REPRESENTATION OF EDGE-LIKE INFORMATION IN IMAGES WITH APPLICATION FOR FACE RECOGNITION

A THESIS

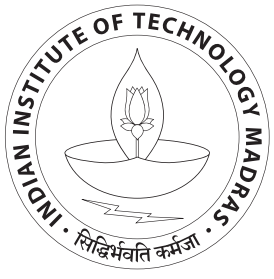
submitted by

ANIL KUMAR SAO

for the award of the degree

of

DOCTOR OF PHILOSOPHY



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

NOVEMBER 2009

To My Parents

Smt. Uma Sao *and* **Sri. G. R. Sao**

Acknowledgments

This thesis is the result of around five precious years of my life whereby I have been accompanied and supported by many people. It is a pleasant opportunity to express my gratitude to all of them. First, I would like to express my deepest respect and sincere gratitude to my guide Prof. B. Yegnanarayana for teaching me both consciously and unconsciously. His devotion and dedication has really motivated me to accomplish my PhD which was a mere dream for me. I am indebted to him for all his patience in rectifying my silly mistakes and making my work experience productive.

I owe debt of gratitude to Dr. C. Chandra Sekhar, my co-guide, for his immense support and encouragement throughout my research work. I am extremely thankful to him for giving his precious time to go through my thesis many times. I also thank him for all the invaluable advice on both personal and professional matters.

I am obliged to Dr. Hema A. Murthy for encouraging me time to time with her very active and dynamic personality. My special thanks to Prof. S. Raman and Prof. Timothy A. Gonsalves, chairpersons of the department during the course of my research work, for providing constant support and excellent facilities to carryout my research. My humble thanks to the members of my doctoral committee, Dr. S. Srinivasan, Dr. S. Mohan, Dr. B. Ravindran, and Dr. P. Sreenivasa Kumar, for sparing their valuable time to evaluate the progress and to give positive feedback about my research work.

I could not have managed at any step without the help and sincere efforts of Dr. Suryakanth, during my stay in IIIT Hyderabad. I am thankful to him for his helping attitude. Sincere thanks to Prof. Rajiv Sanghal, Director, IIIT Hyderabad, for allowing me to stay and pursue my research work at IIIT for the past one year. I also thank Dr. S. R. M. Prasanna, who actually motivated me to pursue research as my career option.

I am fortunate to be blessed with a group of wonderful friends Anand, Dhananjay, Guru and Murty, whose company has been one of the most memorable experiences for me. I thank all of them for updating my knowledge with their expertise in their respective fields. My special thanks to Anand, the computer geek, for his commendable help. Dhanu bhai deserves special appreciation for his composed nature and invaluable suggestions and discussions on every matter. I am grateful to him for his ready to help attitude. I cannot forget Guru for his knowledge in all fields, his jokes (mostly PJs) and fun that make the environment lively. I am also thankful to him for helping me in correcting my papers and thesis. I appreciate Murty's effort and cooperation to clarify my concepts on my research topic and once again I thank him for all his support. Special thanks to Dileep for his selfless help and support to make my stay easier in IIT Madras.

I thank long time friends outside my lab Manish (Jhar bhai), Parikshit, Vinod, Vivek Tiwari, and Vishwam for making my life fulfilled. I also thank basketball team of Microsoft Hyderabad who made me a part of their team, playing with them was always a stress-buster for me and kept me recharged!!! I thank all my old and present lab-mates for the friendly and conducive atmosphere in SVL.

It is time to thank my beautiful wife Reshma, without her support and encouragement this research work would not have been possible. Her good understanding and strong personality made my life easier.

I owe my deepest gratitude to my parents and brothers, Sunil Sao and Gajendra Sao for their affection and encouragement without which I would never have come this far in life. Though this comes at the end it is the most important; I thank the Almighty God who has blessed and watched over me and my loved ones and helped me throughout.

Anil Kumar Sao

ABSTRACT

Keywords: *face recognition, edge, smearing of edges, impulse, 1-D processing of images, zero-frequency filter, Laplacian operation, zero-crossing, analytic image, Fourier transform.*

In this thesis we consider four representations of an image to capture the edge information of the image, and examine their utility in the context of face recognition under variations of pose and illumination. These representations are edginess image (first derivative of an image), second derivative/Laplacian of a smoothed image (filtered image), phase of the analytic image and phase of the Fourier transform of the image. The two main issues that need to be addressed in using edge-based representations are : 1) Edge extraction, i.e., discrimination of spurious and significant edges, and 2) *locality problem*, i.e., slight shift in edges of similar face images leading to poor matching. The proposed four representations do not involve the task of edge extraction directly, but at the same time they preserve the edge information for matching purpose. The locality issue is addressed by smearing the edge information to help improve the matching of face images.

There is a trade-off between noise suppression and the smearing of edges in the computation of the edge information using derivative operation, since it is generally assumed that edge and noise of an image correspond to the high frequency components in the spectrum. On the contrary, processing of signals using zero-frequency filter illustrates that the edge has characteristics of an impulse whose energy is distributed uniformly throughout the spectrum, including around the zero-frequency. This property is exploited in the proposed two methods of realizing the zero-frequency filter for images to capture the edge information in the filtered image representation. The edge information present in the Fourier phase and in the phase of the analytic image are exploited using functions of phase rather than the phase directly. These representations avoid the problem of phase unwrapping. Performance of face recognition is further improved by exploiting the complimentary information present in both the Fourier phase and the analytic phase using Borda count based approach. The performance of the four representations is evaluated using two standard databases covering pose and illumination variations. The representation based on Laplacian of smoothed images performs better than the other representations. The representation derived using the

zero-frequency filtered signal seems to perform better than all other representations under illumination variation. It shows that the edges do not confine to only high frequency components, and that the edge information for face recognition is not only important but also depends how it is computed. The results of this thesis demonstrate that representation, especially the edge information of an image, is crucial in applications such as face recognition using template matching. The major contributions of the thesis are:

- A method for combining evidences using autoassociative neural network.
- Laplacian of a smoothed image for representing edge information.
- Zero-frequency filter for deriving the smoothed image.
- Functions of analytic phase and Fourier phase for representing the edge information of images.
- A method to exploit the complimentary information present in the Fourier phase and analytic phase using Borda count based approach.

Contents

List of Tables

List of Figures

Abbreviations

1-D	- One-dimensional
2-D	- Two-dimensional
DFT	- Discrete Fourier transform
IDFT	- Inverse discrete Fourier transform
PZC	- Positive zero-crossing
NZC	- Negative zero-crossing
FT	- Fourier transform
AANN	- Autoassociative neural network
p.v	- Principal value
MLP	- Multilayer perceptron
PSR	- Peak-to-sidelobe ratio
IIR	- Infinite impulse response
FIR	- Finite impulse response
PCA	- Principal component analysis
ICA	- Independent component analysis
RC-ICA	- Row-column ICA
LDA	- Linear discriminant analysis
HMM	- Hidden Markov model

Chapter 1

Introduction

In many image processing applications, it is essential to represent the information in an image in a form suitable for that application. Since the values of pixels acquired from sensors are affected by lighting conditions and noise, the relevant information needs to be extracted from the image to derive the suitable representation. Deriving a suitable representation from an image is governed by three factors, namely, application, perception and processing. It is difficult to derive a representation which is applicable for all the applications. The representation should contain perceptually significant information of the image, which is specific to the given application. Processing of an image should not add artifacts or enhance noise in the resulting representation. In the context of face recognition (an application where a face image is used to recognize a person), it is difficult to define or articulate the perceptually significant information in a given face image. The information that makes an individual face unique must be found in the subtle variations of the facial features (e.g., eyes, nose and mouth), since all the faces share the same set of features arranged in roughly similar configuration [1].

1.1 Psychophysics issues relevant for representing face

It has been estimated that a normal person knows about 700 people personally, and thousands in general [2]. Humans are also capable of identifying a large number of strange faces with relatively brief exposures. Typically, one sees thousands of faces over one's lifetime. Identi-

fication of known faces or faces of famous people occurs almost instantaneously. However, for both familiar and unfamiliar faces with unusual orientations (example, upside-down or partial exposures) the time taken for identification is usually longer. Face identification by humans is also robust in the sense that humans identify faces under a wide range of transformations.

It has been studied that for every known face there exists a unit in the brain that registers its familiarity under all conditions [3]. More familiar faces need less evidence for identification compared to the less familiar ones [4]. Consequently, familiar faces are less likely to be falsely recognized. Bartlett postulated that all new faces have the same (zero) level of familiarity, and new typical faces produce a stronger impression of familiarity than unusual faces [4]. Increase in the degree of familiarity after a single exposure is higher for unusual faces than typical faces.

Despite the ability of humans to identify faces rather effortlessly, it is difficult to describe a human face. Neuroscientists performed research on human recognition of faces and found many reasons for correct recognition of faces by humans. From psychophysics point of view, the face recognition system in humans utilizes a broad spectrum of stimuli obtained from many, if not all, of the senses (visual, auditory, olfactory (related to the sense of smell), tactile (related to the sense of touch), etc) [5]. These stimuli are used in either an individual or collective manner for both storing and retrieval of face images for the purpose of recognition. There are many instances where contextual knowledge is also applied, for example, recognizing faces in relation to where they are supposed to be located.

1.1.1 Face processing by humans

Evidence for the existence of a dedicated face processing system in human brain comes from three sources:

(a) Faces are more easily remembered by humans than other objects when presented in an upright orientation [6].

(b) Prosopagnosia patients are unable to recognize previously familiar faces, but usually have no other profound agnosia [7]. They recognize people by their voices, hair color,

dresses, etc. Although they can perceive eyes, nose, mouth, hair, etc., they are unable to put these features together for the purpose of identification. It was found that prosopagnosia patients can recognize whether the given object is a face or not, but they have difficulty in identifying the face [7].

(c) It is argued that infants come into the world pre-wired to be attracted by faces. Neonates (a child less than a month old) seem to prefer to look at moving stimuli that have face-like patterns in comparison to those containing no pattern or with jumbled features [8].

1.1.2 Distinctiveness

Studies show that distinctive faces are better retained in memory, and are recognized better and faster than typical faces [4]. For example, a person having a scar on his face can be easily identified than a person having no such distinctive feature. However, if a decision has to be made as to whether an object is a face or not, it takes longer time to recognize an atypical face than a typical face.

1.1.3 Role of race and gender

Humans recognize the faces of person from their own race better than the faces of person from another race [9]. This may be due to the fact that humans may be coding an average face with average attributes, the characteristics of which may be different for different races, making the recognition of faces from a different race harder. The gender of a person enables quick recognition of a face than recognizing a person from a photograph without any gender information [10].

1.2 Perceptual significance of edge information of face image

Psychophysics studies suggest that human visual perception distils and encodes key features or “formless invariants”, regardless of the point of view or style of representation of the

given face image [11]. Interestingly, these key features are present in a caricature or a sketch of a given face image drawn by an artist. Representations of a face image using edge map, sketch, and caricature are discussed below to illustrate the significance of unique information/key features present in the face image.

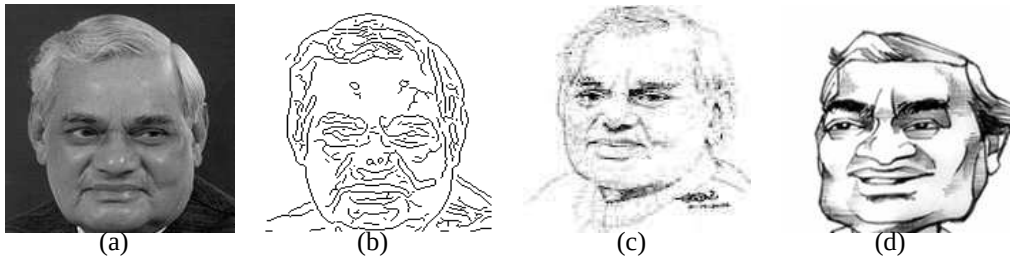
Edge map: The edge of an image is defined as the discontinuity in the image brightness [12]. These discontinuities correspond to discontinuity in depth, discontinuity in surface and variations in illumination. There are several algorithms available in the literature to compute the edges of an image [12–14]. Fig. 1.1(b) shows the edge map of the face image (shown in Fig. 1.1(a)), computed using Canny operator [13].

Sketch: A sketch consists of lines, and perhaps it is the simplest form of drawing. It is also known as line drawing. The sketch of a face shows major contours, and contrast edges plus finer details such as lines and wrinkles of the face [15]. Fig. 1.1(c) shows the sketch drawn by an artist for the person whose face image is shown in Fig. 1.1(a). One can observe from Figs. 1.1(b) and (c) that the information of the sketch is present in the edge map. However all edges are not present in the sketch. Approaches available in the literature to compute the sketch of a face image will be discussed in the next chapter.

Caricature: Caricature is a graphical coding of facial features that seeks, paradoxically, to be more like a face than the face itself [16]. It is a transformation which amplifies perceptually significant information, while reducing less relevant details. The caricature drawn by an artist of the same person (Fig. 1.1(a)) is shown in Fig. 1.1(d). Caricature is a sketch with exaggeration of some significant information.

The images of the same person acquired at different pose, scale, and age are also shown in Fig. 1.2. The following observations can be made from the above mentioned three representations:

1. The exaggerated features of the caricature (such as nose and cheek shown in Fig. 1.1(d)) are independent of pose and scale of the face image. Hence, it is easy to recognize the person using his caricature as compared to the sketch and edge map of the face



(a) (b) (c) (d)

Fig. 1.1: (a) Gray level face image. (b) Edge map of the image (a). (c) Sketch and (d) caricature drawn by an artist for the person shown in (a).

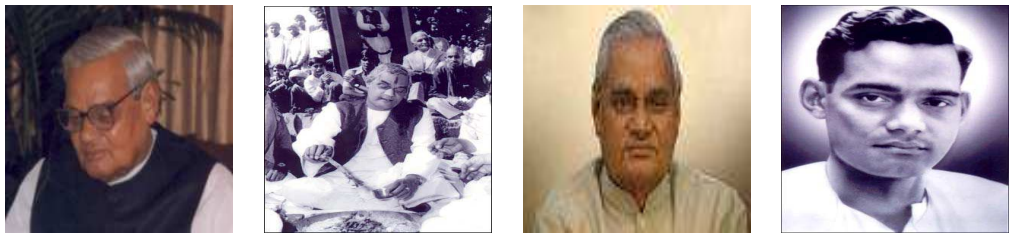


Fig. 1.2: Images of the person in Fig. 1.1 (a) with different pose, scale and age.

image. The exaggerated information is so unique that it helps even in recognizing the person's face image taken at a different age (extreme right in Fig. 1.2).

2. Sketch also contains crucial information about the face image, that is useful to recognize the person. The crucial information of the sketch corresponds to the physical edges (edges perceived by human) that are portrayed by an artist. In this process, an artist may leave some edges of the face image to highlight the uniqueness present in the face image. The crucial information perceived by different artists could be different, but all of them end up in drawing similar sketches of a given face image.

3. Edges are computed as changes in brightness of the face image followed by thresholding. Some spurious edges that do not correspond to physical edges may also appear in the edge map. One way to remove those spurious edges is to choose a high threshold value in the computation of the edge map. However, some physical edges of a face image may also obliterate from the edge map.

4. From the perception point of view, sketch and caricature are more robust representations as compared to the edge map representation of a face image. But it is not easy to develop an algorithm to compute those representations. This is because one needs to define the relevant crucial information of a face image that is difficult to articulate, and also the information may be specific to a given face image.

5. All information of the sketch is present in the edge map of the face image. But the reverse is not true. The reason could be that the definition of edge of an image is different from the edge perceived by human beings.

The edge map has all information of a sketch along with some less significant information. Removing the less significant information (by choosing a high threshold value) may lead to loss of crucial information also. So, there is a trade-off in choosing the threshold value in the edge map based representation of a face image. In this thesis we argue in favor of using a representation that captures the edge information implicitly rather than a representation that requires edge extraction (discrimination between spurious and significant edges). Another issue that needs to be addressed is the *locality problem*, i.e., slight shift in edges of similar face images leads to poor matching.

In this work we discuss four types of representations (listed in Table 1.1) that preserve the edge features of a face image, and compare their performance for face recognition. The first two representations are based on the derivative operation on the face image. An edginess-based representation, obtained using the first derivative of an image, is studied first. It is computed using 1-D processing of images, unlike the traditional 2-D methods of processing. The second is a new representation that associates the edges with the zero-crossings of the second derivative or Laplacian of an image. A smoothing operation is required before the derivative operation is applied, since the Laplacian operation enhances noise in the image. Two smoothing functions are explored, namely, the zero-frequency filter and the standard Gaussian filter. Smoothing using the zero-frequency resonator has an interesting interpretation, in the sense that it preserves the impulse-like characteristics (i.e., edges) of images in the resulting output [17]. The third representation is based on the phase of the analytic image, which also exploits the impulse-like characteristics of edges of an image. It is

well known that the phase of the Fourier transform of an image has perceptually significant information of the image [18], and hence can also be used as a representation capturing the edge information.

Table 1.1: Four types of representations capturing the edge information of an image.

(a)	Edginess-based representation of an image
(b)	Laplacian of a smoothed image
(c)	Phase of the analytic image
(d)	Phase of the Fourier transform of an image

1.3 Objective and scope of the work

The objective of this work is to study the importance of representation of an image in the context of face recognition. The focus is on representation that captures the edge-like information of an image. The effectiveness of representations is examined for face recognition using template matching. It is assumed that all the face images are already cropped and aligned with respect to the location of the eyes. All the face images are rescaled to the same size. The effectiveness of different representations in dealing with variations due to pose and illumination for face recognition is studied in this work.

1.4 Organization of the thesis

The contents of the thesis are organized as follows:

Chapter 2 gives an overview of the existing representations of a face image as well as the approaches for face recognition.

Chapter 3 discusses the edginess-based representation of a face image. It is computed using one-dimensional (1-D) processing of images, which gives multiple partial evidences of the face image. The potential field is derived from the edginess map of a face image to address the locality problem in matching. A novel method is proposed to combine the multiple evidences from different reference face images to address the issues of pose and

illumination in face recognition. An autoassociative neural network (AANN) model based method is proposed for decision making in face recognition.

Chapter 4 discusses a new method of processing an image using zero-frequency resonator. The zero-frequency resonator for two-dimensional signals like images is realized using two methods, namely, spatial domain and Fourier domain. A method based on the zero-frequency resonator is proposed to compute the edge map of a given image. The method is similar to the standard Laplacian of Gaussian (LOG) operation, in the sense that the former approach uses the zero-frequency filter for smoothing, whereas the latter uses Gaussian filter for smoothing. The Laplacian operation is the same in both the cases. We propose a representation based on the Laplacian of the smoothed image (filtered image) that preserves the edge information of a face image around the zero-crossings of the filtered image. The locality problem of the filtered image is addressed by considering only the first few eigenvectors for matching. A method is proposed to highlight the subject-specific unique information present in the eigenvectors derived from the filtered images.

Chapter 5 proposes a representation of face image using analytic image, and discusses issues involved in the definition of an analytic image. This representation exploits the edge information present in the phase of the analytic image. Here also the eigenanalysis based method is employed for matching two face images.

Chapter 6 discusses a representation of a face image using the phase of the Fourier transform. The importance of small spacing or large spacing edges is controlled by selecting the appropriate discrete Fourier transform (DFT) coefficients. A method is proposed to utilize the complimentary information present in the phase of analytic image and Fourier phase. This chapter also compares the proposed four edge-based representations with existing approaches for face recognition.

Chapter 7 gives a summary of the present work by listing the major contributions of the present work, and also suggests some directions for further research in this area.

Chapter 2

Review of representations for face recognition

The problem of automatic face recognition is a composite task that involves the detection of faces from a cluttered background, representation of faces (feature extraction) and matching. A complete face recognition system should address all the above subtasks, and each one of them presents several research issues. This research work focuses on the problem of representation of face images. Different representations of face images proposed in the literature are discussed in Section 2.1. Section 2.2 explains methods used for face recognition. Details of the face databases used for experiments in this research work are given in Section 2.3. Summary of the chapter is given in Section 2.4.

2.1 Representation of face images

The choice of representation significantly affects the performance of a face recognition system. Two types of representations used in the face recognition research are: 2-D array of intensity values and feature vectors. The drawback of representation using 2-D array of intensity values is that it is sensitive to illumination. Also, if the face image is captured using a different sensor, the performance of the face recognition system degrades.

The most widely used representation of face images is based on feature vectors. Two approaches are commonly employed to extract the features: (a) Feature-based approach

and (b) holistic approach. In the former approach the features are derived from the relative positions and measurements of the facial parts. On the other hand, the holistic approach considers the face image as a single unit while deriving the representation.

Features derived using the different facial parts differ depending upon the pose of a face image. For example, size of the eyes, and distances between eyes or lips are clearly visible in the frontal pose. For a face image with side view, a set of characteristic points are employed such as the notch between the brow and nose, the tip of the nose, the notch between the nose and the upper lip, etc.. The features are usually the distances and angles between those characteristic points (facial parts). Table 2.1 lists some commonly used features. In the late nineteenth century Galton devised a system in which portraits were described by numerical formulae [19]. Each formula consists of four groups of figures with five figures in each group. The shapes of the profile at characteristic points were expressed as single numerals. They are determined from a table in which different shapes are represented with an index. Not more than 10 varieties were given for each feature. Fig. 2.1 shows some possible shapes for the ridge of the nose. Each of them was assigned a particular number (0-6).

Though the features derived from the relative positions and measurements between facial parts give a compact representation of a face image, the representation has certain constraints. It depends heavily on the techniques for facial feature detection. It requires precise locations of the facial features, which are difficult to obtain. It also suffers severely from partial face occlusions and image degradations. These problems can be avoided in a rep-

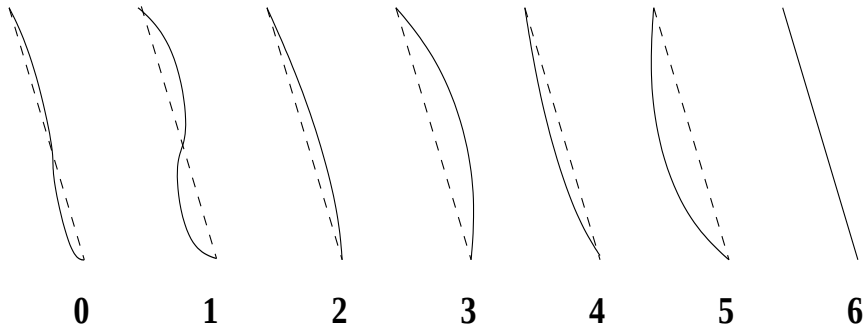


Fig. 2.1: Some possible shapes used by Galton for ridge of the nose.

Table 2.1: Features used for identifying faces (obtained from gray level values) [2]. Here $D(x,y)$ represents the distance between the facial features x and y.

Category	Features
Overall	height, outline, template, area, centroid
Eyes	template, shape, color/shade, area, opening, intensity around pupil, *D(upper eyelid, eyelid-fold)*, difference of ordinates of inner and outer corners, inter-eye distance (center), inter-eye distance (outer corner)
Eyelids	area, length of the bounding curves
Eyebrows	thickness, *D(eye, brow)*
Lips	thickness, shape, template, width
Mouth	template, width, height, area, area of opening
Nose	width, length, template, shape, area of nostrils
Ears	length, shape, area
Hair	intensity/shade/color, amount/coverage, length, texture
Cheeks	intensity
Distances	width of head on the eye-line, *D(chin, eye-line)*, *D(center of lips, chin)*, *D(eye, hairline)*, *D(eye, center of nose)*, *D(tip of chin, center of face)*, *D(left edge, center of nose)*, *D(right edge, center of nose)*, *D(center of face, center of brow)*, *D(center of mouth, end of nose)*
Areas	inner corners of eyes and center of mouth, center of eyelid fold, center of eye and center of face, center of mouth and center of brows
Ratios	*D(center of face, opening of mouth)*/*D(chin, opening of mouth)*, *D(chin, opening of mouth)*/*D(end of nose, opening of mouth)*
Others	chin outline, width and height of philtrum, forehead size, forehead outline

resentation derived by considering the face as a complete unit. This concept is followed in all the subspace based methods, which will be discussed in Section 2.2.5. This approach may give equal importance to all parts of the face image. In other words, contribution of the subject-specific unique information, which may reflect in some specific part of the face image, will not get emphasized during matching. Thus, each approach has its own advantages and disadvantages. Ideally, one should use a combination of these two approaches for representing a face image. These representations include sketch, caricature and edge map, which will be explained below.

2.1.1 Sketch

It is fascinating to see how an artist draws a sketch of a person's face. Sketches are the simplest form of drawings because they consist of mostly lines. The artist distils the unique characteristics of a face, and highlights them with a small number of critical strokes [15].

The use of a sketch representation for face recognition gained attention because of two reasons: (1) Automatic retrieval of suspect's photo from the photo database, and (b) human beings recognize sketches as accurately as gray level face images [15]. Here, photo means the face image with gray level values. Unfortunately, the photo of a suspect is not available in most cases. To deal with such a problem, a simulated sketch is generated with the help of artists and eyewitnesses. Thus, automatically searching through a photo database using a sketch is a very useful application. It will not only help the police to locate a group of potential suspects, but also help the witnesses and the artists to modify drawing of the sketch of the suspect interactively, based on the similar photos retrieved. In addition, a simple sketch can reveal interesting characteristics of a person.

During the past three decades, many face recognition systems have been developed. Nonetheless, very few effective face sketch recognition systems can be found in the literature. In [20, 21] traditional methods of face recognition such as the eigenface and elastic graph matching have been applied on sketches drawn by artists. However, in practice, creating a database of sketches is not easy. So the focus of the problem has shifted towards finding similarities between a photo and a sketch. Sketches and photos are of different modalities, and it is hard to find their similarities. This issue was addressed by shrinking the difference between those two representations and then performing matching. In [20] the objective was achieved by normalizing the sketch in geometry and by blurring using a Gaussian filter, and then recognizing using the eigenface method. This method does not perform well because of large differences between sketches and photos. Thus, the task of matching a photo and a sketch is attempted by transforming those two representations into similar patterns. This is done by transforming a photo to a sketch, or a sketch to a photo, as discussed below:

1. **Photo to sketch transformation:** The objective of these approaches is to derive a sketch from the available photos, and then perform matching with the sketch drawn

by the artist. These approaches can be divided into two categories, namely, feature-based approach and exemplar-based approach.

The feature-based approaches are given in [22, 23]. In [22], the facial feature points such as eyes, nose, and mouth are detected using generalized symmetry operator, rectangle filter and characteristic shapes. Line drawings connecting these feature points give the facial sketch image. The results were shown to be similar to the sketch drawn by an artist. However the representation was not applied for face recognition. The method in [23] proposes a grammatical model based approach for sketch generation. A face image is divided into six zones for detection of the six facial components: two eyes, two eyebrows, nose and mouth. Each component has a set of diverse representations, and their topological configurations such as open and close states. These components are modulated by active appearance model (principal component analysis). The presence of components is controlled by the grammatical rules [24] through hidden variables. The six component zones are further refined with sketch curves corresponding to the subtle differences in eye-lid, eye-shade, nostril and lips. This is performed using prior models and global context derived from the facial components using several face images. Here, the objective was to compute a sketch that can be used for low bit rate communication over wireless platforms. This representation was used to address the issue of expression in face recognition [25]. The performance was evaluated for the three expressions of face, namely, *smiling*, *angry* and *screaming*. The recognition accuracy was better in the case of *angry* as compared to the *smiling* and *screaming expressions*, since the distortion in the sketches is high in the case of those two expressions. A method recently proposed in [26] also uses the grammatical-model based approach for synthesis of a sketch from a given photo.

In exemplar-based approaches, the objective is to learn the mapping between a photo and a sketch through several photo-sketch pairs. In [27, 28], the mapping is assumed to be linear. A new face photo $\mathbf{y}_p$ and sketch $\mathbf{y}_s$ can be represented by a linear combination of training photo-sketch pair samples $\mathbf{X}_p$ and $\mathbf{X}_s$. Each column of $\mathbf{X}_p$ and $\mathbf{X}_s$ represents a training photo sample and the corresponding sketch sample, respectively. The linear relationships are $\mathbf{y}_p = \mathbf{X}_p\mathbf{c}_p$ and $\mathbf{y}_s = \mathbf{X}_s\mathbf{c}_s$, where the linear combination

coefficients $\mathbf{c}_p$ and $\mathbf{c}_s$ are column vectors obtained by principal component analysis. Based on the assumption of linearity, a sketch should have a similar linear reconstruction to its corresponding photo image, i.e., $\mathbf{c}_p \approx \mathbf{c}_s$, so the pseudo-sketch of the photo can be synthesized with $\mathbf{X}_s\mathbf{c}_p$. This method is called eigentransform-based pseudo-sketch synthesis. Three distances based on $\mathbf{c}_p$ and $\mathbf{c}_s$ were defined to recognize the given sketch. The performance was further improved by incorporating the shape and texture information of photo in the pseudo-sketch synthesis [29]. A Bayesian subspace classifier is employed to recognize the pseudo-sketches derived from the probe photo. The assumption of the linear relation of photo-sketch pair is relaxed in the approaches given in [30, 31]. In [30], a local linear embedding (LLE) based method is used for the generation of the pseudo-sketch. Nonlinear discriminant analysis is used to recognize the synthesized pseudo-sketches. This method performed better than the method given in [29]. Learning linear or nonlinear mapping between photo-sketch pair requires many examples. This issue is addressed by the method given in [31], which uses the embedded hidden Markov model (E-HMM) for sketch synthesis. The results are shown only for sketch synthesis, and not for recognition. Recently a method, based on multiscale Markov random field, is proposed to learn the mapping between a photo and a sketch [32]. In all the exemplar-based approaches, the drawing style of artist plays an important role in the synthesis of pseudo-sketch, since they require several photo-sketch pairs drawn by the artists.

2. **Sketch to photo transformation:** Compressing more information into smaller compact representation is a more stable operation than enlarging a compact representation to a full representation [29]. Since photos contain detailed information than sketches, it should be relatively easier to convert a photo to a sketch. As an extreme example, suppose that the sketch contains some simple outlines from the photo of a face. It will be difficult to reconstruct the photo from the line drawings only. The experimental results of the face recognition also give similar conclusion [28]. Nonetheless, synthesis of photo from sketch is attempted in [33, 34].

   A statistical inference method for transformation between a sketch and a photo is given

in [33]. The relationship between photo and sketch is established by using two feature spaces formed by patch-based tensor model. The statistical dependencies between two tensor models hold more precisely and flexibly by using Bayesian tensor interface. This method is compared with eigentransform [28] and method that preserves local geometry preserving [30] for facial photo hallucination, and the results are shown to be better. But this representation is not yet applied for the task of face recognition.

The synthesis of a photo from a sketch given in [34] is similar to the method given in [28]. The only difference is in the computation of the eigenvector matrix. In [34] a hybrid space is derived by appending sketch vector to the corresponding image vector for computation of the eigenvectors. On the other hand, in [28] the eigenvector matrices for image space and sketch space are computed separately. It has been argued that the computed eigenvectors for photo space and sketch space are highly correlated if they are computed using hybrid space rather than separate spaces [34]. The synthesized photos are used in conjunction with advance correlation filter [35] to address the issue of illumination in face recognition.

2.1.2 Caricature/cartoon

"It is not really the perception of likeness for which we are originally programmed, but noticing of unlikeness, the departure from the norm which standsout and sticks to the mind". [36]

These thoughts are kept in mind while drawing caricatures. A caricature is defined as a funny drawing of someone that makes some of his/her distinct features appear exaggerated for entertainment [16]. However, caricature of the same person drawn by different artists can be different, since the drawing style plays an important role. Yet all the caricatures exaggerate similar information, which helps us to recognize the person's face image. Psychological studies also show that caricatured faces yield faster and more accurate recognition than using normal face images [37].

It is not easy to draw the caricature of a given face image using a computer/machine. The reason is that it is difficult to articulate the way an artist portrays a person. There are a few attempts toward this goal. The first attempt to develop computer-assisted caricature

generation system was made by Brennan [16]. The objective of this method was to teach a newcomer about what was going on in an artist's mind.

Available approaches to derive the caricature from a given photo follow two steps: Extraction of the features and exaggeration of those features. A set of facial contour points and manually marked curves are employed in the methods given in [38, 39]. In [40, 41] the features are extracted automatically, and these features are based on valley and edges of the given face image. Koshimizu et al. [42] have used the line sketch drawn by PICASSO system as the initial representation of a face image in the generation of caricature.

For the exaggeration of the extracted features, two approaches are followed in the literature. In the first approach, exaggeration is learned using photos of several persons and the corresponding caricatures drawn by an artist. The learning is performed using neural networks [43], and the partial least-squares based method [41]. These approaches give satisfactory results from the perception point of view, but are sensitive to the drawing style of the artist. In the second approach, the positions of the extracted feature points are compared to the positions of same points in a normed face image. The differences are then exaggerated by a fixed amount to derive the caricature [38–40, 42].

A completely different method is followed in [44] for the caricature generation. Shape features extracted from a face image are expressed in linguistic terms. Fuzziness inherent in the linguistic expressions is expressed by a fuzzy set. A caricature is drawn by varying parameter values of each feature of a face image. This method is interactive, and involves human interaction.

The approaches explained above give a 2-D caricature. The approach is extended to obtain a 3-D caricature which gives more degree of flexibility in mimicking the artistic flavors [45]. Here several 2-D caricatures are derived separately for each view of a given person's face image. The resultant caricatures are combined to derive the 3-D caricature. All the above mentioned approaches focus mainly on the caricature generation. None of the approaches have been applied for face recognition task.

2.1.3 Edge map

The edge map of a face image has many similarities with its sketch and caricature representations. Takacs [46] made use of edge maps to measure the similarity of face images. The faces were encoded into binary edge maps computed using Sobel edge detection algorithm [12]. The modified Hausdorff distance was chosen to measure the similarity of the two point sets, i.e., the edge maps of two faces, because Hausdorff distance can be calculated without an explicit pairing of points in their respective data sets. Recognition accuracy of 92% was achieved in their experiments.

The binary edge map representation was modified in the line edge map (LEM) by grouping pixels of face edge map to line segments [47]. In this approach, a polygonal line fitting process is applied after thinning the edge map. An example of a human frontal face LEM is illustrated in Fig. 2.2. A line segment Hausdorff distance (LHD) measure is used to match the LEMs of face images. Experimental results show that the performance is better than the eigenface and normal edge map matching. These approaches require thresholding to decide whether a given pixel is an edge pixel or a non-edge pixel. The presence of spurious edges or missing of significant edges may lead to poor performance [47]. Another problem with the edge map is that any shift in the edge locations due to small changes in the face image reduces the matching score significantly. This problem is also known as *locality problem*. The matching is improved by using a spread edge profile, where the edge image is smeared using a membrane function [48]. The spread edge profile composes a ghostly face called *hill*. The hills have high values at the edges, and the values decrease as we move away from the edges. The problem of hill representation is that it creates an artificial edginess map of the face.

Recently a representation, namely, Gradientfaces [49], which uses the edge orientation of an image, is proposed to capture the edge information of face images. This method reported recognition performances of 99.83% and 98.96% on illumination variation set of PIE face database [50] and Yale-B face database [51], respectively. But the method does not address the issue of phase unwrapping.

Fig. 2.2: An illustration of a face line edge model (LEM).

2.1.4 Range images

The discussion so far considered representations of a face that use data obtained from 2-D intensity images. Another topic being studied by researchers is face recognition from range image data. A range image contains the depth structure of the object under question. Although such data is not available in most applications, it is important to determine the value added by the information present in range images in terms of its effect on the accuracy of face recognition.

A template-based recognition system involving descriptors based on curvature calculations made on range image is described in [52]. The data is obtained from a rotating laser scanner system with resolution better than 0.4 mm. Surfaces are classified into planar, spherical, and surfaces of revolution. The data is stored in cylindrical coordinate system as $f(r, \theta, y)$. At each point on the surface the magnitude and direction of the minimum and maximum normal curvatures are calculated. Since the calculations involve second-order derivatives, smoothing is required to remove the effects of noise in the image. Smoothing is performed using a Gaussian filter. Surface regions are classified as convex, concave and saddle. Ridges and valley lines are determined by obtaining the maxima and minima of the curvatures. The strategy used for face recognition is as follows:

1. The nose is located.

2. Locating the nose facilitates the search for eyes and mouth.

3. Other features such as forehead, neck, and cheeks are determined by their surface smoothness (unlike hair and eye regions).

4. This information is then used for depth template comparison.

Using the locations of the eyes, nose and mouth, the faces are normalized into a standard position. This position is re-interpolated to regular cylindrical grid, and the volume of the space between the two normalized surfaces is used as a measure of mismatch. This system was tested on a dataset of 24 images of eight persons with three views of each. The data represented four male and four female faces. 97% recognition accuracy was reported for the individual features, and 100% for the whole process. An approach is described in [53] to extract the facial curves from range images. The facial curves are matched using geodesic distance for the task of recognition [54].

2.1.5 Infra-red images

In [55] an initial study comparing the effectiveness of visible and infra-red (IR) imagery for detecting and recognizing faces is described. One of the motivations for this work is that changes in illumination can cause significant degradation in performance for visible image based face recognition. Hence infrared imagery, which is insensitive to illumination variation, can serve as an alternative source of information for detection and recognition. However, the poor resolution of IR images is a drawback. Further, though the IR imagery is insensitive to changes in illumination, it is sensitive to changes in temperature. Three face recognition algorithms were applied to both visible and IR images. The recognition results on 101 subjects show that both visible and IR imagery give a similar performance, and that the fusion of IR and visible imagery is a viable means of enhancing the performance beyond that of the either alone. The similar observation was made in a recently published paper by S. Gundimada and V. K. Asari [56].

2.2 Approaches to face recognition

The earliest work related to face recognition was done by Galton [19] using face profiles. A set of five cardinal points (Fig. 2.3) was derived from the face profile, and features derived from it were used to compare faces. The cardinal points such as the notch between the brow and nose, the tip of the nose, the notch between the nose and the upper lip, parting of the lips and tip of the chin were used.

One of the earliest works in the last century was reported by Bledsoe [57]. In this system, the feature points of the face were located by a human operator, and the located positions were fed into a computer. Given a set of feature point distances of an unknown person, nearest neighbor or other classification rules were used for identifying the test image. Since feature extraction was done manually, this system could accommodate wide variations in pose, tilt, image quality, and contrast. Kelly worked in a framework similar to that of Bledsoe, but his method does not involve any human intervention [58]. He used the body and close up head images for recognition. Once the body and head have been outlined, ten measurements were made for use in recognition. Kanade used distance and angle between points such as eye corners, mouth extremities, nostrils, and chin top to discriminate faces [59]. The facial feature points were located in two stages. A coarse-grain stage is employed to simplify the succeeding differential operation and feature-finding algorithm. Once the eyes, nose and mouth were approximately located, more accurate information was extracted by confining the processing to four smaller regions, scanning at higher resolution and using the best beam intensity for the region. The four regions were the left eye, right eye, nose and mouth. The beam intensity was based on the local area histogram obtained in the coarse-grain stage. A set of 16 facial parameters which are ratios of distances, areas, and angles to compensate for the varying size of the pictures were extracted. To eliminate differences due to scale, the components of the resulting vector were normalized. A distance measure was used to check for similarity between an image of the test set and the image in the reference set. The range of matching accuracies varies from 45% to 75% depending on the parameters used.

