psi_i(a,b{:}\,x,y), f(x,y) \rangle \qquad i \in \{1,2,3\}$$
$$\varphi_{sf}(a,b{:}\,x,y) = \langle \phi(a,b{:}\,x,y), f(x,y) \rangle \tag{3.27}$$

where,
x and y are the spatial coordinates,
a is the scale factor,
b is the spatial translation,
$\varphi_{sf}(a,b{:}x,y)$ is the projection of the image over the scale function $\phi(a,b{:}x,y)$, and
$\varphi_{wf;i}(a,b{:}x,y)$ is the projection of the image over the wavelet functions $\psi_i(a,b{:}x,y)$ for $i=1,2,3$ (see below for explanation).

For ease of implementation, sub-optimal separable bases constructed on 1-D bases are often used [1][48]. In such cases, the scale and wavelet functions may be defined as:

$$\begin{aligned}
\phi(a,b{:}\,x,y) &= \phi(a,b{:}\,x)\phi(a,b{:}\,y) \\
\psi_{0^o}(a,b{:}\,x,y) &= \phi(a,b{:}\,y)\psi(a,b{:}\,x) \\
\psi_{90^o}(a,b{:}\,x,y) &= \psi(a,b{:}\,y)\phi(a,b{:}\,x) \\
\psi_{45^o}(a,b{:}\,x,y) &= \psi(a,b{:}\,y)\psi(a,b{:}\,x)
\end{aligned} \tag{3.28}$$

Figure 3.10 shows the wavelet decomposition of Figure 3.9.

With respect to Figure 3.10, three points are noted. Firstly, the frequency resolution increases as the number of stages increases. Secondly, subbands of a 2-D DWT are orientation selective, and finally, the frequency isolation of subbands is poor.

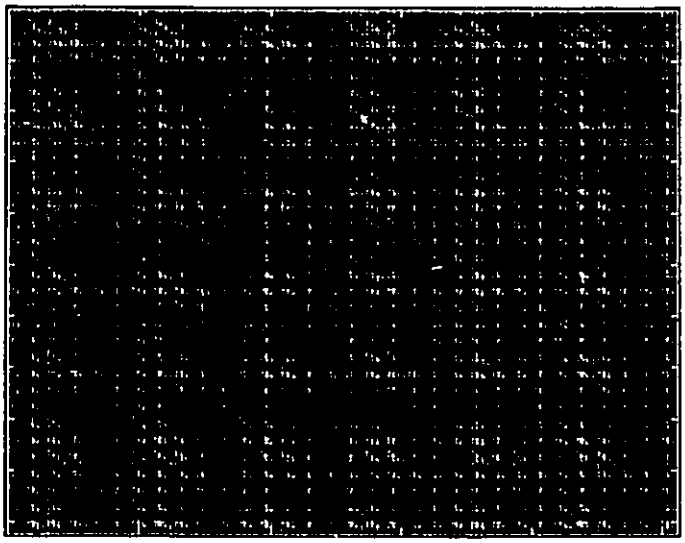


Figure 3.9 Sine Grid
f(x,y)=sin(10x-10y)+sin(50x-50y)+sin(200x-200y)+
sin(10x)+sin(50x)+sin(200x)+
sin(10y)+sin(50y)+sin(200y).

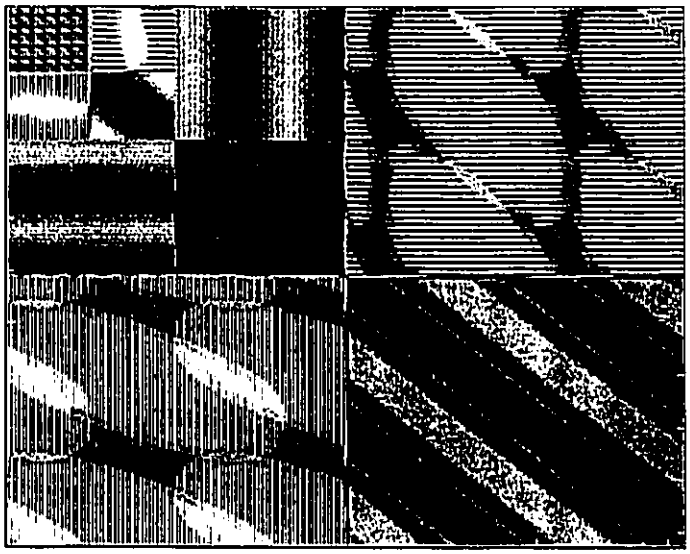


Figure 3.10 Daubechies 12 Taps Filter, 3 Stages Decomposition

3.2.3 Eidos

In this section, a method for face recognition developed by Malsburg *et al.* [8][49] is described. Eidos, as it is called, (Greek for essence) uses non-orthogonal oriented Gaussian wavelets for the purpose of face identification. Eidos is an

attempt to model the response of the primary visual cortex, to achieve comparable performance for face recognition. The goal of the decomposition is to extract a set of information that remains unchanged under facial gestures and illumination changes.

3.2.3.1 Description

One of the most important aspects of modeling the visual cortex is the end-inhibition property. This property refers to the response of visual cells to short lines, edges, line endings, and corners [8]. It has been hypothesized that humans recognize faces based on this set of information and not through the complete image. Furthermore, it has been shown that the characteristic points obtained in such manner correspond to the salient features of face images (eyes, nose, *etc.*). To simulate the end-inhibition property, a multi-resolution approach based on Gaussian CWT has been employed. That is, an image is decomposed at different scales, and orientation. These choices are obtained empirically. Different subbands are compared in order to detect end lines, curvatures, and sharp edges. A simple illustration of this point is shown in Figure 3.12 where the detection of the line break in Figure 3.11 is desired.

Figure 3.11 A Test Image for The End-inhibition Property

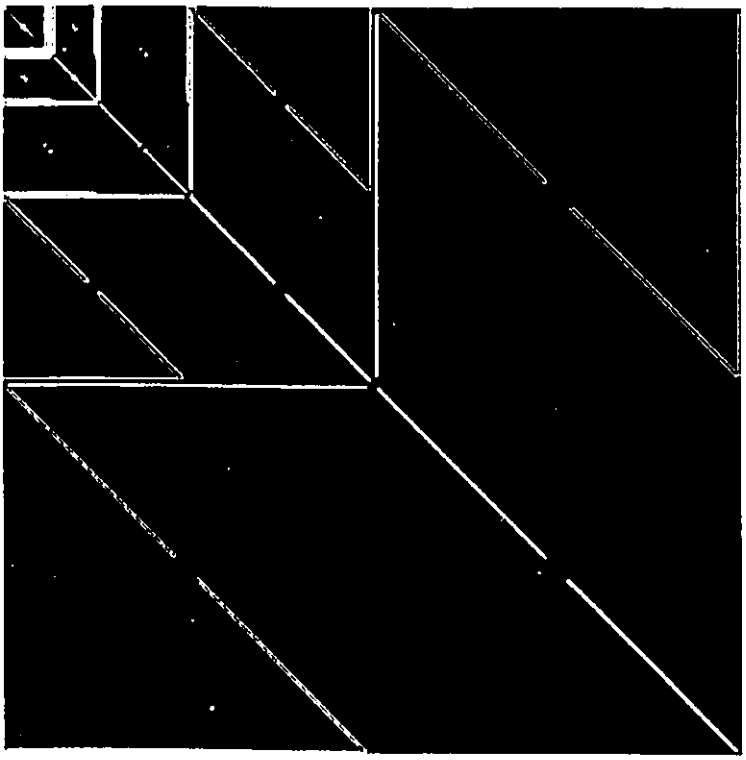


Figure 3.12 Four Stages Decomposition with D_n Filters

To detect the break in the diagonal line, the size adjusted scale-subband (top left) and diagonal element of the fourth subband (bottom right) may be compared. Obviously, to characterize a face image, additional frequency resolutions and orientations are required. After processing small regions of a given face as described above, the characteristic points of the face image are obtained. An example of such a process is shown in Figure 3.13.

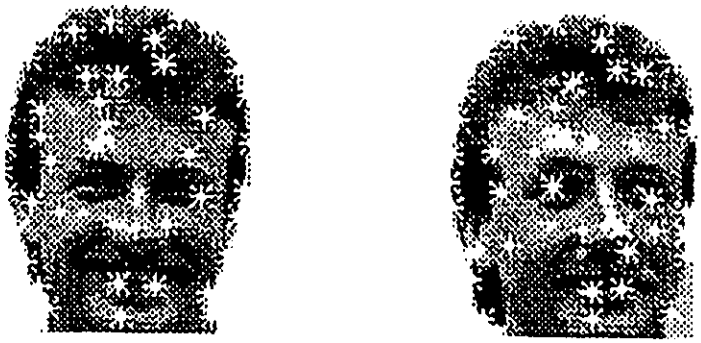


Figure 3.13 Characteristic Points Obtained by Simulation of the End-inhibition Property of Visual Cells [8]

The locations of characteristic points (stars) correspond to the salient features of the face image, and also these points remain unchanged under facial movements. A face image is stored, based upon characteristic points and the Euclidean distance among the points (*i.e.* the geometrical relationship of facial features). When a new face image is obtained, the same set of processes is performed on it and the comparable characteristic points (points from same segments) of the target and reference pair are compared. A pair that minimizes the overall cost function (minimum difference in all comparable points) is considered identical.

Although the system performs well under modest changes in face position, its major drawback is the varying number of characteristic points as a

face image undergoes changes. These changes may be a simple rotation, or change of hair style.

To avoid the pitfalls that follow from the use of end-inhibition, the system can be trained to decompose a specific set of facial features (*e.g.* eyes, nose). To achieve this, a human operator must identify these features on a given training set (about 40-60 face images). To decompose a new face, it is first segmented, then the wavelet transform of each segment is compared with the average wavelet transform of facial features obtained during the training session. If the difference between a segment and a feature is below a threshold (chosen empirically) the segment is recognized as that feature. A face image is stored, based upon recognized features and the Euclidean distance among the features. A pair that minimizes the overall cost function (minimum difference in all comparable features) is considered identical.

3.2.3.2 Shortcomings

The performance of Eidos is hindered by two factors. Firstly, the computational cost of the CWT is high. Secondly, criteria for defining the required subbands, either in their frequency resolution or the preferred orientation, are not well defined. At present, the performance of the system is far

from the real time implementation. Nonetheless, Eidos is among the most popular face recognition systems[1] of today.

---

[1]One of the first two contenders for the best face recognition system in USA (the other one is *photobook* by MIT, see Chapter 4).

3.3 *Karhunen and Loeve transform*

The objective of the Karhunen and Loeve transform (KLT) is to represent a vector x, according to bases generated through observation vectors belonging to the same family of signals as vector x. It can be shown that such bases, under certain constraints, can represent members of the same family optimally in the mean square error (MSE) sense [38]. For an NxN matrix with rank $M \leq N$, there are M eigenvectors and eigenvalues that uniquely identify the matrix. That is,

$$\begin{aligned} R\phi_k &= \lambda_k \phi_k \\ 0 \leq k &\leq M-1 \end{aligned} \tag{3.29}$$

where,
R is the correlation matrix,
ϕ_k is the k_{th} eigenvector, and
λ_k is the k_{th} eigenvalue.

The number of eigenvectors corresponds to the rank of the matrix, and the relative importance of each eigenvector (*i.e.* its contribution in reconstruction) is determined by the magnitude of its eigenvalue [38]. For a wide sense stationary process, the correlation matrix can be approximated by the instantaneous values of the observation vector. That is,

$$R = uu^T \tag{3.30}$$

where u is a Nx1 observation vector.

If the process represents observation points with similar statistical properties, there will be dependencies among rows and columns of the correlation matrix. Thus the rank of the matrix and the number of eigenvectors

would be less than N. Furthermore, R can be best approximated (in the MSE sense) by eigenvectors that contain the highest eigenvalues [37]. In other words, to achieve perfect reconstruction

$$\begin{aligned} v &= \Phi^{*T} u \\ u &= \Phi v = \sum_{k=0}^{N-1} v(k)\phi_k \end{aligned} \tag{3.31}$$

where ϕ_k is the kth column of Φ.