Brunelli and Poggio computed a set of geometrical features such as width and length of nose, mouth position and chin shape [60]. A recognition accuracy of 90% was quoted on a

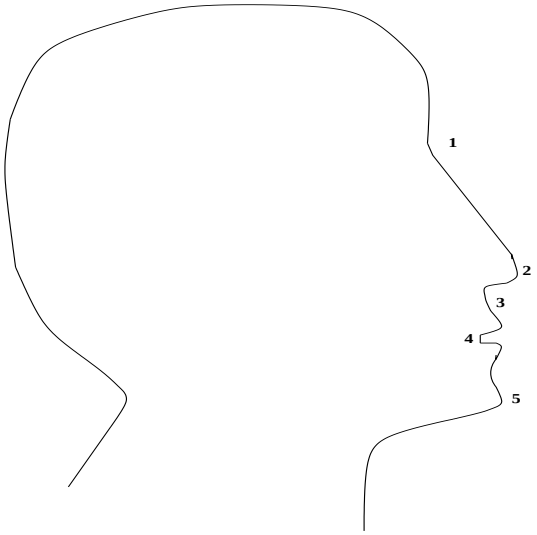


Fig. 2.3: Characteristic points used by Galton simplifies the succeeding differential operation and feature-finding algorithms.

database of 97 persons. However, simple template matching scheme gives 100% recognition for the same database.

In [61] Wong proposed a technique to measure distances between eyes, left eye to middle of the nose, right eye middle of the nose, middle of the nose to left edge of face, and middle of the nose to right edge of face. These measurements were used for recognizing human faces. It works well for a database with small number of face images.

Performance of the above mentioned feature-based matching methods decreases with partial face occlusions and image degradations. Moreover, these methods require reliable and precise detection of facial features. These drawbacks of the feature-based techniques led to the evolution of template-based methods. A simple version of template-based method of matching is to use 2-D intensity values for comparing face images using a suitable metric such as Euclidean distance or cross-correlation. This approach will be sensitive to even small variations in intensity value. Hence, it gives poor matching under the intra-class variation. This issue is addressed by building a complex model using several face images under different conditions. These approaches can be divided into five broad categories: Graph matching, neural network model, hidden Markov model (HMM), 3-D morphable model and subspace methods.

2.2.1 Graph matching

Graph matching based approaches construct a topological graph of each person's face image using the feature points that were computed using Gabor wavelet decomposition of facial image [62–64]. Typically 35-45 points were derived from the given face image. The identification process utilizes the information present in the topological graph representation of the feature points. After compensating for differing centroid locations, two cost values were evaluated, namely, topological cost and similarity cost. The recognized face was the one that has minimum of the combined cost values. The method works on controlled background images like passport and driver's license pictures. Lades et al. [65] presented a dynamic link architecture for distortion invariant object recognition, which employs elastic graph matching to find the closest stored graph. Objects are represented with sparse graphs whose vertices are labeled with a set of complex Gabor wavelet coefficients called a jet. Only the magnitudes of the coefficients were used for matching and recognition. When recognizing a face of a new image, each graph in the model gallery (database) was matched to the image separately, and the best match indicated the recognized person. The variations in pose were compensated by elastic deformation of the graphs. This method gives a good accuracy for a database of 87 persons, consisting of face images with different expressions and poses.

Wiskott et al. extended this system to handle larger galleries, and larger variations in pose, and also to increase the matching accuracy [66]. Firstly, the phase of the complex Gabor wavelet coefficients is used to achieve an accurate location of the nodes, and to disambiguate patterns which would be similar in the magnitudes of their coefficients. Secondly, object-adapted graphs are employed, so that nodes refer to specific facial landmarks, called fiducial points. The correspondence between two faces can then be found across large changes in viewpoint. Thirdly, a new data structure called the bunch graph was introduced, which serves as a generalized representation of faces by combining jets of a small set of individual faces. This allows the system to find the fiducial points in one matching process, which eliminates the need for matching each model graph individually. This reduces computational load significantly. It presented an accuracy of about 98% on FERET face database [67, 68]. The drawback in this approach is that it requires manual intervention to select the fiducial points in the facial image, and it requires precise locations of those points.

Though graph matching gives good results, it has the drawback of preprocessing to be done manually. It requires the exact position of the facial features to build the models. It also fails when there is a partial occlusion of faces, since finding facial positions in such cases is difficult.

2.2.2 Neural network based approach

Artificial neural network (ANN) [69–71] is a powerful tool for pattern recognition problems. Researchers in the field of face recognition used ANNs for the purpose of developing a model, which operates directly on an image representation of faces rather than geometrical codings of faces. Using A NN on faces [1] several problems have been addressed: gender classification, face recognition and classification of facial expressions. One of the earliest demonstrations of A NN for face recall applications is reported in Kohonen's work using associative map [72]. Using a small set of face images, accurate recall was reported even when the input image is noisy, or when portions of an image are missing.

A single layer adaptive NN (one for each person in the database) for face recognition, expression analysis and face verification was reported in [73]. It uses typically 200-400 presentations for training. The training patterns included translation and variation in facial expressions. Sixteen classifiers were used for the database containing face images of 16 persons. Classification was achieved by determining the classifier that gives the highest response for a given input image. Extensions to face verification and expression analysis were also presented.

A hierarchical neural network which grows automatically, and not trained with gradient descent, was used for face recognition [74] . Good results were reported for discrimination of ten distinctive subjects [74]. The ability of the compression networks was demonstrated by Cottrell and Fleming in [75, 76]. The network was trained with the whole image as input, and it was shown that the network can classify face images according to faceness, gender and identity. In [71] linear autoassociative networks, nonlinear autoassociative (or compression), and hetero-associative backpropagation networks are explored for face processing.

In [77] Lin et al. proposed a face recognition technique based on probabilistic decision based neural network (PDBNN). It adopts a hierarchical network structure with nonlinear

basis functions and a competitive credit assignment scheme. It demonstrated a successful application of PDBNN on FERET and ORL face databases [78].

In [79] Lawrence et al. presented a hybrid neural network solution which combines local image sampling, a self-organizing map (SOM) neural network and a convolutional neural network. The SOM provides quantization of the image samples into a topological space, where inputs that are nearby in the original space are nearby in the output space as well, thereby providing dimensionality reduction and invariance to minor changes in the image sample. The convolutional neural network provides for partial invariance to translation, rotation, scale and deformation. The recognizer provides a measure of confidence at its output, and the classification error approaches zero while rejecting as few as 10% of the examples when applied on a database of 400 images of 40 individuals, containing high degree of variability in expression and pose.

In [80] Srinivas et al. described an application of mixtures of experts for gender and ethnic classification of human faces and pose classification. The mixture consists of an ensemble of radial basis function (RBF) networks. Inductive decision trees (DTs) and support vector machines (SVMs) implement the gating network components for deciding which of the experts should be used to determine the classification output, and to restrict the support of the input space. Experiments show good results on gender, ethnic and pose classification, which can also be used effectively in face recognition.

Neural network based methods fail to perform well due to large pattern dimension and lack of sufficient training samples. Moreover, neural network methods need a lot of computational effort to build a face recognition system.

2.2.3 Hidden Markov model (HMM)

Stochastic modeling of nonstationary vector time series using HMM has been successful for processing speech signals [81], and has been applied for face recognition as well [82]. A face image is divided into regions such as the eyes, nose and mouth, etc. These regions can be associated with the states of a HMM. Since HMMs require an observation sequence, the images are converted into 1-D spatial sequence of pixels by concatenating the rows/columns.

In [82, 83] a spatial observation sequence is extracted from a face image by using a band

sampling technique. Here, each face image is represented by a 1-D vector series of pixels. Each observation vector is a block of L lines, and there is an overlap of M lines between successive observations. A test image is first converted to an observation sequence using band sampling technique. Then, it is matched against HMM of each of the faces in the database. The model with the highest likelihood is considered as the best match, and the corresponding face is the identity of the test face image. The recognition rate of the HMM-based approach is 87% on ORL face database consisting of 400 images of 40 individuals. The conversion of a face image into 1-D vector may lead to loss of 2-D structural information of the face image. This issue is addressed in the pseudo 2-D HMM [84]. In the HMM-based approach the classification time and training time are very high. Moreover, the choice of parameters of the HMM is quite specific to the subject.

2.2.4 3-D Morphable model

A morphable face model is based on a vector space representation of faces that is constructed such that any convex combination of shape and texture vectors of a set of examples describes a realistic human face [85]. After fitting a 3-D morphable model to the face image, recognition across different viewing conditions is performed in two ways. (a) Compare the model coefficients which represent the intrinsic shape and texture of faces, and are independent of the imaging conditions. (b) Three dimensional face reconstruction is employed to generate synthetic views from gallery probe images [51, 86, 87]. The synthetic views are then transfered to a second viewpoint-dependent recognition system.

The computer graphics simulation of projection and illumination are combined with 3-D morphable model to automatically estimate 3-D shape, texture, and all the relevant 3-D scene parameters including the head positions and orientation, focal length camera, and illumination directions [88]. A single model is built to address both the issues of pose and illumination. Illumination is not restricted to Lambertian reflection, but takes into account the specular reflections and cast shadows, which have considerable influence on the appearance of human skin. A recognition accuracy of 95% was obtained using PIE face database [89], based on side-view gallery. The 3-D morphable model based approaches may smear the subject-specific unique information of a face image or introduce some artifacts while build-

ing the model.

2.2.5 Subspace based methods

Bayes classifier gives the minimum error in the classification accuracy in pattern recognition provided the estimate of prior and conditional probability density functions (pdf) are accurate. As the dimensionality of the feature vector increases, the number of samples required for accurate estimation of the pdf also increases. This phenomenon is also known as curse of dimensionality [90]. For a face image with size of 30×30 pixels, the dimensionality of image space will be 900. For such large dimensions, it is difficult to get enough samples for accurate estimation of the pdf.

However, much of the surface of a face is smooth and has a regular texture. Hence the pixel values will be highly correlated. Moreover, the appearance of a face is highly constrained. For example, any frontal face is roughly symmetrical. Thus, the natural constraints suggest that the dimensionality of the image space can be reduced. There are several methods available in the literature to compute the subspace or reduced dimensional space, and these methods can be categorized into two types: linear and nonlinear methods.

2.2.5.1 Linear subspace methods

These methods assume linearity while computing the subspace. The methods include principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA). The principal component analysis is also known as Karhunen-Loeve expansion or eigenanalysis [90]. It is a well known method for dimension reduction in pattern recognition literature, and is one of the extensively studied methods for face recognition.

Eigenanalysis exploits the second order dependencies in the multivariate stochastic observation to obtain a compact representation. Let $\mathbf{x} \in \mathbb{R}^N$ be a vector representation of a face image, where N is the dimension of the image space. The vector is formed by concatenating the row/column pixels of a face image. The covariance matrix ($\Sigma_{\mathbf{x}} \in \mathbb{R}^{N \times N}$) of $\mathbf{x}$ is defined as follows:

$$\Sigma_{\mathbf{x}} = \mathrm{E}\{[\mathbf{x} - \mathrm{E}\{\mathbf{x}\}][\mathbf{x} - \mathrm{E}\{\mathbf{x}\}]^t\}, \tag{2.1}$$

where E{.} and t denote the expectation and transpose operators, respectively. In eigenanalysis, the covariance matrix $\Sigma_{\mathbf{x}}$ is factorized as follows:

$$\Sigma_{\mathbf{x}} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^t, \tag{2.2}$$

where $\mathbf{\Psi} \in \mathbb{R}^{N\times N} = [\psi_1, \ldots, \psi_N]$ is an orthonormal eigenvector matrix, and $\mathbf{\Lambda} \in \mathbb{R}^{N\times N} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ is a diagonal eigenvalue matrix with diagonal elements (eigenvalues) in nonincreasing order ($\lambda_1 \geq \lambda_2 \ldots \geq \lambda_N$).

An important characteristic of the PCA is that it is a dimension reduction method, optimal in the sense of minimum square error (MSE), where only a subset of eigenvectors $[\{\psi_1, \psi_2, \ldots, \psi_m\}, m \leq N]$ are used to represent the original image [91]. The low dimensional representation of the original vector is given by $\{a_1, \ldots, a_m\}$, where $a_i = \mathbf{x}^t\psi_i$. An approximation to the original face image can be reconstructed using the subset of eigenvectors as follows:

$$\hat{\mathbf{x}} = \sum_{i=1}^{m} a_i\psi_i. \tag{2.3}$$

The eigenvectors associated with the largest eigenvalues are referred as eigenfaces. In 1990, Kirby and Sivovich proposed the use of PCA for face analysis and representation [92]. This idea was extended by Turk and Pentaland [93], which was the first application of PCA for face recognition. Each face image was projected onto the principal subspace defined by the m principal eigenvectors, giving a m-dimensional representation of the face image. When a test face image is projected onto the subspace, the Euclidean distance between its coefficient vector and the vector representing each person in the database is computed. Depending on the distances to the subjects, the test image is classified as belonging to one of the persons. The basic method ignored the variation modes within the subspace and outside it, which was incorporated later using Bayesian similarity measure [94, 95]. The assumption of equal covariance matrices of all the classes is eliminated in the principal component null space analysis (PCNSA) [96].

In [97], a two-dimensional principal component analysis (2-D PCA) is proposed for face recognition. In contrast to the conventional PCA, the 2-D PCA is based on 2-D matrices

rather than vectors. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix is constructed directly using the original image matrices in (2.1). Subsequently the method is similar to PCA. When a new face image matrix is projected to the eigenvector matrix, we get a feature matrix or feature image instead of feature vector as in PCA. The feature matrix of a test face image is compared with the feature matrices of the reference face images to determine the identity. In contrast to the covariance matrix of PCA, the size of the image covariance matrix obtained using 2-D PCA is much smaller. As a result, 2-D PCA has two advantages over PCA. First, estimation of the covariance matrix is relatively accurate even with small number of training samples. Second, less time is required to determine the corresponding eigenvectors. Experimental results show that the 2-D PCA outperforms PCA and Kernel PCA on the Yale face database [97]. However, the projected feature matrix of the 2-D PCA reflects only correlation between rows of the images, while the correlation between columns is omitted [98]. This issue was addressed by a method called DiaPCA [98], where a diagonal face image derived from the original face image is used in the 2-D PCA. Due to this, the correlation between rows and columns of the image is utilized in the computation of the projective feature matrix. Experimental results on a subset of FERET face database show that the DiaPCA performs better than both the PCA and 2-D PCA. In addition, it is shown that the recognition accuracy can be improved further by combining the DiaPCA and 2-D PCA.

The PCA based methods essentially select a subspace that retains most of the variation of the training face images. It is suitable for reconstruction of the face images, but may not be suitable for discrimination purpose. This limitation was overcome by linear discriminant analysis (LDA), where the subspace is derived using class-specific information of the training samples. This information is useful to develop a method for reducing the dimension of the feature space such that the resulting subspace is more suitable for the task of classification. The LDA determines the projection matrix $\mathbf{\Psi}_{\text{LDA}}$ in such a way that the ratio of the between-class scatter and the within-class scatter is maximized [91, 99], i.e.,

$$\mathbf{\Psi}_{\text{LDA}} = \arg\max_{\mathbf{\Psi}} \frac{\mid \mathbf{\Psi}^T \mathbf{S}_B \mathbf{\Psi} \mid}{\mid \mathbf{\Psi}^T \mathbf{S}_W \mathbf{\Psi} \mid}, \tag{2.4}$$

where $\mathbf{S}_B$ is the between-class scatter matrix, and $\mathbf{S}_W$ is the within-class scatter matrix. They are defined as

$$\mathbf{S}_B = \sum_{i=1}^{N} T_i(\mu_i - \mu)(\mu_i - \mu)^t, \tag{2.5}$$

and

$$\mathbf{S}_W = \sum_{i=1}^{N} \sum_{\mathbf{x}_k \in \mathbf{X}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^t. \tag{2.6}$$

In the above expressions, μ is the mean of all the training samples, T_i is the number of training samples in class i, N is the number of distinct classes, μ_i is the mean vector of samples belonging to class i, and $\mathbf{X}_i$ represents the set of samples belonging to class i. The projection matrix $\mathbf{\Psi}_{\text{LDA}}$ in (2.4) are the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ associated with the largest eigenvalues [99].

In the context of appearance-based object recognition, it is generally observed that algorithms based on LDA are superior to those based on PCA (since LDA directly deals with class discrimination). The PCA might outperform LDA, if the number of training face images per class is small [100]. Especially, this is the scenario in the case of face recognition, where the task is also known as small size problem or singularity problem [91]. Many methods have been proposed to deal with this problem, including pseudo-inverse LDA [101], PCA+LDA [100], regularized LDA [102] and recursive LDA [103]. Some of these methods were applied for face recognition, and the performance using PCA+LDA was found to be better. The issue of the singularity problem is also addressed using the two-dimensional LDA (2-D LDA) [104], which is analogous to 2-D PCA.

The approaches based on PCA and LDA use the second order statistics of the training set, and ignore the higher order statistical dependencies such as the relationships among three or more pixels of the images [105, 106]. Independent component analysis (ICA) [107] separates the higher order moments of stochastic observations in addition to the second order moments, and can be seen as generalization of the PCA. In [105], two architectures are proposed to perform the face recognition using ICA. In the first architecture, images

are considered as random variables and pixels are considered as trials. Here, the projective matrix $\mathbf{\Psi}_{\text{ICA}}$ is derived by making the face images as independent as possible. Two images $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, if when moving across pixels it is not possible to predict the value of the pixel in the image $\mathbf{x}_2$ based on the value of the same pixel in the image $\mathbf{x}_1$. On the other hand, pixels are considered as random variables, and images as trials in the second architecture. Here the $\mathbf{\Psi}_{\text{ICA}}$ is derived by making pixels of images as independent as possible. Pixels *i* and *j* would be independent if when moving across the entire set of images it is not possible to predict the value taken by pixel *i* based on the corresponding value taken by pixel *j* on the same image. In the first architecture, the data matrix $\mathbf{X}$ is organized such that the training face images are in rows and the pixels are in columns. The ICA finds a weight matrix $\mathbf{W}$ such that the rows of $\mathbf{\Psi}_{\text{ICA}} = \mathbf{WX}$ are as statistically independent as possible. The rows of the projective matrix ($\mathbf{\Psi}_{\text{ICA}}$) called source images are used as basis functions to represent the face images. The number of basis images (n) is the same as the number of face images in the training set. A face image can be represented using the basis images as

$$\mathbf{x} = b_1\psi_1 + b_2\psi_2 + \cdots + b_n\psi_n. \tag{2.7}$$

The coefficient vector $\mathbf{b} = (b_1, b_2, \ldots, b_n)^t$ is compared using the cosine similarity measure in the nearest neighbor algorithm to compute the identity of a test face image. The second architecture is realized by repeating the above mentioned procedure using data matrix as transpose of $\mathbf{X}$ rather than $\mathbf{X}$. Both the architectures addressed the issue of singularity by reducing the dimensionality of the image space using PCA. The performance of face recognition is similar using both the architectures, and is superior to the PCA for recognizing face images across days and changes in expression [105]. In [108], experiments were conducted on a face database containing 206 subjects (2060 face images), and the method reported a recognition performance of 79.1%, 81.0% and 88.1% for PCA, LDA and ICA, respectively.

In [109], ICA and its variants, specifically the independent subspace analysis (ISA), and topographic independent component analysis (TICA) are used to learn the view-specific subspace of the face images. The learned subspaces are employed to estimate the pose of a given new test face image. In all the ICA-based approaches an image is transformed into

a vector. On the contrary, row-column independent component analysis (RC-ICA) which worked directly on images without image-to-vector stretching, is given in [110]. The experimental results show that RC-ICA performs better than PCA, 2-D PCA and ICA on Yale-B, AR, and FERET face databases. The performance of the RC-ICA is compared with the representations proposed in this thesis.

2.2.5.2 Nonlinear subspace methods

In these methods, the constraint of linearity is relaxed while computing the subspace (projection matrix). These methods include nonlinear PCA (NLPCA), Kernel PCA (KPCA), kernel LDA (KLDA) and nonlinear ICA. The NLPCA can be performed using autoassociative neural network (AANN) [111] model which is a mulitlayer perceptron network (MLP) in autoassociation mode. In this network, the number of nodes in the hidden layer is less as compared to nodes in the input and output layers (dimension of the feature vector). The same vector $\mathbf{x}$ is used both as input and as the desired output in AANN network. If a linear activation function is employed for each node, then the outputs of the hidden layer give the projected coefficients corresponding to the principal eigenvectors. In the NLPCA, a nonlinear function (sigmoid function) is chosen as an activation function.

The basic methodology of KPCA is to apply a nonlinear mapping to the input $F(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^L$, and then apply a linear PCA in the resulting feature space $\mathbb{R}^L$, where L is larger than N, and possibly infinite. The mapping F($\mathbf{x}$) is made implicit (and economical) by the use of kernel function satisfying Mercer's theorem [112]. The selection of kernel function is an engineering problem. The KPCA does not require nonlinear optimization, is not subject to overfiting, and also does not require prior knowledge of the network architecture as in the case of AANN based method.

Similar to KPCA, the kernel linear discriminant analysis (KLDA) was also proposed in [113]. The experimental results were shown to be better as compared with ICA, PCA, KPCA, and standard Fisherfaces (LDA) using two data sets that contained images of 40 and 11 subjects, respectively, with varying pose, scale and illumination. Each subspace method has its own advantages and disadvantages, and comparison of these methods for face recognition task can be found in [114].

2.3 Face databases

In this thesis, face recognition for various representations is studied using three face databases, namely, FacePix, PIE (Pose, Illumination and Expression) and Yale-B face database. The details of the databases are given below.

2.3.1 FacePix database

FacePix face database is collected at Arizona State University [115, 116]. It consists of 30 persons, and each having two sets of face images: A set with pose angle variation and a set with illumination angle variation. The set with pose angle variation has 181 images (representing angles from -90° to 90° at 1° interval) with varying pose. In this thesis we denote these images by $I^1, \ldots, I^{181}$. The illumination set is captured with the subject looking directly into the camera, while the light source is moved around the subject. The light source is moved at 1° interval from $-90°$ to 90°. These images are denoted by $L^1, \ldots, L^{181}$. Some of the face images for a person are shown in Fig. 2.4.

2.3.2 Pose, Illumination and Expression (PIE) database

Pose, illumination and expression (PIE) face database [89] collected at Carnegie Mellon University contains three variations of face images. The pose variation set consists of 68 persons, each having face images for 13 different poses. We have used three point normalization [117], based on the locations of eyes and nose, to crop the face images. The cropped face images of a person are shown in Fig. 2.5(a). The *C22, C25, C02, C37, C05, C09, C07, C29, C11, C31, C14* and *C34* are the positions of cameras located at different locations to get the face images at different poses. The illumination variation subset of the PIE face database (PIE NL dataset) [50] , has 65 subjects, each having 21 face images, as shown in Fig. 2.5(b).

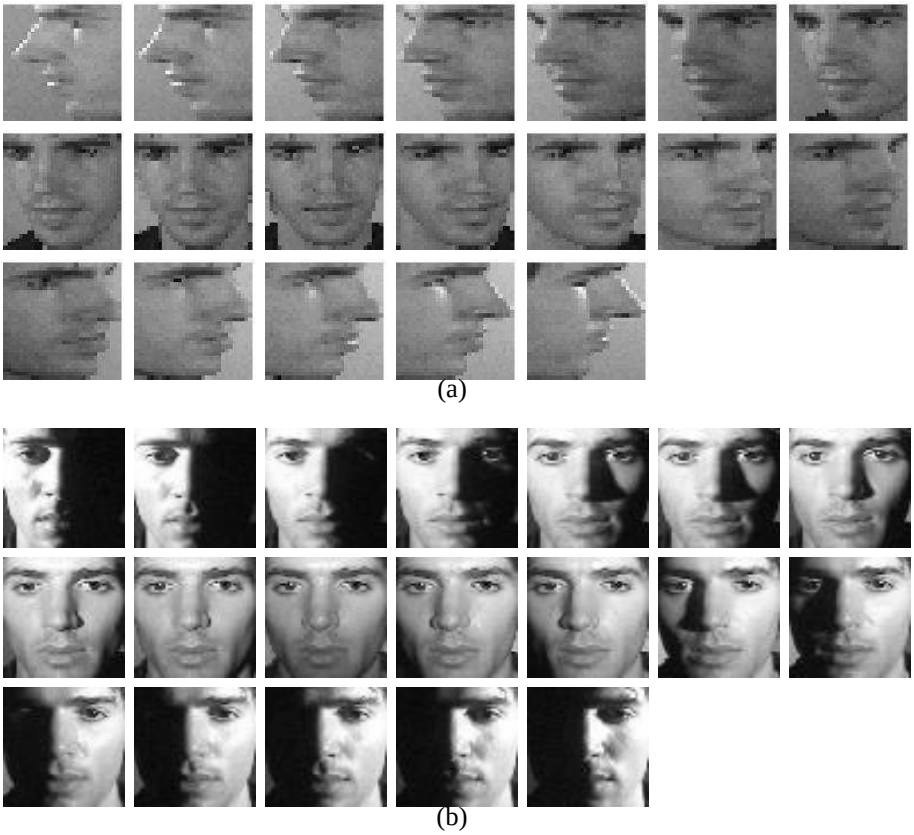


Fig. 2.4: Sample face images of a person with (a) pose variation and (b) illumination variation set of FacePix face database.

2.3.3 Yale-B database

The Yale-B face database [51] contains face images of 10 persons. For each person 576 viewing conditions (nine different poses and 64 different lighting conditions from negative azimuth to positive azimuth) are captured. The 64 frontal face images under different lighting conditions for each person are used in our experiments. Fig. 2.6 shows some samples of face images for a person.

2.4 Summary

In this chapter we have discussed different representations of face images used in face recognition. The gray level values of a face image cannot be used directly for matching in a face

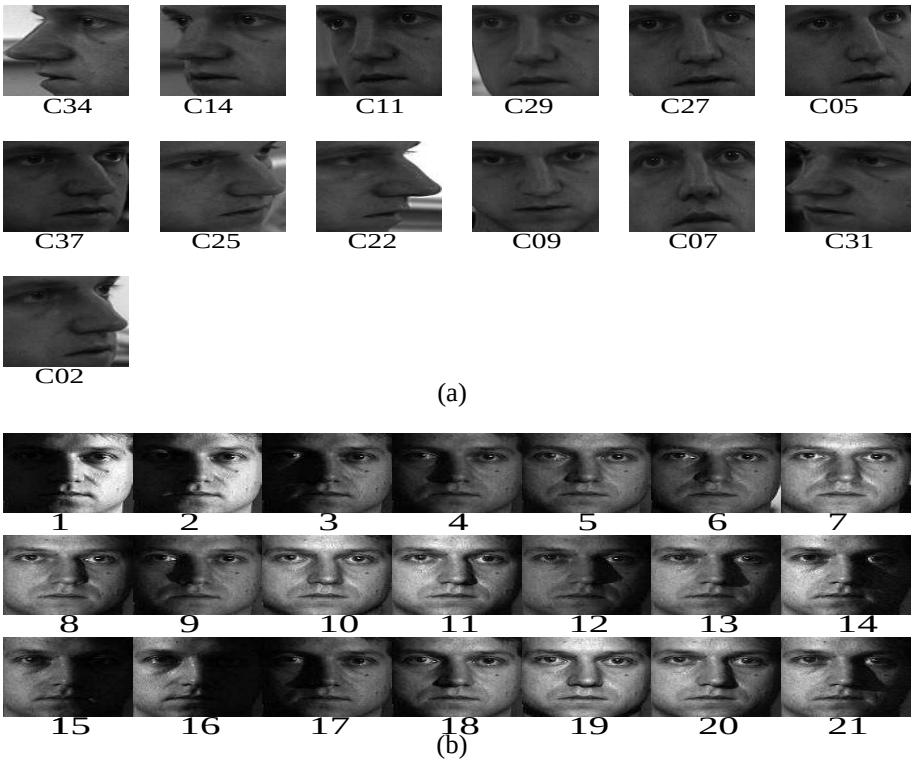


Fig. 2.5: Sample face images of a person with (a) pose variation and (b) illumination variation of PIE face database.

recognition task due to variability in pose and illumination of the face image. It is important to extract suitable representation to address the issues of inter-class and intra-class variations. The available methods for representing a face image use either feature-based approach or holistic approach. The feature-based approach attempts to highlight the unique information, which may reflect in the relative positions of the facial features. This approach gives good performance even in the extreme cases of intra-class variations, as for example, a face image with large variations in the pose. However, extraction of facial features from such extreme cases is not easy. The feature-based approach also suffers from the problem of partial face occlusion. In the holistic approach, representation of a face image is obtained by considering the entire face as a single unit. In this approach, the subject-specific unique information of face image may not get emphasized during matching. Thus, both approaches have their own advantages and disadvantages. Ideally, one should use a combination of these two approaches for representing a face image. Such representations are sketch, caricature

Fig. 2.6: Sample images of a person from illumination variation subset of Yale-B face database.

and edge map of a face image. The sketch and caricature are the best representation from the perception point of view. Though there are a few approaches proposed in the literature to derive the sketch and caricature of a given face image, mimicking the artist's ability is still a challenge.

Edge map is another representation that contains perceptually crucial information of a face image. Moreover, an edge map contains all information present in the sketch and caricature of a face image. However, the edge map representation has two drawbacks in the context of face recognition: a) Selection of threshold value for edge extraction, and b) matching of the edge maps. In this work we have explored four types of representations of face image that capture the edge-like information of a face image, but do not perform the task of edge extraction. The second issue of matching edge-based representations is addressed to some extent by smearing the edge information. Template matching based approach is employed to illustrate the significance of the proposed edge-based representations for face recognition.

Chapter 3

Edginess-based representation

In this chapter, we consider edginess-based representation of a face image. It is computed using one-dimensional processing of the image [118–120], unlike the traditional 2-D processing methods. A potential field is derived from the edginess image to improve matching under intra-class variation of the face image. Template matching is used for matching face images with pose and illumination variation. Evidences from different reference templates are combined, and classified using AANN models.

The chapter is organized as follows: Section 3.1 discusses computation of the edginess image using 1-D processing. The template matching of images using the edginess-based representation is discussed in Section 3.2. The scores obtained by matching with several templates are combined in a selective way as explained in Section 3.3. A neural network model is proposed in Section 3.4 to develop a classifier using the combined scores. Experimental results are discussed in Section 3.5, and a summary of the chapter is given in Section 3.6.

3.1 Edginess-based representation using 1-D processing of images

The continuous gradient computed at every pixel is called edginess of an image [119]. Direct computation of the gradient using a derivative operator enhances noise as well, along

with the edge information. The noise level can be reduced by applying a smoothing operation followed by a derivative operation. The smoothing operation smears the edge information. One-dimensional (1-D) processing of images can be used to reduce the effect of noise without smearing the edges significantly [119, 120]. In 1-D processing of a given image, the smoothing operator is applied along one direction, and the derivative operator is applied along the orthogonal direction. By repeating this procedure of smoothing followed by derivative operation along the orthogonal direction also, two edge gradients are obtained. As the smoothing is done along a direction orthogonal to the direction of the edge extraction, smearing of the edges is reduced. This method differs from the traditional methods [12], in the sense that smoothing is done along one direction, and differential operator is applied along the orthogonal direction.

In 1-D processing of images, a 1-D Gaussian filter is used for smoothing, and the first derivative of the 1-D Gaussian function is used as differential operator. The 1-D Gaussian filter is given by

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \frac{-x^2}{2\sigma_1^2}, \tag{3.1}$$

where σ_1 is the standard deviation of the Gaussian function. The first derivative of a 1-D Gaussian is given by

$$c(x) = \frac{-x}{\sqrt{2\pi}\sigma_2^3} \exp \frac{-x^2}{2\sigma_2^2}, \tag{3.2}$$

where σ_2 is the standard deviation of the Gaussian function. The values of σ_1 and σ_2 decide the spatial extent of these 1-D filters. The edge gradient obtained by applying the derivative operator along a direction θ with respect to a horizontal scan line is denoted as i_θ^g. Fig. 3.1 shows examples of the gradient maps obtained along different directions for the same face image. It shows that they contain complementary information of the face image. In this study, the values of σ_1 and σ_2 are chosen to be 0.6 and 1, respectively, based on experimental studies on several face images. It must be pointed out that the choice of the values of σ_1 and σ_2 is not critical. One of the problems with the edge gradient representation is that they are very sparse, i.e., most of the values are very small, close to 0. Because of this, even a small deviation in the edge contour for the same face image leads to poor scores in template matching. This is called *locality problem*, and it can be reduced by spreading the

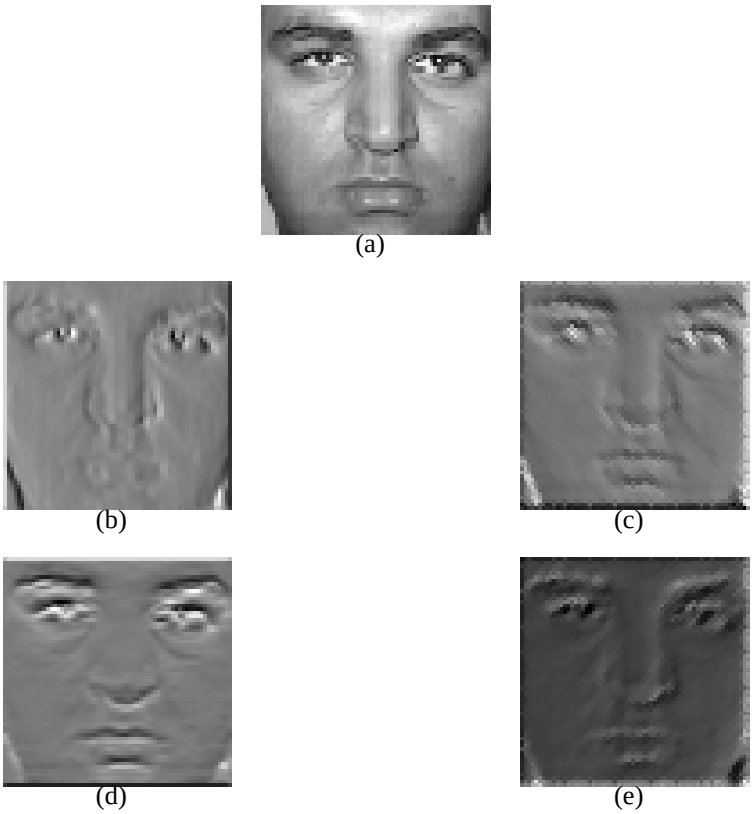


Fig. 3.1: (a) Gray level face image. Edge gradient (i_θ^g) of the face image obtained using $\sigma_1 = 0.6$, $\sigma_2 = 1$, for (b) $\theta = 0°$, (c) $\theta = 45°$, (d) $\theta = 90°$ and (e) $\theta = 135°$.

edge information in the edge gradient representation. The edge information can be spread by either using high values of σ_2, or by removing some of the high frequency DFT coefficients of the edginess image. But spreading the edge information in this manner leads to loss of information.

Alternatively, the edge information can be spread using potential field representation [121]. The potential field u_θ, derived from the edge gradient i_θ^g, is obtained by minimizing the energy E_θ given by [122]

$$E_\theta = \int\int [\mu((u_{\theta_x})^2 + (u_{\theta_y})^2) + |(i_\theta^g)^2 + (i_{\theta+90}^g)^2| \, |u_\theta - i_\theta^g|^2 \,] dx dy, \tag{3.3}$$

where u_{θ_x} and u_{θ_y} are the partial derivatives of u_θ along the x and y axes, respectively. The parameter μ is the scaling factor used to control the amount of smearing. This variational

formulation follows the standard principle that the result must be smooth when there is no data. In particular, we see that when the gradients i^g_θ and $i^g_{\theta+90}$ are small, the energy is dominated by the partial derivatives of the field, yielding a smooth field. This smoothing term (the first term in the integrand of (3.3)) is the same term used by Horn and Schunk in their classical formulation of optical flow [123]. On the other hand, when the gradient is large, the second term dominates in the integrand of (3.3), and it is minimum when $u_\theta = i^g_\theta$. This term is responsible for introducing the gradient information in the potential field. This approach is similar to the one used in [122] for developing a deformation force of snakes. To minimize the energy E_θ given by (3.3), the following Euler equation is a necessary condition [124]

$$\mu \bigtriangledown^2 u_\theta - (u_\theta - i^g_\theta)|(i^g_\theta)^2 + (i^g_{\theta+90})^2|^2 = 0, \tag{3.4}$$

where $\bigtriangledown^2$ is the Laplacian operator. We can see that wherever the gradient is zero, (3.4) will reduce to the Laplacian operator. It can be solved by considering u_θ as a function of the variable t, and the solution is given by [122]

$$\begin{aligned} u_{\theta\, t}(x,y,t) &= \mu \bigtriangledown^2 u_\theta(x,y,t) - [u_\theta(x,y,t) \\ &\quad -i^g_\theta(x,y)][(i^g_\theta(x,y))^2 + (i^g_{\theta+90}(x,y))^2], \end{aligned} \tag{3.5}$$

where $u_{\theta\, t}$ is the partial derivative of $u(x,y,t)$ with respect to t. The above equation is known as generalized diffusion equation, commonly encountered in heat conduction and reactor physics [125]. Let $\triangle x$ and $\triangle y$ be the inter pixel distance along the x and y axes, and $\triangle t$ be the change in t. Then, using the standard approximations for the partial derivatives, (3.5) can be written as

$$\begin{aligned} u^{m+1}_{\theta\, i,j} &= (1 - E^2_{i,j} \triangle t)u^m_{\theta\, i,j} + \frac{\mu \triangle t}{\triangle x \triangle y}(u^m_{\theta\, i,j+1} + u^m_{\theta\, i+1,j} \\ &\quad + u^m_{\theta\, i-1,j} + u^m_{\theta\, i,j-1} - 4u^m_{\theta\, i,j}) + E^2_{i,j}(i^g_{\theta\, i,j} \triangle t), \end{aligned} \tag{3.6}$$

where $E_{i,j} = \sqrt{(i^g_{\theta\, i,j})^2 + (i^g_{\theta+90\, i,j})^2}$, and $u^m_{\theta\, i,j}$ is the potential field at the location (i, j) after m^{th} iteration. Equation (3.6) is stable, provided $E_{i,j}$ and $i^g_{\theta\, i,j}$ are bounded, and the Courant-

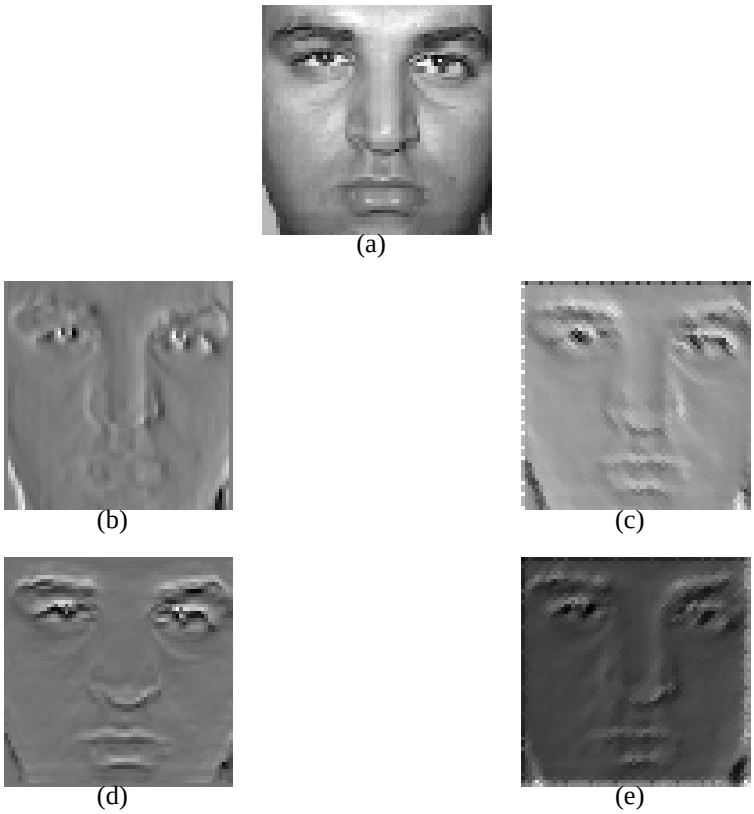


Fig. 3.2: (a) Gray level face image. Potential field (u_θ) developed from the edge gradient of the face image (a) using $\sigma_1 = 0.6$, $\sigma_2 = 1$, $\mu = 0.005$ for (b) $\theta = 0°$, (c) $\theta = 45°$, (d) $\theta = 90°$ and (e) $\theta = 135°$.