Alternatively, using L eigenvectors as

$$\begin{aligned} \tilde{u} &= \sum_{k=0}^{L-1} v(k)\phi_k \\ L &\leq N \end{aligned} \tag{3.32}$$

will give the best reconstruction under the MSE assumption, if the eigenvectors are ordered in decreasing magnitude of their respective eigenvalues.

3.3.1 Properties

Properties of the KLT that particularly relate to image processing applications are described in the following sections.

3.3.1.1 Decorrrelation

It can be easily shown that the KLT coefficients are uncorrelated and have a zero mean [37]. That is,

$$E[v(k)] = 0$$
$$E[v(k)v^*(l)] = \lambda_k \delta(k-l) \tag{3.33}$$

3.3.1.2 Energy compaction

If the eigenvectors are decreasingly ordered with respect to the magnitude of their respective eigenvalues, it can be shown that the KLT packs the maximum average energy in the first $n \leq N$ samples of the transformed coefficient.

3.3.1.3 Minimum mean square error

Under the same condition as the above, it can be shown that the KLT optimally represents a signal in the MSE sense, with a minimum number of coefficients compared to any other block transformation.

The goal of the KLT can be described as shown in Figure 3.14. A member of a WSS process (*i.e.* x) can be represented by a subspace (plane p) defined by the eigenvectors of a correlation matrix, representing the same family of signals as x [38].

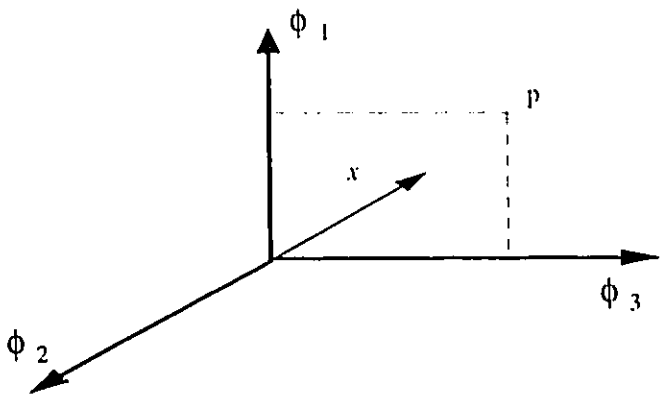


Figure 3.14 Identifying the Best MSE Subspace

It should be noted that the KLT filters, based on the eigenvector with the largest eigenvalue, are the stochastic counterpart of matched filters, *i.e.* both attempt to maximize SNR [38].

In the next three chapters, electronic and optical methods for face recognition are presented. Each method is evaluated with respect to the response time and tolerance towards changes in facial expression.

As mentioned in Chapter 1, two theoretical approaches are considered, namely orthogonal and non-orthogonal transformations. It is worth mentioning that, although a one-to-one relationship between the transformed image and the image itself is required, the transforms themselves need not be invertible. The goal is to achieve a better performance in recognition and not in image compression. Thus there is no preference for orthogonal transformations over non-orthogonal transformations.

In the following three chapters, a face is considered identified if it is among the first three detected images. Unless specified otherwise, identification

of individuals in the database is with respect to modest changes in head position (tilt of 5 degrees), facial expressions (smiling), and removal of glasses.

Chapter 4

Photobook

In this chapter, the performance of a commercial face recognition system, based upon the KLT, for the propose of comparison is provided. The basis of this work is the *photobook* face recognition system developed at MIT [2].

4.1 Description

The properties of the KLT, as mentioned in Chapter 3, are desirable in both image recognition and compression. However, unlike an image compression application, in face recognition the eigenvector bases are not required. The coefficients, obtained by the projection of a face image over eigenvectors, are sufficient to identify a face image uniquely. In this context the eigenvectors are referred to as eigenfaces. The main difficulty in applying the KLT is the computational complexity of the eigenvectors. An NxN image can be considered as a $1xN^2$ vector; therefore the corresponding correlation matrix is of size N^2xN^2. The computational complexity for obtaining the eigenvectors of this matrix is of order N^6. Assuming that the correlation matrix is separable, the complexity is reduced to N^3 [37], which is still prohibitively large for typical image sizes (*e.g.* 640x480).

However, as shown in [2], this complexity can be reduced to the order of the number of training sets ($M<<N^3$). Given the samples faces $\Gamma_1,\Gamma_2,...,\Gamma_M$, the following training set is defined,

$$\phi_i = \Gamma_i - \Psi$$
$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{4.1}$$

where Ψ is the average image.

The advantage in defining the *differences* as the training set, versus the original images, is that the DC bias component will be minimized. Let

$$A = [\phi_1, \phi_2, ..., \phi_M] \tag{4.2}$$

be the training set. Then the correlation matrix C will be defined as

$$C = \frac{1}{M}\sum_{n=1}^{M}\phi_n\phi_n^T \tag{4.3}$$
$$= AA^T$$

The eigenvectors v_i of A^TA (MxM) are

$$A^TAv_i = \mu_i v_i \tag{4.4}$$

where μ_i is the respective eigenvalue.

Premultiplying both sides by A, the following is obtained:

$$AA^TAv_i = \mu_i Av_i \tag{4.5}$$

Thus Av_i are the eigenvectors of C (N^2xN^2) matrix .

Intuitively, if there are M training points, there will only be M meaningful eigenvectors rather than N^2. The remaining vectors will have their respective

respective eigenvalues equal to zero [2]. A typical set of eigenvectors for face objects is shown in Figure 4.1.

Figure 4.1 Typical Eigenface Space [2]

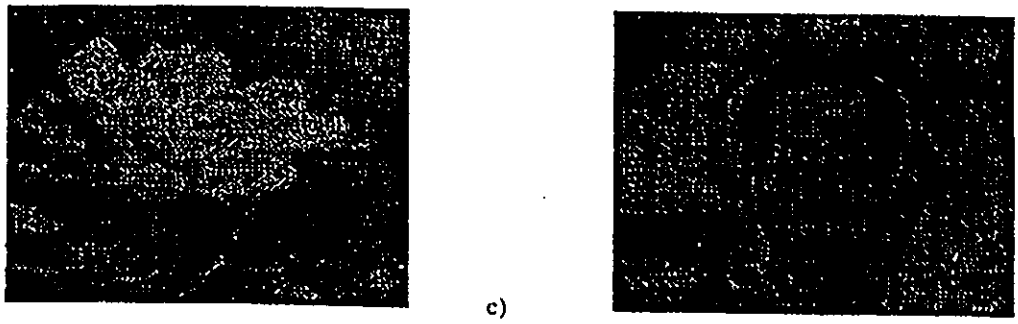
Once the main eigenfaces are identified (*i.e.* those with the highest eigenvalues), the task of recognition becomes that of selecting target and reference pairs in the transformed domain with the minimum Euclidean distance. Images in the database are projected over the space defined by the eigenfaces, and their respective coefficients are saved. When a target image is captured, it is projected over the space defined by the same eigenfaces, and the

transformed coefficients are compared. If the Euclidean distance is below a given threshold, then the image is identified.

Since the space of N^2 span has been reduced to M (number of training sets), it is possible that a non-face object may give a false alarm (*i.e.* it produces the same coefficients as a typical face object in the database). To ensure the validity of the transform image, the Euclidean distance between the mean adjusted image and its synthesis version should be below a given threshold. Figure 4.2 illustrates this point.

a)

b)



c)

Figure 4.2 Identification of Proper Input Objects Using the KLT
Right Hand Side Are the Original Images, On the Left, the Synthesis Images
The Relative Distances, Top to Bottom Are
a) 29.8
b) 58.5
c) 5217.4 [2]

Images on the right hand side are the original images. These images are decomposed and reconstructed using eigenvectors defined during the training session. The reconstructed images are shown on the left hand side. As shown in Figure 4.2, a non-face object (flower) is far away from the face space. That is, its Euclidean distance from the original mean adjusted image is far greater than a typical face image.

4.2 Performance

The scheme described in the last section has been implemented by the MIT Vision and Modeling Group [2][39]. The code was obtained from their FTP site. To evaluate the performance of this approach, sets of five, ten, and fifteen images were used as the training sets. Fifteen eigenfaces and less were used as the bases of the face space. Processing time on a DEC-ALPHA machine was about 900 ms. Even though individuals in the database are identified uniquely, the system performed poorly under changes of illumination and facial expression. Only 5 out of 15 were identified under such conditions.

4.3 Shortcomings

To investigate the poor performance of this approach, two sets of tests were conducted with respect to the suitability of the trained eigenfaces and with regard to facial expression tolerance.

To evaluate the former, linear combinations of the training sets (also available in the database) were introduced to the system . That is, image i_{mn} was defined as

$$\begin{aligned} i_{mn} &= ki_m + (1-k)i_n \\ 0 &\le k \le 1 \end{aligned} \tag{4.6}$$

A three dimensional example of i_{mn} evolution in the spatial domain is depicted in Figure 4.3.

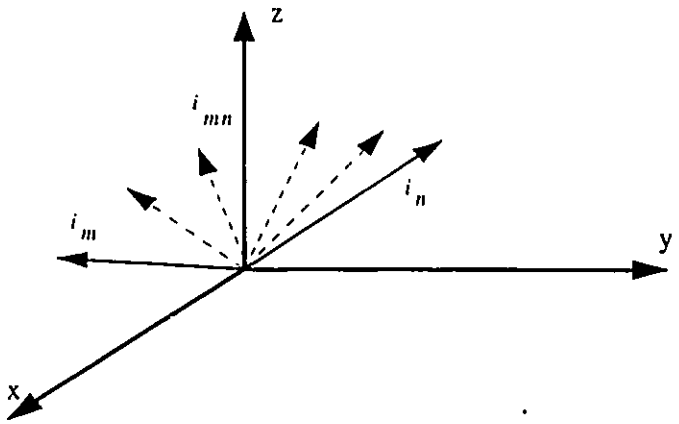
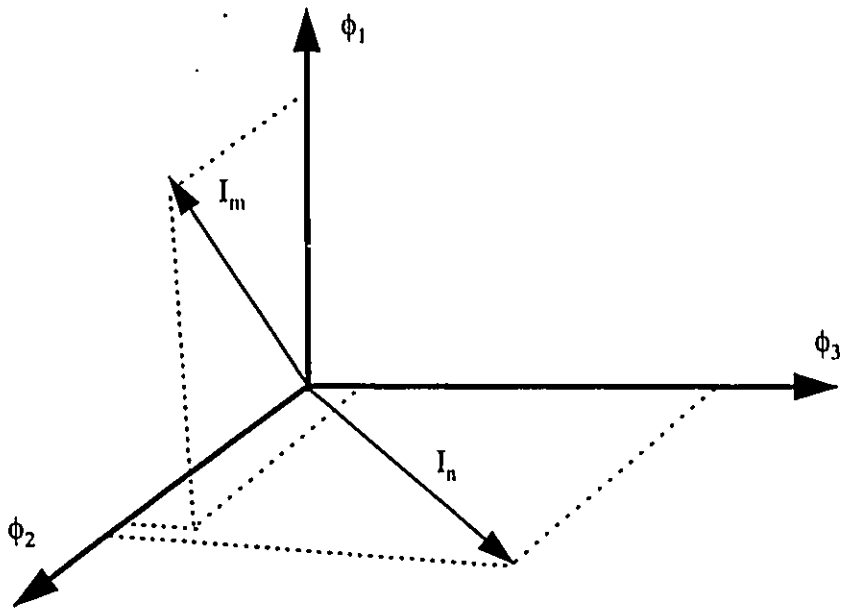


Figure 4.3 Evaluation of the KLT Subspace
Evolution of $i_{\_}$ For Different k (dashed lines)

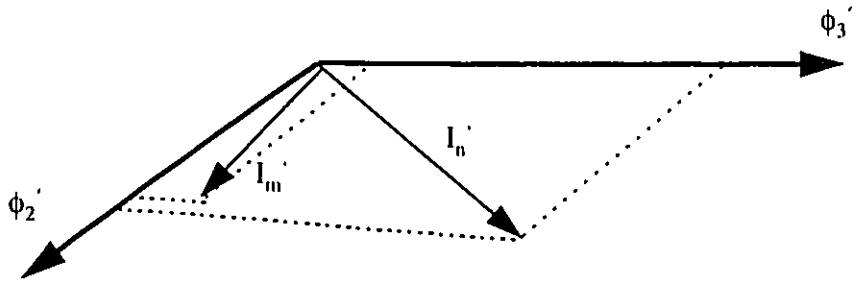
Under ideal conditions (*i.e.* closed, orthogonal bases), it is expected that I_{mn} (the transform of i_{mn}) is to be described as a linear combination of the i_m and i_n transformations. That is, the Euclidean distance between i_{mn} to either i_m or i_n should be the same as the Euclidean distance of I_{mn} to either I_m or I_n (Parseval's

Theorem). However, this is not the case. Most often a preference is given to one of the images (i_m or i_n), or i_{mn} causes a complete false alarm (neither i_m nor i_n is detected). The reason for this error cannot be lack of a sufficient number of bases, since the error persisted when the number of bases equaled the rank of the correlation matrix.