Friedrich-Lewis restriction $\frac{\mu \triangle t}{\triangle x \triangle y} \leq \frac{1}{4}$ is maintained [124]. The potential fields obtained by (3.6) for different edge gradient images are shown in Fig. 3.2. It shows that the edges are smeared in the potential field representation, and thus it improves the matching in the case of deviation in edge contours between test and reference face images of the same person.

3.2 Template matching using edginess-based representation

Template matching is one of the approaches proposed in the literature to address the issue of inter-class variation [60], because it takes into account the unique information of a person's face image. But this approach has the drawback that it gives poor performance

under intra-class variation [126]. The problem of intra-class variation can be overcome using an approach which synthesizes a 3-D model of the face image from a given sample [51, 86, 87, 127, 128]. But synthesis of face image may result in some artifacts, and also some loss of the unique information. Thus synthesis-based approach may degrade the performance of the face recognition system. Another way to address the issue of intra-class variation is to consider several reference face images which capture variations in the face images such as due to different poses or due to different lighting conditions. These multiple reference face images can be used to build a model for that person's face image. The resulting model is used to identity a test face image. Such methods are discussed in [96, 129–133]. In these cases the model may be averaging out some of the information that is unique for that person. In this work we propose a template matching based method, which neither synthesizes the face image, nor derives a model for the person's face. We use the reference face images (at different poses or at different lighting condition) separately for template matching.

Template matching is performed using a correlation-based technique. Correlation between reference face image $r(x,y)$ and test face image $i(x,y)$ is computed as follows:

$$\begin{aligned} c(\tau_x, \tau_y) &= i(x,y) \odot r(x,y) \\ &= \int\int i(x,y) r(x+\tau_x, y+\tau_y) dxdy \\ &= \int\int I^*(u,v) R(u,v) \exp(j2\pi(u\tau_x + v\tau_y)) dudv, \end{aligned} \tag{3.7}$$

where $I(u,v)$ and $R(u,v)$ are the Fourier transforms of $i(x,y)$ and $r(x,y)$, respectively, and $\odot$ denotes the correlation operator. The correlation output $c(\tau_x, \tau_y)$ should have a high value at the origin, when the test and reference face images are similar. On the other hand, if the test and reference face images are not similar, then the correlation output should have a relatively low value even at the origin. Here, the origin refers to the center of the correlation function. Figs. 3.3(a) and (b) show the normalized correlation outputs for a true class face image and for a false class face image, respectively. The correlation output of a true class image (Fig. 3.3(a)) has a high peak value near the origin, whereas for the false class image (Fig. 3.3(b)) the correlation output is generally low. The relative heights of the values at the

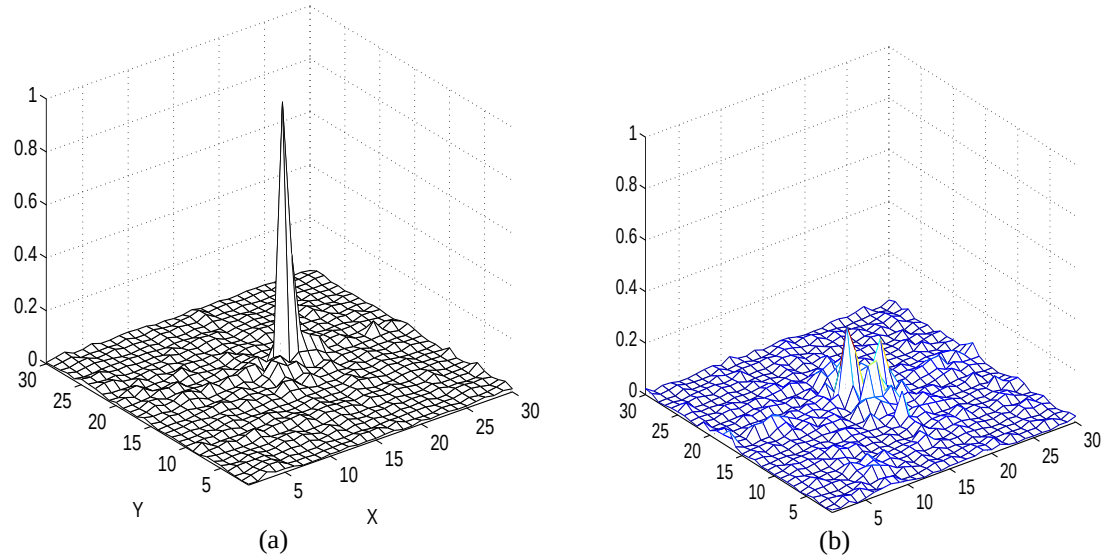


Fig. 3.3: A typical correlation output for (a) true class face image and (b) false class face image.

origin determine whether the test and reference face images are similar or not. The sharpness of the peak in the correlation output is quantified using peak-to-sidelobe ratio (PSR) [134]. The PSR (P) is defined as

$$P = \frac{p - \mu}{\sigma}, \tag{3.8}$$

where p is the value of the maximum peak in the correlation output, μ is the mean of the correlation output within a window (of size 19×19 pixels) around the peak, and σ is the standard deviation of the values in this window. In practice, we leave out a region of size 7×7 pixels around the center of the window while computing μ and σ. The choice of the sizes of these windows is arrived at empirically. Other choices can also be made to highlight the peak to sidelobe characteristics. The PSR measures the sharpness of the highest peak in the correlation output. For a similar face images, the peak will be sharp and high relative to the values in the neighborhood. Otherwise the peak will be low and blunt.

The edge gradients (i_θ^g) for different directions (θ) give different information of a face image. Hence, we have performed correlation between partial evidence (u_θ) of the given test and reference face images. Let $\mathbf{c}_\theta$ be the correlation output obtained when the correlation between the partial evidence along θ direction of the test and reference face images

is performed. The resultant correlation output is used to compute the PSR (P_θ) using (3.8). Ideally the P_θ should be high if the given test face image is similar to the reference image. In our experiment we have computed the partial evidence (u_θ) along the four directions ($\theta = 0^\circ$, 45°, 90° and 180°). Hence, for a given test face image a four dimensional feature vector (i.e., containing four PSR values) is obtained. Fig. 3.4(a) shows the scatter plot obtained from the PSR vectors of the true class and false class face images for a person using pose variation set of the FacePix face database. For visualization we show the plot using three ($\theta = 0^\circ$, 45° and 90°) of the four dimensions. In this example we have used I^1 as a training (reference) face image of a person. The remaining 180 face images of the given person form examples of the true class image, and the corresponding PSR values are denoted by the diamond ($\diamond$) symbol in the plot. For the false class, $29 \times 181 = 5249$ face images are available, and the PSR values are shown by the point ($\cdot$) symbol in the scatter plot. Though the separation between the true and false class face images are not decisive, but one can observe from the plot that high scores are given by the face images I^2, I^3, I^4, I^5 and I^6 of the true class. These face images have pose that is close to the pose of the training (reference) face image. One can also see that none of the face images of the false class gives high scores. It means that the chances of matching face images of two different persons even with the same pose, is less. Similar observations can be made from the scatter plot shown in Fig. 3.4 (b), which is obtained using I^{46} as the reference face image. This behavior is utilized to recognize a given face image.

3.3 Combining scores from different templates

One can conclude from the previous section that if a test face image of the true class has a pose that lies between poses of two reference face images, then the test image will give high scores with respect to both the reference face images. It is better to combine these scores rather then use them separately for taking decision. One way to combine the scores is as follows. Let $P_\theta^{t,l}$ is the similarity score (PSR) obtained when the potential field representation along θ direction of the test face image I^t is correlated with the corresponding representation of reference image I^l. The combined similarity score for two reference images I^l and I^m is

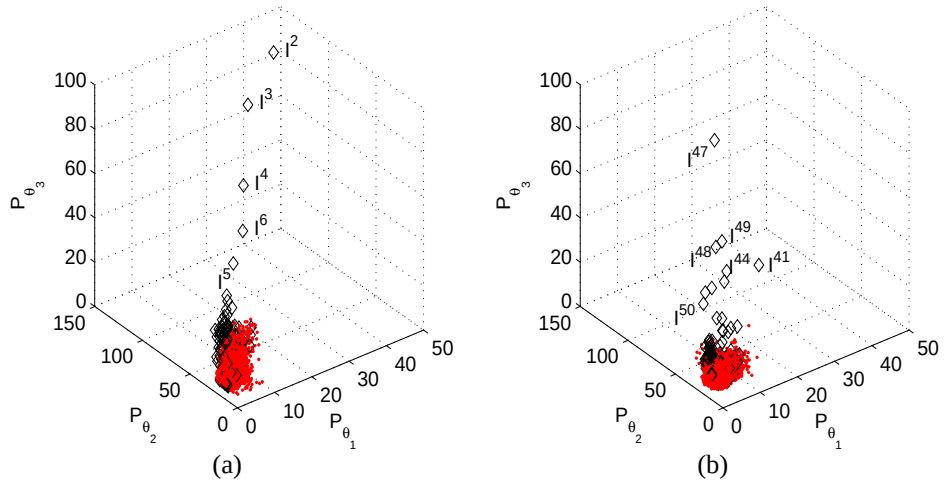


Fig. 3.4: Scatter plots of PSRs for a person's face image using potential field representation, obtained from $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$, and using (a) I^1 and (b) I^{46} as the training (reference) face image.

given by

$$P_\theta^{t,l,m} = \left[\frac{1}{2}\left[\left(P_\theta^{t,l}\right)^\alpha + \left(P_\theta^{t,m}\right)^\alpha\right]\right]^{\frac{1}{\alpha}}, \tag{3.9}$$

where the parameter α decides the weights associated with the scores. For $\alpha \leq 1$, $\min[P_\theta^{t,l}, P_\theta^{t,m}] \leq P_\theta^{t,l,m} \leq \frac{P_\theta^{t,l}+P_\theta^{t,m}}{2}$, and for $\alpha \geq 1$, $\frac{P_\theta^{t,l}+P_\theta^{t,m}}{2} \leq P_\theta^{t,l,m} \leq \max[P_\theta^{t,l}, P_\theta^{t,m}]$. A low value of α is suitable for the case of false class, and a high value of α for true class. One has to choose a suitable value of α such that the separation between true and false classes is enhanced. We have found empirically that $\alpha = 3$ is a good choice. Fig. 3.5 illustrates the effect of α on PSR vector through scatter plots. Figs. 3.5 (a) and (b) are the same scatter plots as shown in Figs. 3.4 (a) and (b), respectively, but with a different view. In this example we have shown only the scores obtained from the true class face images I^t for $2 \leq t \leq 45$. Figs. 3.5 (c), (d), (e) and (f) are the scatter plots obtained after combining the PSR scores in Figs. 3.5 (a) and (b) using (3.9), for $\alpha = 0.4$, 1, 3, and 5, respectively. One can see that as α increases the points due to true and false classes move away from the origin. Similarly, we can combine the scores obtained from other reference face images of adjacent poses.

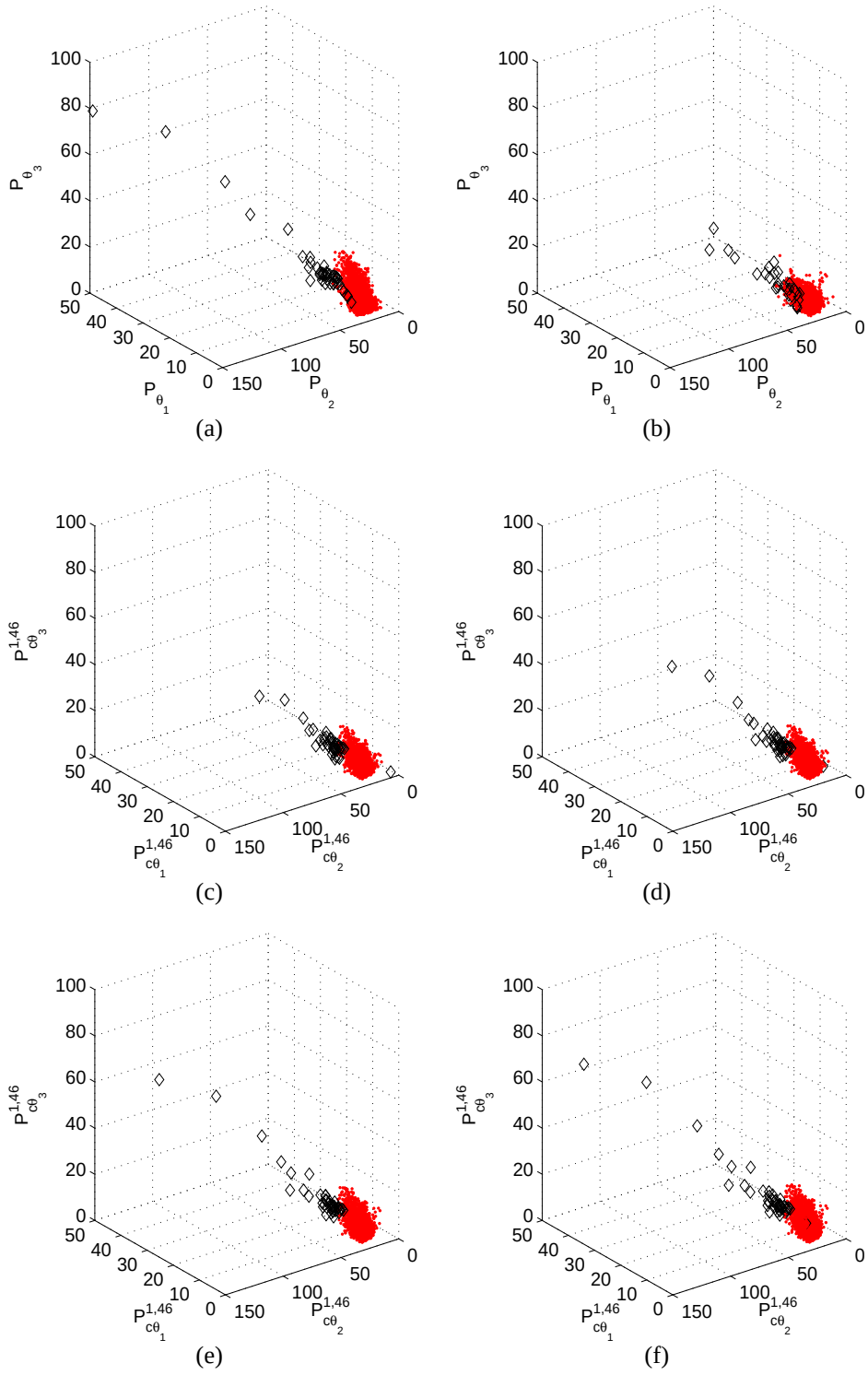


Fig. 3.5: Scatter plots of PSRs for a person's face image using potential field representation, obtained from $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$, and using (a) I^1 and (b) I^{46} as the training (reference) face image. Combining (a) and (b) using (c) $\alpha = 0.4$, (d) $\alpha = 1$, (e) $\alpha = 3$ and (f) $\alpha = 5$.

3.4 AANN based classification using distribution of feature vectors

The next task is to classify a given test face image using the feature vectors consisting of the combined score $P_{\theta}^{t,l,m}$ for the four different values of θ. One can employ a classifier based on multilayer perceptron (MLP) [69] neural network model or support vector machine (SVM) [70]. But these models require samples from both the true and false classes. Though we can have a large number of false class images for a given person, but we cannot afford to have many face images of the true class. This problem can be overcome by using an approach based on autoassociative neural network (AANN) model [135]. There are two reasons for adopting this approach: Firstly, one can have many false class face images for a given person. Secondly, the feature points due to false class are more dense as compared to the feature points due to the true class in the scatter plots. The distribution of these closely spaced points of the false class can be modeled using the distribution capturing ability of an AANN model. The distribution capturing ability of an AANN is illustrated in Fig. 3.6(b) for the 2-D data shown in Fig. 3.6(a). The structure of the AANN model is 2L 10N 10N 1N 10N 10N 2L, where L denotes a linear unit, and N denotes a nonlinear unit. The network is trained using the coordinates of the data as input and output. Fig. 3.6(b) is obtained using the confidence value derived from the error between the input and output for each test point (i) in the 2-D plane in the range (-1 to +1, -1 to +1). The confidence value is computed for the squared error (E_i) of the i^{th} point as

$$C_i = \exp(-\beta E_i), \tag{3.10}$$

where β is constant. It can be seen that the confidence value is high when the density of points is high in the input data.

The distribution of points due to false class could be different for different reference face images. Hence, separate AANN models are obtained for each reference face image. The AANN model can be used for accepting or rejecting a claim. When a test face image belonging to the true class is given to the AANN model, the resulting score vector does not

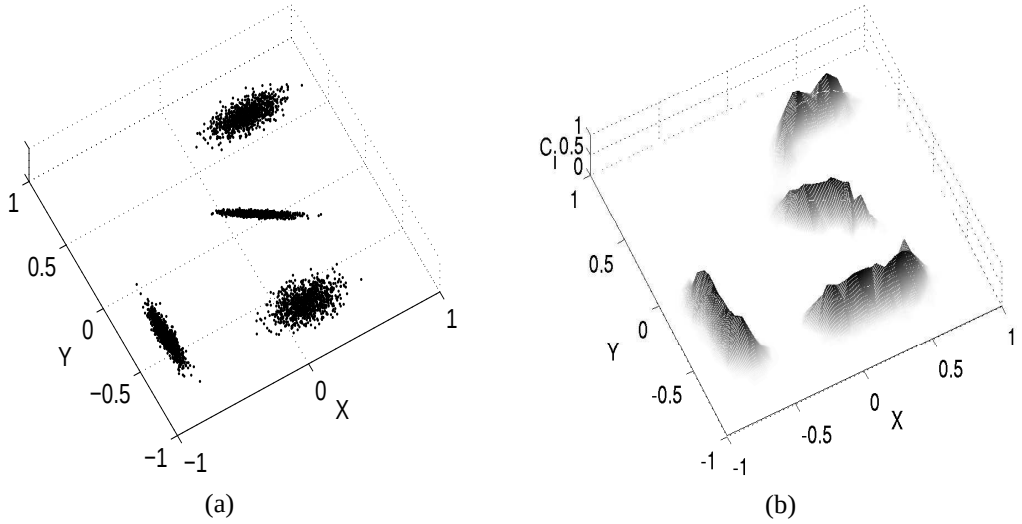


Fig. 3.6: Distribution capturing ability of AANN model. (a) Artificial two dimensional data. (b) Confidence surface realized by the network structure $2L$ $10N$ $10N$ $1N$ $10N$ $10N$ $2L$.

fall into the cluster of points belonging to the false class. Thus using a suitable threshold for the output of the AANN model, a decision can be made whether to accept the claim of the test input or not.

3.5 Experimental results

Here, we give a brief summary of our experiments. The results on the pose variation set of the FacePix face database is explained first. The block diagram of the training phase is shown in Fig. 3.7(a). In this block diagram we have shown training with five reference face images I^{1}, I^{46}, I^{91}, I^{136} and I^{181}. This process can be generalized for any number of reference face images. The reference face images are chosen, in such a way that their poses are uniformly spaced over the span of $0^{\circ} - 180^{\circ}$. Several false class images are chosen, and their potential field representations (u_{θ}) are computed along the four directions ($\theta = 0^{\circ}$, 45°, 90° and 135°). These representations are correlated with the corresponding representation of each reference face image. The resulting correlation output is used to compute the similarity score using (3.8). Hence, five sets of four dimensional feature vectors (4 PSR

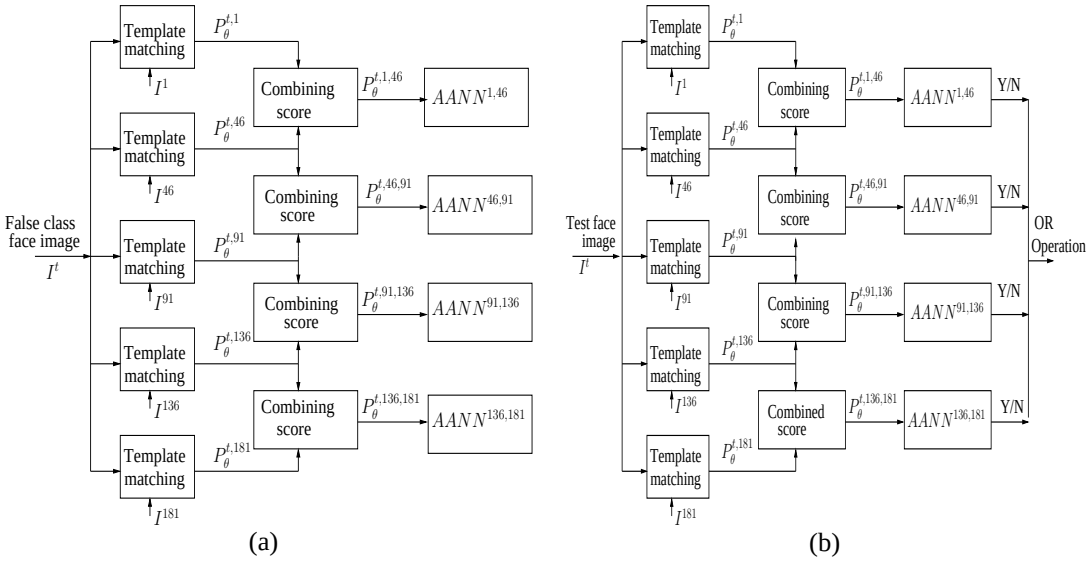


Fig. 3.7: Block diagram of face verification system for (a) training and (b) verification.

values) are obtained for each false class image. The similarity scores obtained from reference face images having adjacent poses are combined using (3.9), as shown in Fig. 3.7(a). These combined scores are presented to an AANN model for training. Let $AANN^{1,46}$ denote the AANN model trained with the combined similarity scores ($P_\theta^{t,1,46}$) obtained using the reference face images I^1 and I^{46}. The structure of the AANN model is $4L$ $8N$ $2N$ $8N$ $4L$. The AANN model is trained using backpropagation algorithm for about 3000 epochs. Similarly, we have designed $AANN^{46,91}$, $AANN^{91,136}$ and $AANN^{136,181}$ using the same false class face images. The block diagram of the testing phase of face verification system is shown in Fig. 3.7(b). For a given test face image, the potential field representation u_θ is computed along the four directions (θ = 0°, 45°, 90° and 135°). These representations are correlated with the corresponding representation of each reference face image of the claimed identity. The resulting similarity scores are combined as in the training phase, and are presented to AANN models as shown in Fig. 3.7(b). The combined similarity score (4 dimensional feature vector) is used to compute the error in associating the vector with the AANN models corresponding to the reference face images. If the error is above a threshold in *any* one of the AANN models, the claim is accepted. Here, the threshold value for each AANN model could be different. False acceptance ratio (FAR) and false rejection ratio (FRR) are two er-

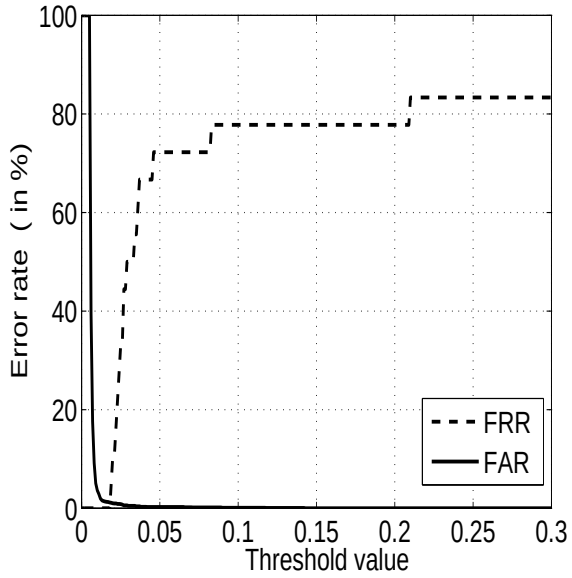


Fig. 3.8: ROC curves for a person using $AANN^{1,46}$.

ror metrics that are used to evaluate a face verification system. The trade off between FAR and FRR is a function of the decision threshold. Equal Error Rate (EER) is the value for which the error rates FA and FR are equal. Here, we will explain the computation of EER for a single person. We have used I^1, I^{46}, I^{91}, I^{136} and I^{181} as reference face images. The remaining $176 = (181 - 5)$ face images form the examples of the true class. For false class, $5259 = (29 \times 181)$ face images are available. Out of these, 3000 face images are used to train the $AANN^{1,46}$, $AANN^{46,91}$, $AANN^{91,136}$ and $AANN^{136,181}$ models. The remaining 2249 $= (5349 - 3000)$ false class face images are used for testing. But the true class sample will be different for each of the AANN models, because the objective of AANN is to verify if the true class face image has a pose in a specific range. Thus for $AANN^{1,46}$, the true class samples will be the images between I^1 to I^{46}. By varying the threshold value of $AANN^{1,46}$ the resultant receiver operating characteristics (ROC) curve is obtained as shown in Fig. 3.8. The ROC characteristics show that the FAR curve is steep, indicating that the corresponding combined PSR values are clustered around low values. On the other hand, the FRR curve is slowly varying, indicating that the corresponding PSR values are more scattered. The intersection point of FAR and FRR curves gives the EER for this model. Likewise, the EERs

Table 3.1: Performance (average recognition rate in %) using edginess-based representation for different sets of training (reference) face images under pose variation of FacePix face database.

Approaches		Set of reference face images		
		I^{91}	I^{1}, I^{91} and I^{181}	I^{1}, I^{46}, I^{91}, and I^{136} and I^{181}
Principal component analysis	[115, 116]	20.74	50.53	71.66
Linear discriminant analysis		20.7	56.9	78.7
Hidden Markov model		31.7	41.3	63.5
Bayesian information criteria		18.5	45.2	69.5
Edginess-based approach		**46.1**	**74.4**	**92.2**

for the other AANN models are computed, and the average EER is obtained for that person. The experiment is repeated by building verification model for each person, and the corresponding value of EER is used as a measure of performance. The average EERs obtained for pose variation set of FacePix face database for one, three and five reference templates are 51%, 34.55% and 14.17%, respectively. Likewise, the average EERs obtained for illumination variation set of FacePix face database [115, 116] for one, three and five reference templates are 33%, 16.24% and 3.5%, respectively. For pose subset of the PIE database [89], an average EER of 43.40% was obtained using single reference templates.

For comparison of results with other studies on these databases, an identification system was developed using the verification models. The identification is done by verification on a closed set of test samples. We call the set of all test samples accepted by any of the verification models as the closed set data. Table 3.1 shows the performance of the identification system in comparison with other systems for different sets of reference images. The proposed method seems to perform better than the existing methods. The reason could be that the proposed method may be preserving some unique information of a person's face image for a given pose. The results in Tables 3.1 shows that it may be better to use reference face images separately, rather than building a single model from them. The proposed approach is also evaluated for the pose subset of the PIE face database. The average recognition rate using one training (reference) template is shown in Table 3.2, along with the performance obtained using some existing approaches [117]. The results show that the proposed method

Table 3.2: Performance (average recognition rate in %) using edginess-based representation on pose variation subset of PIE face database using a single face image for training.

	Eigenfaces	FaceIt	Eigen Light-Fields	Proposed approach
Average recognition	16.6	24.3	52.5	**57.3**

performs better than the existing methods. The performance can be improved further by increasing the number of reference templates. But selection of the reference template is crucial. When three reference templates (frontal view, left profile and right profile) are used, the average recognition rate increases to **74**.**69**%.

The experiments were repeated with the set of images corresponding to the variation of illumination angle in FacePix and Yale-B face databases. The performance is summarized in Table 6.3 and Table 6.4 in Chapter 6, along with the performance of the other proposed representations.

3.6 Summary

In this chapter we have discussed the edginess-based representation that highlight the edge-like information of an image. The edginess-based representation is very sparse, and hence gives poor scores in template matching based approach. This problem was overcome by spreading the edge information using potential field representation. The representation was derived using 1-D processing of images, to obtain multiple partial evidences for a given image. An approach was proposed to combine the scores obtained by matching multiple partial evidences of different face images to address the issues of pose and illumination in face recognition. This approach preserves the unique information of face images while matching. It uses the given reference face images (at different poses or illumination conditions) separately for template matching, rather than building a model or synthesizing a face image. The resulting combined scores were used to verify the identity of the person using AANN models. The proposed AANN model based approach has the advantage that it does not require training images of the true class. Experimental results show that the proposed approach is a promising approach for dealing with the problem of pose in face recognition.

Chapter 4

Zero-frequency resonator based representation

This chapter proposes an approach based on the zero-frequency resonator [17,136] to extract the edge information of images. The proposed approach is counterintuitive to the concept that edges correspond to the high frequency components of an image. The impulse-like characteristics of edges in an image, distribute the energy uniformly over all frequencies of the spectrum, including around the zero-frequency. This property is exploited in processing a signal using the zero-frequency resonator, for extracting the locations of impulses [17,136]. Spatial domain and Fourier domain methods are employed to realize the zero-frequency resonator for two-dimensional signals. The Laplacian of the Gaussian (LOG) and the proposed approach are similar in the sense that the former approach uses a Gaussian filter for smoothing, whereas a zero-frequency resonator is used for smoothing in the proposed approach. The output of the resonator is processed using a Laplacian operator to extract the edge information. In the resulting filtered image the edge information is preserved around the zero-crossings, and the edges are extracted using sign correspondence principle [137] to identify the zero-crossings corresponding to edges.

Since the information of the edges is preserved around the zero-crossings of the filtered image, the Laplacian of the smoothed image is proposed as a representation of the face image for face recognition task. The range of values of the filtered image are normalized for matching using a nonlinear transformation, without affecting the locations of the zero-

crossings. The *locality problem* is addressed by considering only the first few eigenvectors derived from the filtered face images of the training set. To emphasize the unique information of a face image, the eigenvectors are weighted differently in the distance computation for matching. Two smoothing filters, namely, the zero-frequency filter and the standard Gaussian filter are used to demonstrate the results. The zero-frequency filter for smoothing performs better than the Gaussian filter.

This chapter is organized as follows: Section 4.1 explains the processing of 1-D signals using the zero-frequency resonator. The zero-frequency filter is an infinite impulse response (IIR) filter, and it is realized for 2-D signals using two approaches, namely, the spatial domain and the Fourier domain. These are explained in Section 4.2. Section 4.3 discusses comparison of the proposed approach with the LOG operation using some illustrations of edge extraction from clean and noisy images. Issues involved in using the Laplacian of the smoothed image as a representation are discussed in Section 4.4. Section 4.5 discusses the *locality problem* of the filtered image using eigenanalysis. Experimental results are given in Section 4.6, and a summary of the chapter is given in Section 4.7.

4.1 Zero-frequency resonator for 1-D signals

An ideal impulse has, in principle, equal amplitude/energy at all frequencies, including around the zero-frequency. Any other zero mean signal, with the same strength or energy of the unit impulse, has its energy distributed unevenly both in the time and frequency domains. It has been shown that the characteristics of an impulse in a signal are preserved in the output obtained by filtering the signal using a zero-frequency resonator [17].

An ideal digital zero-frequency resonator is a second order system whose system function is given by (see Appendix A)

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - 2z^{-1} + z^{-2}}, \tag{4.1}$$

where $Y(z)$ and $X(z)$ are the z-transforms of the output ($y[n]$) and input ($x[n]$) of the zero-frequency resonator, respectively. The unit sample response of the filter is given by the

recurrence relation

$$y[n] = 2y[n-1] - y[n-2] + \delta[n], \tag{4.2}$$

where $\delta[n]$ is the unit sample sequence. Note that (4.2) can be interpreted as a cumulative sum of the input performed twice. That is,

$$\begin{aligned} y_1[n] &= \sum_{m=0}^{n} x[m] \\ y[n] &= \sum_{m=0}^{n} y_1[m]. \end{aligned} \tag{4.3}$$

Here, the output $y[n]$ at each instant depends only on the past two output values, and the current input value.

Because of the cumulative summation, the output of the zero-frequency resonator for a sequence of impulses grows approximately as a polynomial function of time [17]. But the fluctuations in the output contain the information of the impulses and their locations. In order to extract this information, the local mean is subtracted from the output signal. The resulting signal called, *filtered signal*, is given by

$$\hat{y}[n] = y[n] - \bar{y}[n], \tag{4.4}$$

where the local mean $\bar{y}[n]$ is computed as

$$\bar{y}[n] = \frac{1}{2N_1 + 1} \sum_{i=-N_1}^{N_1} y[n+i],$$

using a window size of $(2N_1 + 1)$ samples. Here, the size of the window is not very critical as long as it is in the range of 0.5 to 1.5 times the average interval between the impulses [17]. The positive zero-crossings (PZCs) in the filtered signal, where the signal changes its sign from negative to positive, correspond to the locations of the impulses. Generally, it is preferable to pass the signal through two or more 0-Hz resonators to suppress the effects of high frequency components [17]. Fig. 4.1(a) shows a sequence of impulses with varying

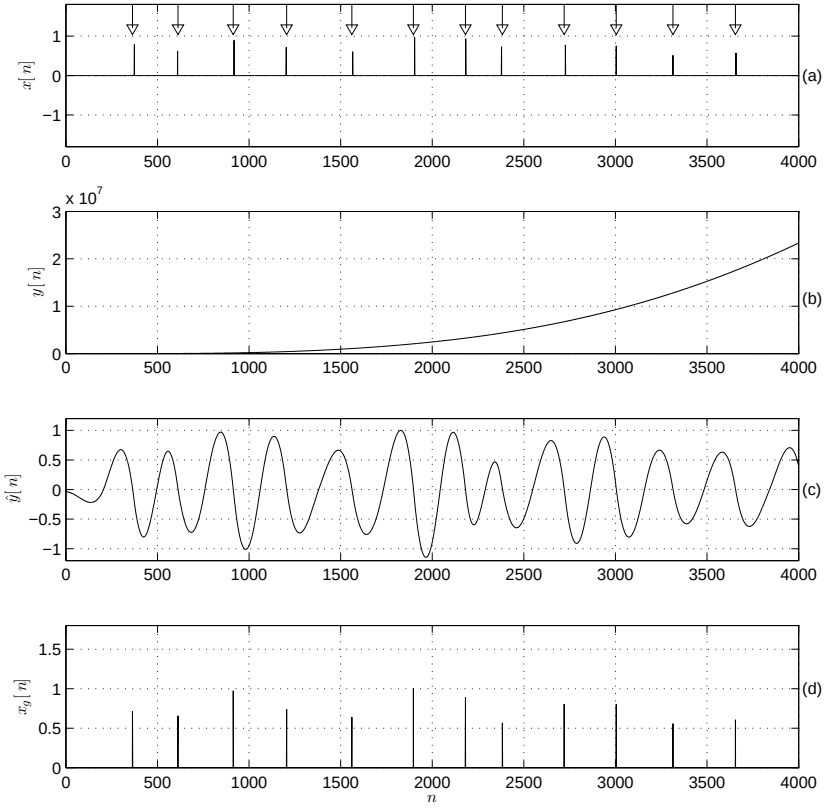


Fig. 4.1: (a) Impulses at random locations with varying magnitudes. (b) Output of cascade of two 0-Hz resonators. (c) Filtered signal. (d) Estimated magnitudes of the impulses from the filtered signal.

strengths. The corresponding filtered signal is shown in Fig. 4.1(c). The locations of the impulses derived from the filtered signal are indicated by arrow marks in Fig. 4.1(a). The range of the values of the output of the zero-frequency resonator is very large because of the cumulative sum, as shown in Fig. 4.1 (b). The strengths of the impulses at each PZC are given approximately by the slope of the filtered signal around the PZCs. The slope is obtained by computing the difference between samples on either side of the PZCs in the filtered signal. Thus, both the locations and strengths of the impulses can be obtained from the filtered signal, as can be seen from Fig. 4.1(d). If the sequence of randomly located impulses is passed through a time varying filter with some resonance or antiresonance characteristics, the resulting signal does not show the excitation information clearly. But the excitation information, i.e., the locations and strengths of the impulses, can still be derived

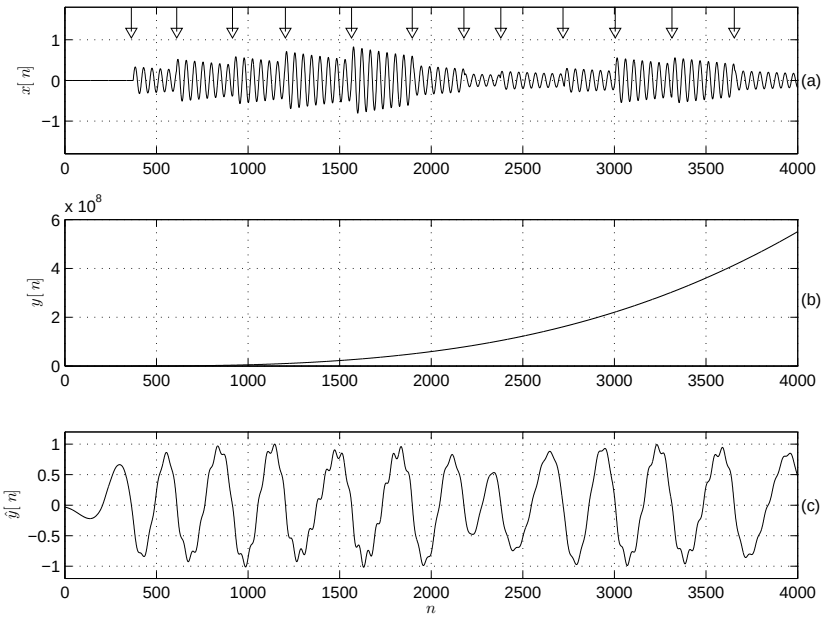


Fig. 4.2: (a) Response of a time varying system with unit impulses located at random locations. (b) Output of cascade of two 0-Hz resonators. (c) Filtered signal.

from the filtered signal. The filtered signal is obtained by passing the signal through a zero-frequency resonator, and then subtracting the local mean as in (4.4). Fig. 4.2 illustrates the signal and the filtered signal, along with the locations of the excitation impulses (shown by arrow marks).