To investigate the source of this error, the suitability of the training sets was analyzed. Given a well behaved correlation matrix (a diagonal matrix with an eigenspread of unity), let I_m and I_n represent the projections of i_n and i_m on the eigenvector space defined by this correlation matrix, as depicted in Figure 4.4*a)*.



a)



b)

Figure 4.4 Eigenspaces of White Noise (a) and a Colored Noise (b)

Under the conditions depicted in 4.4*a*) (eigenspread of unity) Parseval's theorem holds, and the MSE in the spatial domain equals that of the transform domain.

For a correlation matrix with a large eigenspread (Figure 4.4*b*)), projection of the same i_m and i_n will generate different vectors. I_n', the counterpart of I_n, has remained the same as before; however I_m' has been modified by this new projection. Under such conditions, it is obvious that the error energy in the transform domain is not equal to that in the spatial domain. The MSE surfaces of these spaces are illustrated in Figure 4.5.

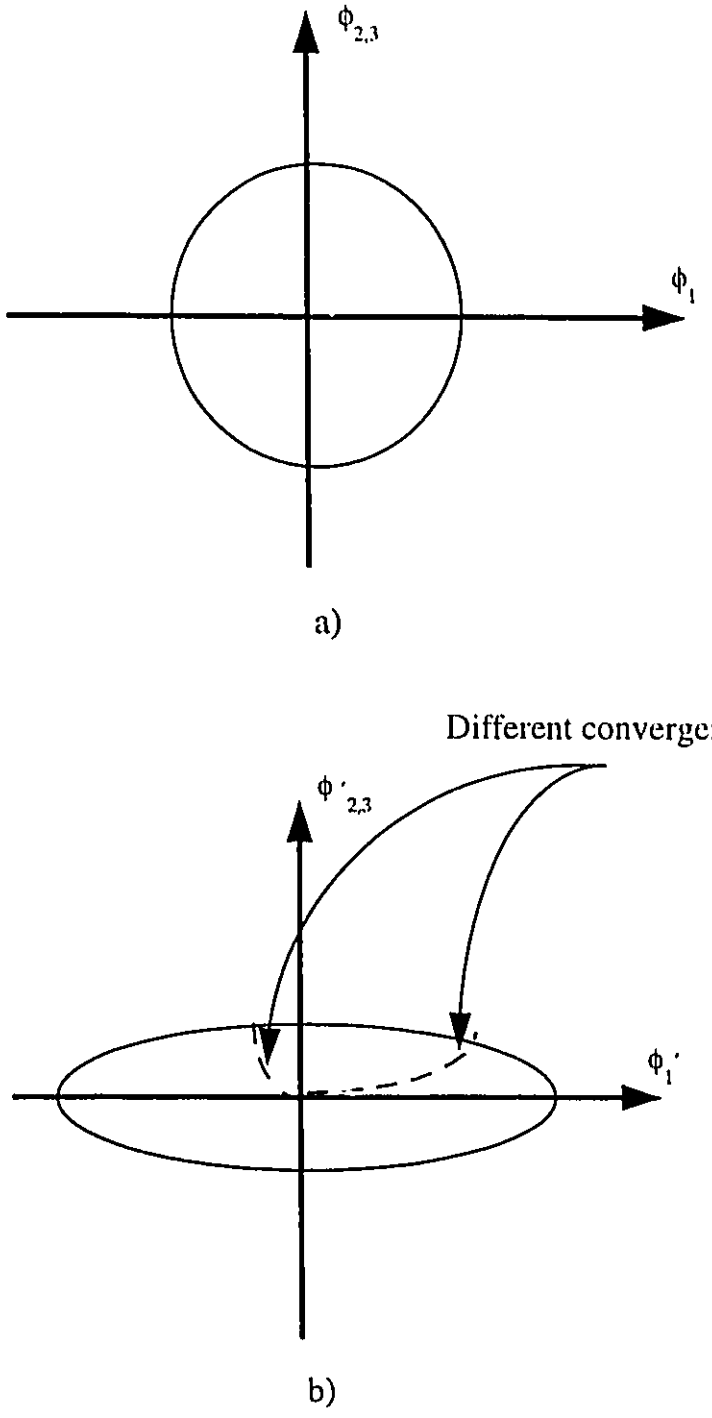


Figure 4.5 Error Surface Contours, for an Eigenspread of Unity (a), and Larger (b)

In the first case Figure 4.5*a*), all the points on the error surface contour are equidistant from the minimum error point (the origin); whereas in Figure 4.5*b)*, points residing on eigenvectors with larger eigenvalues ($\phi_{2,3}$) will have a smaller error compared to those residing on other eigenvectors (ϕ_1). It then follows that the face images that reside on eigenfaces with larger eigenvalues will always yield to a smaller error, regardless of the similarities between the target and reference image. To rectify this situation, all the sample points should reside on eigenvectors with the highest energy (plane ϕ_2,ϕ_3).

Even though situations depicted in Figures 4.4*a)* and 4.5*a)* are ideal, *i.e.* they do not discriminate among different samples, these conditions are not attainable when the training points belong to the same family. These conditions correspond to uncorrelated white noise [38].

4.4 Proposed methods for improvement

To ensure the suitability of eigenfaces, a number of solutions has been offered. One of the proposed methods suggests defining eigenspaces for different races [2]. The assumption is that the statistical behaviours of the face images within the same race are similar. Alternatively, spaces for different features [39] may be defined. The motivation is that similar features (*e.g.* nose, eyes) would have the same statistical properties.

Statistical similarity is also crucial in implementation, because the approximation $R = uu^T$ will only hold if the process of interest is WSS.

To ensure the suitability of eigenfaces, a better solution is to define a family member not based on subjective measures (race, feature), but on the statistical properties of the sample point. To do so, the training sets should be grouped based on the closeness of their averages and variances, hence ensuring a properly defined correlation matrix, and defining different spaces based on statistical similarities.

The performance of the KLT was also measured with respect to facial gestures. As mentioned in [40], the performance degrades due to statistical changes in the whole image, initiated by local changes such as smiling or removal of glasses. To overcome this problem, a feature based approach may be considered, since the contribution due to each feature may be selectively eliminated. For example, to tolerate smiling, the contribution of the mouth space is eliminated [39].

In summary, even though the KLT possesses many feasible properties in face recognition, care must be taken during the training sessions to ensure that statistically related observation points are used to generate the different eigenspaces. This is to ensure, firstly, the validity of the $R = uu^T$ approximation and secondly, the suitability of the generated bases. Thus during the recognition process, the target object (face or feature) must first be categorized based on its statistical behaviour (mean and variance) and then projected on the eigenspace with similar statistical properties.

Chapter 5

Novel methods for face recognition

The propose of this chapter is to describe novel statistical methods that can tolerate changes in facial expressions and identify an individual uniquely. This is accomplished under the conditions of real time response and implementation feasibility of the software and hardware components.

5.1 Histogram-moment mapping

In this section, a histogram-moment based approach in face recognition is evaluated. The purpose of this new technique is to reduce the effect of differential noise described in Chapter 3.

5.1.1 Description

The improvement gained in histogram comparison, by the windowing method, leads to the conjecture that a better performance could be obtained if the window of interest were the whole histogram and characterizing measures were the statistical properties of the histogram. Hence, to characterize the intensity variations of a face image, the first three moments of a histogram are used. The histogram of a face image can be approximated by a Gaussian distribution, as observed in Chapter 3, thus three moments are sufficient to describe a face histogram (*i.e.* a Gaussian distribution can be uniquely identified by its mean and

variance). The same observation was made by, independently, in [45]. Simulations are performed based upon the first three centralized moments. The use of centralized moments (mean centered) is advantageous, since such moments are translation invariant. That is, the moments remain the same if the histogram is shifted.

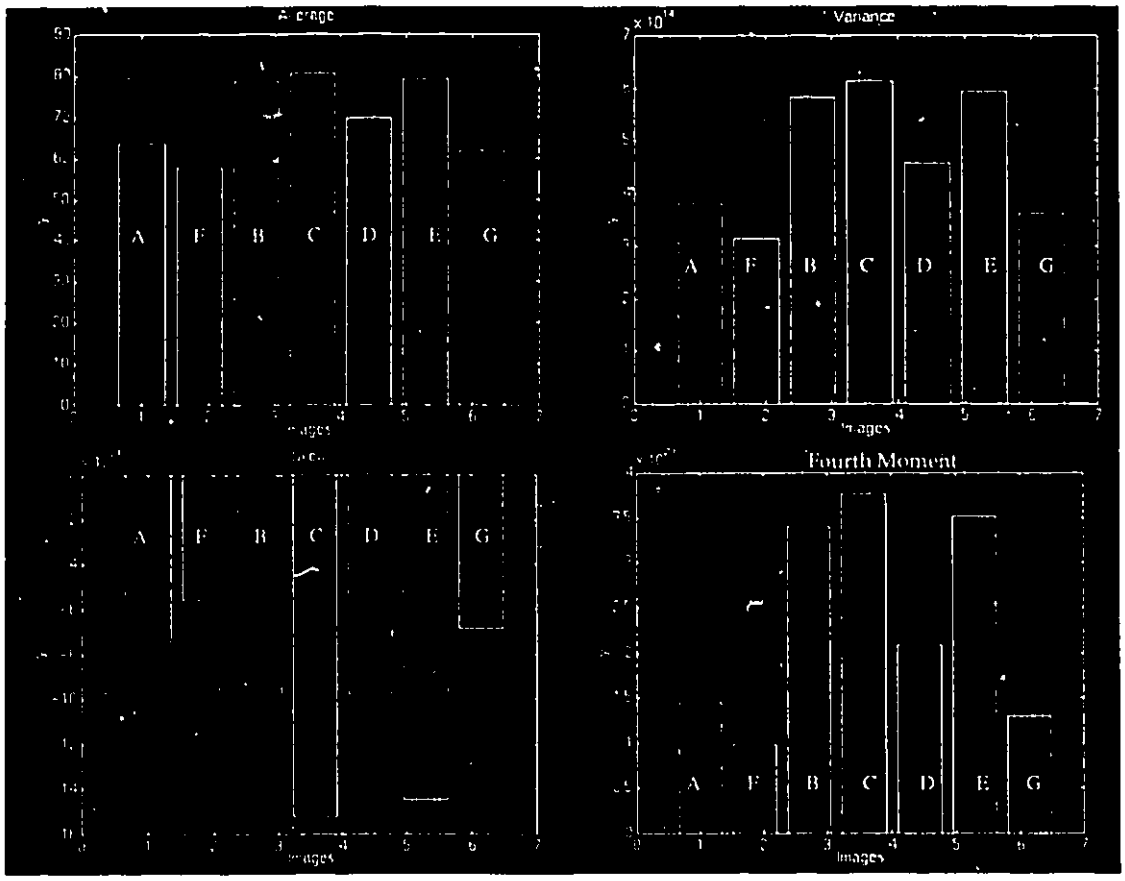


Figure 5.1 Unique Identification of a Face, Based on the First Three Centralized Moments of Its Histogram

Indices on the X Axis Are With Respect to Images in Section 3.1

the Y Axis refers to Average, Variance, Skew, and Fourth Moment Magnitude of Histograms of Face Images

Figure 5.1 shows the first four moments of face images depicted in Chapter 3. Based on extensive set of simulations, it was observed that a face

image can be uniquely identified using the first three moments of its histogram. It was noted that the fourth moment, bottom right, does not add any new information to the decision making process, whereas the first three moments are essential to unique identification.

The statistics of the histogram and its moments, based on the stationary nature of the intensity distribution of a face image, are invariant to facial expressions. To validate this assumption, the first three moments of the images in Figure 3.2 (images of the same person with different facial expressions) are shown in Figure 5.2.

It is noted that the first three moments of different facial expressions and head positions, of the same person, are the same unless the statistics of a histogram are significantly altered, due to major changes in the gray level distribution of a face image (*e.g.* A_4),

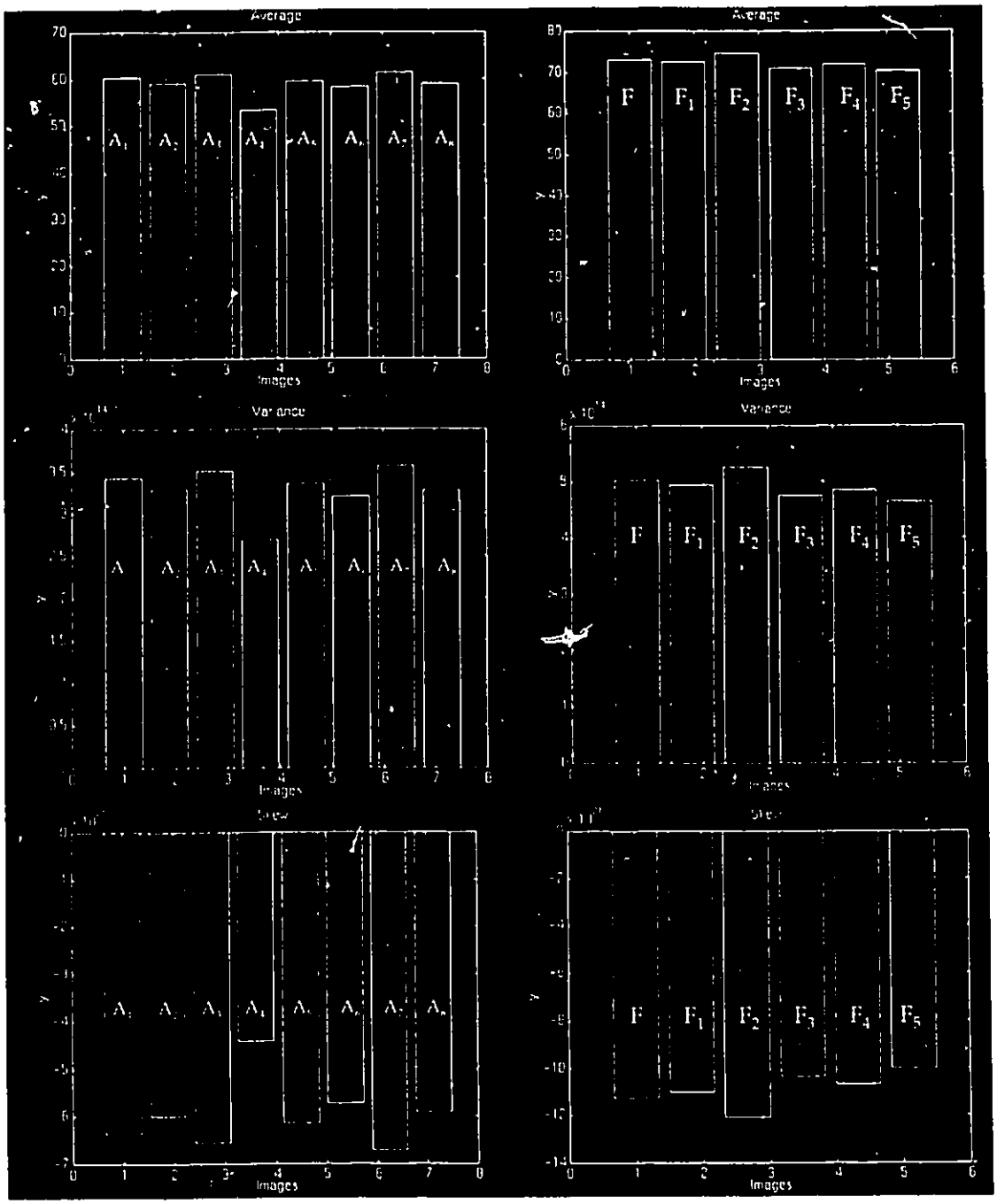


Figure 5.2 Moment Invariance With Respect to Facial Expression Variations and Head Movement for Different Faces. Indices Refer to Figure 3.2

Two major problems are encountered with this approach. Firstly, there is a difference in the order of magnitude of different moments, and secondly, the histogram is expanded and dilated due to changes in illumination. The first problem is relatively well known in different areas of vectorial decision making [43], where one component of a vector dominates the overall behavior of that vector, (*e.g.* energy, Hamming distance). A number of methods is commonly used to address this problem. By far the most popular approach is normalization with respect to the sum of the variances of the training set [44]. In this method, each moment is normalized with respect to the total energy of the moments of the training set. Noting that the histogram moments vary within the same order of magnitude (Figures 5.1 and 5.2), an approximation to this method is used. That is, normalization with respect to the order of magnitude of each moment, multiplied by some arbitrary factor (close to one) is performed. By employing this approximation, the need to re-normalize all moments, as new images are added to the database, is removed.

The second problem, which is related to the illumination changes, is shown in Figure 5.3.

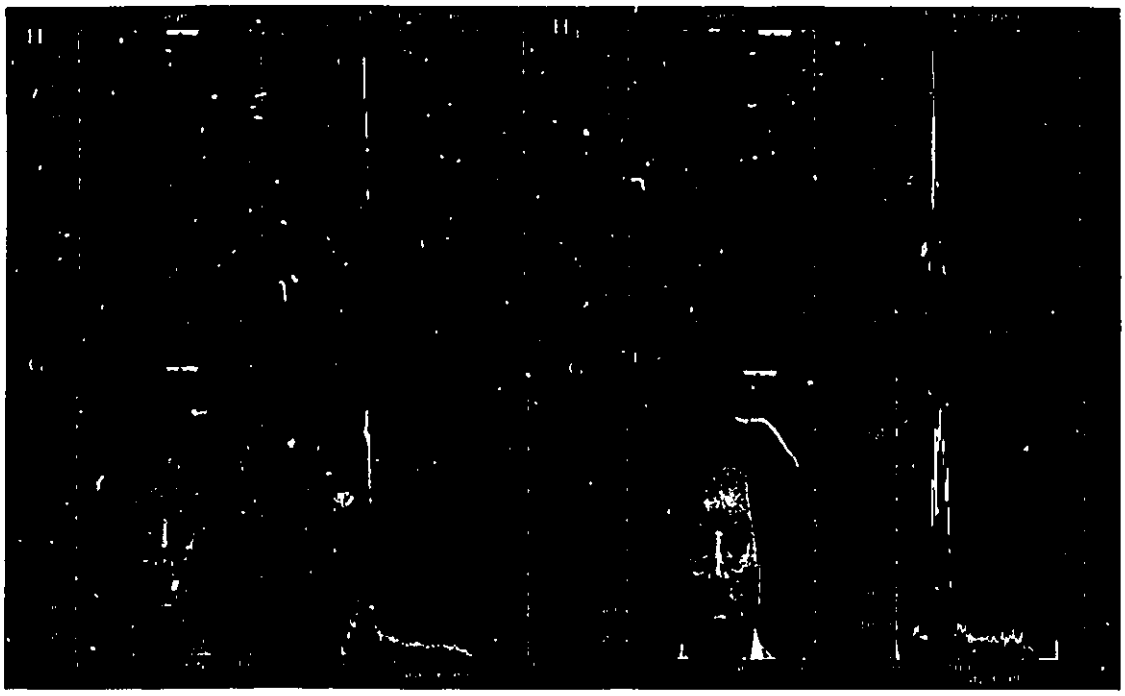


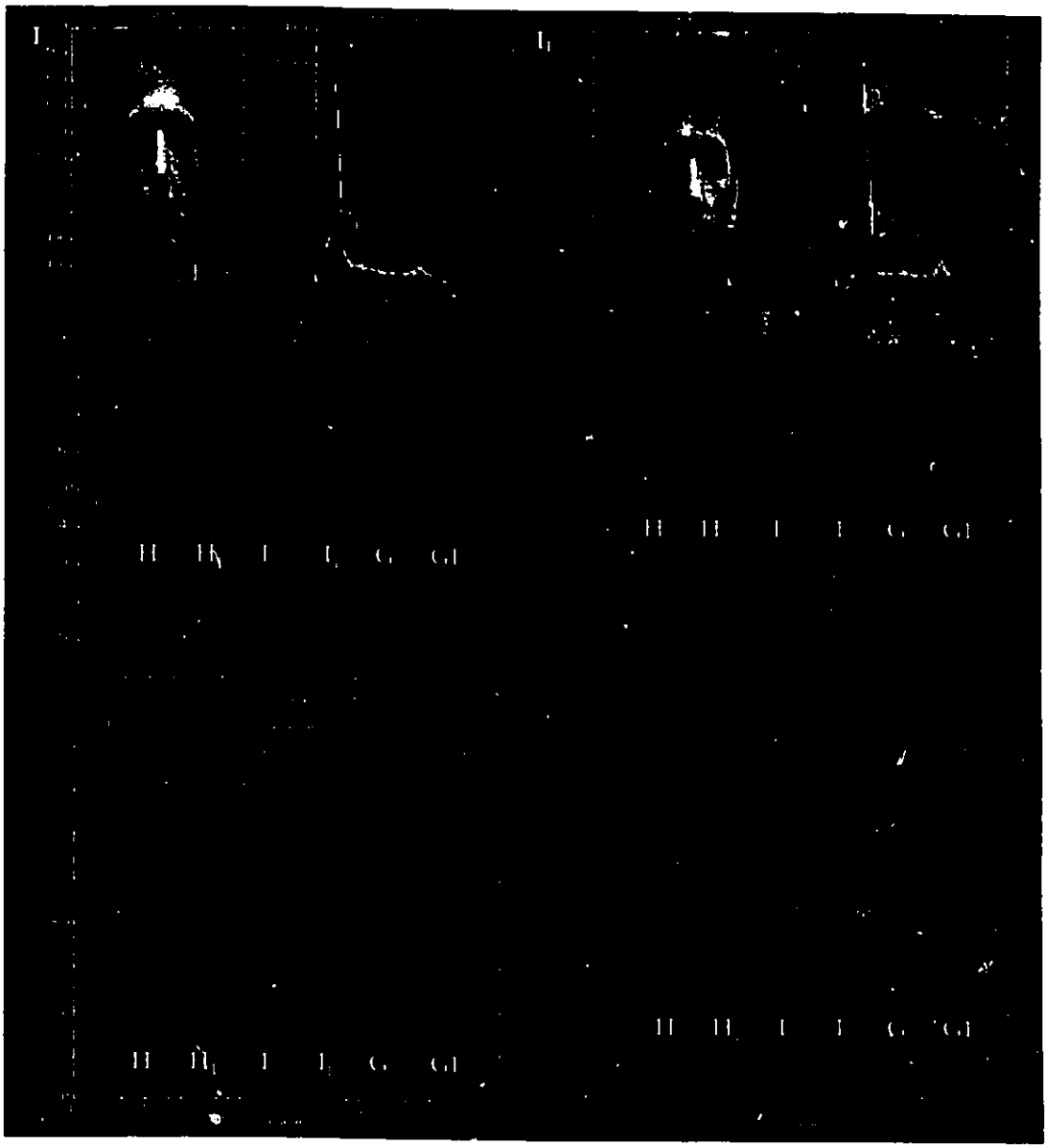
Figure 5.3 Changes in Histogram Due to Changes in Illumination

The histograms on the right side are expanded compared to those on the left side, as the illumination increases. From basic statistics, it is known that if a random variable (*e.g.* intensity variation) is multiplied by a constant (change in illumination), the respective moments change as follows:

$$m_k' = \alpha^k m_k \tag{5.1}$$

where,
α is the multiplication constant,
m_k is the kth order moment before multiplication, and
m_k' is the kth order moment after multiplication.

Application of Equation (5.1) in the decision making process will account for changes in illumination. The improved results are shown in Figure 5.4.