The sequence of impulses can be viewed as edges in 1-D. However, an edge in 1-D signal is not an impulse, but a step function. The strength of the edge is given by the gradient of the step. In addition, the step could be positive or negative. The objective here is to determine the locations and strengths of the steps from a given noisy 1-D signal. It is also necessary to know whether the edge is due to positive step or negative step. Thus both positive zero-crossings (PZCs) and negative zero-crossings (NZCs), where the filtered signal changes its sign from positive to negative, are needed to mark the edge locations in a noisy 1-D signal [137].

Extraction of the locations of positive and negative step edges from a noisy 1-D signal using zero-frequency resonator is illustrated in Figs. 4.3 and 4.4, respectively. Fig. 4.3(b) shows a noisy version (overall SNR of 0 dB) of the 1-D signal shown in Fig. 4.3(a). The

negative and positive step edges of the original signal correspond to NZCs and PZCs of the filtered signal, respectively. All the NZCs of the filtered signal are shown by arrow marks in Fig. 4.3(c). There are some NZCs (see Fig. 4.3(c) for $n > 1000$ and $n \approx 160$) that do not correspond to the negative step edges. These can be eliminated using slope thresholding, since the slopes of these NZCS are relatively low. The NZCs retained after slope thresholding are shown by arrow marks in Fig. 4.3(d). One can observe that the slope of one NZC at $n = 375$ is high, but the NZC does not correspond to a negative step edge. Such NZCs are removed using the "zero-crossing sign correspondence principle" (see Appendix B). According to this principle, if a step edge has a positive first derivative, it must correspond to a PZC of the filtered signal. On the other hand, if the step edge has a negative first derivative, it must correspond to a NZC in the filtered signal. A zero-crossing violating this principle cannot correspond to a step edge. Thus, the resulting NZCs (after slope thresholding followed by application of the sign correspondence principle) correspond to the genuine step edges of the original signal, as shown in Fig. 4.3(e). Similar observations can be made from Fig. 4.4, where the elimination of false PZCs is demonstrated. The sign correspondence principle helps in suppressing the false zero-crossings which can not be removed using simple slope thresholding. The significance of various parameters involved in realizing the zero-frequency resonator will be discussed in the next section in the context of edge map extraction.

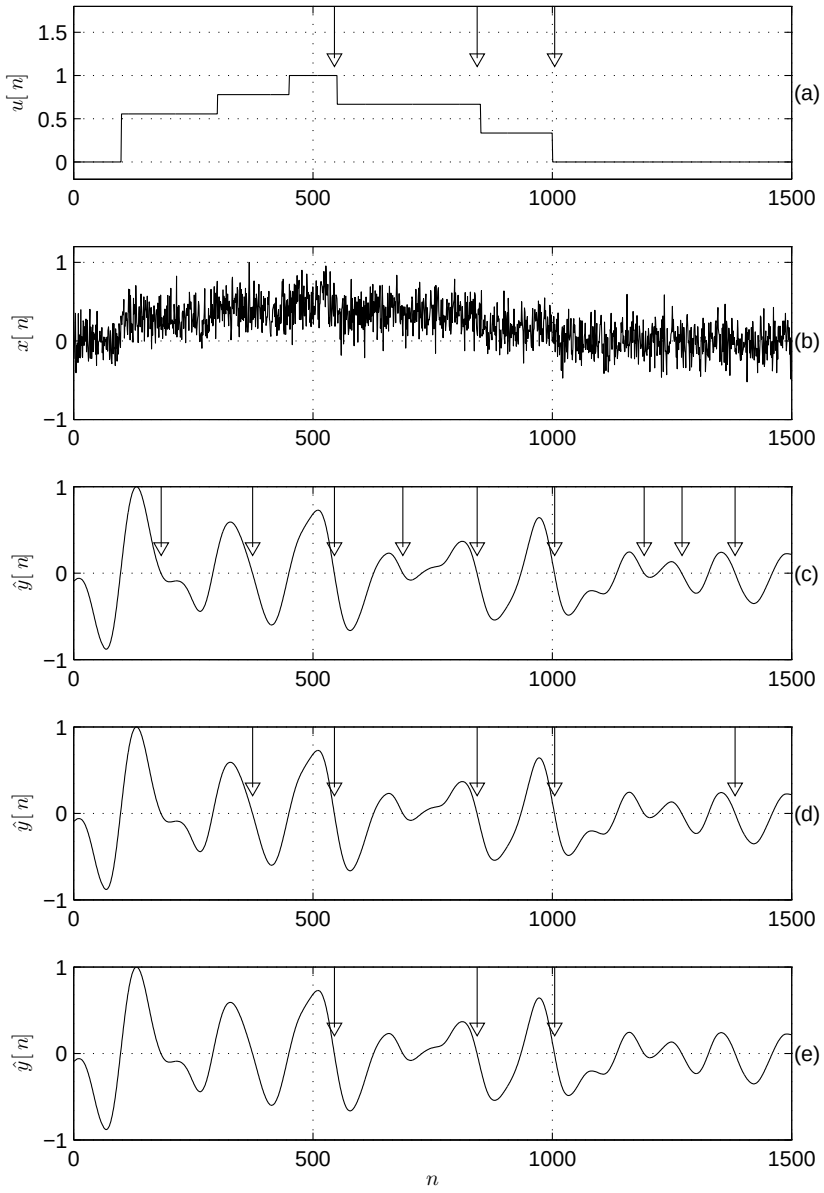


Fig. 4.3: Extraction of locations of the negative step edges in a noisy step signal. (a) Step signal. (b) Noisy step signal. The filtered signal ($\hat{y}[n]$) derived from $x[n]$ is shown in (c), (d) and (e). The downward arrow marks in (c) indicate all the NZCs. The NZCs retained after slope thresholding operation are shown in (d). The NZCs retained after slope thresholding and application of zero-crossing sign correspondence principle are shown in (e). The NZCs retained in (e), also shown in (a), correspond to the genuine negative step edges.

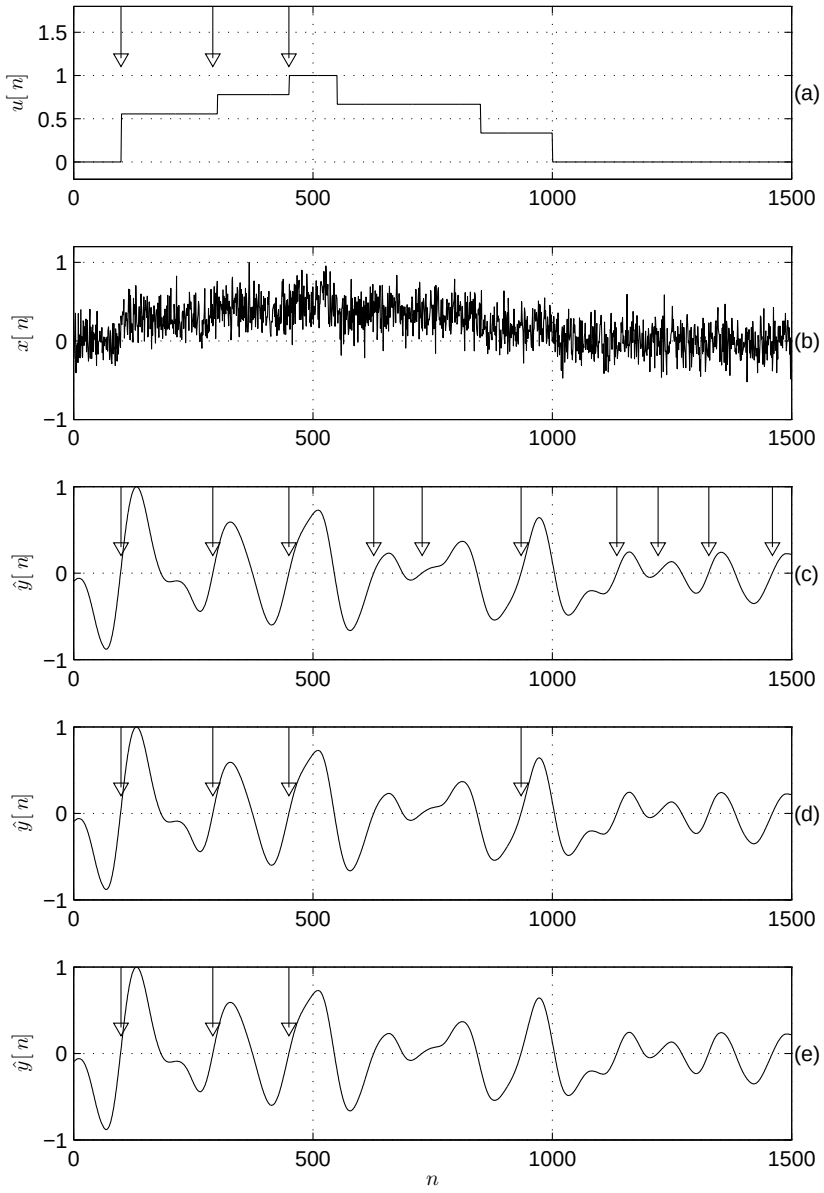


Fig. 4.4: Extraction of locations of the positive step edges in a noisy step signal. (a) Step signal. (b) Noisy step signal. The filtered signal ($\hat{y}[n]$) derived from $x[n]$ is shown in (c), (d) and (e). The downward arrow marks in (c) indicate all the PZCs. The PZCs retained after slope thresholding operation are shown in (d). The PZCs retained after slope thresholding and application of zero-crossing sign correspondence principle are shown in (e). The PZCs retained in (e), also shown in (a), correspond to the genuine positive step edges.

4.2 Zero-frequency resonator for 2-D signals

This section presents a method for extracting the edge information by processing an image using a zero-frequency resonator. The realization of the zero-frequency resonator using multiple integration of 2-D array of pixel values in the spatial domain is not straightforward. It can be realized approximately either in the spatial domain or in the Fourier domain. These methods are discussed in the following sections.

4.2.1 Spatial domain realization

In the spatial domain, the 1-D zero-frequency resonator is applied along each dimension of the image separately. Here, the dependency of pixel values along one dimension does not play a role while computing the response of the zero-frequency resonator along the other dimension. The dependency can be brought up to some extent by applying a smoothing operator along one dimension followed by application of the zero-frequency resonator along the orthogonal direction. The algorithm for realizing the zero-frequency resonator along the horizontal direction is given in Table 4.1. Fig. 4.5(c) shows the edge map obtained for a simulated image (shown in Fig. 4.5(a)), using the algorithm given in Table 4.1. It gives the vertical edges of the given image. Similarly, the horizontal edges of the image are obtained by changing the operations along row to column, and vice versa in the algorithm. The resulting edge map is shown in Fig. 4.5(e). The complete edge map is obtained by

Table 4.1: Algorithm to compute the vertical edges of an image using spatial domain realization of the zero-frequency filter.

1. Apply a smoothing operator (1-D Gaussian filter of very small variance (0.2-0.3))) along each column of the image.
2. Implement (4.3) and (4.4) along each row of the smoothed image obtained from Step 1. The resultant image is called filtered image along the horizontal direction.
3. Extract the locations of the edges along each row of the filtered image obtained in Step 2, using the zero-crossing sign correspondence principle.

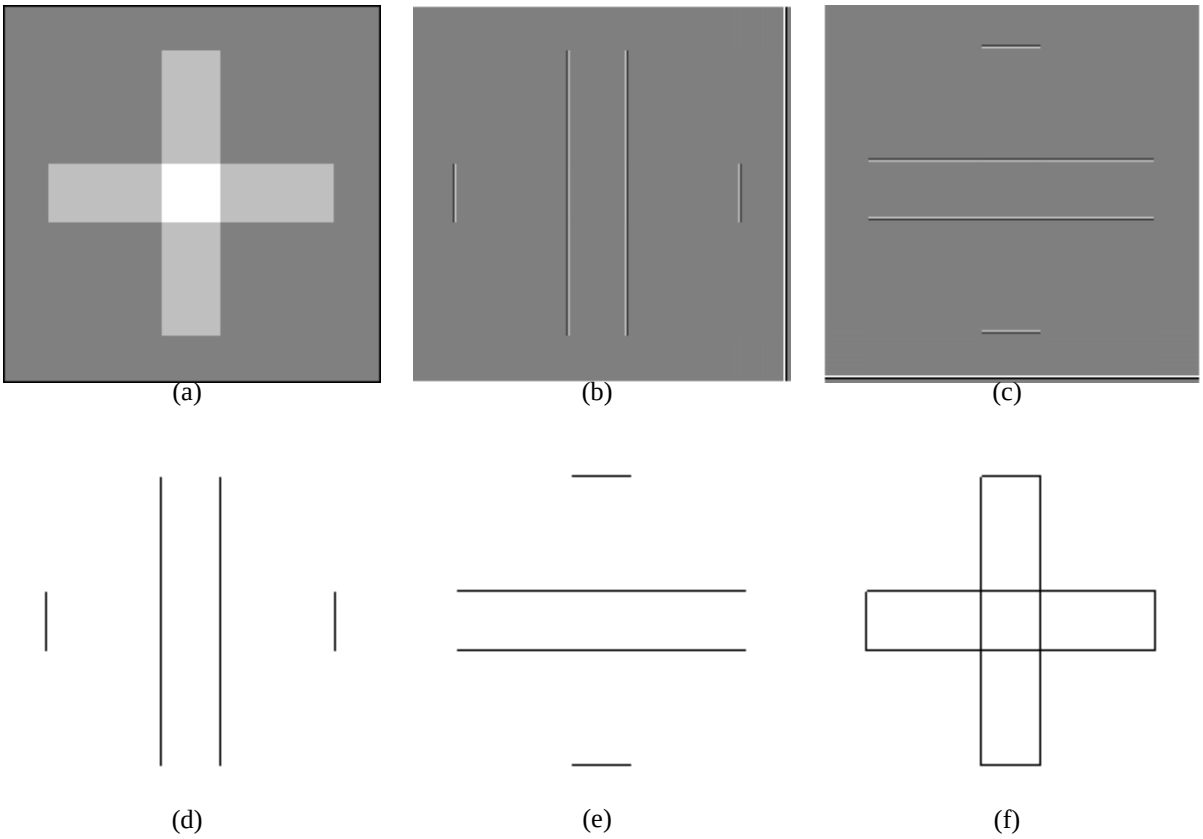


Fig. 4.5: (a) Gray level image. Filtered images along (b) horizontal, and (c) vertical directions obtained using spatial domain realization of the zero-frequency filter. (d) Edge map obtained from the filtered image in (b) using zero-crossing sign correspondence principle. (e) Edge map obtained from the filtered image in (c) using zero-crossing sign correspondence principle. (f) Edge map obtained using 'OR' operation on the partial edge map evidences shown in (d) and (e).

combining the two partial edge map evidences using OR operation, as shown in Fig. 4.5(f). Results for another example are shown in Fig. 4.6.

The output of a zero-frequency resonator is obtained by integrating the input signal twice. In practice, more than one zero-frequency resonator may be needed to suppress the noise present in the image. The magnitude response of a cascade of N zero-frequency resonators is given approximately by $\frac{1}{f^{2N}}$, as shown in Fig. 4.7 for different values of N, where f denotes the frequency variable. As N increases, the decay in the magnitude spectrum with respect to f becomes sharper. Thus the suppression of the high frequency components of the signal will be better for large N.

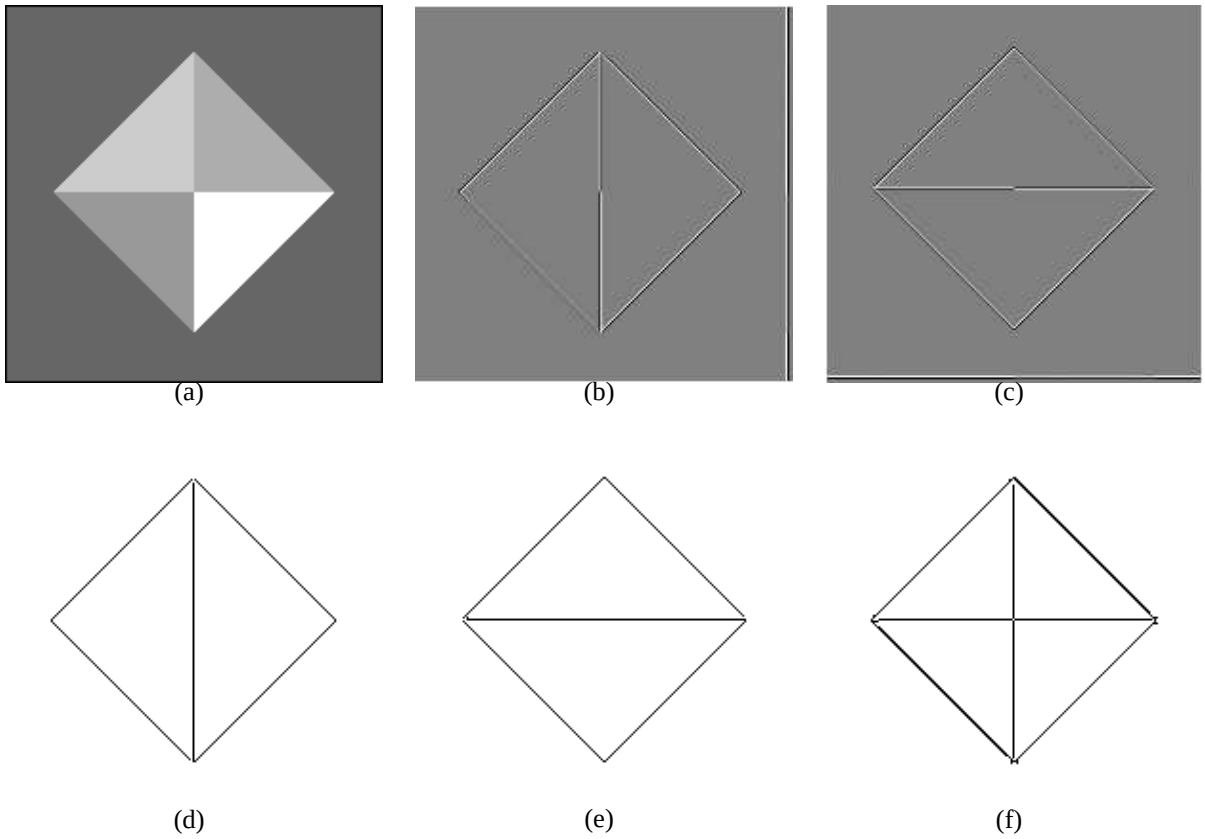


Fig. 4.6: (a) Gray level image. Filtered images along (b) horizontal, and (c) vertical directions obtained using spatial domain realization of the zero-frequency filter. (d) Edge map obtained from the filtered image in (b) using zero-crossing sign correspondence principle. (e) Edge map obtained from the filtered image in (c) using zero-crossing sign correspondence principle. (f) Edge map obtained using 'OR' operation on the partial edge map evidences shown in (d) and (e).

The length of the window ($w = 2N_1 + 1$) for computing the local mean for trend removal also plays a significant role in extracting the edge information. Fig. 4.8 shows the edge maps obtained using different values of w for a noisy image. For small values of w, closely spaced edges are emphasized in the edge map. This can be explained in the following way. The operation of subtracting the local mean is equivalent to the Laplacian operation (will be explained in the next section), which works as a high pass filter, because the impulse response of the Laplacian operator is proportional to f^2. So the magnitude response of the complete system (zero-frequency resonator followed by the local mean subtraction) will be a narrow bandpass filter. The center frequency of the narrow bandpass filter increases as w

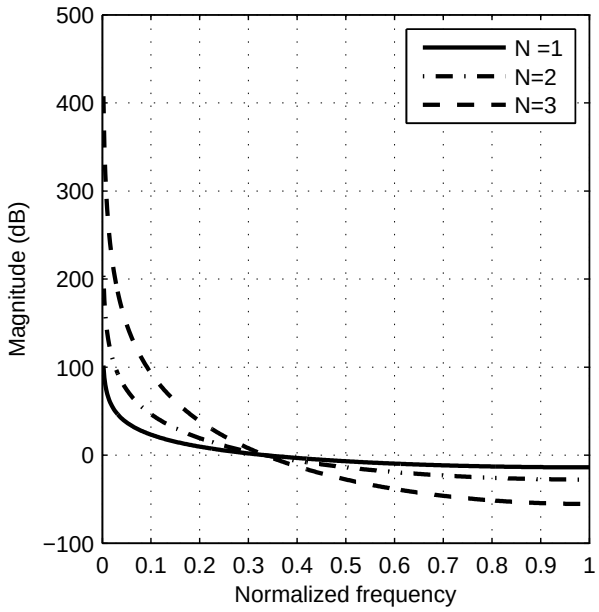


Fig. 4.7: Magnitude response of a cascade of *N* zero-frequency resonators.

decreases. Thus, for small values of *w* relatively closely spaced edges are emphasized in the edge map.

The information of the edges of an image may not appear in the edge map if the local means are subtracted only once. This is because the range of the values of the output of the zero-frequency resonator is very large. The local means may have to be subtracted more than once to derive the information of the edges. At the same time, excessive subtraction of the local means may emphasize spurious edges in the edge map, because the magnitude response corresponding to *M* times local mean subtractions (Laplacian operation) is proportional to f^{2M}. So for larger values of *M*, the emphasis on the higher frequency components will be more. The edge maps obtained for different values of *M* are shown in Fig. 4.9.

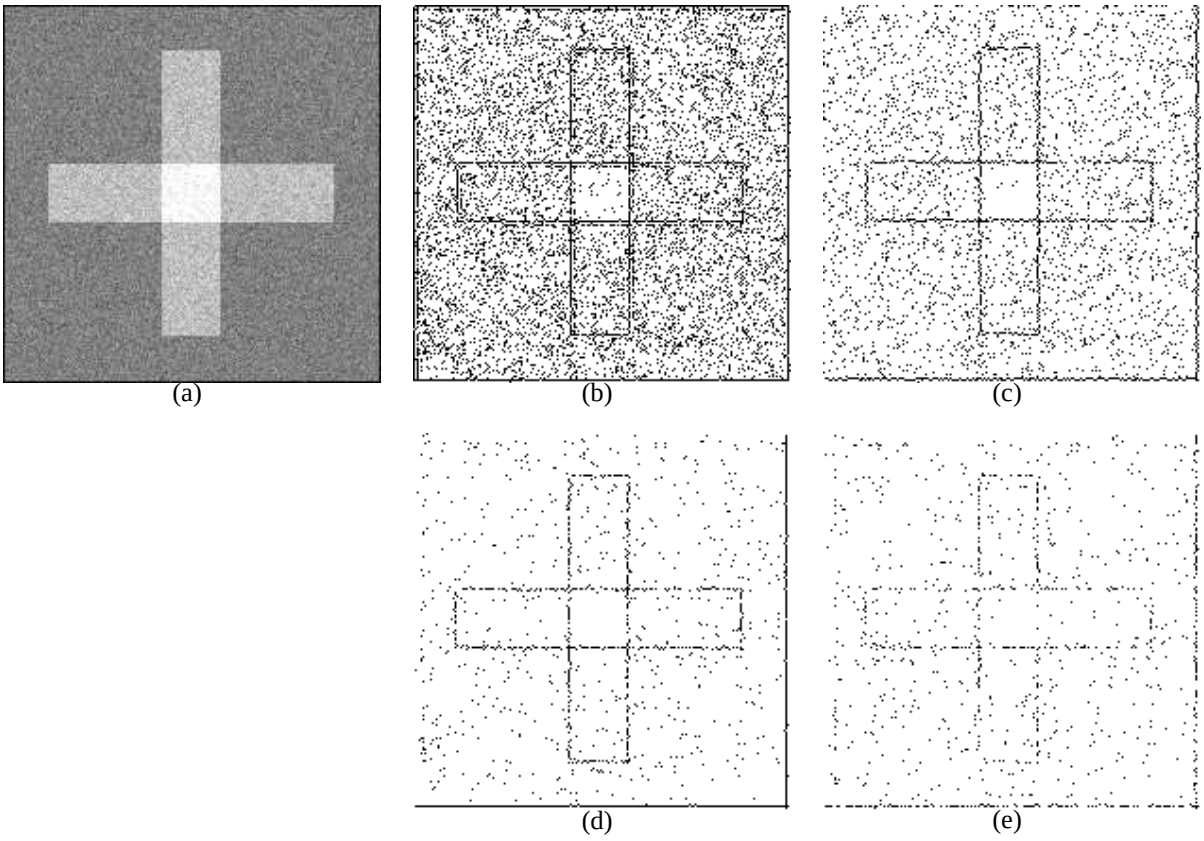


Fig. 4.8: (a) Noisy image. Edge maps obtained using spatial domain realization of the zero-frequency filter for $N = 3$, $M = 4$, and (b) $w = 3$, (c) $w = 5$, (d) $w = 7$ and (e) $w = 9$.

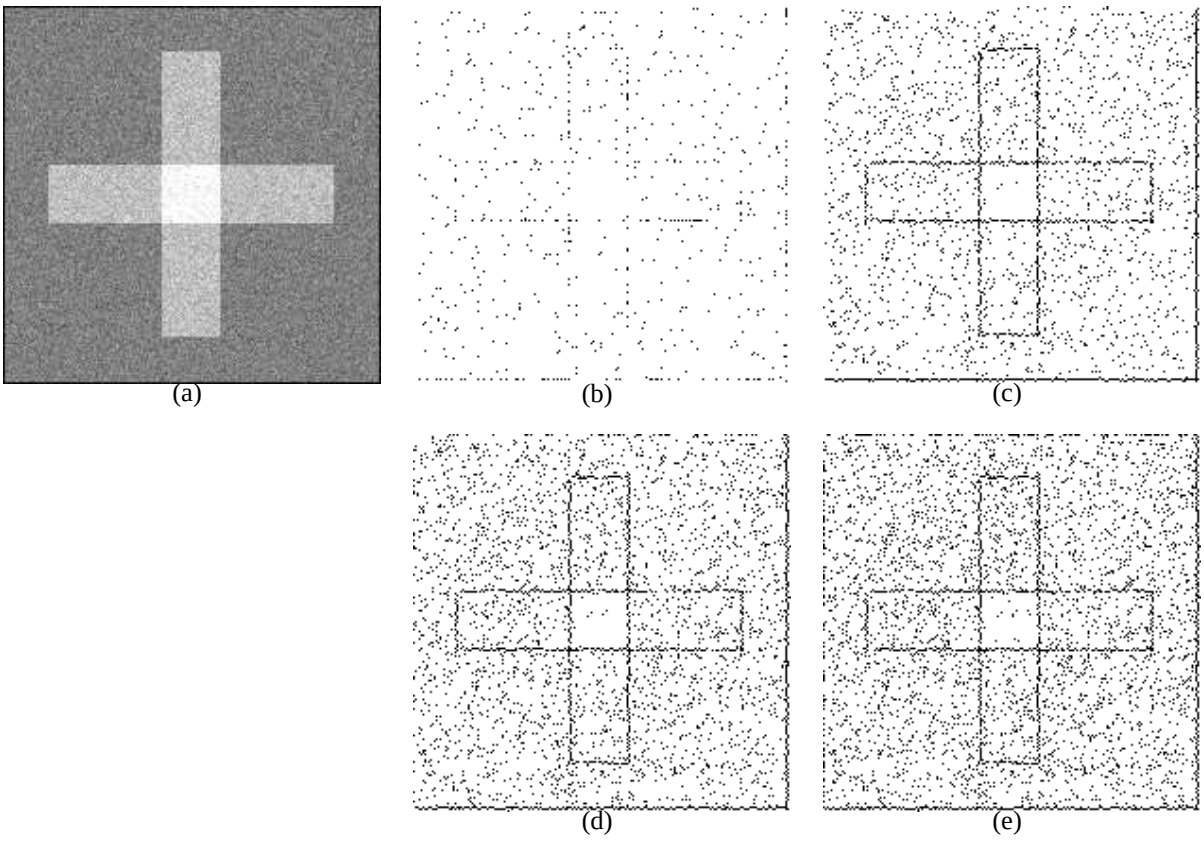


Fig. 4.9: (a) Noisy image. Edge maps obtained using spatial domain realization of the zero-frequency filter for $N = 3$, $w = 5$, and (b) $M = 2$, (c) $M = 3$, (d) $M = 4$ and (e) $M = 5$.

4.2.2 Fourier domain realization

In the spatial domain realization, the two dimensions of an image are considered separately while computing the filtered image, even though smoothing is performed along the orthogonal direction. In the Fourier domain approach, the zero-frequency filtering is performed on the 2-D DFT (discrete Fourier transform) of the entire image. Normally in the case of 1-D signals, the DFT needs to be computed over segments of short data record to capture the time varying characteristics of the system. In such a case the truncating window severely affects the components around the zero-frequency also. But implementation of the zero-frequency resonator in the time domain is straightforward in the case of 1-D, whereas conceptually it is more difficult in the case of 2-D.

The zero-frequency filter in the frequency domain corresponding to the single integration in the spatial domain is given by

$$W[f_1, f_2] = \frac{1}{\sqrt{f_1^2 + f_2^2}}, \tag{4.5}$$

where f_1 and f_2 denote the indices along the two spatial frequencies. Fig. 4.10 shows the magnitude response of the zero-frequency filter. Note that $W[0, 0]$ is set to zero. The 2-D

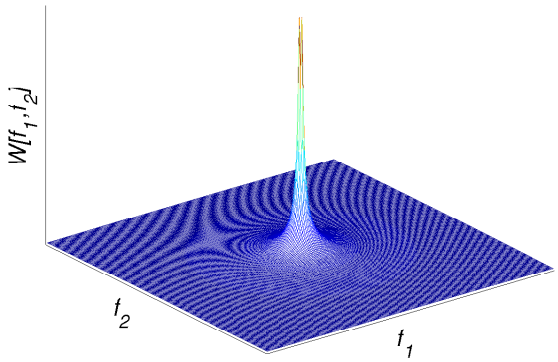


Fig. 4.10: Magnitude response of the 2-D zero-frequency filter.

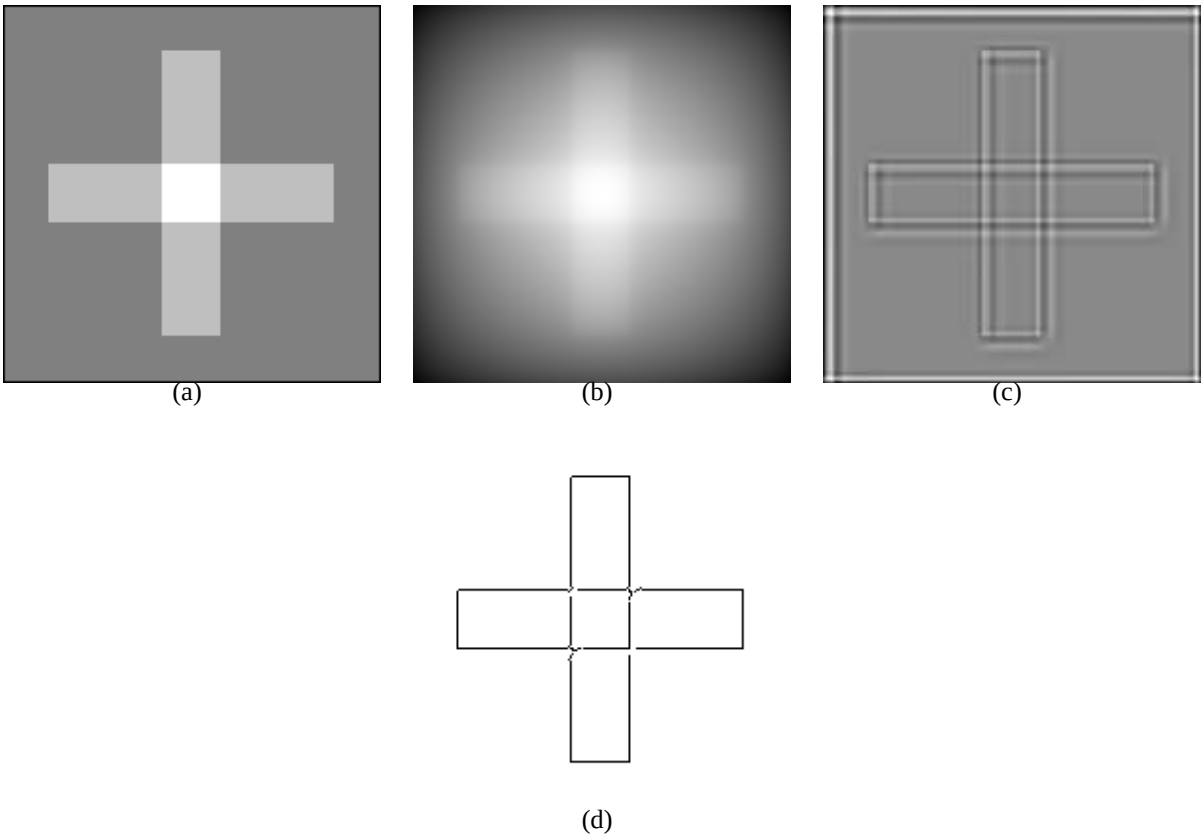


Fig. 4.11: (a) Gray level image. (b) Output of 2-D zero-frequency filter. (c) Filtered image obtained after local means are subtracted using $w = 5$. (d) Edge map obtained from the filtered image in (c) using zero-crossing sign correspondence principle.

DFT of an image is multiplied with $W[f_1, f_2]$, and the inverse discrete Fourier transform (IDFT) of the product gives the response

$$y[n_1, n_2] = \text{IDFT}\{X[f_1, f_2]W[f_1, f_2]\}, \tag{4.6}$$

where $X[f_1, f_2]$ denotes the 2-DFT of $x[n_1, n_2]$. Fig. 4.11(b) shows the output of the zero-frequency filter ($W[f_1, f_2]$) for an image shown in Fig. 4.11(a). The range of values in the output is very large in the spatial domain as expected, due to the effect of equivalent cumulative sum in the spatial domain. This is not exactly same as the cumulative sum in the spatial domain, since only the magnitude of the zero-frequency resonator is used in (4.5). Moreover, the value of $W[f_1, f_2]$ is set to zero at the origin $f_1 = f_2 = 0$.

As in the case of 1-D, it is necessary to remove the trend in the output to bring out the

features corresponding to the edge information. The trend is removed by subtracting the mean of $(2N_1+1)\times(2N_1+1)$ pixels around each pixel of the output. Thus the filtered image is given by

$$\hat{y}[n_1, n_2] = y[n_1, n_2] - \bar{y}[n_1, n_2], \tag{4.7}$$

where

$$\bar{y}[n_1, n_2] = \frac{1}{(2N_1+1)(2N_1+1)} \sum_{i=-N_1}^{N_1} \sum_{j=-N_1}^{N_1} y[n_1+i, n_2+j]. \tag{4.8}$$

The filtered image is shown in Fig. 4.11(c). In the filtered image each genuine edge could be either at a positive or at a negative zero-crossing, depending on whether the edge is from white to black pixels or vice versa in the original image. It is not possible to derive and interpret the zero-crossing locations in 2-D easily, as in the case of 1-D. The edge map from the filtered image is computed by finding the zero-crossings along each row and column, separately. Spurious edges are removed using the sign correspondence principle as explained earlier. The resulting edge map is shown in Fig. 4.11(d). Here, the estimation of the edge locations at the corners is not accurate. This is because the 2-D zero-crossings of the filtered image are interpreted as 1-D zero-crossings.

A cascade of N zero-frequency filters can be realized by multiplying the 2-D DFT of the input image with the filter $[W[f_1, f_2]]^N$. The effects of the parameters, such as the number of zero-frequency filters in the cascade, the number of times the local means are subtracted, and the length of the window used for computation of the local means are similar as in the case of the spatial domain realization of the zero-frequency resonator.

The following observations can be made in the realizations of the zero-frequency resonator in the Fourier domain and in the spatial domain. The Fourier domain method uses a finite impulse response (FIR) approximation of the zero-frequency resonator, which may introduce artifacts due to finite window length [138]. In addition, the phase information of the zero-frequency filter is not used in the Fourier domain approach, which is not the case in the spatial domain realization. The spatial domain realization does not exploit the relations between pixels along the rows and columns of an image. Fig. 4.12 shows edge

maps obtained for noisy synthetic images using both the spatial domain and Fourier domain approaches. The spatial domain realization gives perceptually better results as compared to the Fourier domain realization.

4.3 Comparison with Laplacian of Gaussian (LOG)

The Laplacian of Gaussian (LOG) operation was proposed by Marr and Hildreth [14] to detect the edges in an image. According to the principle of LOG, zero-crossings of the second derivative (Laplacian) of an image give the locations of the edges. Since the derivative operation is sensitive to noise, the effect of noise is reduced by smoothing the image before applying the derivative operator. Smoothing is performed using a Gaussian filter, given by

$$g[n_1, n_2] = \frac{1}{2\pi\sigma^2} \exp\left\{\frac{-(n_1^2 + n_2^2)}{2\sigma^2}\right\}, \tag{4.9}$$

where σ denotes the variance. A high value of σ is chosen for noisy image, but this leads to smearing of the edges also. The frequency response of the Gaussian filter is given by

$$G[f_1, f_2] = 2\pi\sigma^2 \exp\left\{\frac{-\sigma^2(f_1^2 + f_2^2)}{2}\right\}. \tag{4.10}$$

The response of the Gaussian filter for a given image $x[n_1, n_2]$ is computed as

$$y[n_1, n_2] \quad = \quad \text{IDFT}\{X[f_1, f_2]G[f_1, f_2]\}. \tag{4.11}$$

The Laplacian of the resultant output is computed as [139]

$$\nabla^2[y[n_1, n_2]] = \nabla^2[x[n_1, n_2] \otimes g[n_1, n_2]], \tag{4.12}$$

where $\otimes$ and ∇^2 denote the convolution and Laplacian operations, respectively. The computational complexity of the ∇^2 operation is high because it is image dependent [139]. This issue is addressed in two ways: In the first method the linearity and shift invariance proper-

ties of convolution and Laplacian operations are exploited [139]. That is

$$\nabla^2[x[n_1, n_2] \otimes g[n_1, n_2]] = \quad x[n_1, n_2] \otimes \nabla^2 g[n_1, n_2]. \tag{4.13}$$

Therefore, Gaussian filtering followed by Laplacian is same as filtering with Laplacian of Gaussian. The right hand side of (4.13) is more efficient, since $\nabla^2 g[n_1, n_2]$ can be generated in advance, as it is independent of the image. In the second method [140], an approximate mask of the Laplacian operator is derived, as shown in Fig. 4.13 (c), and this mask is used to compute the Laplacian of the Gaussian smoothed image. In fact, this mask can be derived from two parts, one is an all-pass filter and the other is a mean filter, whose masks are shown in Figs. 4.13 (a) and (b), respectively. Based on this decomposition, the Laplacian of an image can also be obtained by subtracting the outputs obtained after applying the masks shown in Figs. 4.13 (a) and (b) on the image. This operation is equivalent to the local mean subtraction (4.7), which is proposed to extract the edge information from the output of the zero-frequency filter.

One can observe from (4.6) and (4.11) that in the proposed approach, the smoothing of image is performed using the zero-frequency filter, instead of using the Gaussian filter as in the case of LOG. The Laplacian operation is common in both the approaches. Fig. 4.12 shows the edge map obtained using the LOG operation, and the edge maps obtained using the two approaches for realizing the zero-frequency resonator. Here, the parameters for each approach are chosen in such a way that the resultant edge map is perceptually good. Suppression of noise and estimation of the edge locations are better in the proposed approaches as compared to the LOG operation. This is due to the fact that the decay in the magnitude spectrum of the zero-frequency filter with respect to f is sharper as compared to that in the case of Gaussian filter, as illustrated in Fig. 4.14. In this example the magnitude spectrum of the Gaussian filter is computed using $\sigma = 3$. It is possible to obtain the spectrum of the Gaussian filter similar to that of the zero-frequency filter by choosing a very large value of σ. In practice large values of σ result in large error in the edge localization.

The proposed approaches are also used to derive the edge maps for several real world images under noisy condition. Additive white Gaussian noise with zero mean and variance

of .002 is used to generate the noisy images. Figs. 4.15 and 4.16 show the results for the "peppers" and "car" images, respectively. Estimation of the edges due to the peduncles and boundaries of the peppers is better in the spatial domain realization of the zero-frequency resonator for both the original and the noisy images. Similar observations can be made from the estimated edges due to the mud guard and rim of a tyre, as shown in Fig. 4.16(b). Though the estimation of edges is similar in the both the Fourier domain realization of the zero-frequency filter and the Gaussian filter, the former approach suppresses the noise better than the latter. The results also show that the edge localization is poor in the case of the edge map obtained using the LOG operation as compared to the proposed approaches. This is due to the fact that the error in the edge localization due to Gaussian filter increases as the standard deviation of the Gaussian increases. On the other hand, even if we emphasize the components of the image near the zero-frequency, the edges are still detected accurately while suppressing the noise. This is because the edges of an image have characteristics of an impulse, whose effect is spread uniformly over all frequencies of the spectrum.

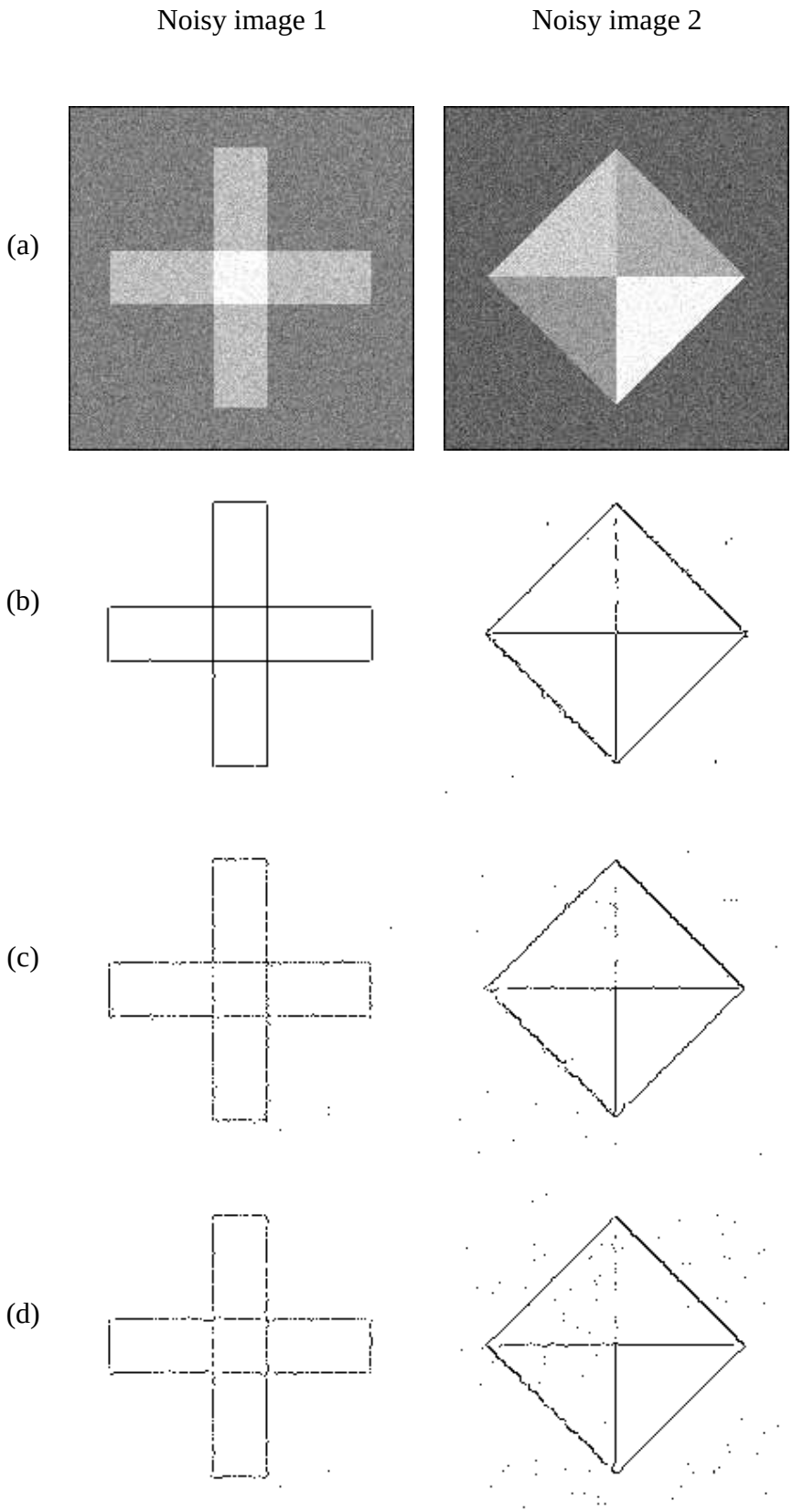


Fig. 4.12: (a) Noisy images. Edge maps obtained using (b) spatial, and (c) Fourier domain realization of the zero-frequency filter. (d) Edge maps obtained using LOG operation.

0	0	0
0	1	0
0	0	0

(a)

−

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

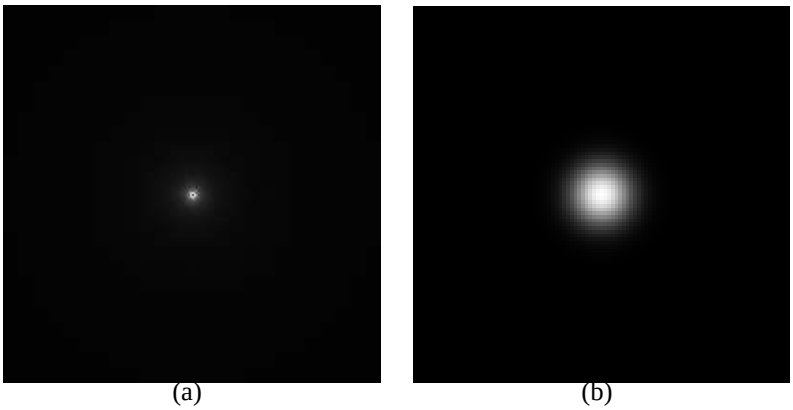
(b)

=

−1/9	−1/9	−1/9
−1/9	8/9	−1/9
−1/9	−1/9	−1/9

(c)

Fig. 4.13: Mask for Laplacian operation



(a) (b)

Fig. 4.14: Magnitude response of (a) 2-D zero-frequency filter and (b) Gaussian filter. The large and small values of the magnitude spectrum are denoted as white and back in the image, respectively.

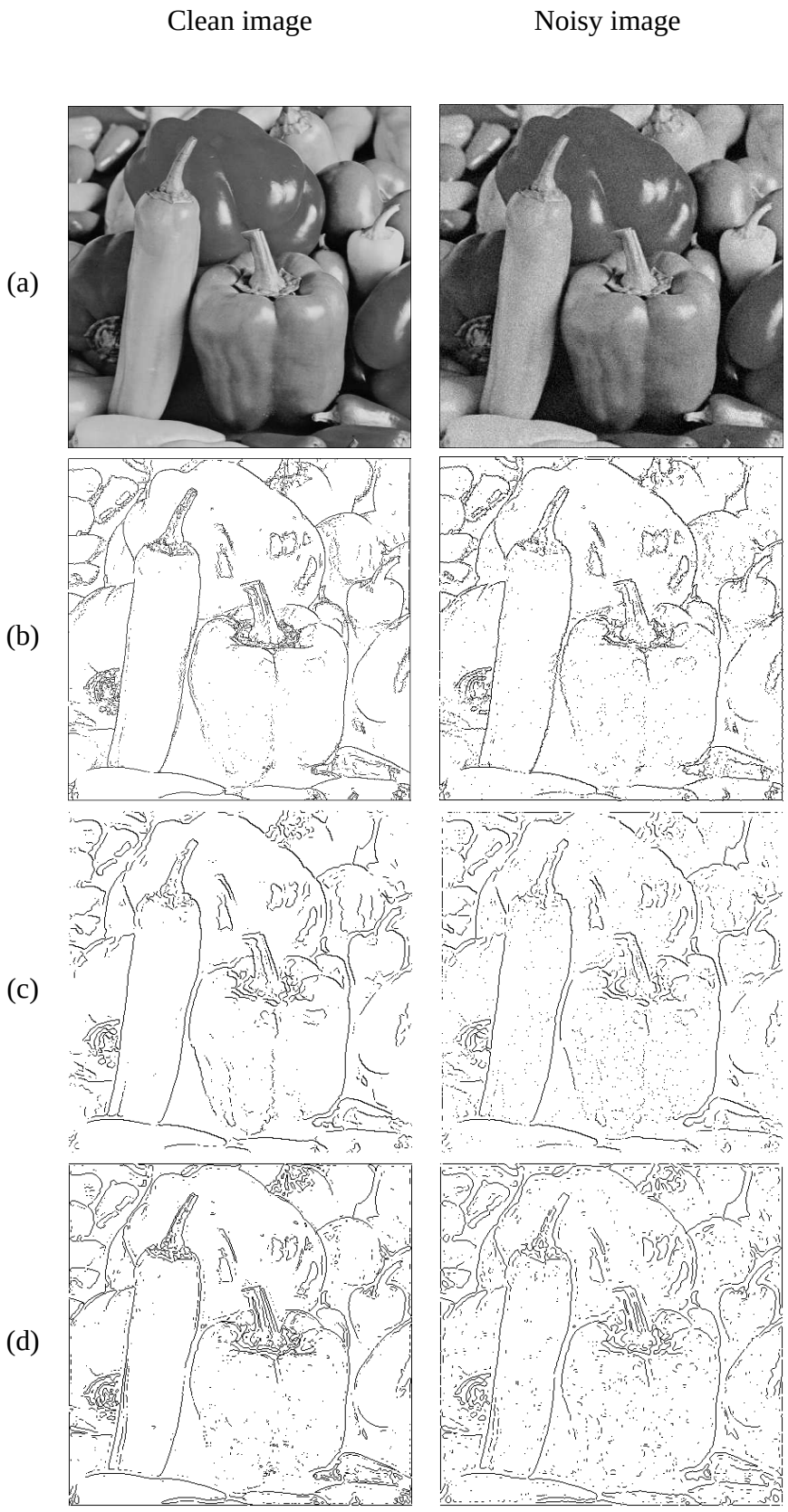


Fig. 4.15: (a) Original and noisy images of “peppers”. Edge maps obtained using (b) spatial, and (c) Fourier domain realization of the zero-frequency filter. (d) Edge maps obtained using LOG operation.

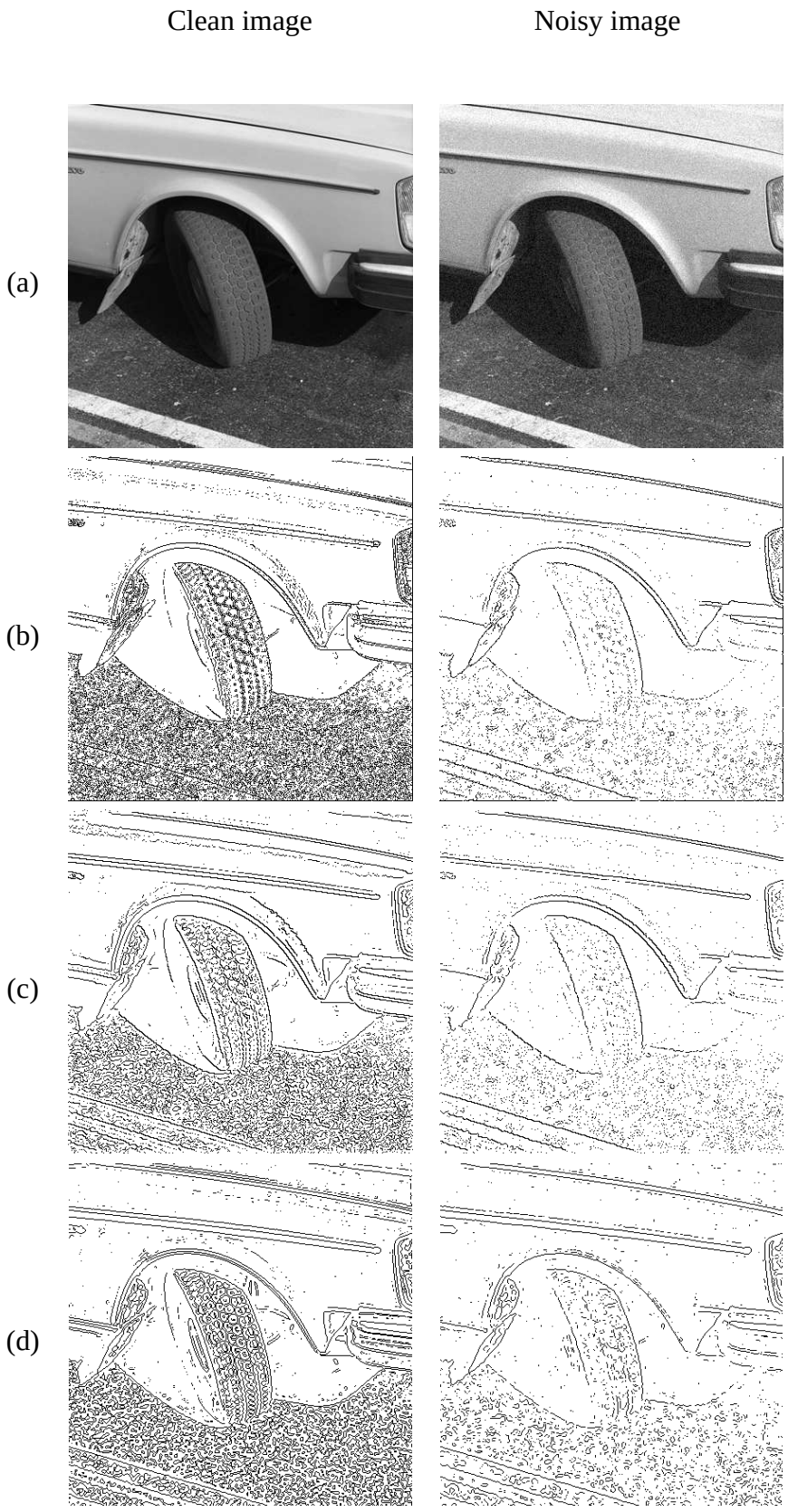


Fig. 4.16: (a) Original and noisy images of "car". Edge maps obtained using (b) spatial, and (c) Fourier domain realization of the zero-frequency filter. (d) Edge maps obtained using LOG operation.

4.4 Laplacian of smoothed image as a representation of face image

Here we propose the use of filtered images obtained using the zero-frequency filter and the LOG operation for representation of face images. We show that these representations help in reducing the effects of illumination in face recognition. Fig. 4.17 shows the filtered images obtained using the Gaussian filter, and the two methods of realizing the zero-frequency filter for 2-D signals. In the spatial domain method, two filtered images are derived, one along the horizontal direction and the other along the vertical direction. These filtered images capture the vertical and horizontal edge information of the face image as shown in Figs. 4.17(b) and (c), respectively. In the case of Fourier domain realization of the zero-frequency filter or in the LOG operation, only one filtered image is obtained for a given face image, as shown in Figs. 4.17(d) and (e). The filtered images obtained using different smoothing filters can not be compared perceptually, since the information of the edges is present around the zero-crossings of the filtered images.

For matching two filtered images, it is necessary to have the images within the same dynamic range, at the same time preserving the information of the edges present around the zero-crossings. Note that normalization of the filtered image using a linear transformation between the maximum and minimum values shifts the level of the zero-crossings, and hence the resultant normalized filtered images cannot be used for matching. A nonlinear transformation is proposed which truncates the positive and negative values of the filtered image beyond certain threshold values, and which has a linear transformation within the range of the positive and negative threshold values. The nonlinear transformation $T(x)$ is shown in Fig. 4.18. The value of the threshold amplitude α for the transformation function is derived from the filtered image $\hat{y}[n_1, n_2]$ as follows:

$$\alpha = m \times \frac{p_c}{100},$$

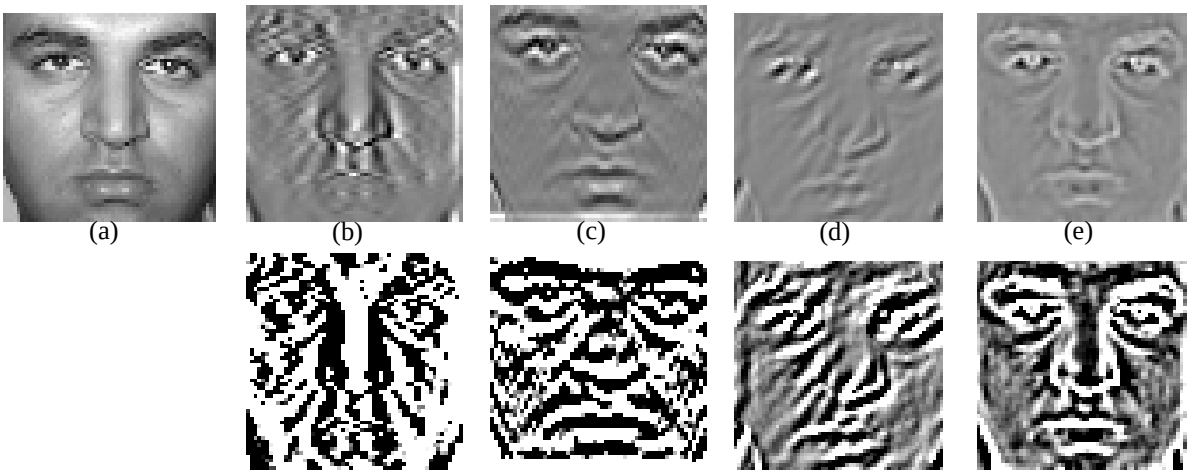


Fig. 4.17: (a) Gray level image. Filtered images along (b) horizontal and (c) vertical directions obtained using spatial domain realization of the zero-frequency filter. (d) Filtered image obtained using Fourier domain realization of the zero-frequency filter. (e) Filtered image obtained using Gaussian filter. The second row gives the corresponding nonlinearly transformed filtered images.

where

$$m = \min\{|\max(\hat{y}[n_1, n_2])|, |(\min(\hat{y}[n_1, n_2]))|\}. \tag{4.14}$$

The value of the parameter p_c is chosen between 0 and 100. Note that that the locations of the zero-crossings are preserved. For very small values of p_c (close to zero), the transformation function gives only two outputs -1 and +1, which is not desirable, as it results in a binary image. Experimentally we have observed that a value of p_c in the range of 70 to 90 is reasonable for face recognition. The transformed filtered images obtained from the zero-frequency filter and the Gaussian filter are shown in the second row of Fig. 4.17 for $p_c = 80$.

4.5 Locality problem: Smearing of the filtered images

One of the problems in using any representation based on the edge information for face recognition is the *locality problem*. Due to this, matching of filtered images results in poor

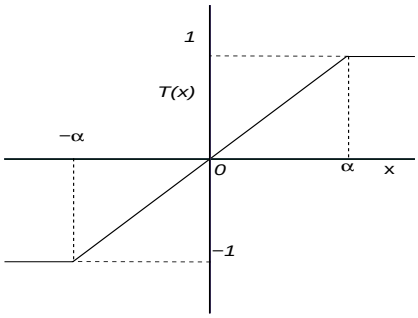


Fig. 4.18: Nonlinear transfer function to preserve the zero-crossings of filtered images.

score under intra-class variation of the face image. One way to improve the performance is to smear the edge information and then perform matching. But excessive smearing of edges may result in loss of information, leading to poor performance for inter-class variation of the face image.

Smearing is affected by removing the coefficients of the eigenvectors corresponding to low variances. Eigenanalysis using two partial filtered images, obtained with the spatial domain realization of the zero-frequency filter, will be considered first. Let D_i denote the set of training face images for the i^{th} person. For all the available training face images, the two partial filtered images are computed. Let $\mathbf{x}^v$ and $\mathbf{x}^h$ denote the vector representations of the filtered images along the vertical and horizontal directions for the image $x[n_1, n_2]$, respectively. The eigenvector matrix $\mathbf{\Psi}^v \in \mathbb{R}^{p\times p} = [\psi_1^v, \ldots, \psi_p^v]$, and the diagonal eigenvalue matrix $\mathbf{\Lambda}^v \in \mathbb{R}^{p\times p} = \text{diag}\{\lambda_1^v, \lambda_2^v, \ldots, \lambda_p^v\}$ with the diagonal elements (eigenvalues) arranged in decreasing order ($\lambda_1^v \geq \lambda_2^v \ldots \geq \lambda_p^v$), are obtained using $\mathbf{x}^v$ representations of the training face images. Similarly $\mathbf{\Psi}^h \in \mathbb{R}^{p\times p}$ and $\mathbf{\Lambda}^h \in \mathbb{R}^{p\times p}$ are obtained using $\mathbf{x}^h$ representations of the training face images. Here $p = R \times C$, where R and C are the number of rows and columns of the given face image, respectively. A face image $x[n_1, n_2]$ can be represented in terms of the eigenvectors as [91]

$$\begin{aligned} \mathbf{a}_{\mathbf{x}}^{h,p} &= (\mathbf{\Psi}^h)^t \mathbf{x}^h, \\ \mathbf{a}_{\mathbf{x}}^{v,p} &= (\mathbf{\Psi}^v)^t \mathbf{x}^v, \end{aligned} \tag{4.15}$$

where $\mathbf{a}_{\mathbf{x}}^{h,p}$ and $\mathbf{a}_{\mathbf{x}}^{v,p}$ are the projected coefficients. Here, t denotes the transpose operation. The filtered face image can be reconstructed using the projected coefficients $\mathbf{a}_{\mathbf{x}}^{v,p}$ as follows:

$$\begin{aligned} \mathbf{x}^v &= \mathbf{a}_{\mathbf{x}}^{v,p} \boldsymbol{\Psi}^v \\ &= \sum_{i=1}^{p} a_{\mathbf{x},i}^{v,p} \psi_i^v, \end{aligned} \tag{4.16}$$

where $\mathbf{a}_{\mathbf{x}}^{v,p} = [a_{\mathbf{x},1}^{v,p}, \ldots, a_{\mathbf{x},p}^{v,p}]$. The eigenvectors associated with the largest eigenvalues are also referred to as eigenfaces [141]. Removing the last few projected coefficients in (4.16) leads to smearing of edges.

The significance of smearing of edges of the filtered face image in template matching can be observed using the scatter plot of distances ($d_{i,\mathbf{y}}^{h,n}$ and $d_{i,\mathbf{y}}^{v,n}$) for a person's face image. The distances are defined for a given test face image ($y[n_1, n_2]$) with respect to the reference face images of the i^{th} person using the first n projected coefficients as follows:

$$\begin{aligned} d_{i,\mathbf{y}}^{h,n} &= \min_{\mathbf{x} \in D_i} \|\mathbf{a}_{\mathbf{y}}^{h,n} - \mathbf{a}_{\mathbf{x}}^{h,n}\|_2, \\ d_{i,\mathbf{y}}^{v,n} &= \min_{\mathbf{x} \in D_i} \|\mathbf{a}_{\mathbf{y}}^{v,n} - \mathbf{a}_{\mathbf{x}}^{v,n}\|_2. \end{aligned} \tag{4.17}$$

Ideally, the distances should be low if the test face image is of true class, and high for a false class face image. The maximum value of n is $(R \times C) = 2500$, as each face image was rescaled to 50×50 pixels in our experiments. Fig. 4.19 shows the scatter plots for a person for different values of n (number of projected coefficients), using one face image (L^{91}) as training (reference) image for each person in the illumination variation set of the FacePix face database. The remaining 180 face images of the person form the test samples of the true class, and the matching distances are shown by the symbol '$\diamond$' (diamond) in the scatter plots. The number of test samples of the false class is $29 \times 181 = 5249$, and the corresponding distances are shown by the symbol '$\cdot$' (dot) in the the scatter plots. Ideally, the points due to true class should be close to the origin, and the points due to false class should be farther from the origin. The plots show that the points are more dense for false class, and most of the points are far from the origin in Fig. 4.19(d) for both true and false classes as compared to Figs. 4.19(a)-(c). Since the *locality problem* for the true class images

is severe when all the projected coefficients (n = 2500) are employed for matching, the chances of matching of any test face image with the reference face images is less. Removing a few coefficients corresponding to the small eigenvalues leads to smearing of edges of the filtered face image, and thus reduces the locality problem. But excessive smearing reduces discrimination between true and false class images as can be seen in Fig. 4.19(a), where the true class points are closer to the origin, due to better matching, but at the same time the false class points are more spread out and are brought closer to the origin. Thus there is a trade-off in the smearing of the edges, which affects the performance of face recognition.

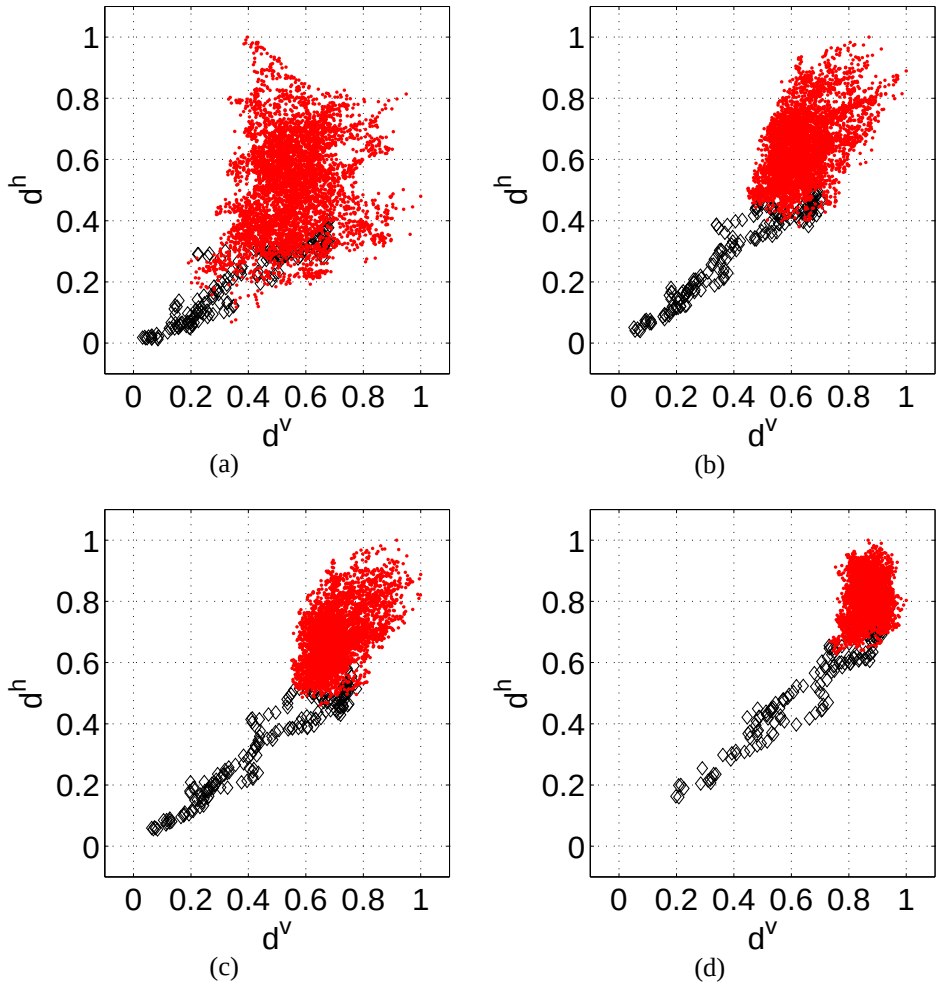


Fig. 4.19: Scatter plots of distances for a person's face image using vertical and horizontal filtered face images and I^{91} as training (reference) face image. (a) n = 5, (b) n = 20, (c) n = 30 and (d) n = 2500. Here n denotes the number of eigenvectors used to compute the similarity measures ($d^{h,n}_{i,\mathbf{y}}$ and $d^{v,n}_{i,\mathbf{y}}$).

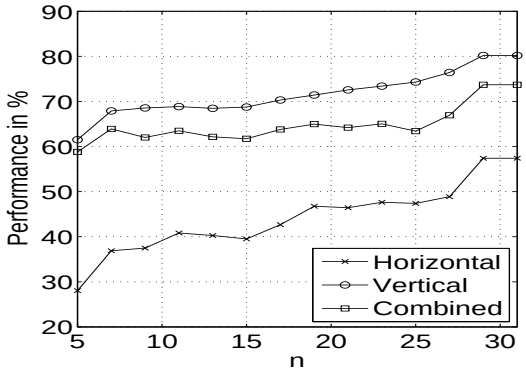


Fig. 4.20: Comparison of the performance of vertical and horizontal filtered images for different number (n) of eigenvectors and using I^{91} as training (reference) face image.

The identity (i^*) of a given test face image $y[n_1, n_2]$ using the first n projected coefficients is computed as follows:

$$i^*(n) = \arg\min_i [(d_{i,\mathbf{y}}^{h,n})^2 + (d_{i,\mathbf{y}}^{v,n})^2]^{1/2}. \tag{4.18}$$

The performance ($\eta(n) = \frac{\text{Number of Correctly identified face images}}{\text{Total number of available test face images}} \times 100$) for different values of n is shown in Fig. 4.20. The performances obtained with the two partial evidences ($d_{i,\mathbf{y}}^{h,n}$ and $d_{i,\mathbf{y}}^{v,n}$) separately are also shown in the figure. The performance using the filtered image along the vertical direction is better than the performance using the filtered image along the horizontal direction. This is because the edges along the vertical direction seem to capture more discriminative information of face images. Moreover, the performance does not improve by combining the two evidences, where they are combined by simple norm-2.

We can observe from (4.18) that the identification performance (η) depends on the distance values of the test face image with the reference face images of all the persons. Note that, the scatter plot shows similarity scores of the test face image with respect to reference face images of only one person. Thus the scatter plots are not a direct indication of the identification performance, but they are used to examine the behavior of the true and false class samples.

4.5.1 Weighted eigenvectors

The eigenvalues indicate the spread of the training face images along the corresponding eigenvectors. The eigenvector corresponding to the largest eigenvalue gives information common to all the training samples. On the other hand, the noise or unwanted information is present in the eigenvectors corresponding to small eigenvalues. The unique information of a face image can be highlighted by removing the common and noisy components of the given filtered face image. Following this idea, (4.16) can be divided into three terms as

$$\mathbf{x}^v = \sum_{i=1}^{l_1} a_{\mathbf{x},i}^{v,p} \psi_i^v + \sum_{i=l_1+1}^{l_2} a_{\mathbf{x},i}^{v,p} \psi_i^v + \sum_{i=l_2+1}^{p} a_{\mathbf{x},i}^{v,p} \psi_i^v, \tag{4.19}$$

where l_1 and l_2 are indices. The coefficients $\{a_{\mathbf{x},i}^{v,p}, \quad i = 1, \ldots, l_1\}$ defining the first term, correspond to the information which is common to most of the training face images. The second term corresponds to the unique information present in the given face image, and is defined by the coefficients $\{a_{\mathbf{x},i}^{v,p}, \quad i = l_1 + 1, \ldots, l_2\}$. The third term corresponds mostly to noise. The second term will be more useful for discrimination as compared to the first and third terms. These characteristics were observed with the gray level representation of the face image also [142, 143]. The performance of face recognition can be improved by assigning more weight to the coefficients of the second term as compared to the coefficients of the first and third terms during matching. One of the issues that need to be addressed is, how to decide the values of the indices l_1 and l_2 that divide the given representation of the face image into the three terms as discussed above. The indices l_1 and l_2 may be specific to a given face image.

In this work, the inverses of the eigenvalues are used as weights to the coefficients in the matching task. This will give more weightage to those projected coefficients which correspond more to the unique information of a given face image. Following these ideas, the weighted distance measure is given by

$$\begin{aligned} d_{i,\mathbf{y}}^{h,n} &= \min_{\mathbf{x} \in D_i} \|\mathbf{a}_{\mathbf{y}}^{h,n} \mathbf{W}_n^h - \mathbf{a}_{\mathbf{x}}^{h,n} \mathbf{W}_n^h\|_2 \\ d_{i,\mathbf{y}}^{v,n} &= \min_{\mathbf{x} \in D_i} \|\mathbf{a}_{\mathbf{y}}^{v,n} \mathbf{W}_n^v - \mathbf{a}_{\mathbf{x}}^{v,n} \mathbf{W}_n^v\|_2, \end{aligned} \tag{4.20}$$

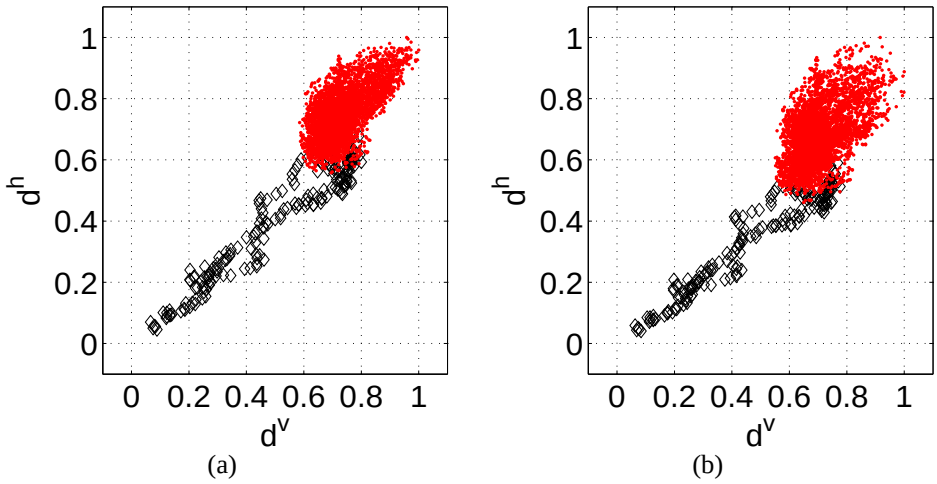


Fig. 4.21: Scatter plots of distances for a person's face image using the horizontal and vertical filtered images, I^{91} as reference face image and 30 projected coefficients. (a) With weighted eigenvectors. (b) Without weighted eigenvectors.

where $\mathbf{W}_n^v = \text{diag}\{\frac{1}{\sqrt{\lambda_1^v}}, \ldots, \frac{1}{\sqrt{\lambda_n^v}}\}$ and $\mathbf{W}_n^h = \text{diag}\{\frac{1}{\sqrt{\lambda_1^h}}, \ldots, \frac{1}{\sqrt{\lambda_n^h}}\}$ are diagonal matrices. The advantage of using the weighted distance measure can be seen from the scatter plots shown in Fig. 4.21. The points due to false class samples are more dense and farther from the origin in Fig. 4.21(a) as compared to Fig. 4.21(b). This is due to the fact that the coefficients which have discriminative information get more weight than the coefficients which have common or less discriminative information, and hence the performance is improved by ≈4% as shown in Fig. 4.22. In the computation of the weights, to avoid giving large weightage to the coefficients which have less discriminative information, a small positive value is assigned to the corresponding eigenvalues.

4.5.2 Effect of choice of parameters for face recognition

It is to be noted that the parameters involved in realizing the zero-frequency filter affect the recognition performance to some extent. Hence, it is necessary to study the effect of such parameters. The information about the edges of the face image is captured around the zero-crossings of the filtered image. Closely spaced edges generally correspond to noise in the filtered image. The edges with large spacing correspond the global information of the face

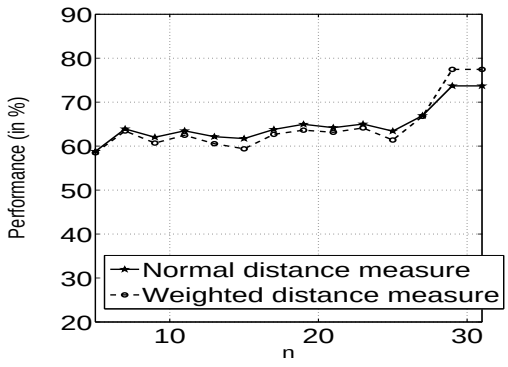


Fig. 4.22: Comparison of the performance under weighted and normal similarity measures obtained using the horizontal and vertical filtered face images for different number (n) of eigenvectors.

image that is similar across the face images of different persons. The importance of edges of small or large spacing in matching can be controlled by the parameters N, w and M used in realizing the zero-frequency filter. Fig. 4.23 shows the performance of face recognition obtained using (4.18) for different values of the parameters and one training (reference) face image (I^{91}) of each person. The following observations can be made:

1. The complete system of smoothing (low pass filter) followed by the local mean subtraction/Laplacian operation (high pass filter) is a narrow bandpass filter near the zero-frequency. The center frequency of the bandpass filter increases as w decreases. Thus, closely spaced edges are highlighted in the filtered image for smaller values of w.

2. When more than one zero-frequency filter is employed to suppress the noise present in the image, the range of the resultant smoothed image becomes very large. It is necessary to remove the local mean more than once to bring out the edge information. But excessive subtraction of the local mean may emphasize noise (closely spaced edges) also.

3. For a fixed N and small window ($w = 3$), as M increases, the performance decreases, because the relatively closely spaced edges get highlighted in the filtered image.