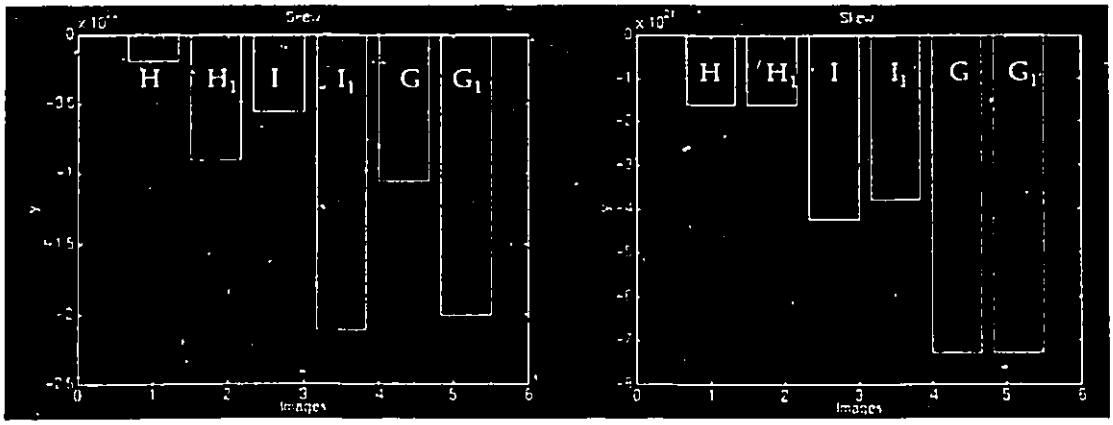


Figure 5.4 Moment Adjusted Test Images

Right Hand Side No Adjustment, Left Hand Side

Application of Equation (5.1)

Comparing the adjusted (right) and non-adjusted (left) moments, it is noted that the first three moments have been adjusted so that they are tolerant towards changes in illumination.

However, before applying Equation (5.1), the adjustment factor α must be determined. Adjustment with respect to the total energy of the image, average of a known segment of the image, or mean of the whole image may be chosen. To ensure the correct scaling of the final adjustment factor (*i.e.* histogram in range of 0-255) the measured quantity has to be normalized with respect to a constant. The constant quantity may be the total expected energy, expected background average, or expected mean of the image. In our system, we have chosen to define the adjustment factor as the ratio of the average value of a pre-defined segment to the expected value of that segment under normal illumination.

5.1.2 Performance

As expected, this method achieves real time performance. On a DEC-ALPHA machine, a 30 ms response time was obtained. Among a database of 15 individuals, 10 identifications were made.

5.2 *Wavelet transform and histogram mapping*

To further improve the performance of the histogram-moment based technique, the geometrical similarity of face images, as discussed in Section 3.1, must be guaranteed. To ensure the validity of this assumption a solution based on the 2-D DWT is proposed.

5.2.1 Description

The motivation for the 2-D DWT approach is that, regardless of the structural geometry of individual faces, there are common orientations among all faces. For example, all noses are oriented vertically and all eyes are oriented horizontally. Thus by extracting commonly oriented features in a face, the requirement for geometrical similarity of a given set of images may be satisfied. Since the 2-D DWT is orientation selective, the subbands of interest are those that capture the most common characteristics among all faces. That is, the subband or the group of subbands should provide maximum distinction among different faces, and minimum difference among the same face images with different poses. The method described below is used to identify the subbands of interest. 2-D Daubechies filters with horizontal, vertical, and diagonal orientation are used for this purpose. An example of such a decomposition is shown in Figure 5.5.

Figure 5.5 D_{II}, 3 Stages Decomposition of Figure 3.1-A

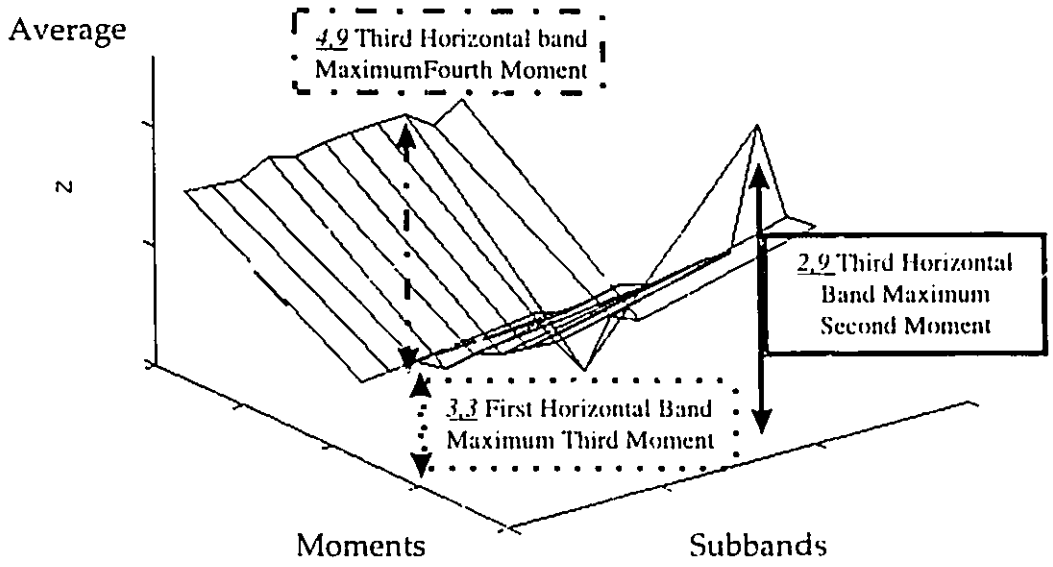
The objective is to identify subbands that will produce a maximum variation in moments for different faces, and a minimum variation in moments for the same face in different poses. This measure is based upon the rationale that an appropriate subband is minimally disturbed by the same set of face images (its moments are insignificant) and is maximally disturbed by different sets of faces (its moments are significant). To identify the subbands of interest, the following procedure is proposed:

1. The 2-D DWT of face images is obtained.
2. Histogram of each subband is constructed.
3. First three moments of all histograms are calculated.
4. All averages, variances, and skews of different images with similar subbands are grouped together.

5. To quantify the disturbance in each subband, the first four moments of each group are calculated.
6. To identify the most suitable bands that tolerate facial expressions, those bands which give the minimum moments are chosen.
7. To identify the most suitable bands that discriminate different faces, those bands which give the maximum moments are chosen.
8. A weighted combination of the bands from steps 6) and 7) is used to satisfy the maximal-minimal disturbance criteria.
9. When a new face is presented, its 2-D DWT is calculated, and the proper bands (6 7) are weighted accordingly (8). The average, variance, and skew obtained in this manner are used for the purpose of the identification.

A sub-optimal solution, by considering only one band, which gives satisfactory results in steps 6) and 7), may be considered. With this in mind, 3 and 5 decomposition levels were performed over a training set. Figures 5.6 and 5.7 show the results of minimum-maximum disturbance analysis for only 3 levels of decomposition. The reason is that as the number of subbands increases, finding a common subband that will satisfy steps 6) and 7) becomes more difficult. This is to be expected because, as the number of stages in a dyadic structure increases, the cross-talk among bands increases as well (see Figure 3.10). The x axis in Figures 5.6 and 5.7 refers to subbands as follows: the first band refers to the original image (column 1), the second band refers to the scale subband, and thereafter, each subband is presented by horizontal, vertical, and

diagonal components of their respective subbands. The y axis refers to the second moment, third moment, and fourth moment. For example, index *3,5* means third row (third moment) fifth column (diagonal component of the first subband). In Figure 5.6, the indices refer to the maximum values of moments in each subband. Thus, for example, index *2,9* in Figure *5.6-Average* means that the second moment of the horizontal component of the third subband has the maximum value among all other subbands (original image included). In Figure 5.7, the indices refer to the minimum value of the moments. Thus index *2,4* in Figure *5.7-Variance* means that the second moment of the vertical component of the first subband has the minimum value among all other subbands (original image included).



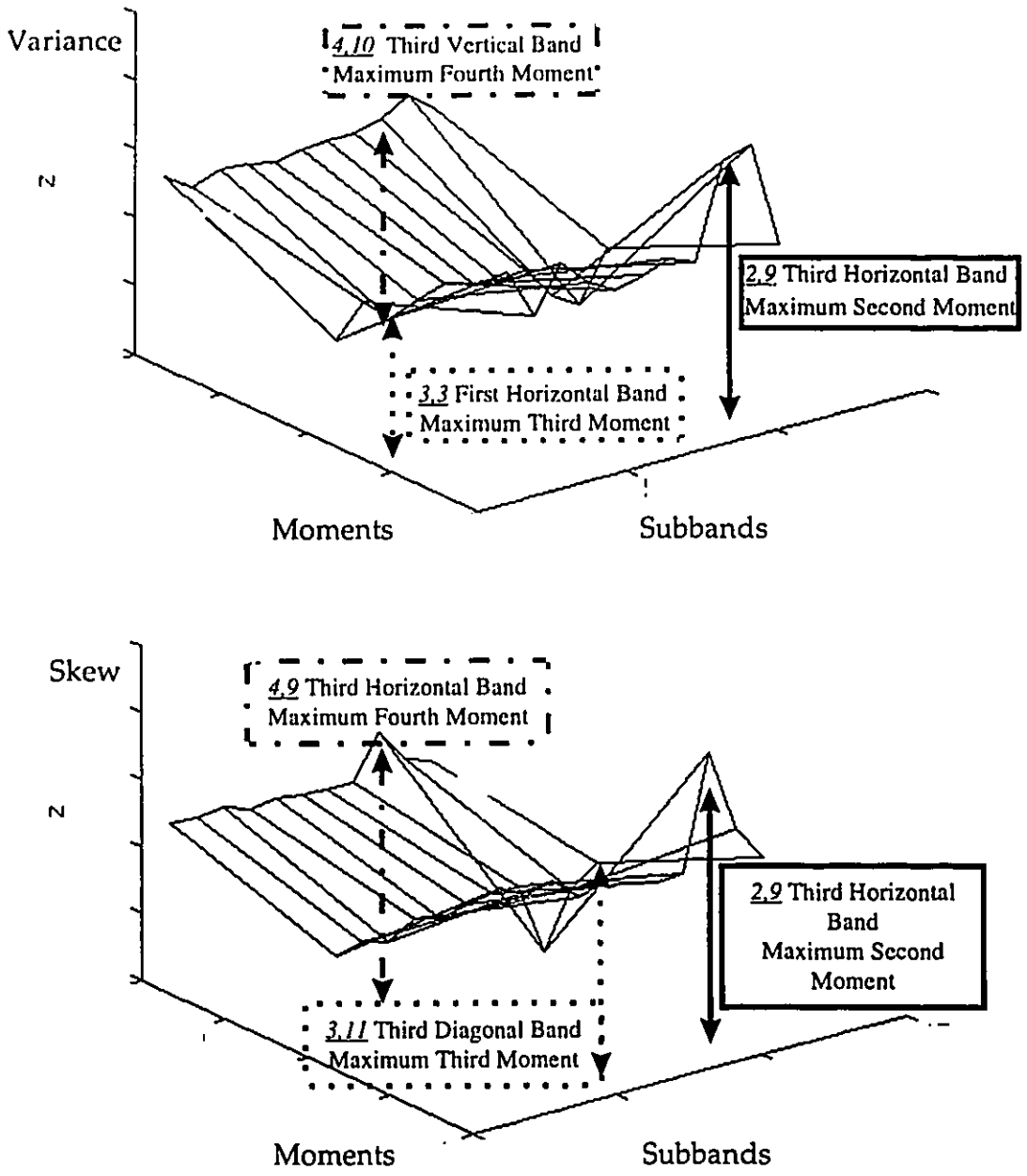
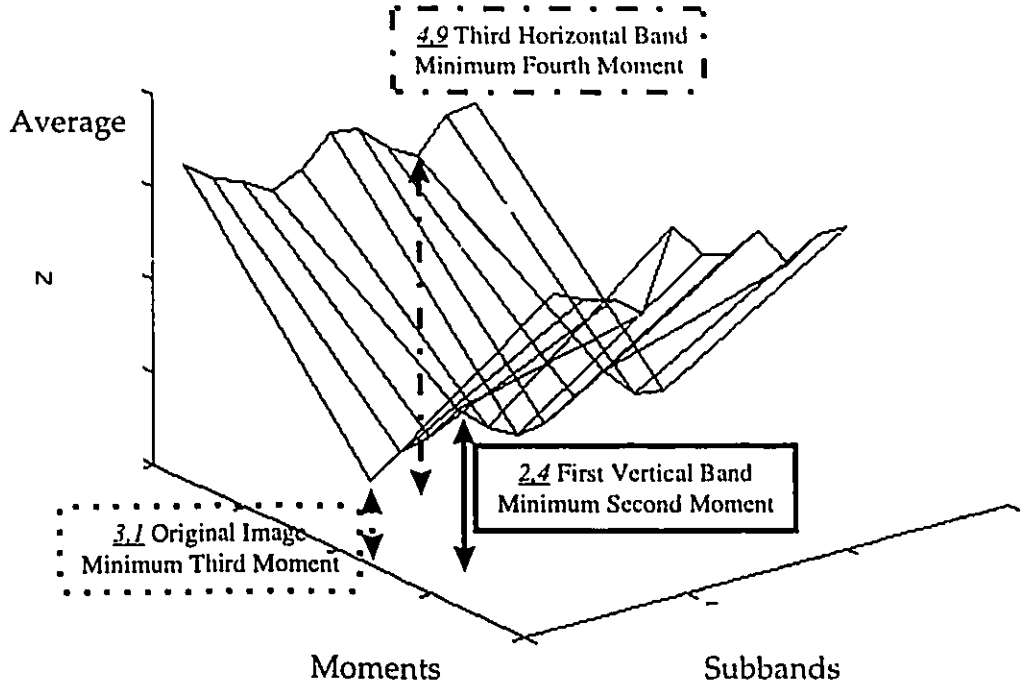
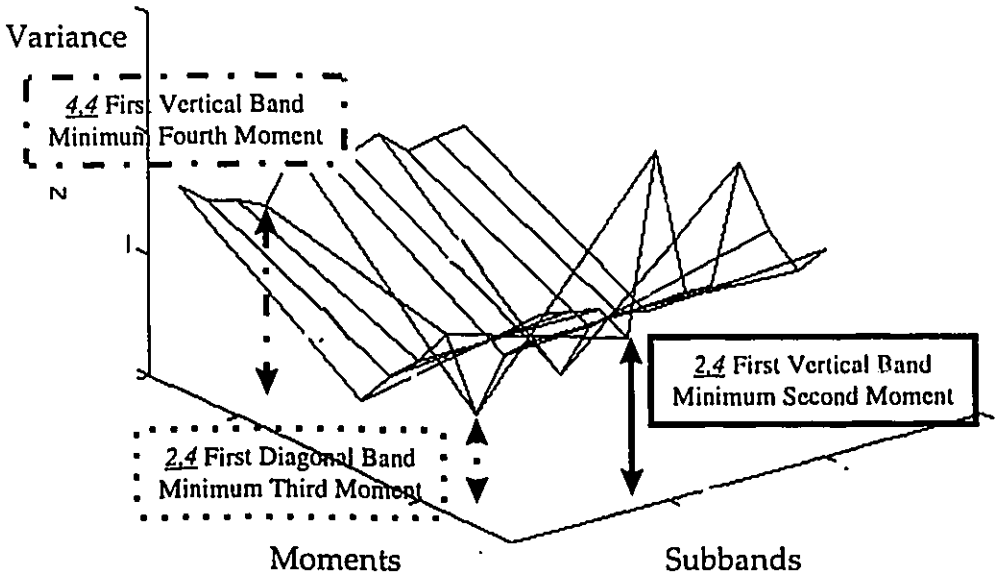