4. For large values of the window length (w = 7 to 9) and fixed N, as M increases the

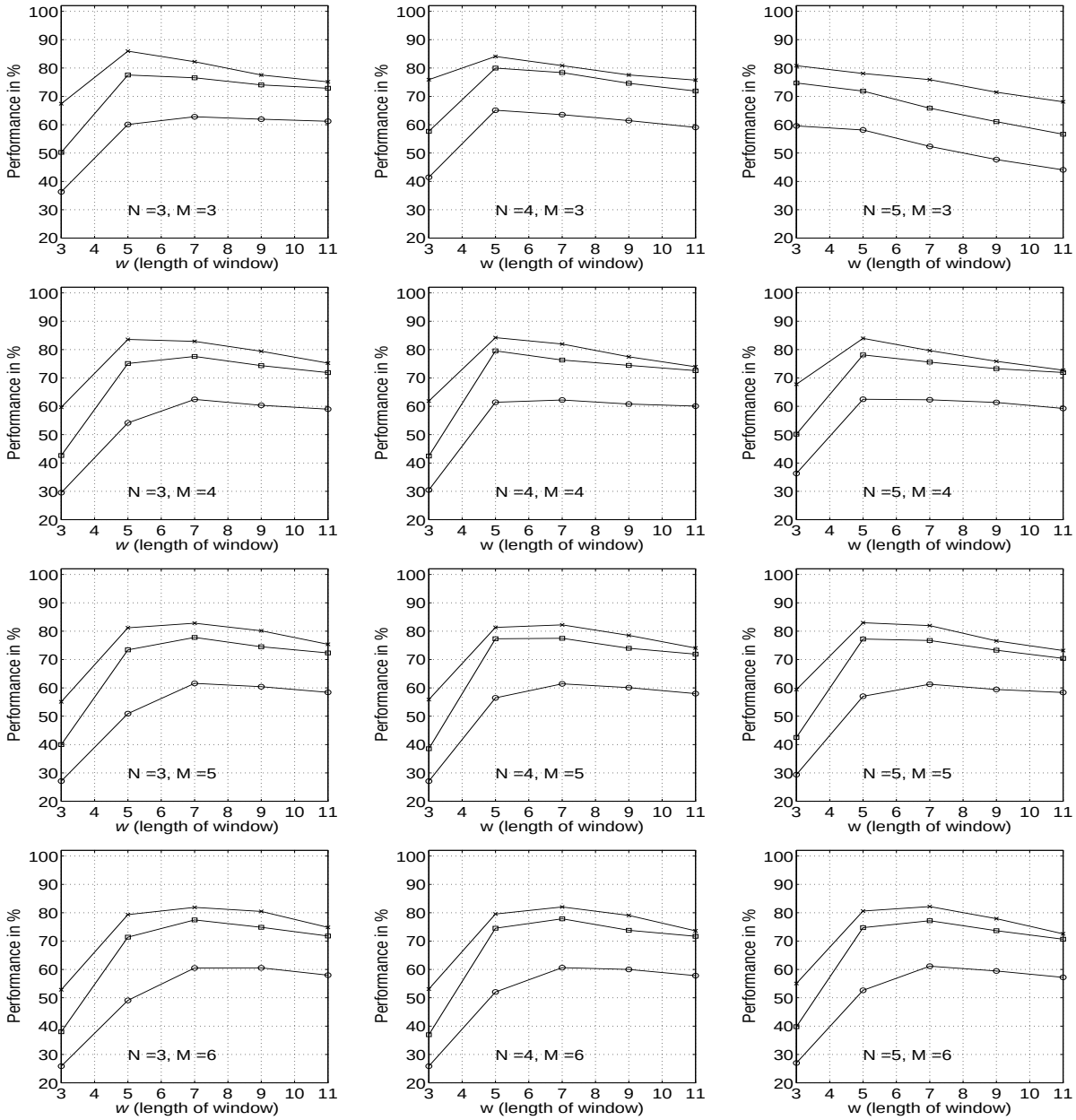


Fig. 4.23: Significance of the parameters used to compute the filtered image using spatial domain realization of the zero-frequency filter, in the context of face recognition. N = number of integrations, M = number of times local mean is subtracted and w = length of the window to compute the local mean. Performances obtained using the vertical evidence, horizontal evidence, and combined evidence are shown using $\times$, $\circ$ and $\square$, respectively.

performance increases. The reason could be that large window highlights the widely spaced edges for small M. With increase in M, relatively closely spaced edges also get highlighted.

5. The set of parameters needs to be chosen suitably, such that neither very closely-spaced edges nor widely spaced edges are highlighted in the filtered image.

4.6 Experimental results

Performance of the proposed filtered image representations using different smoothing filters is compared for different sets of training images from FacePix and Yale-B face databases. The results are summarized in Table 4.2 and Table 4.3. For Yale-B face database two sets of training data are formed: Set 1 containing $35 \times 10 = 350$ images of negative azimuth (with 35 images for each person), and set 2 containing $29 \times 10 = 290$ images of positive azimuth (with 29 images for each person). The eigenvectors are derived using the data set 1, and the performance is evaluated using the data set 2. The performance is also obtained by interchanging the training and test data sets. The experiments are repeated with filtered images derived from the spatial domain and Fourier domain realization of the zero-frequency filter. In the case of spatial domain realization, the evidences using vertical and horizontal directional smoothing are obtained separately. For both the zero-frequency filtered image and the Gaussian filtered image, the parameters (N, w, M and σ) are varied to obtain the best results. Among the representations obtained using directional smoothing filters, the filtered image derived along the vertical direction in the spatial domain realization of the zero-frequency filter gives the best results. This is because the edges along the vertical direction seem to contain more discriminatory information of the face image compared to the edges along the horizontal direction. The filtered image obtained using the Fourier domain realization of the zero-frequency filter gives good performance, although it is slightly lower compared to the performance of the spatial domain realization of the filtered image. Note that the 1% improvement over Gaussian filter is significant due to the fact that the performance has improved to 99.6% from 98.5% (see column 2 in Table 4.3). These results will be compared with other representations as well as with some existing approaches in Chapter 6.

Table 4.2: Performance (average recognition rate in %) of filtered images obtained using the zero-frequency filter and Gaussian filter for different sets of training (reference) face images under illumination variation of FacePix face database.

Approaches		Set of reference face images		
		L^{91}	L^{1}, L^{91}, and L^{181}	L^{1}, L^{46}, L^{91}, L^{136} and L^{181}
Spatial domain realization of the zero-frequency filter	Vertical filtered image	**85.9**	**98.6**	99.8
	Horizontal filtered image	60.1	94	95.6
	Combined	77.4	98.5	**99.9**
Fourier domain realization of the zero-frequency filter		75.2	97	99.3
LOG		76	96	99.2

Table 4.3: Performance (average recognition rate in %) of filtered face images obtained using the zero-frequency filter and the Gaussian filter for different sets of training (reference) face images under illumination variation of Yale-B face database.

Approaches		Set of reference face images	
		Set 1	Set 2
Spatial domain realization of the zero-frequency filter	Vertical filtered image	**99.6**	99.5
	Horizontal filtered image	94	93
	Combined	99.1	**99.7**
Fourier domain realization of the zero-frequency filter		98	98.9
LOG		98.5	98

4.7 Summary

In this chapter we have discussed processing of images using the zero-frequency resonator for two applications, namely, computation of edge map and representation of face image. The idea behind the zero-frequency resonator is that the energy of unit impulse is distributed equally at all frequencies, including the zero-frequency. But this is not true for other zero mean signals of the same strength or energy as that of the unit impulse. The output of the zero-frequency resonator captures the impulse-like characteristics of the signal. The results were demonstrated using several one-dimensional signals. The edges of an image are expected to possess the characteristics of an impulse. We have proposed two approaches based on spatial domain and Fourier domain to realize the zero-frequency resonator for two-dimensional signals. The spatial domain realization does not exploit the relation between two dimensions (between rows and columns) of an image. The relation between rows and columns is utilized in the Fourier domain approach, but it is a FIR approximation of the zero-frequency resonator. Moreover, the frequency domain realization does not specify any phase information.

The locations of the edges are detected by finding the zero-crossings of the filtered image, which is obtained after removing the trend from the output of the zero-frequency filter. In the case of Fourier domain realization, the zero-crossings in the filtered image are two-dimensional. Since it is not easy to detect the zero-crossings in 2-D, it is performed by finding the zero-crossings along each dimension separately. This could be the reason that the detection of edges at the corners is not accurate in the case of Fourier domain realization of the zero-frequency filter. The operation of removing the trend from the output of the zero-frequency filter is equivalent to the Laplacian operation. Thus the proposed approach can be viewed as smoothing using the zero-frequency filter followed by the Laplacian operation. It has similarity with the LOG operation where smoothing is performed by a Gaussian filter. The results show that suppression of noise and edge localization is better in the case of the proposed two approaches.

We have also shown that the filtered images, derived from the Laplacian of the smoothed images, can be used directly as a representation for face recognition, without explicitly de-

riving the edge information. The zero-crossings of the filtered image preserve the edge information of the image. A nonlinear transformation is proposed to normalize the values of the filtered image, by preserving the information around the zero-crossings. The normalized filtered images are used to compute the distance between two face images for face recognition. The *locality problem* of the filtered face image is addressed by smearing the edge information, which is realized by removing the eigenvectors corresponding to low variances. Excessive smearing of the edges may give importance to edges with large spacing, which may correspond to global information of a face image. This may result in poor performance under inter-class variation. The performance can be improved by giving more weightage to the subject-specific unique information.

The zero-frequency filter and the Gaussian filter are used for smoothing the images. The spatial domain realization of the zero-frequency filter gives two filtered images that capture the vertical and horizontal edge information separately. The filtered image derived along the vertical direction seems to capture well the discriminatory information of the face image. This study demonstrates that the edge information need not be confined only to the high frequency components in the spectrum. The edge information is distributed evenly throughout the spectrum, including the region around the zero-frequency.

Chapter 5

Analytic image based representation

Another representation that preserves the impulse-like characteristics of a signal is based on the concept of analytic image, as discussed in this chapter. The theory of analytic function underpins many concepts of signal analysis such as amplitude-frequency demodulation (AM-FM) [144], instantaneous frequency [145] and interferometry [146]. Moreover, it has applications in geophysics [147, 148], narrow-band communication [149] and Radar [150].

The analytic signal is a complex signal and can be broken into two components, namely, magnitude and phase. It is well known that the phase of the analytic signal, derived from a real 1-D signal, preserves the information of the edges of the original signal [151]. Computation of the phase using arctan function gives wrapped phase. It can be avoided by using functions of the analytic phase rather than phase directly. In this work, the trigonometric functions of the analytic phase are used to exploit the edge information present in the face image to address the issue of illumination in face recognition. Here also the *locality problem* is addressed by smearing the edges of the function of analytic phase by removing the eigenvectors corresponding to low variances before matching. The performance is further improved by giving more weightage to the eigenvectors correspond to the unique information present in the face image.

The chapter is organized as follows: Section 5.1 gives the definition of the analytic function for two-dimensional signals. Section 5.2 discusses the information present in the magnitude and phase of the analytic signal. The proposed functions of the analytic phase are explained in Section 5.3. Eigenanalysis of these functions is made for use in the face

recognition task, as explained in Section 5.4. Experimental results for face recognition are given in Section 5.5. Section 5.6 summarizes the chapter.

5.1 Analytic image

The analytic signal of a real one-dimensional (1-D) signal was proposed by Gabor in 1946 [152]. Since then the analytic signal has been used as an important tool in processing of 1-D signals [153]. The analytic signal is derived by suppressing all negative frequencies of the real signal. This results in a complex signal that is a sum of the given real 1-D signal and an imaginary component that is Hilbert transform [153] of the given signal.

Let $x(t)$ be a real 1-D signal. Then the analytic signal $x_A(t)$ is defined as

$$x_A(t) = x(t) + jx_H(t), \tag{5.1}$$

where $x_H(t)$ is the Hilbert transform (HT) of $x(t)$, given by [153]

$$x_H(t) = \mathcal{H}\{x(t)\} = \frac{1}{\pi}\,\text{p.v}\int \frac{x(\tau)}{t-\tau}d\tau. \tag{5.2}$$

Here, p.v denotes the Cauchy principal value of the integral. In the Fourier domain, the relation between the analytic signal and the original signal is given by

$$X_A(\omega) = (1 + \text{sgn}(\omega))X(\omega), \tag{5.3}$$

where signum function (sgn) is defined as

$$\text{sgn}(\omega) = \begin{cases} 1 & \omega>0 \\ -1 & \omega<0. \end{cases} \tag{5.4}$$

The Fourier transform of $x_A(t)$ and $x(t)$ are denoted by $X_A(\omega)$ and $X(\omega)$, respectively. One can observe from (5.3) that the spectrum of the analytic signal is zero for the negative frequencies.

The complex signal obtained using (5.1) satisfies the Cauchy-Riemann condition for

differentiability, and has been traditionally called analytic signal [153]. This definition is extended to 2-D signals as well. The 2-D analytic signal has been applied for feature extraction and classification in image processing [154]. There are three definitions proposed for analytic signal of 2-D signals [144]. Two of them define the analytic image as a sum of the original image and its Hilbert transform, as in the case of 1-D. These are based on the total Hilbert transform [155] and the partial Hilbert transform [156]. The third definition is based on the principle of suppressing some frequencies in the spectrum of the original image. These definitions are as follows:

Definition I: Here, the definition of the Hilbert transform (5.2) is extended to two dimensions. The analytic image of an image $x(t_1, t_2)$ is written as [144]

$$\begin{aligned} x_A(t_1, t_2) &= x(t_1, t_2) + j\mathcal{H}\{x(t_1, t_2)\} \\ &= x(t_1, t_2) + jx_H(t_1, t_2), \end{aligned} \tag{5.5}$$

where

$$x_H(t_1, t_2) = \mathcal{H}\{x(t_1, t_2)\} = \text{p.v} \int \int \frac{1}{\pi^2} \frac{x(\tau_1, \tau_2)}{(t_1 - \tau_1)(t_2 - \tau_2)} d\tau_1 d\tau_2.$$

Applying Fourier transform on both side of (5.5), we get

$$X_A(\omega_1, \omega_2) = (1 - \text{sgn}(\omega_1)\text{sgn}(\omega_2))X(\omega_1, \omega_2), \tag{5.6}$$

where $X_A(\omega_1, \omega_2)$ and $X(\omega_1, \omega_2)$ are the Fourier transforms of $x_A(t_1, t_2)$ and $x(t_1, t_2)$, respectively.

Definition II: In this definition, the 2-D Hilbert transform is considered as the successive applications of the 1-D Hilbert transform along the two dimensions separately. The partial Hilbert transform of an image $x(t_1, t_2)$ along t_1 direction is written as

$$x_H^{t_1}(t_1, t_2) = \text{p.v} \int \frac{1}{\pi} \frac{x(\tau_1, t_2)}{t_1 - \tau_1} d\tau_1. \tag{5.7}$$

The corresponding analytic image is given by

$$x_A^{t_1}(t_1, t_2) \quad = \quad x(t_1, t_2) + j x_H^{t_1}(t_1, t_2), \tag{5.8}$$

and is called horizontal analytic image. Similarly, the vertical analytic image can be computed by using the partial Hilbert transform along t_2 direction.

Definition III: This definition of analytic signal was proposed by S. L. Hahn [156]. In analogy to 1-D analytic signal, the spectrum of the image is made zero everywhere except in one quadrant of the frequency domain. The analytic image, whose spectrum is suppressed in all the three quadrants except in the first quadrant, is given by

$$X_A^1(\omega_1, \omega_2) \quad = \quad (1 + \text{sgn}(\omega_1) + \text{sgn}(\omega_2) + \text{sgn}(\omega_1)\text{sgn}(\omega_2)) X(\omega_1, \omega_2). \tag{5.9}$$

The inverse Fourier transform is given by

$$x_A^1(t_1, t_2) = x(t_1, t_2) - x_H(t_1, t_2) + j(x_H^{t_1}(t_1, t_2) + x_H^{t_2}(t_1, t_2)). \tag{5.10}$$

Similarly by preserving the values of the spectrum in any one of the remaining three quadrants in the Fourier domain, one can derive analytic images as [156]

$$\begin{aligned} x_A^2(t_1, t_2) &= x(t_1, t_2) + x_H(t_1, t_2) - j(x_H^{t_1}(t_1, t_2) - x_H^{t_2}(t_1, t_2)), \\ x_A^3(t_1, t_2) &= x(t_1, t_2) + x_H(t_1, t_2) + j(x_H^{t_1}(t_1, t_2) - x_H^{t_2}(t_1, t_2)), \\ x_A^4(t_1, t_2) &= x(t_1, t_2) - x_H(t_1, t_2) - j(x_H^{t_1}(t_1, t_2) + x_H^{t_2}(t_1, t_2)). \end{aligned} \tag{5.11}$$

In the first two definitions, the original image can be recovered by considering the real part of the analytic image. But these definitions do not satisfy the property of having zero spectrum in the negative frequencies. The third definition satisfies this property, however reconstruction of the original image requires one more analytic image constructed from any of the other quadrants. This is one of the issues in the analytic image based representation. In addition, all the three definitions have a degree of directionality associated with them due to product of the 1-D signum functions [150, 157]. The product of the 1-D signum functions

results in a quadrant signum function. Such products are highly anisotropic owing to the directional line discontinuities, resulting in the anisotropic definitions [157]. One way to address this issue is to maintain the point discontinuity of 1-D signum function in 2-D also. A point is a non-directional discontinuity in two or more dimensions [148, 158]. In this work, the first two definitions are used to represent an image.

5.2 Significance of phase of analytic image

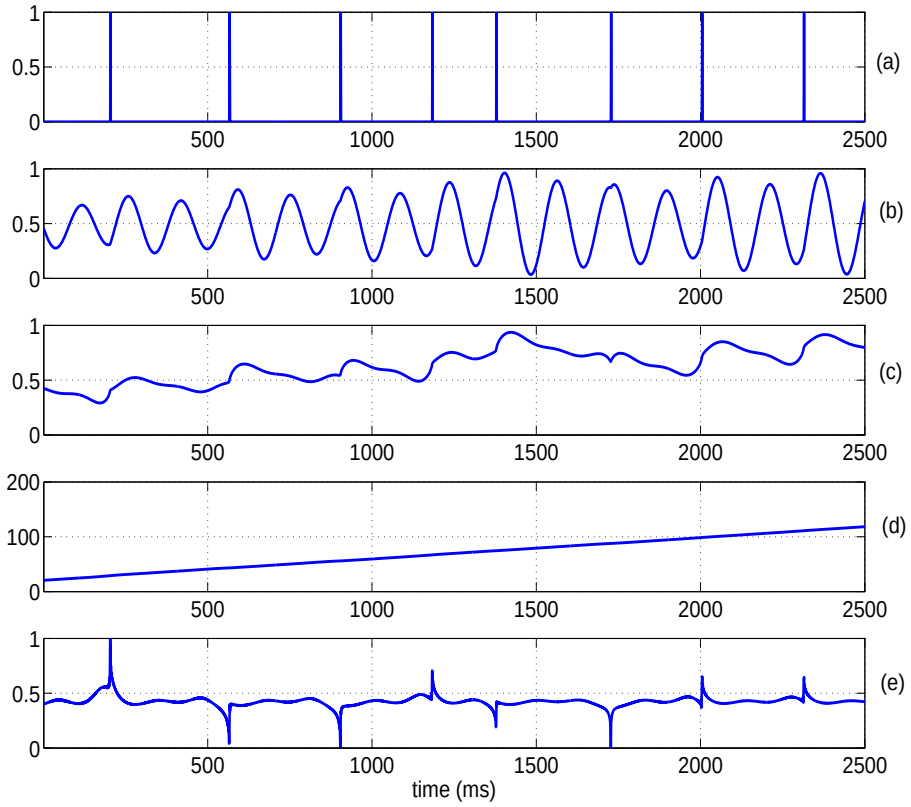
The analytic image computed with any one of the definitions discussed in the previous section can be written as

$$x_A(t_1, t_2) \quad = \quad |x_A(t_1, t_2)| \exp(j\phi(t_1, t_2)), \tag{5.12}$$

where $x_A(t_1, t_2) = \sqrt{(x(t_1, t_2)^2 + x_H(t_1, t_2)^2}$, and $\phi = \arctan\{\frac{x_H(t_1,t_2)}{x(t_1,t_2)}\}$. It is difficult to visualize how the information in these two components are related, because the magnitude and phase are not directly comparable.

The significance of the phase and magnitude of the analytic signal is illustrated using 1-D signal in Fig. 5.1. An aperiodic sequence of impulses with unit strengths is shown in Fig. 5.1(a). The sequence is filtered by a 200 Hz resonator, and the filtered output is shown in Fig. 5.1(b). The magnitude and phase of the analytic signal of the filtered output are shown in Figs. 5.1(c) and (d), respectively. The phase of the analytic signal of the filtered output can be expressed as $\phi(t) = \omega_0 t + \theta(t)$, where ω_0 is the frequency of the resonator. It is difficult to see the information present in the phase of the analytic signal because of the linear term ($\omega_0 t$) present in equation. The information can be extracted by computing the derivative of the phase of the analytic signal, as shown in Fig. 5.1(e). Here, the derivative of the phase is not computed by simply differentiating the phase of the analytic signal. It is computed directly from the analytic signal [17]. This approach avoids unwrapping of the phase function. One can observe from Figs. 5.1(c) and (e) that the information of the impulses of the original signal is preserved better in the phase of the analytic signal. This observation leads us to use the phase of the analytic image to characterize the impulse-like

characteristics (i.e., edge information) of the face image.



(a)

Fig. 5.1: Significance of magnitude and phase of an analytic signal. (a) Aperiodic sequence of impulses with unit strengths. (b) Output of a resonator (whose resonant frequency is 200 Hz) for the sequence of impulses shown in (a). (c) Magnitude and (d) phase of the analytic signal obtained from the filtered output. (e) Derivative of the phase of the analytic signal.

5.3 Proposed functions of analytic phase image

The phase of the analytic image contains the information of the edges of a given image. It can be computed as

$$\phi_w(t_1, t_2) = \arctan\left\{\frac{\text{imag}\{x_A(t_1, t_2)\}}{\text{real}\{x_A(t_1, t_2)\}}\right\}, \tag{5.13}$$

where $x_A(t_1, t_2)$ can be defined using any one of the definitions explained in the Section 5.1. This computation gives only the principal value of the phase in the range of $(-\pi, \pi]$. Any value outside this interval will be wrapped around, producing a wrapped phase function. The actual phase $\phi(t_1, t_2)$ and the wrapped phase are related by [159]

$$\phi(t_1, t_2) = \phi_w(t_1, t_2) \pm l(t_1, t_2)2\pi, \tag{5.14}$$

for some integer $l(t_1, t_2)$. The objective of phase unwrapping is to compute the actual phase $\phi(t_1, t_2)$ either from the wrapped phase $\phi_w(t_1, t_2)$ or from the analytic image $x_A(t_1, t_2)$.

Unwrapped phase in two dimensions has potential applications. In synthetic aperture radar (SAR) interferometric imaging, the phase values correspond to the terrain elevation heights [159–161]. A similar application is in optical interferometry, where the phase values are used to obtain information such as shape and displacement of the surface [162]. The unwrapped phase values of the magnetic resonance (MR) images contain information about flow or inhomogeneities in the magnetic field [163]. Also, computation of the complex cepstrum requires phase unwrapping as one of the processing steps [138]. Other applications of unwrapped phase are in compensated imaging [164] and speckle imaging [165].

It can be observed from (5.14) that the difference between the wrapped phase and the actual phase is multiples of 2π. Due to this, the wrapped phase will have many discontinuities in multiples of 2π. A simple approach to unwrap the phase is to first detect the discontinuities in the wrapped phase using a predetermined threshold, and then remove them by adding or subtracting multiples of 2π. This method works well for noiseless signals. In general, detection of the discontinuities may be erroneous, resulting in inaccurate estimate of the unwrapped phase [166].

There are several approaches proposed in the literature to compute the unwrapped phase [146, 159, 166–175]. Each approach has its own advantages and disadvantages. The unwrapping of phase is still an open research problem. In this work, we propose to use functions of the phase, instead of the phase itself, to avoid the problem of phase unwrapping. One can rewrite (5.12) as

$$\begin{aligned} \exp(j\phi(t_1,t_2)) &= \cos(\phi(t_1,t_2) + j\sin(\phi(t_1,t_2)) \\ &= \frac{x_A(t_1,t_2)}{|x_A(t_1,t_2)|}, \end{aligned} \tag{5.15}$$

where

$$\cos(\phi(t_1,t_2)) = x^c(t_1,t_2) = \frac{x(t_1,t_2)}{\sqrt{x^2(t_1,t_2) + x_H^2(t_1,t_2)}},$$

and

$$\sin(\phi(t_1,t_2)) = x^s(t_1,t_2) = \frac{x_H(t_1,t_2)}{\sqrt{x^2(t_1,t_2) + x_H^2(t_1,t_2)}}. \tag{5.16}$$

The trigonometric functions of the analytic phase computed using the 2-D analytic image defined in (5.5) and the vertical and horizontal analytic images defined in (5.8) are shown in Figs. 5.2, 5.3 and 5.4, respectively. The following observations can be made from the images

1. Computation of the vertical and horizontal analytic images considers the two dimensions of the image independently. Although the dependencies between the two dimensions are exploited in the computation of the 2-D analytic image, its definition is ambiguous. For example, the spectrum of the 2-D analytic image is nonzero for negative frequencies.

2. The cosine function of the analytic phase is similar to the original gray level image ($x^c(t_1,t_2) = \frac{x(t_1,t_2)}{\sqrt{x^2(t_1,t_2)+x_H^2(t_1,t_2)}} \approx x(t_1,t_2)$) for all the definitions of the analytic image.

3. The values of the trigonometric functions are bound between -1 to +1. Due to this

most of the values in cosine and sine functions of the analytic phase are limited to that range. Therefore the functions of the analytic phase appear like a binary image.

4. Each function of the phase of the analytic image gives different information of the face image.

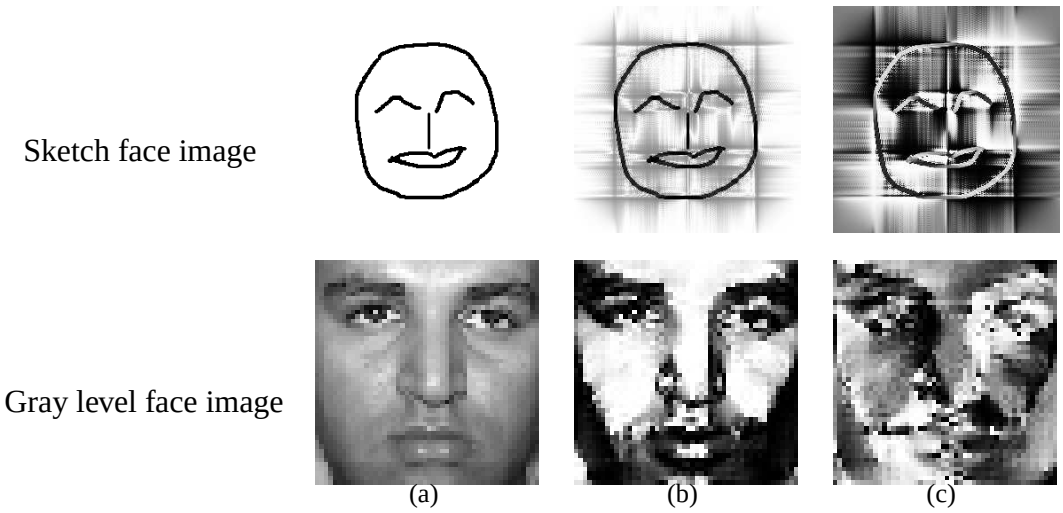


Fig. 5.2: Illustration of the trigonometric functions of analytic phase for two images using 2-D analytic image. (a) Original image. (b) Cosine and (c) sine of the analytic phase.

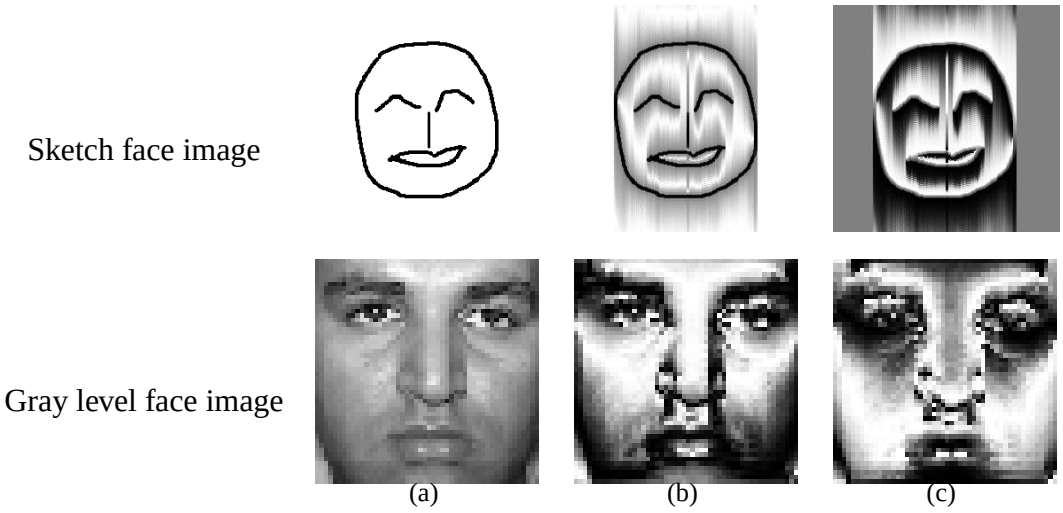


Fig. 5.3: Illustration of the trigonometric functions of analytic phase for two images using vertical analytic image. (a) Original image. (b) Cosine and (c) sine of the analytic phase.

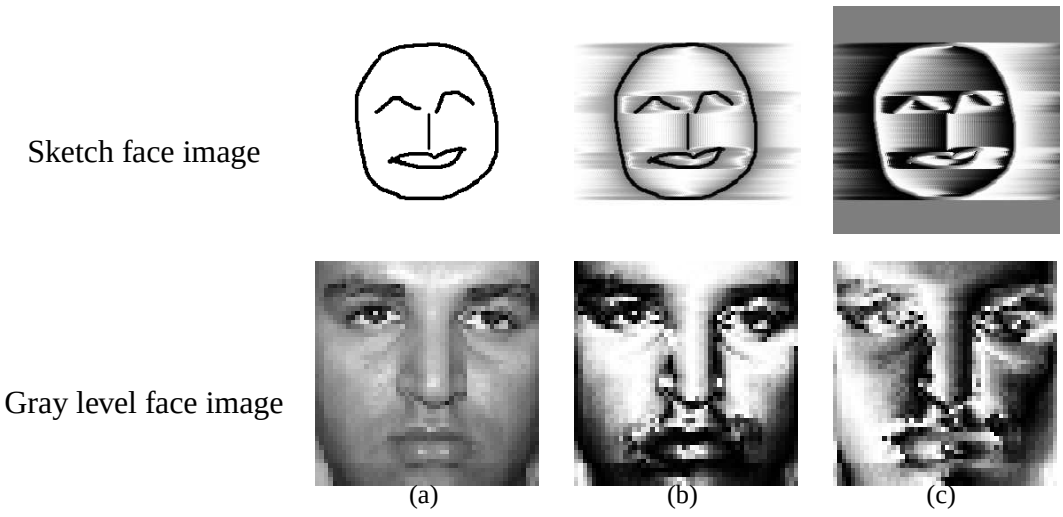


Fig. 5.4: Illustration of the trigonometric functions of analytic phase for two images using horizontal analytic image. (a) Original image. (b) Cosine and (c) sine of analytic phase.

5.4 Locality problem of functions of phase of analytic image

In the case of template matching based face recognition, the functions of analytic phase give poor matching under intra-class variations. The matching can be improved by smearing the edge information of the function of the analytic phase. The smearing of edges is performed using eigenanalysis based approach as explained in Section 4.5.

Let $d_{i,\mathbf{y}}^{s,n}$ and $d_{i,\mathbf{y}}^{c,n}$ be the two distance measures (Equation (4.17)) between a test face image $y[n_1, n_2]$ and the training (reference) face images of the i^{th} person obtained using first n projected coefficients of sine and cosine functions of the analytic phase, respectively. Fig. 5.5 shows the scatter plots obtained using the two distance measures for a person using different values of n with one training face image (L^{91}) of each person in the illumination variation set of the FacePix face database. It can be observed from the plots that the behavior of the true and false class samples using the functions of the analytic phase for different values of n is similar to the behavior of the horizontal and vertical filtered images obtained using the spatial domain realization of the zero-frequency filter (as explained in Section 4.5, Fig. 4.19). For small values of n, matching of any two face images is easier because the global information of the face image (large spacing edges) plays an important role in matching due to excessive smearing of the edges. On the other hand, the *locality problem* of the trigonometric functions of the analytic phase is severe when all the projected coefficients are used for matching.

Another observation from the scatter plots is that the samples of the true class are following a trajectory. The face images ($L^1, \ldots, L^{181}$) of illumination variation set of the FacePix face database were captured by moving the light source at 1° interval from $-90°$ to 90°. The scatter plots are obtained using L^{91} (light source at frontal position) as the reference face image. The distance measures will be large for the face image L^{181} or L^1 as compared to the face image L^{92} or L^{89}. The distances with other true class face images follow a trajectory.

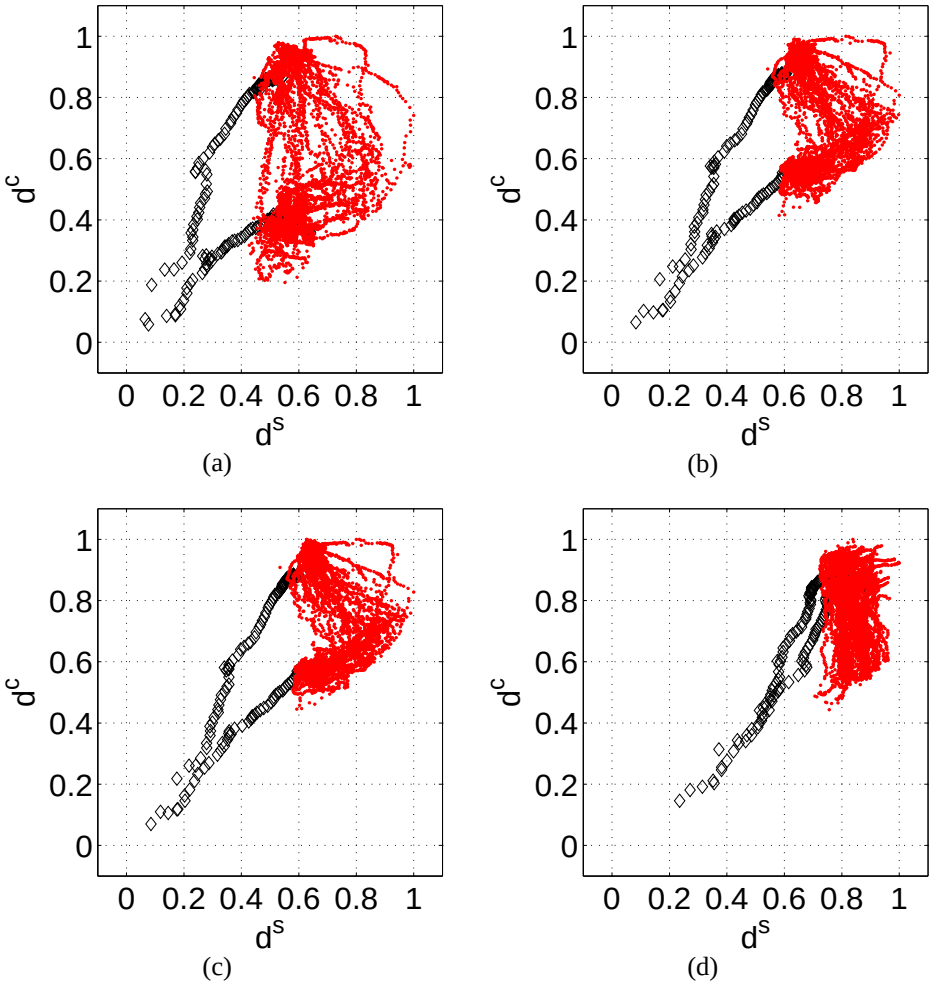


Fig. 5.5: Scatter plots of distances for a person's face image using trigonometric functions of the analytic phase and I^{91} as the reference face image. (a) $n = 5$, (b) $n = 20$, (c) $n = 30$ and (d) $n = 2500$. Here n denotes the number of eigenvectors used to compute the similarity measures ($d^{c,n}_{i,\mathbf{y}}$ and $d^{s,n}_{i,\mathbf{y}}$). The definition of 2-D analytic image given in Equation 5.5 is used to compute the trigonometric functions of the analytic phase.

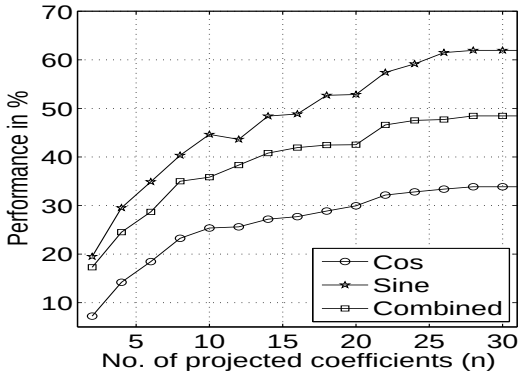


Fig. 5.6: Comparison of the performance of trigonometric functions of the analytic phase (computed using 2-D analytic image) for different number (n) of eigenvectors and using (I^{91}) as training (reference) face image.

Fig. 5.6 shows the performance obtained using the combined scores (Equation (4.18)) for different values of n. The performance obtained using the cosine and sine functions of the analytic phase separately are also shown in the same plot. The sine function seems to perform better than the cosine function of the analytic phase. The reason could be that the edge information of a face image is captured better in the sine function as compared to the cosine function of the analytic phase. The cosine function of the analytic phase is similar to original image, as can be observed from the images shown in Figs. 5.2 (a) and (b). Furthermore, the performance does not seem to improve by combining the evidences from the two functions when combined using norm-2. The performance is improved by using weighted measure (as discussed in Section 4.5.1), in which more weightage is given to the subject specific unique information of the face image present in the function of the analytic phase. The scatter plots using normal and weighted distance measures are shown in Fig. 5.7. It shows that the points due to false class samples are more dense and farther from the origin in the case of weighted distance as compared to normal distance measures. In the case of weighted distances, the chances of matching false class face image is less, a desirable property to address the issue of inter-class variation in face recognition. The improvement in the performance using combined similarity scores can be seen in Fig. 5.8. In addition, the performance of the individual evidences (sine and cosine functions of the analytic phase) is

also improved using the weighted measure.

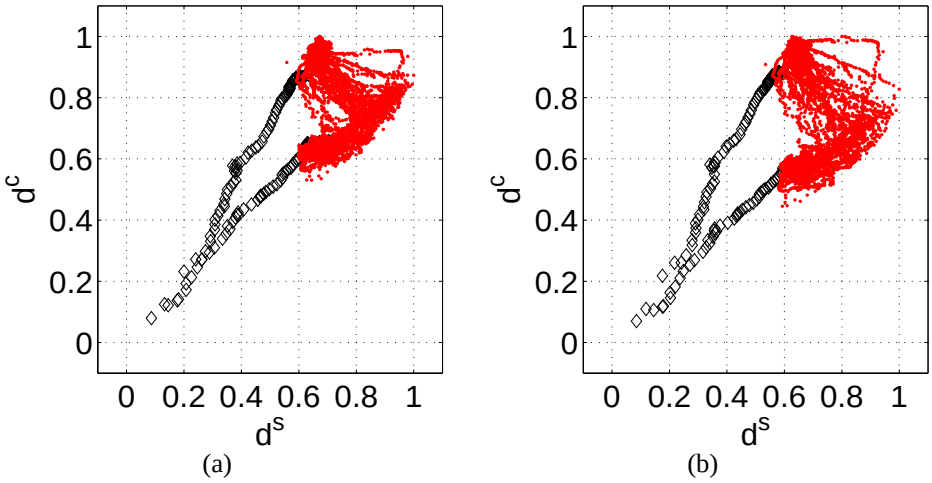


Fig. 5.7: Scatter plots of distances for a person's face image derived using trigonometric functions of analytic phase (obtained using 2-D analytic image), I^{91} as training image and 30 projected coefficients. (a) With weighted eigenvectors. (b) Without weighted eigenvectors.

5.5 Experimental results

Performance of the functions of the analytic phase for face recognition is obtained using the first two definitions (Definitions I and II) of the analytic image on the illumination variation sets of FacePix and Yale-B face databases. The results are shown in Table 5.1 and Table 5.2, respectively. The performance is computed using weighted distance measures for different training sets. For comparison, we have also shown the performance using sine and cosine functions of the analytic phase. The following observations can be made:

1. The sine function of the analytic phase gives better performance than the cosine function of the analytic phase for all the definitions of the analytic image.

2. Among the different definitions and the corresponding functions of the analytic phase, the sine function of the analytic phase obtained using the vertical analytic image gives the best performance for the two face databases and for different sets of training face images. It is due to the fact that the vertical analytic image captures the edges along

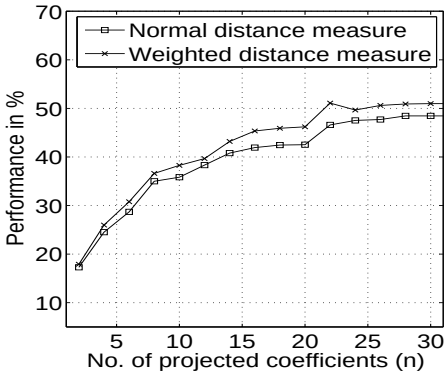


Fig. 5.8: Comparison of the performance under weighted and normal similarity measures obtained using trigonometric functions of the analytic phase (computed using 2-D analytic image) for different number (n) of eigenvectors.

the vertical direction of the face image. The edges along the vertical direction of a face image are more useful for discrimination as compared to the edges along the horizontal direction. The poorer performance using the 2-D analytic image may be due to ambiguity in its definition.

3. The evidences of sine and cosine functions of the analytic phase are combined using norm-2. This approach does not assure that the performance using the combined evidence is better than the performances using individual evidences.

The performance of the representations discussed in this chapter is compared with the existing approaches in Chapter 6.

Table 5.1: Performance (average recognition rate in %) of trigonometric functions of the analytic phase derived using two definitions of the analytic image and different sets of training (reference) face images under illumination variation of FacePix face database.

Representations		Sets of training face image		
		L^{91}	L^{1}, L^{91}, and L^{181}	L^{1}, L^{46}, L^{91}, L^{136} and L^{181}
2-D analytic image	Cos	39.2	92.7	99.1
	Sine	62	95.6	99.3
	Combined	51	96.7	99.3
Vertical analytic image	Cos	42.5	93.1	99.2
	Sine	**68.3**	**97.6**	99.2
	Combined	54.4	97.6	**99.3**
Horizontal analytic image	Cos	39.7	86.3	97
	Sine	41.9	91.2	98.7
	Combined	40.6	91	98.8

Table 5.2: Performance (average recognition rate in %) of trigonometric functions of the analytic phase derived using two definitions of the analytic image and different sets of training (reference) face images under illumination variation of Yale-B face database.

Representations		Sets of training face image	
		Set 1	Set 2
2-D Hilbert analytic image	Cos	63.1	58.6
	Sine	67.9	67.8
	Combined	81	74.5
Vertical analytic image	Cos	73.5	61.5
	Sine	84	**78.5**
	Combined	**84.3**	75.7
Horizontal analytic image	Cos	62.4	58
	Sine	73	70.5
	Combined	68	66

5.6 Summary

In this chapter we have explored a representation based on the phase of the analytic image to capture the edge (impulse-like) information of a face image to address the issue of illumination in face recognition task. The significance of phase and magnitude of the analytic signal was demonstrated using 1-D signal. The definition of analytic signal for 2-D signals is not as straightforward as in the case of 1-D signal. It is still a research issue that needs to be addressed. The two definitions of the analytic image available in the literature have been used to show our results. Computation of the analytic phase using arctan function leads to the problem of phase unwrapping. It is avoided by using trigonometric functions of the analytic phase. These functions of the analytic phase are used separately in template matching for face recognition task. The edge information present in the functions of the analytic phase is smeared to improve the matching of face images under the intra-class variations. The smearing was done by considering only the first few eigenvectors for template matching. The performance is further improved by using weighted distance measure. The idea behind the weighted distance measure is to give more weightage to the projected coefficients which have more discriminative information as compared to the coefficients with less discriminative information of a face image. Another observation is that the edges along the vertical direction are more useful for discrimination as compared to the edges along the horizontal direction.

Chapter 6

Fourier transform based representation

The phase of the Fourier transform (FT) preserves the locations of events such as lines and edges of an image [18, 176]. It can be used as a representation of a face image that captures the edge-like features of the face image. As discussed in the previous chapter, the computation of the phase of the analytic signal using arctan function leads to the problem of phase unwrapping. The problem of unwrapping exists in the computation of the Fourier phase also. In this chapter, we consider a representation based on the phase of the Fourier transform [138], to address the issue of illumination in face recognition. The effect of noise (closely spaced edges) is enhanced in the proposed representations, but can be reduced using the eigenanalysis based approach explained in Chapter 4. The use of phase of the Fourier transform of a given face image for face recognition was studied in [177], where the phase was computed directly using the arctan function. The method proposed in this chapter does not use the phase of the Fourier transform directly, but uses the information present in the phase. The performance of the face recognition is improved by combining evidences from both the phase of Fourier transform and the phase of the analytic image, as both of them contain the information of face image in a complimentary fashion.

The organization of the chapter is as follows: Section 6.1 discusses the information present in the phase and magnitude spectrum of an image. The proposed FT phase representation of a face image is described in Section 6.2. Eigenanalysis on the proposed representation is discussed in Section 6.3 to address the *locality problem*. Section 6.4 gives the results of experimental studies on face recognition. Evidences from the phase of the ana-

lytic image and the phase of the Fourier transform are combined using Borda count based approach [178], as explained in Section 6.5. Section 6.6 gives comparison of the edge-based representations proposed in this thesis. A summary of the chapter is given in Section 6.7.

6.1 Significance of phase of Fourier transform of image

The representations discussed so far in this thesis were based on the information in spatial domain, i.e., a 2-D array of positive numbers, corresponding to the gray levels of the pixels. Here, a representation based on the discrete Fourier transform (DFT) of the 2-D array of pixels [138] is proposed. The Fourier representation involves complex numbers, i.e., magnitude and phase components. The relative importance of the DFT magnitude and phase of a signal/image under different conditions was studied in [18, 176]. Though these two components cannot be compared, the information present in them can be visualized by synthesizing images using only magnitude and using only phase separately. The DFT of an image $x[n_1, n_2]$ is given by

$$\begin{aligned} X[k_1, k_2] &= \text{DFT}\{x[n_1, n_2]\} \\ &= X_r[k_1, k_2] + jX_i[k_1, k_2] \\ &= |X[k_1, k_2]| \exp[j\theta[k_1, k_2]], \end{aligned} \tag{6.1}$$

where $|X[k_1, k_2]| = \sqrt{X_r[k_1, k_2] + X_i[k_1, k_2]}$ and $\theta[k_1, k_2] = \arctan\left\{\frac{X_i[k_1,k_2]}{X_r[k_1,k_2]}\right\}$ are the magnitude and phase of the DFT, respectively. The real and imaginary parts of the DFT are denoted by $\mathbf{X}_r$ and $\mathbf{X}_i$, respectively. The information contained in the magnitude and phase of the DFT can be visualized using magnitude-only synthesis of the face image $x_m[n_1, n_2]$ and the phase-only synthesis of the face image $x_p[n_1, n_2]$, respectively. These are defined as follows:

$$x_m[n_1, n_2] = \text{IDFT}\{|X[k_1, k_2]|\} \tag{6.2}$$

$$x_p[n_1, n_2] = \text{IDFT}\{\exp[j\theta[k_1, k_2]]\}. \tag{6.3}$$

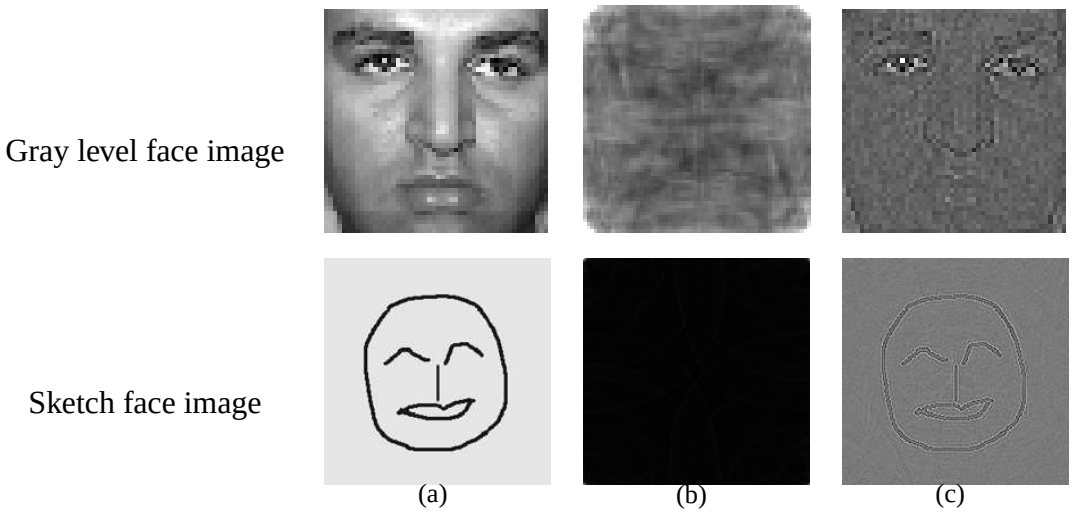


Fig. 6.1: Significance of magnitude and phase of the DFT, illustrated using two images. (a) Original image. (b) Magnitude-only and (c) phase-only synthesis images.

Here IDFT refers to inverse discrete Fourier transform. The images $x_m[n_1, n_2]$ and $x_p[n_1, n_2]$ are shown in Fig. 6.1. The phase-only face image retains many of the useful features of the original face image. The phase-only synthesis image is also shown for a line diagram of a face which is equivalent to the sequence of impulses in 1-D. It shows that the information of impulses in 2-D is preserved better in the phase of the FT as compared to the magnitude. This can be explained as follows. The phase-only synthesis of a face image can be written as

$$\begin{aligned} x_p[n_1, n_2] &= \text{IDFT}\left\{\frac{X[k_1, k_2]}{|X[k_1, k_2]|}\right\} \\ &= \text{IDFT}\left\{\frac{1}{|X[k_1, k_2]|}\right\} \otimes \text{IDFT}\left\{X[k_1, k_2]\right\} \\ &= h[n_1, n_2] \otimes x[n_1, n_2], \end{aligned} \tag{6.4}$$

where $h[n_1, n_2] = \text{IDFT}\left\{\frac{1}{|X[k_1,k_2]|}\right\}$, and $\otimes$ is a convolution operator. Since the spectral magnitude $|X[k_1, k_2]|$ of a face image tends to fall off at high frequency, the phase-only face image $x_p[n_1, n_2]$ has a high frequency emphasis, which will accentuate lines, edges and other spatially narrow events (noise) without changing their positions [18]. The edges have characteristics of an impulse whose energy is distributed uniformly over all frequencies in

the spectrum (as discussed in Chapter 4). Therefore, due to emphasis of the high frequency components, the edges will appear in the phase-only synthesis face image. In addition, noise (spatially narrow events or closely spaced edges) will also appear in $x_p[n_1, n_2]$, because it corresponds to high frequency components of the spectrum. Although, these arguments provide a general basis for the information present in the phase-only face image, the processing method to obtain the phase-only signal is non-linear, whereas the above mentioned interpretation assumes a linear operation [18].

6.2 Functions of phase of Fourier transform

The phase spectrum of an image contains edge information of the image. But computation of the phase spectrum using arctan function leads to the problem of phase wrapping [138]. This issue was addressed in the literature using phase unwrapping and group-delay processing [138, 166, 179, 180].

One way to address this issue is to use a function of the phase spectrum (as discussed in the previous chapter), instead of the phase spectrum directly. We can write

$$\begin{aligned} \exp[\, j\theta[k_1, k_2]] &= \cos[\theta[k_1, k_2]] + j \sin[\theta[k_1, k_2]] \\ &= \frac{X[k_1, k_2]}{|X[k_1, k_2]|}. \end{aligned} \tag{6.5}$$

That is,

$$\sin[\theta[k_1, k_2]] = X^s[k_1, k_2] = \frac{X_i[k_1, k_2]}{|X[k_1, k_2]|}$$

and

$$\cos[\theta[k_1, k_2]] = X^c[k_1, k_2] = \frac{X_r[k_1, k_2]}{|X[k_1, k_2]|}. \tag{6.6}$$

We can use the sine and cosine functions of the phase to avoid the phase wrapping problem. These components may contain complementary information of the face image, and hence they are analyzed separately for template matching based approach for face recognition.

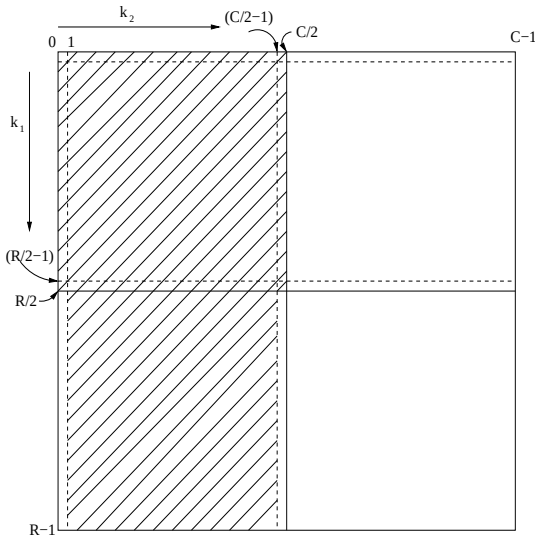


Fig. 6.2: 2-D spatial frequency domain. DFT coefficients $X[k_1, k_2]$ in the shaded area determine the remaining coefficients.

6.3 Locality problem of functions of phase spectrum

The sine and cosine functions of the phase spectrum accentuate the noise (closely spaced edges) also, which leads to poor matching under the intra-class variation of the face image in template matching. This is equivalent to the *locality problem* of edge-based representation. The matching can be improved using eigenanalysis based approach as explained in Section 4.5. The DFT of a real image exhibits conjugate symmetry, i.e.,

$$X[k_1, k_2] = X^*[R - k_1, C - k_2], \qquad 0 \leq k_1 \leq R - 1, \quad 0 \leq k_2 \leq C - 1. \tag{6.7}$$

Hence, only the non-redundant coefficients $RC/2 + 2$ (the points in the shaded region in Fig. 6.2), are used for eigenanalysis. Due to this, the proposed representations ($\mathbf{X}^s$ and $\mathbf{X}^c$) have an advantage in the context of eigenanalysis, because the size of resulting covariance matrix $((\frac{RC}{2} + 2) \times (\frac{RC}{2} + 2))$ is approximately one fourth of the size of the covariance matrix ($RC \times RC$) obtained using gray level values of the face image directly. Thus, estimation of the covariance matrix (eigenvectors) may be more accurate for the same number of training

face images.

Let $d_{i,\mathbf{y}}^{c,m}$ denote the minimum Euclidean distance obtained for a given test face image $y[n_1, n_2]$ using cosine function of phase spectrum (first m coefficients) of the available training face images of the i^{th} person. Similarly, the minimum Euclidean distance is computed using sine function of the phase spectrum of the test and training face images, and is denoted by $d_{i,\mathbf{y}}^{s,m}$. The effect of closely spaced edges or noise can be reduced by considering only the first m projected coefficients in the matching process. In Fig. 6.3, the significance of eigenvectors is illustrated with the scatter plots obtained using $d_{i,\mathbf{y}}^{c,m}$ and $d_{i,\mathbf{y}}^{s,m}$ for different values of m. The points are found to be more dense and far away from the origin in Fig. 6.3(d) as compared to Figs. 6.3(a)-(c). The reason for this is that matching of any face image is poor due to the effect of noise if all the coefficients ($m = RC/2+2=1252$) are used for matching. By removing the last few projected coefficients, the effect of noise can be reduced in the matching process. Removing many coefficients leads to excessive smearing of the edges, which results in retaining only the global information of the face image. This results in matching with some of the false class face images also. Thus there is a trade-off in reducing the effect of noise and smearing of the edges.

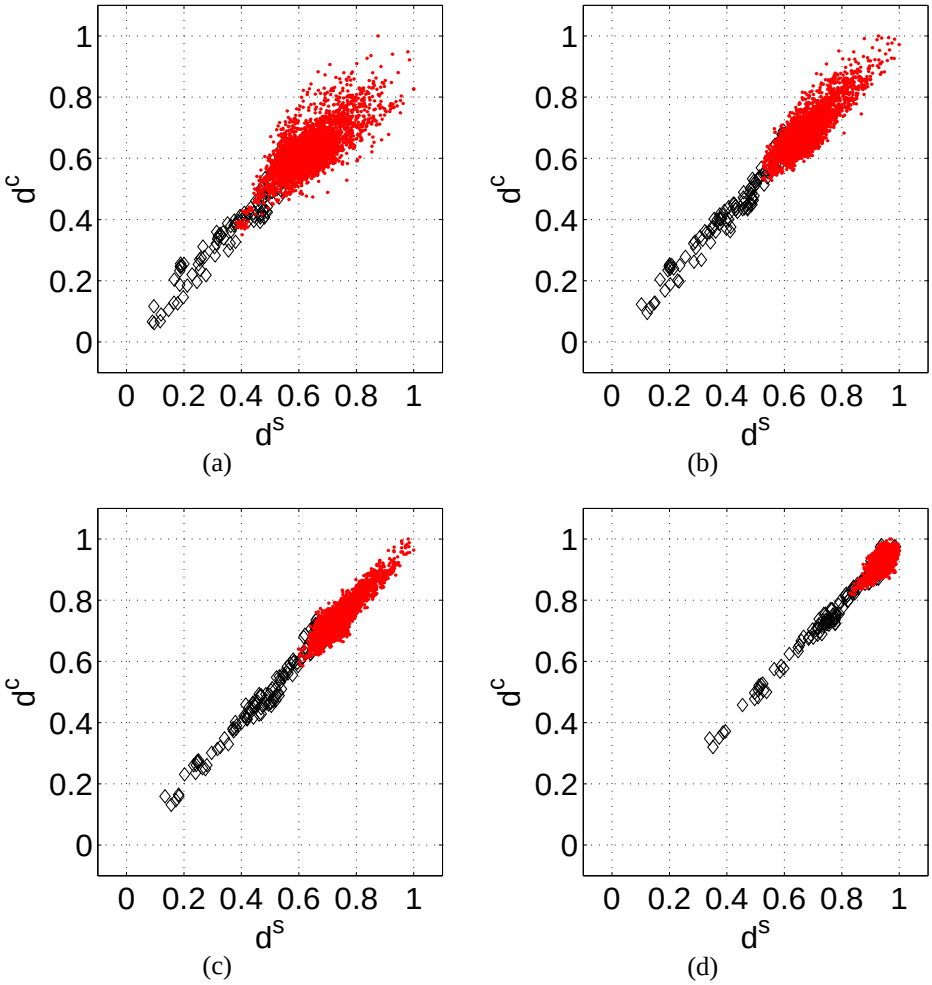


Fig. 6.3: Scatter plots of distances for a person's face image using trigonometric functions of the Fourier phase and I^{91} as the reference face image. (a) $m = 10$, (b) $m = 20$, (c) $m = 100$ and (d) $m = 1252$. Here m denotes the number of eigenvectors used to compute the similarity measures ($d_{i,\mathbf{y}}^{c,m}$ and $d_{i,\mathbf{y}}^{s,m}$).

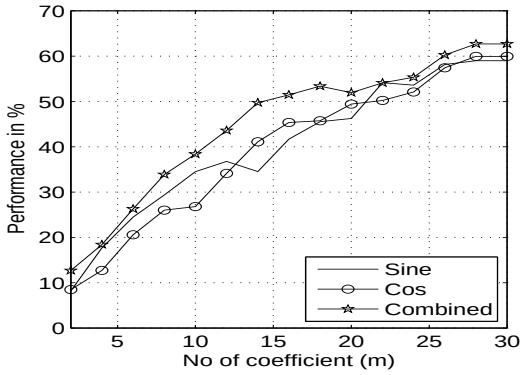


Fig. 6.4: Comparison of the performance of trigonometric functions of the Fourier phase for different number (m) of eigenvectors and using I^{91} as training (reference) face image.

Performance is computed using combined distance scores $d^{c,m}_{i,\mathbf{y}}$ and $d^{s,m}_{i,\mathbf{y}}$ (Equation (4.18)) for different values of m, as shown in Fig. 6.4. The performance is also shown using cosine and sine functions of the phase spectrum separately in the same plot. The edge information of a face image is distributed uniformly in the two functions of the Fourier phase. Hence one cannot say that any one of the functions performs better, as observed in the case of functions of the analytic phase. In general, the performance increases with m, but after some value of m the performance reaches a saturation point. The performance of the proposed representations of the Fourier phase does not improve if weighted eigenvectors (explained in Section 4.5.1) are used in template matching. This is because the edge information of the face image is distributed uniformly in all the projected coefficients obtained from the eigenanalysis of $\mathbf{X}^c$ and $\mathbf{X}^s$ representations of the face image. Nonetheless, the same task can be performed by selecting appropriate DFT coefficients while computing the FT phase representation of the face image.

6.3.1 Significance of DFT coefficients

Information of the relative locations of the edges or lines is preserved in the phase of the Fourier transform. The spacing of the edges will be inversely proportional to the frequency in the phase of the Fourier transform. Thus the low frequency DFT coefficients correspond

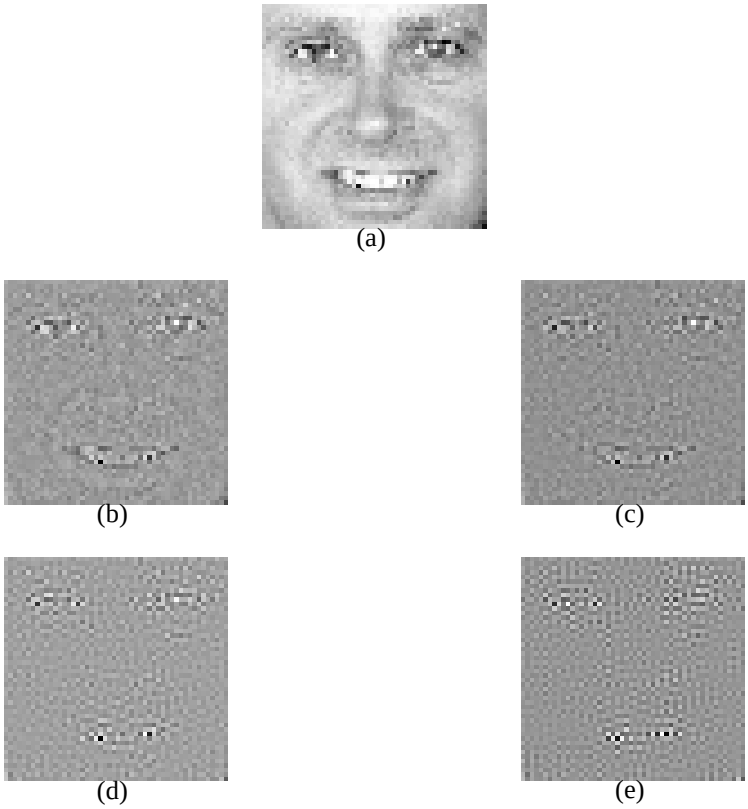


Fig. 6.5: (a) Gray level face image. Phase-only synthesis face image for (b) $l = 5$, (c) $l = 10$, (d) $l = 15$ and (e) $l = 20$.

to events separated by large spacing, and the high frequency DFT coefficients correspond to events separated by small spacing. The effect of the DFT coefficients can be seen in the phase-only synthesis face image, by making the first l DFT coefficients zero along both the axes in the frequency domain, and preserving the remaining DFT coefficients. For different values of l, the phase-only synthesis face images are shown in Fig. 6.5, and the features in the images justify the above-mentioned arguments.

Matching between true class face images can be improved by removing the high frequency DFT coefficients. This is an advantage because noise and events with closely spaced edges are given less importance. Face recognition experiments were repeated by taking only the first k DFT coefficients along both the axes of the $\mathbf{X}^s$ and $\mathbf{X}^c$ representations of the given training face image. Only non-redundant coefficients (the shaded region in Fig. 6.6) were used in the eigenanalysis. Since the number of eigenvectors depends on the size of covari-

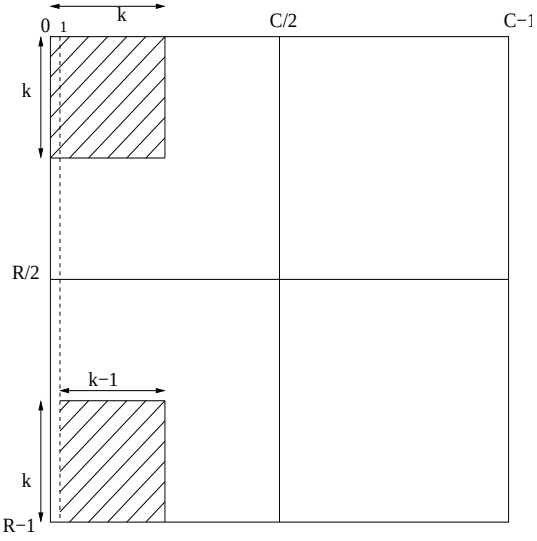


Fig. 6.6: 2-D spatial frequency domain. The shaded region in the figure determines the non-redundant coefficients if only the first k DFT coefficients along both the axes of $X[k_1, k_2]$ are considered for computation of the DFT phase.

ance matrix, which in turn depends on the value of k, the optimal number of eigenvectors (m^*) for which maximum identification performance is obtained is different for each case. Fig. 6.7 shows the scatter plots for different values of k using the optimal number of eigenvectors (m^*) for each k, with one training (reference) face image (L^{91}) of each person. It shows that the points are more scattered for low values of k as compared to high values of k. This is because, for high values of k, events/edges with small spacing get importance in matching process. Chances of the matching with the true class face images is also poor. On the other hand, for low values of k, only the events/edges with large spacing take part in the matching process. This is equivalent to shape matching or matching of global information of face image, hence the chances of matching of false class face images becomes high. Performance of the proposed approach for different values of k using different sets of training face images is discussed in Section 6.4.

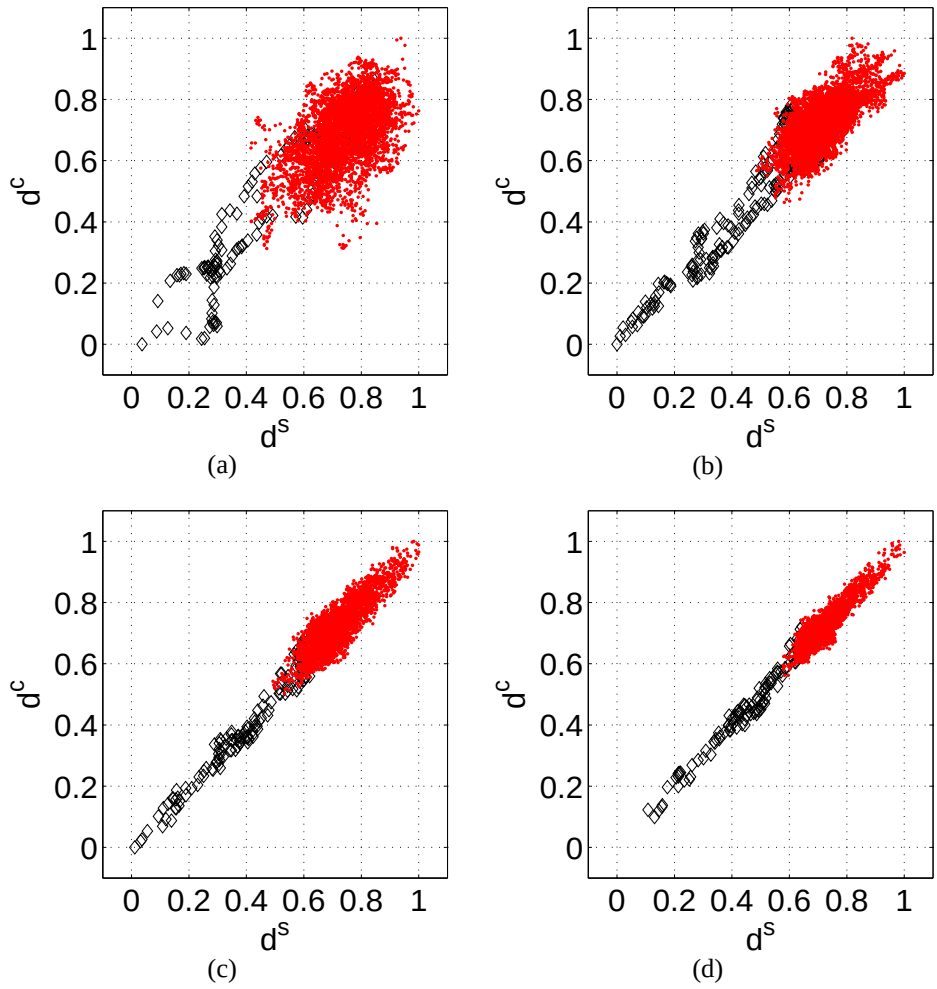


Fig. 6.7: Scatter plots of distances for a person's face image using trigonometric functions of the Fourier phase, I^{91} as training (reference) image, for (a) $k = 5$, (b) $k = 10$, (c) $k = 18$ and (d) $k = 25$. Here k DFT coefficients along both the axes of $\mathbf{X}^c$ and $\mathbf{X}^s$ representations are employed to derive the eigenvectors.

6.3.2 Zero-padding in the computation of Fourier transform phase

One of the properties of DFT is that zero-padding in one domain results in increased number of samples in the other domain [138]. Hence zero-padding in the spatial domain results in increased number of samples in the frequency domain. A simple way of zero-padding is to append strings of zeros at the end of the 2-D sequence. After appending, the new 2-D sequence attains the size $fR \times fC$, where f is a factor parameter, and R and C are the number of rows and columns in the original face image, respectively. Fig. 6.8 shows the scatter plots for different values of f using one training (reference) face image for each person. It shows that the points due to false class are dense, and are farther away from the origin for high values of f, as compared to the low values of f. For high values of f the finer resolution of the phase spectrum is captured, and this helps in improving the discrimination. Performance of the proposed approach for different values of f is discussed in the next section.

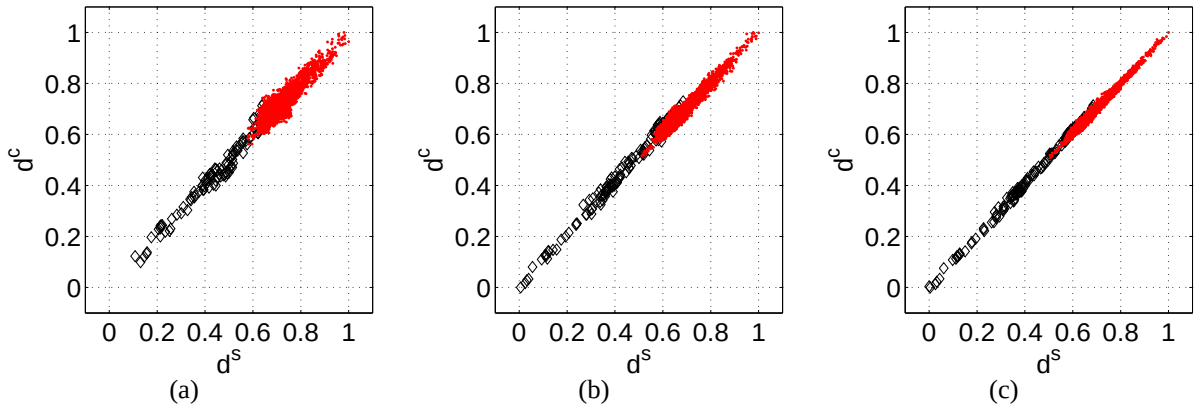


Fig. 6.8: Scatter plots of the distances of a person's face image using trigonometric functions of the Fourier phase, I^{91} as training (reference) image and for parameter (a) f = 1, (b) f = 1.5 and (c) f = 2. Here the parameter f governs the zero-padding in the DFT computation.

6.4 Experimental results

Performance of the proposed approach on illumination variation set of FacePix face database for different values of k and f (symbols are explained in Section 6.3) using different sets of

Table 6.1: Performance (average recognition rate in %) of trigonometric functions of the phase spectrum for different values of k and f, using different sets of training (reference) face images of the FacePix face database.

f	k	Set of training face images		
		L^{91}	L^{1}, L^{91} and L^{181}	L^{1}, L^{46}, L^{91}, L^{136} and L^{181}
1	5	48.1	80	96.8
	10	68.2	97.1	99.2
	15	67.1	96.9	99.2
	18	65.7	95.7	99.2
	20	64	94.5	99.1
	25	62.6	93.5	98.6
1.5	10	62	92.8	99.1
	15	71.2	97.3	99.2
	20	**72.6**	97.1	99.3
	25	71.1	96.4	99.2
	30	69.4	94.8	98.9
	37	68.1	93.5	98.7
2	10	52.5	87.3	98.1
	20	72.47	**97.9**	99.3
	30	71.6	97.5	**99.3**
	40	69.6	95.3	98.9
	45	68.6	94.3	98.7
	50	67.6	93.4	98.6

training (reference) face images is given in Table 6.1. It shows that for a fixed value of f, the performance is improved by removing some high frequency DFT coefficients. But removing many coefficients leads to loss of information, and hence reduction in the performance. The performance can be improved by increasing the parameter f in the DFT computation by padding with zeros. The performance was found to be the best for $f = 2$ and $k = 20$. Table 6.2 compares the performance of the proposed approach with the method discussed in [177] which uses wrapped phase of the Fourier transform, using different sets of training face images of PIE-NL face database. The results in the table show that the proposed approach uses the phase information effectively, and avoids phase unwrapping as well. Comparison with the performance of other approaches using different representations of the face image is discussed in Section 6.6

Table 6.2: Performance (average recognition rate in %) of trigonometric functions of the phase spectrum for different sets of training (reference) face images of PIE NL dataset. Here I^7, I^{10}, and I^{19} are face images with frontal lighting. I^3 and I^{16} are face images with left shadow and right shadow respectively.

Approaches	Set of training face images		
	I^7	I^7, I^{10} and I^{19}	I^3, I^{16} and I^7
Eigenphase [177]	-	97	100
Proposed	**97.7**	**99.1**	**100**

6.5 Analytic phase and Fourier phase

The edge information is preserved in a complementary fashion in the analytic phase and the Fourier transform phase of an image, in the sense that the large spacing of impulses or edges in the spatial domain are reflected as closely spaced impulses or edges in the frequency domain and vice versa. This can be observed from the following basic Fourier transform relation for a sequence of uniformly spaced impulses in time domain:

$$\sum_m \delta(t - mT) \overset{FT}{\Longleftrightarrow} \frac{2\pi}{T} \sum_k \delta(\omega - \frac{2\pi}{T}k). \tag{6.8}$$

Here, $\delta(.)$ denotes the impulse function, t and ω denote the time and frequency variables, respectively. T denotes the spacing between the impulses in the time domain. The performance of the face recognition can be improved by combining the evidences from both the representations. We have used Borda count [181] based approach to combine the rank given by the individual evidences. The Borda count is a simple yet effective method of combining the rankings.

In the literature of pattern recognition, several approaches are proposed to combine the ranks given by multiple classifiers [181]. The simplest way of combing classifiers is to let the classifier cast the vote by forwarding the class they prefer the best (first rank). The identity of a test sample will be the class which obtain the first rank by a majority of the classifiers. It is simple and quite effective, but uses only the top ranking classes. This approach does not exploit the information given by the second or third rank classes, which may be more useful in the context of face recognition. The distance scores of the face images of the top few

ranks assigned by different classifiers are more or less similar. So considering only the first rank class of different classifiers for final decision may not lead to the correct result. This issue is addressed in the Borda count [178, 182, 183] based approach of combining the ranks of different classifiers. Here, the highest ranked class (for N classes) of a classifier gets N points, and each subsequent class gets one point less (so the 2nd ranked class gets *N-1* points and 3rd ranked class gets *N-2* points and so on). All the points for each class obtained from different classifiers are added up, and the class with the highest number of points is assigned as identify of the test sample.

In this work we have combined the evidences given by four representations (two functions of the Fourier phase and two functions of the analytic phase) of an image to exploit the complementary information present in both the Fourier phase and analytic phase. Each representation can be seen as a classifier, and the person refers to the class. For a given test face image ($\mathbf{y}$) the distance scores using all the classifiers are computed with all the available person's reference face images. The rank of each person is computed using these distance scores. Let r be the rank given to the i^{th} person using the cosine function of the analytic phase. The resulting score for the i^{th} person will be

$$p^{ca}_{i,\mathbf{y}} = N - r. \tag{6.9}$$

Here N is the number of persons in the face database. Similarly, $p^{sa}_{i,\mathbf{y}}$, $p^{sf}_{i,\mathbf{y}}$ and $p^{cf}_{i,\mathbf{y}}$ are computed using the sine function of the analytic phase (Here only 2-D analytic image definition is used), sine function of the Fourier phase and cosine function of the Fourier phase, respectively. The identity (i^*) of a test face image $\mathbf{y}$ is computed as

$$i^* = \arg\max_i \left\{ p^{ca}_{i,\mathbf{y}} + p^{sa}_{i,\mathbf{y}} + p^{cf}_{i,\mathbf{y}} + p^{sf}_{i,\mathbf{y}} \right\}. \tag{6.10}$$

The performance obtained using this approach is given in Table 6.3 and Table 6.4.