Figure 5.6 A Minimum-maximum Analysis For Different Faces
The Indices Refer to Maximum Moments, For Different Images

Figure 5.6 refers to a minimum-maximum disturbance analysis for different faces (Figure 3.1). In this case, subbands with maximum values of moments are of interest (distinction). It is noted that the third horizontal band satisfies this requirement.


4,9 Third Horizontal Band
Minimum Fourth Moment
Average
N
2,4 First Vertical Band
Minimum Second Moment
3,1 Original Image
Minimum Third Moment
Moments
Subbands


Variance
4,4 First Vertical Band
Minimum Fourth Moment
N
2,4 First Vertical Band
Minimum Second Moment
2,4 First Diagonal Band
Minimum Third Moment
Moments
Subbands

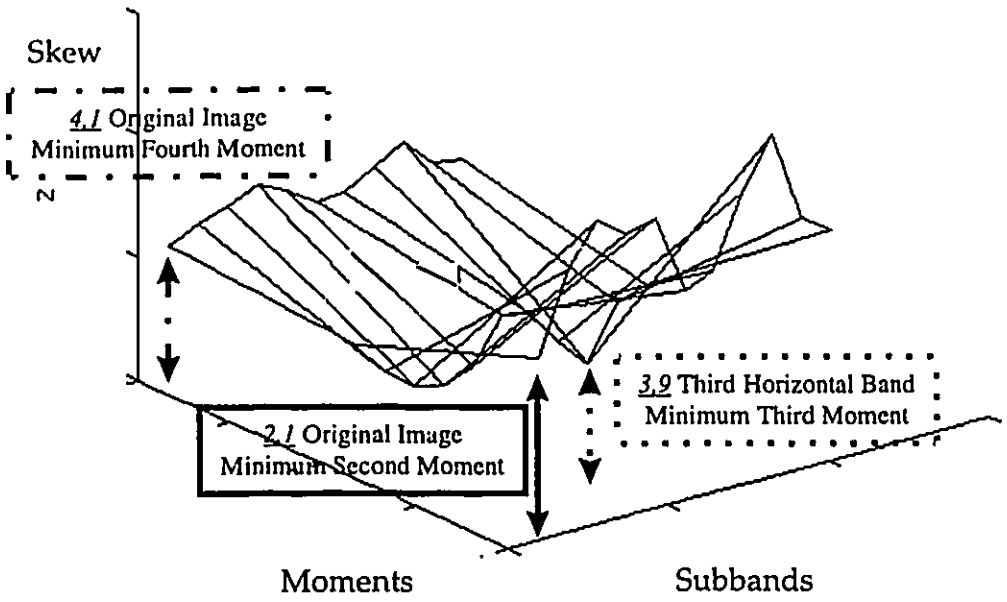


Figure 5.7 A Minimum-maximum Analysis For the Same Face With Different Poses The Indices Refer to Minimum Moments, For the Same Images

Figure 5.7 refers to a minimum-maximum disturbance analysis of the same face with different poses (Figure 3.2), thus subbands with minimum moments are of interest(facial-expression tolerance). It is noted that the first vertical band satisfies this requirement.

In our implementation, the first vertical band (indices *{2,3,4},4*) was used as the discriminating band. That is, after an image was captured, only its first vertical band was used in the histogram-moment analysis. This particular case makes intuitive sense, since this subband captures the vertical component of a face image, which are minimally disturbed for different poses and facial expressions.

5.2.2 Performance

On a DEC-ALPHA machine, a 500 ms response time was obtained. Among a database of 15 individuals, 12 identifications were made. The higher number of identification is at the cost of increased processing time.

In summary, it has been shown that the stationary characteristics (illumination and geometrical similarities) of the operating environment can be exploited to obtain a compact representation of a face image. Furthermore, to ensure the geometrical similarities among all face images the 2-D DWT has been employed to selectively choose the proper orientation in all face images.

Chapter 6

Optical correlator

In this chapter a detailed description of an optical JTC is presented. The main motivation in constructing optical correlators is their fast response time and ease of implementation. The first step in identifying a suitable application for optical correlators is to identify the flexibilities and limitations of such systems. Face database searches, as described in Chapter 1, are among applications with a simple optical solution. The reason is that control can be exercised over the operating environment. Since both the reference and target images are captured within a well defined confinement, performance degradation, as described in Chapter 2, due to scale and rotation are minimized. Similarly, noise, introduced due to varying illumination, can be minimized. Furthermore, the background is conditioned such that its intensity is uniform and close to that of an average face intensity. With this condition, three objectives are satisfied. Firstly, Horner efficiency will be high, since the scenery itself is a low frequency signal. Secondly, overlapping noise requirement is achieved. Finally, the need for an adaptive approach to approximate the background noise is alleviated. To avoid the use of holographic media, the JTC is the preferred approach in realizing optical correlators.

6.1 Description

The setup shown in Figure 6.1 was designed, implemented, and tested for the purpose of optical database searches.

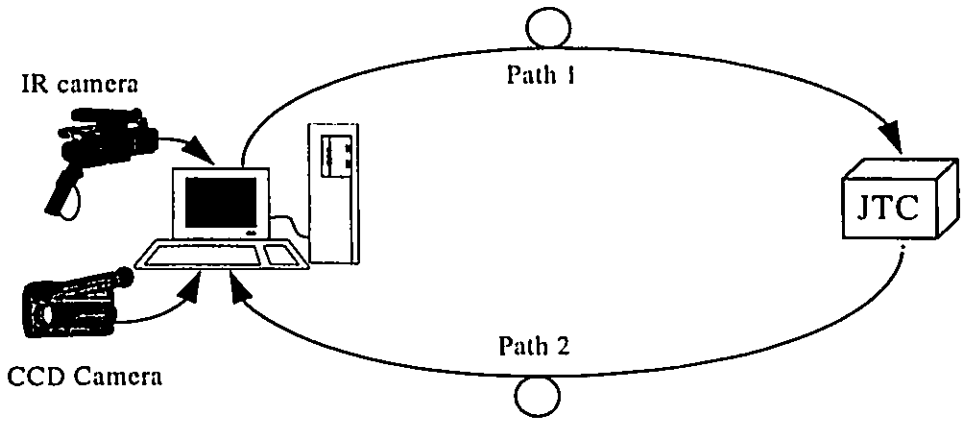


Figure 6.1 A Realization of the Optical JTC Based Face Recognition System

The visible images are obtained by the PULNIX TM-7CN camera, and the infra-red images are obtained by the MITSUBISHI IR-AM500 camera, both at 640x480x8 resolution.

The multi-mode optical links (Path 1 and Path 2) are used to transfer information (images and correlation peak) to and from the optical JTC.

The JTC subsystem is as shown in Figure 6.2. The write-beam is at the wavelength of 830nm, and the read-beam is at wavelength of 852nm (refer to Chapter 2 for the operational characteristics of the SIMQW devices). The CCD camera is the PULNIX TM-7CN camera at 640x480x8 resolution.

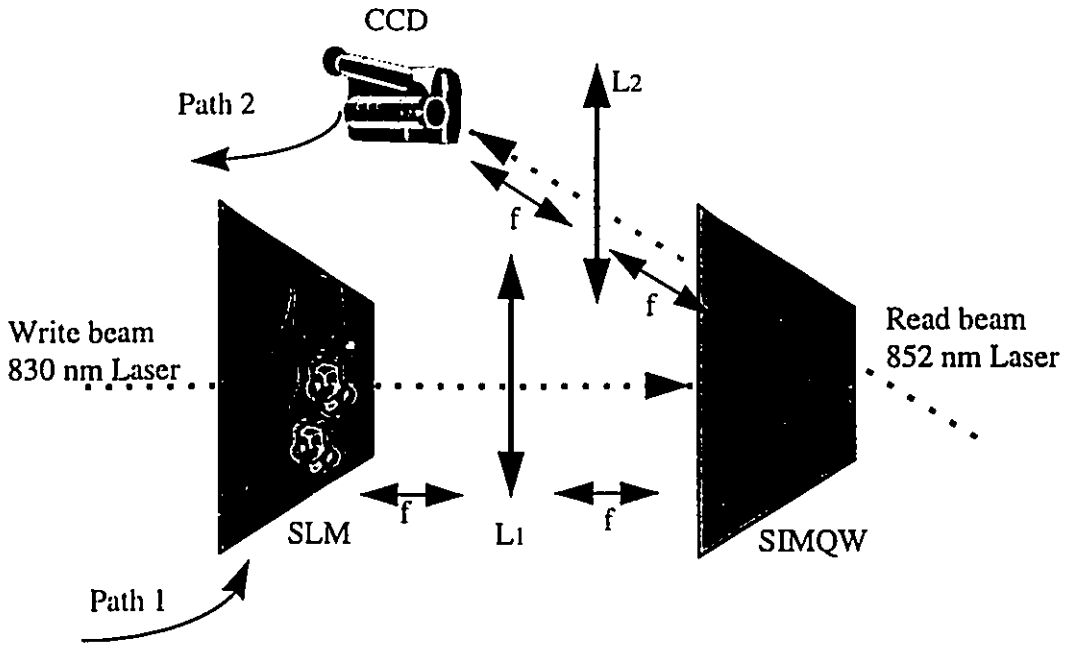


Figure 6.2 An Implementation of JTC

Two modes of operation are considered, image and information acquisition, and image identification. During data acquisition, personal information along with visible and infra-red images of the subject are stored. The system maintains an internal database to uniquely identify an individual with his or her respective personal data. Furthermore, image frames containing faces of four individuals are created at this time. During the identification process, the target image is pasted to the frames and is displayed on the SLM, which complies to NTSC standard, via an optical fiber link (path 1). Figure 6.3 represents a typical input to the JTC.

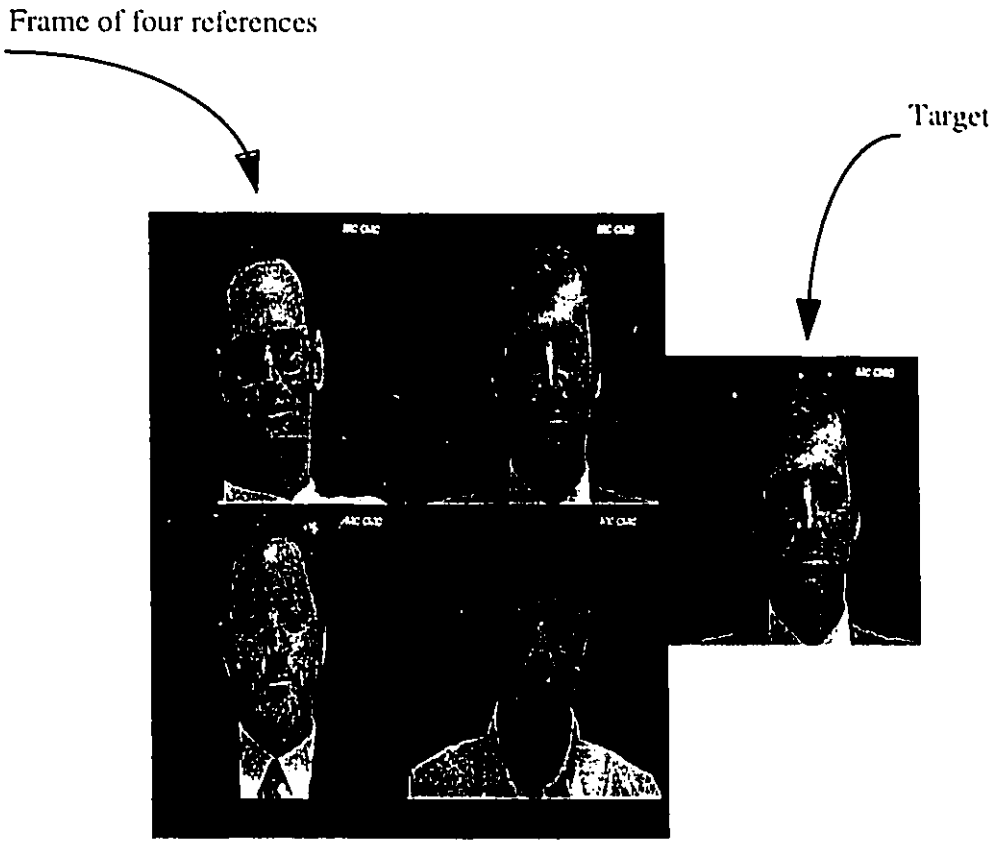


Figure 6.3 Typical JTC Frame

The output of the JTC is captured with a CCD camera, and transferred via an optical fiber link (path 2) to the PC. The output is searched at small specific regions for possible correlation peaks. Figure 6.4 depicts the corresponding JTC output of Figure 6.3.

To increase the overall throughput of the system, four faces at a time are compared (frame of four images). The main criterion, in limiting the number of images per frame, is the sensitivity of the JTC with respect to the size of the face images. That is, for a given display area (SLM), as the number of images increases, the resolution of each face image decreases.

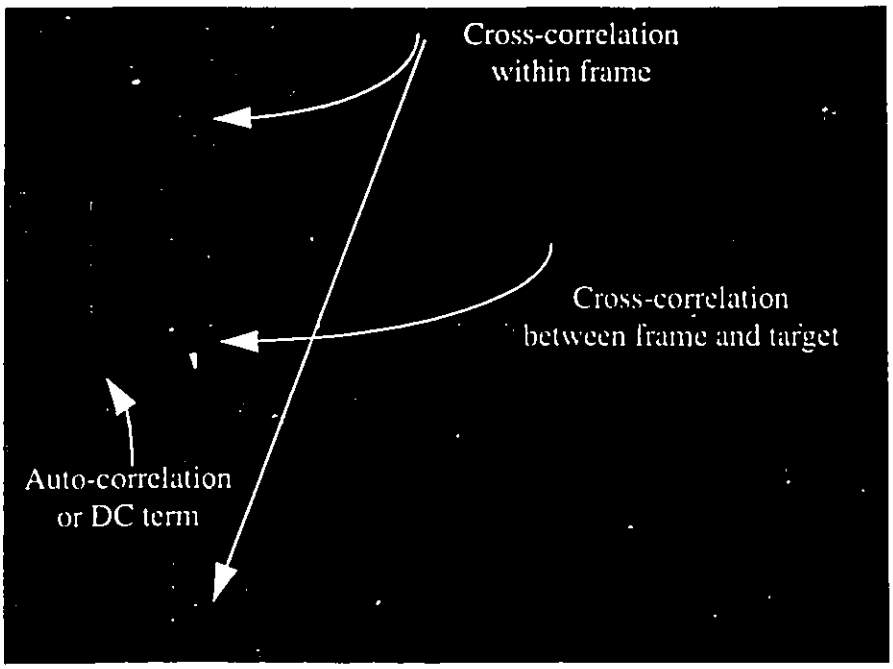


Figure 6.4 Typical Output of an Optical JTC

Empirically, it was determined that face images of size 420x480 pixels, which corresponds to the frame size 1260x960, provide a sufficient resolution for the JTC.

6.2 Performance

The upper time limit, for processing each individual, is 1/120 sec/person. That is, 1/30 of second is spent for four persons where 1/30 of second is the NTSC frame rate. However, this upper limit is not reached, due to a bottleneck during disk access time (*i.e.* frame retrieval time).

All 15 individuals in the database were identified correctly. The system also shows a good tolerance towards variations in facial expressions such as smiling and removal of glasses [41] as shown in Figure 6.5.

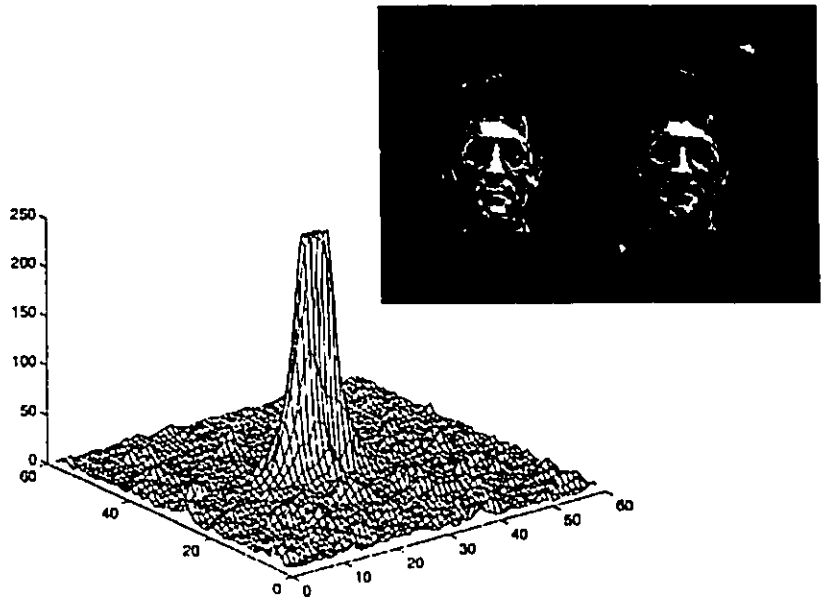


Figure 6.5 High Correlation Peak Obtained Under Facial Changes

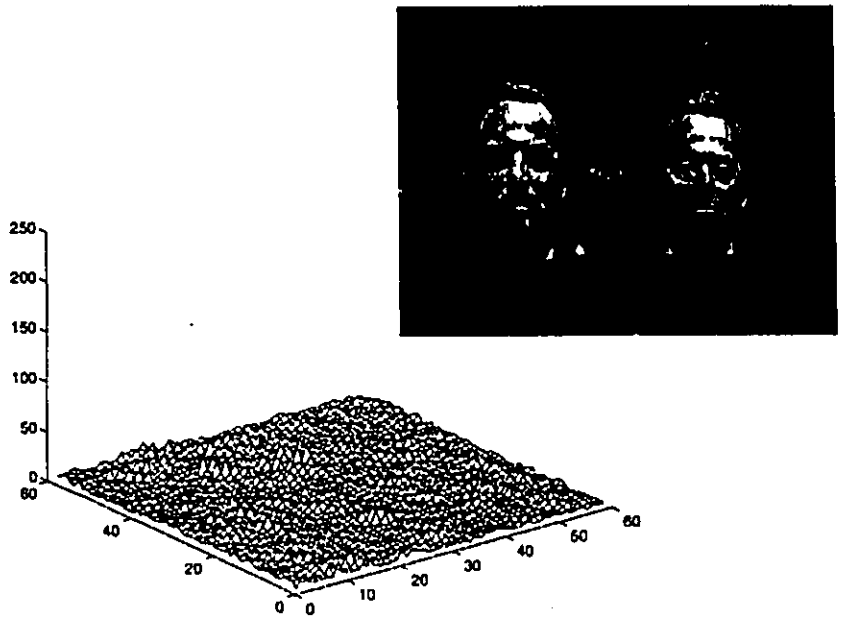


Figure 6.6 Negligible Correlation Peak for Dissimilar Faces

The system also shows a low correlation peak for dissimilar faces, as illustrated in Figure 6.6.

With respect to Figures 6.5 and 6.6, a threshold above 150 is used for the purpose of identification and discrimination.

6.3 Shortcomings

Even though the system performs well under the conditions of real time response and facial expression tolerance, its power to discriminate could be enhanced if a preprocessing filter were applied to both the target and reference images. Such a filter may simply be a DC-block that is used to enhance the discrimination among faces by eliminating the common DC components. In the following section, an optical solution to this problem is offered that simultaneously preprocesses both the target and reference images.

6.4 Proposed methods for improvement

Let us consider the application of a DC-block filter, on both the target and reference images, in order to enhance the discriminating power of the optical JTC. In the following derivations, 1-D signals have been employed for ease of notation. The result also holds for 2-D signals.

In the input plane the following function is observed:

$$i(x) = t(x - x_0) * h_{DC}(x) + r(x + x_0) * h_{DC}(x) \tag{6.1}$$

where,
$i(x)$ is the function describing the input plane,
$t(x)$ and $r(x)$ are functions describing the target and reference images, respectively,
x_0 is the distance with respect to the center of the input plane, and
$h_{DC}(x)$ is the DC-block filter.

The Fourier transform of $h_{DC}(x)$ is defined as:

$$H_{DC}(w) = \begin{cases} 0 & w = 0 \\ 1 & 0 < w < 2\pi \end{cases} \qquad (6.2)$$

The Fourier transform of Equation (6.1) is

$$I(w) = T(w)H_{DC}(w)e^{-jwx_0} + R(w)H_{DC}(w)e^{jwx_0} \qquad (6.3)$$

which is the amplitude of the field before the SIMQW device. The magnitude of the field (*i.e.* its intensity on the SIMQW device), is as follows:

$$I(w)I^*(w) = \left|T(w)H_{DC}(w)\right|^2 + \left|R(w)H_{DC}(w)\right|^2 + T(w)R^*(w)\left|H_{DC}(w)\right|^2 e^{-2jwx_0} + T^*(w)R(w)\left|H_{DC}(w)\right|^2 e^{2jwx_0} \qquad (6.4)$$

Factoring out the common magnitude of the filter, the following is obtained.

$$\left|I(w)\right|^2 = \left|H_{DC}(w)\right|^2 \left[\begin{matrix} \left|T(w)\right|^2 + \left|R(w)\right|^2 + \\ T(w)R^*(w)e^{-2jwx_0} + T^*(w)R(w)e^{2jwx_0} \end{matrix} \right] \qquad (6.5)$$

With respect to the definition of $H_{DC}(w)$, it follows that

$$H_{DC}(w) = \left|H_{DC}(w)\right|^2 \qquad (6.6)$$

From Equations (6.5) and (6.6), it is concluded that in order to filter both the target and reference images, as described by (6.1), it is sufficient to block the

DC component at the Fourier plane (*i.e.* SIMQW). Optically, this is achieved by the application of an opaque point in the center of the Fourier plane.

Chapter 7

Conclusions and future direction

In this thesis, novel and improved methods in face recognition have been presented, analyzed, and evaluated. Each method has been evaluated based upon the following requirements (*i*) facial expression tolerance (*ii*) real time response, and (*iii*) implementation feasibility.

To ensure tolerance towards changes in facial expression, statistical, wavelet, correlation, and eigen based approaches have been considered. These approaches are the most successful face recognition systems of today [40][41].

A detailed analysis of a KLT based approach, the *photobook*, was presented. It was shown that even though the KLT offers many beneficial properties in image processing applications; care must be taken during the training phase of such a system. Otherwise, severe degradation in the system performance could arise, under the requirement of facial expression tolerance. A simple quantitative method, based upon statistical properties of the images, was suggested as a solution to this problem.

Statistical behaviour of the operating environment has been exploited to devise simple algorithms, amenable to electronic implementation, while satisfying facial expression tolerance and real time response requirements. A histogram-moment based approach has been offered as a solution that exploits

the stationary behaviour of the operating environment, thus achieving all three requirements of a face recognition system. It was shown that such a system provides a reasonable accuracy and very high throughput. The 2-D DWT was another statistical technique presented. The main objective in employing the 2-D DWT is its orientation selectivity. It was observed that application of the histogram-moment analysis on a selected group of subbands would improve the performance of the system. A method for identifying the appropriate group of subbands has been offered.

To achieve real time response, optical architectures for correlation based systems were implemented. It was observed that the optical JTC performed well under the facial expression tolerance and real time response requirements. Furthermore, an optical method for further improvement was offered. However, the application of optical techniques may be hindered due to the unavailability of optical components and the specialized expertise needed for constructing such systems.

7.1 Future work

With regards to the statistical methods proposed in this thesis a number of possible improvements can be envisioned. To improve the performance of the histogram-based system, the fundamental assumption of uniqueness based on intensity variation must be ensured. This means that the geometrical symmetry of all face images must be guaranteed. The simplest way to achieve geometrical symmetry, aside from using 2-D DWT, is to categorize the face images based on

their similarities. Wire-frame templates may be employed to group face images based on the similarity of facial structure (round faces, narrow faces, *etc.*). Thereafter, histogram decision making is applied over the family of faces with similar structure. Although the wavelet-histogram method was originally proposed as an improvement over the histogram-moment based approach, it still may be modified to produce a better result. An adaptive procedure such as LMS may be employed to determine the proper weighting of each subband. Furthermore, subbands oriented at angles other than 0, 45, and 90 degrees may be considered.

The class of optical approaches will achieve a higher throughput and faster response time, given the rapid improvements in the optical technology. However, two issues must be addressed concerning the feasibility of optics; firstly, interface cost and secondly, analog accuracy.

All optical systems must interface with an electronic subsystem. Thus the limited accuracy of an analog optical system (about 8 bits [10]) is an issue that must be considered before employing an optical engine. To maximize the feasibility of an optical system, its interface cost must be minimized by interfacing at points where accuracy is not critical.

To elaborate on this point, in Chapter 6 it was shown that interfaces to the electronic subsystem were made at the points of raw-input (target-reference plane) and final output (correlation plane). The electronic subsystem was not concerned with the intermediate operation of the optical black box. On the other

concerned with the intermediate operation of the optical black box. On the other hand, if we were to overextend the capabilities of an optical correlator by using it as a pure correlation engine (*i.e.* any image, any filter), then the performance of the system would degrade considerably on two accounts. Firstly, IO cost will increase between the decision making electronic subsystem and the optical engine and secondly, the limited accuracy of an optical engine will degrade the overall performance of the system.

The future work in electronic approaches, mainly lies in integrating different methods to address new applications. For example, tracking an individual in an uncontrolled environment is at present gaining importance. To satisfy the requirements of such a system, two paths may be envisioned. The first is the construction of an algorithm that, at minimum complexity, will extract the maximum relevant amount of data. The second is the construction of a system that will employ existing methods at different operational points.

Although implementation of highly complex algorithms is possible, a closer study of an application may point to a hierarchical approach where not one, but varieties of approaches are employed during different phases of the operation. As with any design problem, an accomplishment of a large task can be achieved by completing manageable small tasks. For example, considering the problem of a moving face image, the face entity itself can be identified by use of a crude KLT. After isolating the face, the individual may be recognized by a simpler approach such as correlation or histogram based approaches.

References

[1] A. N. Akansu, and R. A. Haddad, Multiresolution signal decomposition transforms, subbands, wavelets, Academic Press Inc., 1992

[2] M. Turk, and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, Vol. 3 No. 1, pp. 71-86, 1991

[3] I. Daubechies, pp. 38 to 65 of [4]

[4] J. M. Combes, A. Grossmann, and Ph. Tchamitchian, *Wavelets time-frequency method and phase space*, Second Edition, Spring-Verlag, 1990

[5] A. Vanderlugh, *Optical processing*, JohnWiley & Sons, Inc. , 1992

[6] M. K. Hu, "Visual pattern recognition by moment invariance", IRE Transactions on Information Theory, IT-8, pp. 179-187, 1962

[7] J. G. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields", IEEE Transactions on Biomedical Engineering, Vol. 36, No. 1, pp. 107-114, Jan 1989.

[8] B.S. Manjunath, R. Chellappa, and C. V. Malsburg, "A feature based approach to face recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceeding: pp. 373-378, June 15-18, 1992, Champaign, Illinois

[9] F. L. Pedrotti, and L. S. Pedrotti, *Introduction to optics*, Prentice Hall, New Jersey,1993

[10] D.G. Feitelson, *Optical computing*, MIT Press, 1988

[11] M. Francon, *Optical image formation and processing*, Academic Press, New York, 1971

[12] J. W. Goodman, *Introduction to Fourier optics*, McGraw-Hill, San Francisco, 1968

[13] S. E. Shields, Chapter 12 of [26]

[14] D. A. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Electrical field dependence of optical absorption near band gap of quantum-well structures", The American Physical Society, pp. 1043-1060, July 15, 1985

[15] Shin-Tson Wu, Chapter 1 of [26]

[16] H. H. Aresnault, and Y. Sheng, *An introduction to optics in computers*, SPIE Press, Vol. TT8, 1992

[17] L. Esaki, Chapter 1 of [27]

[18] L. Migilo, C. Molteni, and L. Colombo, pp. 279-307 of [28]

[19] M. Jaros, *Physical and application of semiconductor microstructures*, Oxford Science Publications, 1989

[20] D. A. B. Miller, and D. S. Chemla, and S. Schmit, Chapter 13 of [29]

[21] D. S. Chemla, pp. 423-432 of [27]

[22] U. Efron, and G. Livescu, Chapter 5 of [26]

[23] A. M. Glass, D. D. Nolte, D. H. Olson, G. E. Doran, D. S. Chemla, and W. H. Knox, "Resonant photodiffractive four-wave mixing in semi-insulating

GaAs/AlGaAs quantum wells", Optics Letters, Vol. 15, No. 5, pp. 264-266, March 1, 1990

[24] A. Partovi, A. M. Glass, T. H. Chiu, and D. T. H. Liu, "High-speed joint-transform optical image correlator using GaAs/AlGaAs semi-insulating multiple quantum wells and diode lasers", Optics Letters, Vol. 18, No. 11, pp. 906-908, June 1, 1993

[25] D. D. Nolte, D. H. Olson, G. E. Doran, W. H. Knox, and A. M. Glass, "Resonant photodiffractive effect in semi-insulating multiple quantum wells", Optical Society of America, Vol. 7, No. 11, pp. 2217-2225, November 1, 1990

[26] U. Efron, *Spatial light modulator technology material, devices and application*, Marcel Dekker, Inc. , 1995

[27] E. E. Mendez, and K. V. Klitzing, *Physics and applications of quantum wells and superlattices*, NATO ASI Series, 1987

[28] L Migilo, A. Stella, *Semiconductor superlattice and interface*, Proceeding of the International School of Physics, 1993

[29] H. Haug, *Optical non-linearities and instabilities in semiconductors*, Academic Press, Inc. , 1988

[30] B. Javidi, P. Refregier, J Wang, and P. Willett, Chapter 1 of [36]

[31] C. S. Weaver, and J. W. Goodman, " A technique for optically convolving two functions", Applied Optics, Vol. 5, No. 7, pp. 1248-1249, Jul. 1966

[32] J. A. Davis, M. A. waring, G. W. Bach, R. R. Lilly, and D. M. Cotter, " Compact optical correlator design", Applied Optics, Vol. 28, No. 1, pp. 10-11, Jan. 1989

[33] K. H. Fielding, J. L. Horner, and C. K. Makeau, " Modified two focal length correlator", Applied Optics, Vol. 29, No. 21, pp. 4332-4333, Oct. 1990

[34] B. V. K. V. Kumar, Chapter 2 of [36]

[35] D. Casasnet, D. Psaltis, " New optical transforms for pattern recognition" Proceeding of IEEE, Vol. 65, No. 1, pp. 77-84, Jan. 1977

[36] B. Javidi, and J. L. Horner, *Optical information processing*, Academic Press, Inc. , 1994

[37] A. K. Jain, *Fundamentals of digital image processing*, Prentice Hall, 1989

[38] S. Haykin, *Adaptive filter theory*, Prentice Hall, 1991

[39] B. Moghaddam, and A. Pentland, " Face recognition using view-based and modular eigenspaces", Automatic Systems for the Identification and Inspection of Humans, SPIE Vol. 2277, July 1994

[40] Y. Yacoob, and H. Lam, and L. S. Davis, "Recognizing faces showing expressions", To appear in international workshop on automatic face and gesture recognition, Zurich, 1995

[41] S. A. Boothroyd, A. David, S. Chang, P. Palacharla, and J. Chrostowski, "Nonlinear joint transform correlator with multiple quantum well photorefractive device", OC'96, Proceedings of the International Optical Meeting on Optical Computing, Sendai, Japan, April 21, 1996

[42] R. C. Gonzales, and P. Wintz, *Digital image processing*, Addison-Wesley Publishing Company, Second Edition, 1987.

[43] R. Mukundan, and K. R. Ramakrishnan, "Fast computation of Legendre and Zernike moments", Pattern Recognition, Vol. 28, No. 9, pp. 1433-1442, 1995

[44] A. Khotanzad, and Y. H. Hong, "Rotation invariant image recognition using feature selected via a systematic method", Pattern Recognition, Vol. 23, No. 10, pp. 1089-1101, 1990

[45] M. Stricker, and M. Orengo, "Similarity of color images", SPIE Vol. 2420, pp. 381-392, 1995

[46] H. Hsin, and C Li, "An experiment on texture segmentation using modulated wavelets", Submitted to IEEE Transactions on Systems, Man, and Cybernetics, April 1994 (Based on 1993 IEEE International Conference On Systems, Man, and Cybernetics, Le Touquet, France)

[47] C. S. Burrus, and R. A. Gopinath, *Introduction to wavelets and wavelet transforms*, Rice University, Houston, April 26, 1993.

[48] S. G. Mallat, pp. 313 to 327 of [4]

[49] NRC-Institute for Information Technology, Video Tapes Archive, Jan. 1995, Talk by Dr. M. Lynos