6.6 Comparison of the proposed representations

Performance of all the proposed edge-based representations are compared along with some existing approaches on the illumination variation set of FacePix face database and Yale-B face database, and the results are given in Table 6.3 and Table 6.4, respectively. The first four rows in Table 6.3 and the first seven rows in Table 6.4 correspond to the results from some existing approaches. These methods use gray level representation of images, and seem to perform poorly compared to the proposed edge-based representation, indicating the importance of the edge information for face recognition under varying illumination conditions. The tables also show the results using the standard Gabor filter based representation [184], which also captures the edge information of face images. In the Gabor filter based approach, a face image is represented by concatenating the magnitude of the response of the Gabor filters obtained for five different scales and eight different orientations [184]. Matching is performed using eigenanalysis based method. The results show that the edge information seems to have been captured better in the proposed representations as compared to the Gabor filter based representation. In case of one training (reference) face image per person (second column of Table 6.3), the performance of the edginess-based representation is better than the other edge-based representations (except vertical evidence of the directional zero-frequency filter). This is because smearing of the edge information in the former representation is independent of the number of training (reference) images, whereas in the other representations the smearing of the edge is realized using the first few eigenvectors, which are obtained using eigenanalysis on only a small number (30) of the training face images. When the number of training images is increased, the performance of the edge-based representations has improved significantly compared to the edginess-based representation, as can be seen from the results in the third and fourth columns of Table 6.3, where the number of training images are 90 and 150, respectively. In fact this can also be seen in the results for both these types of representations in the second and third columns of Table 6.4, where the number of training face images in each set is large.

Among the representations obtained using directional smoothing filters, the filtered image derived along the vertical direction using the spatial domain realization of the zero-

frequency filter gives the best results. This is because the edges along the vertical direction contain more discriminatory information of the face image. The filtered image obtained using the Fourier domain realization of the zero-frequency filter gives good performance, although it is slightly lower compared to the performance of the spatial domain realization of the filtered image. The Gaussian filter and the Fourier domain realization give similar performances.

Another representation that attempts to capture the impulse-like characteristics of an image is based on the phase of the analytic image. The performance is relatively low, may be because the definition of the analytic image in 2-D is ambiguous. The performance can be improved when the phase of the analytic image is combined with the phase of the Fourier transform using the Borda count based approach.

Representation emphasizing the information of edges appears superior to direct gray level representation. Instead of direct representation of the edge information through edginess image, if the edge information is preserved around the zero-crossings, as in the case of the filtered images, the performance of face recognition improves significantly. For matching it is not necessary to extract the edge information from the filtered images. All the filtered images seem to perform better by more than 5% in absolute performance over the edginess-based representation. Even among the filtered images, the representation derived using the spatial domain realization of the zero-frequency filter gives recognition performance over 99.5%. This clearly demonstrates the significance of representation of images for face recognition task. In addition, one can conclude that the representation based on the edge information is not only crucial but also depends how it is computed.

Table 6.3: Performance (average recognition rate in %) of four proposed edge-based representations along with the existing approaches for different sets of reference face images under illumination variation of FacePix face database.

Approaches				Set of reference face images		
				L^{91}	L^{1}, L^{91} and L^{181}	L^{1}, L^{46}, L^{91}, L^{136} and L^{181}
Principal component analysis			[115, 116]	48.8	71.7	90.3
Linear discriminate analysis				53	79.5	94.9
Hidden Markov model				19.3	37.4	59.4
Bayesian information criteria				49.8	79.1	93.5
Gabor filter				61	88	95
Edginess-based representation				81.4	94.3	99.7
Laplacian of smoothed image	Zero-frequency filter	Spatial domain	Vertical filtered image	**85.9**	**98.6**	99.8
			Horizontal filtered image	60.1	94	95.6
			Combined	77.4	98.5	**99.9**
		Fourier domain		75.2	97	99.3
	Gaussian filter			76	96	99.2
Analytic phase	Cos			39.2	92.7	99.1
	Sine			62	95.6	99.3
	Combined			51	96.7	99.3
Fourier phase	Cos			71.3	97.3	99.2
	Sine			70.2	96.8	99.4
	Combined			72.6	97.1	99.3
Combining Fourier phase and analytic phase				73.1	97.8	99

Table 6.4: Performance (average recognition rate in %) of the four proposed edge-based representations along with the existing approaches for different sets of reference face images under illumination variation of Yale-B face database.

Approaches				Set of reference face images	
				Set 1	Set 2
(PCA) Principal component analysis			[110]	79.3	80.3
(2DPCA) 2-D principal component analysis				82.8	86.9
(BDPCA) Bidirectional PCA				82.8	88.6
(W-BDPCA) Whitened BDPCA				87.6	91.1
(ICA) Independent component analysis				82.4	84.6
(EICA) Enhance ICA				86.9	85.1
(RC-ICA) Row column ICA				91.4	92.6
Gabor filter				83	85.2
Edginess-based representation				92.2	94.4
Laplacian of smoothed image	Zero-frequency filter	Spatial domain	Vertical filtered image	**99.6**	99.5
			Horizontal filtered image	94	93
			Combined	99.1	**99.7**
		Fourier domain		98	98.9
	Gaussian filter			98.5	98
Analytic phase	Cos			63.1	58.6
	Sine			67.9	67.8
	Combined			81	74.5
Fourier phase	Cos			90.6	94.4
	Sine			89.2	95.7
	Combined			93.6	96
Combining Fourier phase and analytic phase				95	97

6.7 Summary

In this chapter, a representation based on the phase of the Fourier transform is proposed to capture the edge-like information of face images. The proposed approach avoids computation of the unwrapped phase by computing two functions of the phase spectrum rather than using the phase directly. These two functions give different information of the face image, and are used separately for matching. Computation of the proposed representation emphasizes the closely spaced edges which are mostly noise. The effect of noise is reduced using selected number of eigenvectors in the eigenanalysis of the phase representations for template matching. The effect of small spacing or large spacing edges in matching of two face images can be controlled by selecting appropriate DFT coefficients in the eigenanalysis. The dimension of the functions of the phase spectrum can be reduced using the conjugate symmetry property of the DFT. This helps in better estimation of the eigenvectors from the same number of training face images.

The present study also demonstrates the complementary nature of the edge information present in both the phase of the analytic image and in the phase of the Fourier transform of the image. The relative performances of all the edge-based representations of face images were discussed in this chapter.

Chapter 7

Summary and conclusions

Edge is one of the important features of the sketch and caricature, which contain perceptually the most informative part of a face image in a concise manner. In this work the effectiveness of the four representations, that capture the edge information of images, was investigated in the context of face recognition.

Derivative operation is employed in the first two representations of a face image. The first representation, namely, edginess (first derivative of an image), is computed using 1-D processing of images. Potential field is derived from the edginess image to address the *locality problem*. The multiple partial evidences, derived using 1-D processing of a face image, are used separately for template matching, because each of them contains some complimentary information of the face image. In the proposed template matching based approach, the scores obtained from matching of partial evidences of a face image with different reference face images (at different poses or different lighting conditions) are combined. The combing of scores was based on the observation that if a test face image of the true class has pose that lies between poses of two references face images, then the test face image will give high scores with respect to both the reference face images. Moreover, the chances of matching of face images at the same pose for different persons will be less. Thus the distribution of feature points due to imposter will be more dense as compared the distribution of feature points of the true class. This behavior was exploited in the proposed AANN-based method to identify a test face image. This method does not require the samples from the true class face image to take decision.

The second representation associates edges with the zero-crossings of the second derivative or Laplacian of an image, called filtered image. For matching, the values of the filtered image are brought into the same range using a nonlinear function while preserving the locations of the zero-crossings. The *locality problem* is addressed by considering only the first few eigenvectors derived from the filtered face images. A weighted eigenvector scheme is employed in the process of matching to give more importance to the unique information present in a face image.

Two smoothing filters, namely, zero-frequency filter and the standard Gaussian filter are used to demonstrate the significance of the filtered image based representation. Smoothing of images using the zero-frequency filter is counter intuitive to the traditional smoothing filters, where it is assumed that noise and edge of an image contribute to the high frequency components in the spectrum. On the contrary, the edge information is present throughout the frequency spectrum as illustrated in the processing of a signal using the zero-frequency filter. The zero-frequency filter for 1-D signal is an infinite impulse response (IIR) filter (recursive filter), and its realization for a 2-D signal is not straightforward. It is approximately realized for images either in the spatial domain or in the Fourier domain. Though the zero-frequency filter is realized as an IIR filter in the spatial domain method, it considers the two dimensions of an image independently. This is not an issue in the Fourier domain method, where a finite impulse response (FIR) approximation of the zero-frequency filter is employed to smooth the image. Experimental results of edge map extraction show that estimation of the edge locations and suppression of noise are better in the case of the zero-frequency filter as compared to the Gaussian filter.

The two filtered images, derived using spatial realization of the zero-frequency filter, emphasize the horizontal and vertical edges of a face image separately. The performance of the filtered image along the vertical direction is better than the horizontal direction, as the edges along vertical direction seem to contain more discriminatory information of face images.

The spatial domain method performs better than the Fourier domain. The performance of the filtered face image obtained using the zero-frequency filter realized in the Fourier domain and the standard Gaussian filter is similar.

Another representation is based on the phase of an analytic signal which preserves the impulse-like characteristics of the signal. It is utilized efficiently using the proposed trigonometric (sine and cosine) functions of the phase of an analytic image. The proposed representations bypass the unwrapping problem of the analytic phase. The *locality problem* in these representations is addressed by the method similar to the one used in the case of filter image based representations. Experimental results show that the edge information is preserved better in the sine function as compared to the cosine function of the analytic phase. Two definitions of the analytic images were used to demonstrate the experimental results. The definition (vertical analytic image) which emphasizes the edges along the vertical direction seems to perform better. It reinforces the statement as mentioned earlier that the edges along the vertical direction of a face image contain more discriminatory information of a face image.

Similar methodology is used in the representation of edge information using the phase of the Fourier transform. Here, the importance of the small and large spacings edges is controlled by selecting appropriate DFT coefficients for eigenanalysis based template matching. The performance is improved by combining the evidences from the phase of the analytic image and the phase of the Fourier transform using Borda count based method to exploit the complementary nature of the edge information present in both of them.

In the case of pose variation, the edginess-based representation performs well when the partial evidences obtained from different reference face images are combined using an AANN model. The representation based on the zero-frequency resonator performs best in comparison with other proposed representations under illumination variation of face image. It also shows that the Laplacian of the smoothed image can be directly used for face recognition, without explicitly extracting the edge information from the zero crossing of the Laplacian of the smoothed image. The smoothing of images using the zero-frequency filter shows that the edge information is not just a high frequency information, but it is present at all frequencies, including around the zero frequency. Moreover, it was shown that the complementary nature of the edge information present in both the phase of the analytic image as well as in the phase of the Fourier transform of the image can be exploited for improving the performance of face recognition under illumination variation. It is to be noted that the

performances of three representations, namely, filtered image, phase of analytic image, and phase of the Fourier transform were not evaluated for the pose variation set of face database. Computation of these representations requires eigenanalysis which in turns governed by the pose of the training (reference) face images. Hence, some other processing is required in order to use these representations to address the issue of pose variation in face recognition.

While the work addressed the issue of representation of the edge information for face recognition under variations of pose and illumination, the real issue in practical face recognition is to identify and extract the relevant features, and then use them for matching. Also, in practice, the features used by a human being for face recognition are different for different persons. That is, there is no common feature space, as is assumed in all the face recognition methods. Therefore, we are nowhere near realization of the sophisticated pattern processing used by human beings for face recognition, even though we mentioned that edge information is an important perceptual information in image processing.

7.1 Major contributions

In this thesis we have discussed four representations of face image that capture the edge information implicitly. The major contributions of this thesis are as follows:

- A method to combine the evidences obtained from matching a test face image with different reference face images.
- AANN-based method to identify a given test face image.
- Laplacian of a smoothed face image to capture the edge information of the face image.
- Spatial and Fourier domain methods to realize the zero-frequency filter for 2-D signals.
- A nonlinear normalization function to preserve the locations of the zero-crossings.
- A weighted eigenvector scheme to incorporate the subject-specific unique information of the face image in matching.

- Functions of analytic phase and Fourier phase to exploit the edge information present in the analytic image and Fourier transform of the image.
- The complimentary information present in both the phase of the analytic image and phase of the Fourier transform is exploited to improve the performance of face recognition.

7.2 Directions for further work

The research work can be extended in the following directions:

- In the spatial domain realization of the zero-frequency filter, a zero-frequency resonator is realized along the two directions of an image separately. In principle, it can be realized along any other directions of the image to characterize the edge information of the image efficiently.
- In the proposed two methods of realizing the zero-frequency filter for images, there is a trade-off between IIR realization of the filter and the relationship of pixels values along all directions. This issue may be addressed by recursive realization of the zero-frequency filter in two dimensions.
- The performance of the analytic phase based representation can be improved by using a definition of analytic image which extends the concept of 1-D signum function to 2-D.
- The approach employed in this thesis to make use of the complimentary information present in both Fourier phase and analytic phase needs to be improved.
- The proposed representations may be useful to reduce the differences between gray level face image and the corresponding sketches drawn by the artist.

Appendix A

Zero-frequency resonator

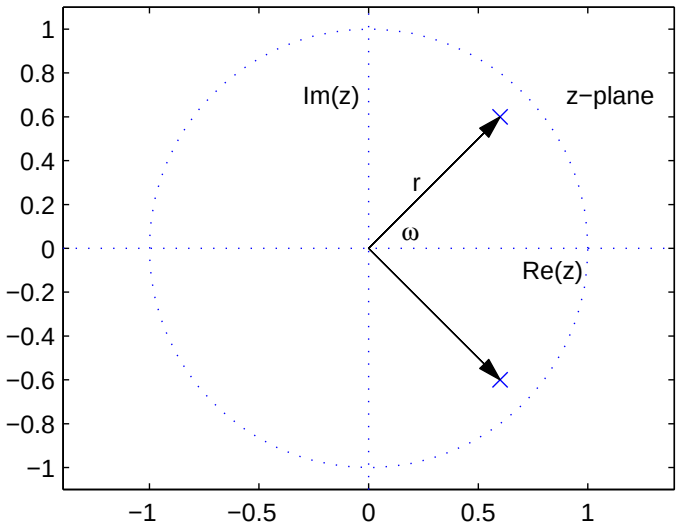


Fig. A.1: Pole-zero configuration for a digital resonator whose resonant frequency is ω.

A digital resonator is an infinite impulse response (IIR) linear system having a complex conjugate pair of poles located inside the unit circle of the z-plane. The angle ω of the poles with abcicssa (Re(z)) of the z-plane decides the resonant frequency of the resonator, while the distance of the poles with the unit circle sets the bandwidth. The closer they are to the unit circle, smaller is the bandwidth. The transfer function ($H(z)$) of a linear system having a pair of complex conjugate poles at $r\cos(\omega) \pm jr\sin(\omega)$ in the z-plane, as illustrated in

Fig. A.1, is given by [138]

$$\begin{aligned} H(z) &= \frac{1}{(1-(r\cos(\omega)+jr\sin(\omega))z^{-1})(1-(r\cos(\omega)-jr\sin(\omega))z^{-1})} \\ &= \frac{1}{1-2r\cos(\omega)z^{-1}+r^2z^{-2}}. \end{aligned} \tag{A.1}$$

For an ideal resonator (minimum bandwidth) with zero resonant frequency, the values of r and ω should be one and zero, respectively. Thus

$$H(z) = \frac{1}{1-2z^{-1}+z^{-2}}. \tag{A.2}$$

Appendix B

Sign gradient principle

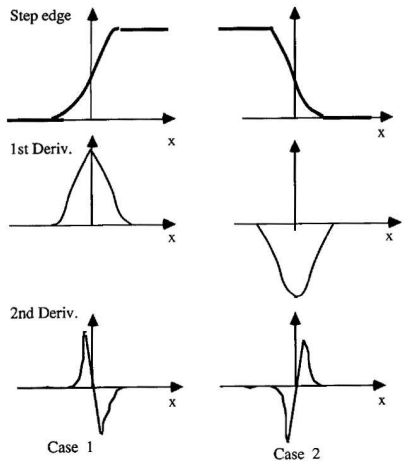


Fig. B.1: Sign correspondence between the first and second derivative.

The step edge can be detected using the zero-crossings of the Laplace operation. Some false zero-crossings, that are difficult to remove by slope thresholding, can be suppressed using the sign correspondence principle. This principle is demonstrated in Fig. B.1, where two different cases of the smoothed steps edges and its first and second order derivatives are shown. In case 1, the first derivative is positive and the second derivative changes its sign from positive to negative. Such zero-crossings are defined as negative zero-crossing

(NZC). In case 2, the first derivative is negative and the second derivative changes its sign from negative to positive. This corresponds to a positive zero-crossing (PZC) in the second derivative. Thus for a step edge, if it has positive first derivative, it must correspond to a PZC in the second derivative. On the other hand for negative first derivative, it must correspond to NZC in the second derivative. The converse should also hold. This principle is called ”zero-crossing sign correspondence principle [137]”.

References

[1] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, no. 9, pp. 1209–1230, Sep. 1994.

[2] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, Jan. 1992.

[3] A. W. Young and H. D. Ellis, *Handbook of Research on Face Processing*. New York, USA: Elsevier, 1989.

[4] J. C. Bartlett, S. Hurry, and W. Thorley, "Typicality and familiarity of faces," *Memory and Cognition*, vol. 12, no. 3, pp. 219–228, Mar. 1984.

[5] J. Sergent, S. Ohta, B. Macdonald, and E. Zuck, "Segregated processing of facial identity and emotion in the human brain: A pet study," *Visual Cognition*, vol. 1, no. 2-3, pp. 349–369, Apr. 1994.

[6] R. K. Yin, "Looking at upside-down faces," *J. Experimental Psychology*, vol. 81, no. 1, pp. 141–145, Jul. 1969.

[7] J. Kurucz and J. Feldmar, "Prospo-affective agnosia as a symptom of cerebral organic brain disease," *J. American Geriatrics Society*, vol. 27, no. 5, pp. 225–230, May 1979.

[8] A. Slater and P. C. Quinn, "Face recognition in the newborn infant," *Infant and Child Development*, vol. 10, no. 1-2, pp. 21–24, Apr. 2001.

[9] A. O'Toole, K. D. Dominique, and D. Valentine, "Structural aspects of face recognition and the other race effect," *Memory and Cognition*, vol. 22, no. 2, pp. 208–224, Feb. 1994.

[10] A. J. O'toole, K. A. Deffenbacher, D. Valentin, K. Mckee, D. Huff, and H. Abdi, "The perception of face gender: The role of stimulus structure in recognition and classification," *Memory and Cognition*, vol. 26, pp. 146–160, 1997.

[11] J. J. Gibson, "On the concept of formless invariants in visual perception," *LEONARDO*, vol. 6, no. 1, pp. 43–45, Jan. 1973.

[12] R. C. Gonzalez and R. E. woods, *Digital Image Processing*. Boston, USA: Addison-Wesley Publishing Company Inc., 1992.

[13] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 11, pp. 679–698, Nov. 1986.

[14] D. C. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, pp. 187–217, 1980.

[15] V. Bruce, E. Hanna, N. Dench, P. Healey, and M. Burton, "The importance of mass in line drawings of faces," *Appl. Cognitive Psychology*, vol. 6, no. 7, pp. 619–628, Dec. 1992.

[16] S. E. Brennan, "Caricature generator: The dynamic exaggeration of faces by computers," *LEONARDO*, vol. 18, no. 3, pp. 170–178, 1985.

[17] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE. Trans. Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, Nov. 2008.

[18] A. V. Oppenheim and I. S. Lim, "The importance of phase in signals," *Proc. of the IEEE*, vol. 69, no. 5, pp. 529–562, May 1981.

[19] S. F. Galton, "Personal identification and description," *Nature*, pp. 173–188, Jun. 1888.

[20] R. G. U. Jr. and N. da Vitoria Lobo, "A framework for recognizing a facial image from a police sketch," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Franciso, CA, USA, 1996, pp. 586–593.

[21] W. Konen, "Comparing facial drawings with gray-level images: A case studies on PHANTOMAS," in *Proc. Int. Conf. Artificial Neural Networks*, Bochum, Germany, 1996, pp. 727–734.

[22] Y. Li and H. Kobatake, “Extraction of facial sketch image based on morphological processing,” in *Proc. Int. Conf. Image Processing*, Santa Barbara, CA, USA, 1997, pp. 316–319.

[23] Z. Xu, H. Chen, and S.-C. Zhu, “A high resolution grammatical model for face representation and sketching,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005, pp. 470–477.

[24] S. P. Abney, “Stochastic attribute-value grammars,” *Computational Linguistics*, vol. 23, no. 4, pp. 597–618, 1997.

[25] Z. Xu and J. Luo, “Face recognition by expression-driven sketch graph matching,” in *Proc. Int. Conf. Pattern Recognition*, vol. 3, Hong Kong, China, 2006, pp. 1119–1122.

[26] Z. Xu, H. Chen, S.-C. Zhu, and J. Luo, “A hierarchical compositional model for face representation and sketching,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 6, pp. 955–969, Jun. 2008.

[27] X. Tang and X. Wang, “Face photo recognition using sketch,” in *Proc. Int. Conf. Image Processing*, New York, USA, 2002, pp. 257–260.

[28] ——, “Face sketch recognition,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[29] ——, “Face sketch synthesis and recognition,” in *Proc. Int. Conf. Image Processing*, Nice, France, 2003, pp. 687–694.

[30] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, “A nonlinear approach for face sketch synthesis and recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, USA, 2005, pp. 1005–1010.

[31] J. Zhong, X. Gao, and C. Tian, “Face sketch synthesis using E-HMM and selective ensemble,” in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, 2007, pp. 485–488.

[32] X. Wang and X. Tang, “Face photo-sketch synthesis and recognition,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[33] W. Liu, X. Tang, and J. Liu, “Bayesian tensor inference for sketch-based facial photo hallucination,” in *Proc. 20th Int. Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, 2007, pp. 2141–2146.

[34] Y. Li, M. Savvides, and B. V. K. V. Kumar, “Illumination tolerant face recognition using novel face from sketch synthesis approach for advanced correlation filters,” in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Toulouse, France, 2006, pp. 357–360.

[35] A. Mahalanobis, B. V. K. V. Kumar, and D. Casasent, “Minimum average correlation energy filters,” *Appl. Opt.*, vol. 26, no. 17, pp. 3633–3640, Sep. 1987.

[36] E. H. Gombrich, *Art and Illusion*. Princeton University Press: Princeton, New Jersey, 1960.

[37] G. Rhodes, *Superportraits: Caricature and Recognition*. Psychology Press, 1996.

[38] P. Y. Chiang, W. H. Lio, and T. Y. Li, “Automatic caricature generation by analyzing facial features,” in *Proc. Asian Conf. Computer Vision*, Jeju, KOREA, 2004.

[39] A. K. Dewdney, “Computer recreations,” *Scientific American*, pp. 20–28, Jul. 1986.

[40] D. E. Pearson and J. Robinson, “Visual communication at very low data rates,” *Proc. of the IEEE*, vol. 73, no. 4, pp. 795–816, Apr. 1985.

[41] L. Liang, H. Chen, Y.-Q. Xu, and H.-Y. Shum, “Example-based caricature generation with exaggeration,” in *Proc. Pacific Cong. Computer Graphics and Applications*, New York, USA, 2002, pp. 386–393.

[42] H. Koshimizu, M. Tominaga, and T. Fujiwara, “On KANSEI facial image processing for computerized facial sketching system PICASSO,” in *Proc. IEEE Int. Conf. System Man,and Cybernetics*, Tokyo, 1999, pp. 294–299.

[43] K. H. Lai, P. W. H. Chung, and E. A. Edirisinghe, “Novel approach to neural network based caricature generation,” in *Proc. IET Int’l Conf. on Visual Information Engineering*, Bangalore, India, 2006, pp. 88–93.

[44] S. Iwashita and T. Onisawa, “A study on facial caricature drawing by fuzzy theory,” in *Proc. sixth IEEE Int. Conf. Fuzzy System*, Barcelona, Spain, 1997, pp. 933–938.

[45] Y.-L. Chen, W.-H. Liao, and P.-Y. Chiang, “Generation of 3D caricature by fusing caricature images,” in *Proc. IEEE Int. Conf. System Man, and Cybernetics*, Taipei, Taiwan, 2006, pp. 866–871.

[46] B. Takacs, "Comparing face images using modified Hausdorff distance," *Pattern Recognition*, vol. 31, no. 12, pp. 1873–1881, Dec. 1998.

[47] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 6, pp. 765–779, Jun. 2002.

[48] A. Yilmaz and M. Gokmen, "Eigenhill vs. eigenface and eigenedge," *Pattern Recognition*, vol. 34, no. 1, pp. 181–184, Jan. 2000.

[49] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using Gradientfaces," *IEEE Trans. on Image Processing*, vol. 18, no. 11, pp. 2599–2606, Nov. 2009.

[50] T. Sim, S. Baker, and M. Bsat, "The CMU pose illumination, and expression PIE database of human faces," *Tech. Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University*, January 2001.

[51] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[52] G. Gordon, "Face recognition based on depth maps and surface curvature," in *Proc. SPIE Geometric Computer Vision*, vol. 1570, San Diego, CA, USA, 1991, pp. 234–247.

[53] C. Samir, A. Srivastava, and M. Daoudi, "Three-dimensional face recognition using shapes of facial curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 11, pp. 1858–1863, Nov. 2006.

[54] E. Klassen, A. Srivastava, W. Mio, and S. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 3, pp. 372–383, Mar. 2004.

[55] J. Wilder, P. J. Philips, C. H. Jiang, and S. Wiener, "Comparison of visible and infrared images for face recognition," in *IEEE Int. Conf. Face and Gesture Recognition*, Killington, Vermont, USA, 1996, pp. 182–187.

[56] S. Gundimada and V. K. Asari, "Facial recognition using multisensors images based on localized kernel eigen spaces," *IEEE Trans. on Image Processing*, vol. 18, no. 6, pp. 1314–1325, Jun. 2009.

[57] W. W. Bledsoe, “The model method in facial recognition,” in *Parametric Research Inc. Tech Rep. PRI: 15*, Palo Alto, CA, 1964.

[58] M. D. Kelly, “Visual information of people by computer,” in *Tech Rep. AI-130, Stanford AI Proj*, CA, 1970.

[59] T. Kanade, “Computer recognition of human faces,” in *Tech Rep. AI-130, Stanford AI Proj*, Birkhauser Verlog, Stuttgart, Germany, 1977.

[60] R. Brunelli and T. Poggio, “Face recognition: Features versus templates,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.

[61] H. M. L. K. H. Wong and P. W. M. Tsang, “A system for recognizing human faces,” in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Glasgow, Scotland 1989, pp. 1638–1641.

[62] L. Wang and Chin-Tu Chen and W. Lin, “An efficient algorithm to compute set of discrete Gabor coefficients,” *IEEE Trans. on Image Processing*, vol. 1, no. 1, pp. 87–92, Jan. 1994.

[63] T. S. Lee, “Image representation using 2D Gabor wavelets,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[64] B. S. Manjunath, R. Chellappa, and C. V. D. Malsburg, “A feature based approach to face recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Champaign, IL, USA, 1992, pp. 373–378.

[65] M. Lades, J. C. Vorbrüggen, J. M. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” *IEEE Trans. Computers*, vol. 42, no. 3, pp. 300–311, Mar. 1993.

[66] L. Wiskott and C. von der Malsburg, “Recognizing faces by dynamic link matching,” *NeuroImage*, vol. 4, no. 3, pp. S14–S18, Dec. 1996.

[67] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[68] http://www.nist.gov/humanid/colorferet/home.html.

[69] B. Yegnanarayana, *Artificial Neural Networks*. New Delhi: Prentice-Hall, 1999.

[70] S. Haykin, *Neural Networks - A Comprehensive Foundation*. New York, USA: Macmillan College Publishing Company Inc., 1994.

[71] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press, 1995.

[72] T. Kohonen, *Self-Organization and Associative Memory (Third Edition)*. Berlin, Heidelberg: Springer, 1989.

[73] T. J. Stonham, "Practical face recognition and verification with WISARD," in *Proc. Aspect of face processing*, Boston, USA, 1984, pp. 426–441.

[74] J. J. Weng, N. Ahuja, and T. S. Huang, "Learning recognition and segmentation using the cresceptron," *Int. J. Comput. Vision*, vol. 25, no. 2, pp. 109–143, Nov. 1997.

[75] G. W. Cottrell and M. K. Flemming, "Face recognition using unsupervised feature extraction," in *Proc. Int. Conf. Neural Network*, Paris, France, 1990, pp. 322–325.

[76] ——, "Categorization of faces using unsupervised feature extraction," in *Proc. Int. Conf. Neural Network*, San Diego, USA, 1990, pp. 65–70.

[77] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. on Neural Networks*, vol. 8, no. 1, pp. 114–131, Jan. 1997.

[78] http://www.cam-orl.co.uk/facedatabase.html.

[79] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Trans. on Neural Networks*, vol. 8, no. 1, pp. 98–112, Jan. 1997.

[80] S. Gutta, J. R. J. Huang, and H. Wechsler, "Mixture experts for classification of gender, ethnic origin and pose of human faces," *IEEE Trans. on Neural Networks*, vol. 11, no. 7, pp. 948–959, Jul. 2000.

[81] L. Rabiner, B.-H. Juang, and B. Yegnanarayana, *Fundamentals of Speech Recognition*. Dorling Kindersley, India: Pearson Education, 2009.

[82] F. Samaria and F. Fallside, "Face identification and feature extraction using hidden Markov models," in *Image Processing: Theory and Applications*. Elsevier, 1993, pp. 292–302.

[83] F. Samaria and A. Harter, "Parameterization of a stochastic model for human face identification," in *Proc. Second IEEE Workshop Applications of Computer Vision*, Sarasota, FL, USA, 1994, pp. 138–142.

[84] V. Bevilacqua, L. Cariello, G. Carro, D. Daleno, and G. Mastronardi, "A face recognition system based on pseudo 2D HMM applied to neural network coefficients," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 22, no. 7, pp. 615–621, 2007.

[85] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 733–742, Jul. 1997.

[86] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in *Proceedings of Int. Conf. Automatic Face and Gesture Recognition*, France, 2000, pp. 285–292.

[87] T. Vetter and V. Blanz, "Estimating colored 3D face models from single images: An example based approach," in *Proc. European Conf. Computer Vision ECCV'98*, Germany, 1998, pp. 499–513.

[88] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[89] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, Washington, USA, 2002, pp. 46–51.

[90] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York, USA: John Wiley and Sons, 1973.

[91] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, USA: Academic Press, 1972.

[92] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[93] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[94] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.

[95] ——, "Probabilistic visual learning for object detection," in *Proc. Int. Conf. Computer Vision*, Cambridge, MA, 1995, pp. 786–793.

[96] N. Vaswani and R. Chellappa, "Principal component null space analysis for image and video classification," *IEEE Trans. on Image Processing*, vol. 15, no. 17, pp. 1816–1830, Jul. 2006.

[97] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[98] D. Zhang, Z.-H. Zhou, and S. Chen, "Diagonal principal component analysis for face recognition," *Pattern Recognition*, vol. 39, no. 1, pp. 140–142, Jan. 2006.

[99] D. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[100] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[101] S. Raudys and R. P. W. Duin, "Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix," *Pattern Recognition Letters*, vol. 19, no. 5, pp. 385–392, Apr. 1998.

[102] D.-Q. Dai and P. C. Yuen, "Regularized discriminant analysis and its application to face recognition," *Pattern Recognition*, vol. 36, no. 3, pp. 845–847, Mar. 2003.

[103] C. Xiang, X. Fan, and T. Lee, "Face recognition using recursive Fisher linear discriminant," *IEEE Trans. on Image Processing*, vol. 15, no. 8, pp. 2097–2105, Aug. 2006.

[104] H. Kong, L. Wang, E. K. Teoh, J. G. Wang, and R. Venkateswarlu, "A framework of 2-D Fisher discriminant analysis: Application to face recognition with small number of training samples," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, USA, 2005, pp. 1083–1088.

[105] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, Jun. 2002.

[106] M. Bartlett, H. Lades, and T. Sejnowski, "Independent component representations for face recognition," in *Proc. SPIE*, vol. 3299, San Jose, CA, USA, 1998, pp. 528–539.

[107] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, USA: John Wiley Sons, 2001.

[108] X. Lu, Y. Wang, and A. Jain, "Combining classifiers for face recognition," in *IEEE Int'l Conf. Multimedia and Expo.*, vol. 3, Baltimore, 2003, pp. 13–16.

[109] S. Z. Li, X. Lu, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," *IEEE Trans. on Image Processing*, vol. 14, no. 6, pp. 705–712, Jun. 2005.

[110] Q. Gao, L. Zhang, and D. Zhang, "Sequential row column independent component analysis for face recognition," *Neurocomputing*, vol. 72, no. 4-6, pp. 1152–1159, Jan. 2009.

[111] M. A. Kramer, "Nonlinear principal components analysis using autoassociative neural networks," *AIChe Journal*, vol. 32, no. 2, pp. 233–243, Feb. 1991.

[112] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, United Kingdom: Cambridge university press, 2004.

[113] M.-H. Yang, "Kernel eigenfaces vs kernel Fisherfaces: Face recognition using kernel methods," in *IEEE Int. Conf. Face and Gesture Recognition*, Washington, DC, 2002, pp. 215–220.

[114] S. Z. Li. and A. K. Jain, *Handbook of Face Recognition*. New York, USA: Springer-Verlag, 2005, ch. 7.

[115] J. Black, M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan, "A framework for performance evaluation of face recognition algorithm," in *ITCOM Internet Multimedia System*, Boston, 2002.

[116] G. Little, S. Krishna, J. Black, and S. Panchanathan, "A methodology for evaluating robustness of face recognition algorithms with respect to change in pose and illumination angle," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Philadelphia, USA, 2005, pp. 89–92.

[117] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.

[118] A. K. Sao and B. Yegnanarayana, "Face verification using correlation filter and auto associative neural networks," in *Proc. Int. Conf. Intelligent Sensing and Information Processing*, Chennai, India, 2004, pp. 364–368.

[119] S. Ramesh, S. Palanivel, S. Das, and B.Yegnanarayana, "Eigenedginess vs. eigenhill, eigenface and eigenedge," in *European Signal Processing Conference*, Toulose, France, 2002, pp. 559–562.

[120] P. K. Kumar, S. Das, and B. Yegnanarayana, "One-dimensional processing of images," in *Proc. Int. Conf. Multimedia Processing and Systems*, Chennai, India, 2000.

[121] A. K. Sao, "Significance of Image Representation for Face Recognition," Master's thesis, Department of Computer Science and Engineering, Indian Institute of Technology Madras, India, Sep. 2003.

[122] C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 66–71.

[123] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, Aug. 1981.

[124] R. Courant and D. Hilbert, "Methods of mathematical physics," *Interscience, New York*, vol. 1, 1953.

[125] A. H. Charles and T. A. Porsching, *Numerical Analysis of Partial Differential Equations*. Engelwood Cliffs, New Jersey: Prentice-Hall, 1990.

[126] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[127] C. Sanderson, S. Bengio, and Y. Gao, "On transforming statistical models for non-frontal face verification," *Pattern Recognition*, vol. 39, no. 2, pp. 288–302, Feb. 2006.

[128] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 733–742, Jul. 1997.

[129] J. Zhang, Y. Yan, and M. Lades, "Face recognition: Eigenface, elastic graph matching and neural network," *Proc. of the IEEE*, vol. 85, no. 9, pp. 1423–1435, Sep. 1997.

[130] O. Deniz, M. Castrilon, and M. Hernandwez, "Face recognition using independent component analysis and support vector machine," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2153–2157, Sep. 2003.

[131] K. I. Kim, K. Jung, and K. Kim, "Face recognition using support vector machine and with local correlation kernels," *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 1, pp. 97–111, Feb. 2002.

[132] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Face verification using correlation filters," in *Proc. Third IEEE Automatic Identification Advanced Technologies*, Tarrrytown, NY, 2002, pp. 56–61.

[133] ——, "Robust shift-invariant biometric identification from partial face images," in *Biometric Technologies for Human Identification OR51, SPIE Defense and Security Symposium*, Aug. 2004, pp. 124–135.

[134] B. V. K. V. Kumar, M. Savvides, K. Venkataramani, and C. Xie, "Spatial frequency domain image processing for biometric recognition," in *IEEE Int. Conf. Image Processing*, New York, USA, 2002, pp. 53–56.

[135] B. Yegnanarayana and S. P. Kishore, "AANN: An alternative to GMM for pattern recognition," *Neural Networks*, vol. 15, pp. 459–469, Apr. 2002.

[136] B. Yegnanarayana, K. S. R. Murty, and S. Rajendran, "Analysis of stop consonants in Indian languages using excitation source information in speech signal," in *Proc. ISCA-ITRW Workshop on Speech Analysis and Processing for Knowledge Discovery*, Aalborg university, Denmark, 2008.

[137] J. Shen and S. Castan, "An optimal linear operator for step edge detection," *Computer, Vision, Graphics, and Image Processing: Graphical Models and Understanding*, vol. 54, no. 2, pp. 112–133, Mar. 1992.

[138] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. New Delhi: Prentice Hall of India, 1997.

[139] A. Bovik, *Handbook of Image and Video Processing*. New York: Academic Press, 2000, ch. 4.11.

[140] X. Wang, "Laplacian operator-based edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 6, pp. 886–890, May 2007.

[141] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Seattle, USA, 1994, pp. 84–89.

[142] A. O'Toole, H. Abdi, K. Deffenbacher, and D. Valentin, "Low-dimensional representation of faces in higher dimensions of the face space," *J. Opt. Soc. Am. A*, vol. 10, no. 3, pp. 405–410, Mar. 1993.

[143] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.

[144] T. Bulow and G. Sommer, "Hypercomplex signals- A novel extension of the analytic signal to the multidimensional case," *IEEE Trans. on Signal Processing*, vol. 29, no. 11, pp. 2844–2852, Nov. 2001.

[145] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - part 1: Fundamentals," *Proc. of the IEEE*, vol. 80, no. 4, pp. 520–538, Apr. 1992.

[146] J. Leitao and M. A. Figueiredo, "Interferometric image reconstruction as a nonlinear Bayesian estimation problem," in *Proc. Int. Conf. Image Processing*, Washington, DC, USA, 1995, pp. 453–456.

[147] M. N. Nabighian, "The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: Its properties and use of automated anomaly interpretation," *Geophysics*, vol. 37, no. 3, pp. 507–517, Jun. 1972.

[148] ——, "Toward a three-dimensional magnetic automatic interpretation of potential field data via generalized Hilbert transform: Fundamental relations," *Geophysics*, vol. 49, no. 6, pp. 780–786, Jun. 1984.

[149] S. Haykin, *Communication Systems*. New Delhi: Wiley Eastern Limited, 1994.

[150] K. G. Larkin, D. J. Bone, and M. A. Oldfield, "Natural demodulation of two-dimensional fringe patterns. I-general background of the spiral phase quadrature transform," *J. Opt. Soc. Am. A*, vol. 18, no. 8, pp. 1862–1870, Aug. 2001.

[151] M. S. Pattichis and A. C. Bovik, "Analyzing image structure by multidimensional frequency modulation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 5, pp. 753–766, May 2007.

[152] D. Gabor, "Theory of communication," *Journal of IEE*, vol. 93, no. 26, pp. 429–457, Nov. 1946.

[153] L. Cohen, *Time-Frequency Analysis*. New Jersey: Prentice Hall, 1995.

[154] A. Bovik, *Handbook of Image and Video Processing*. Academic Press, 2000, ch. 4.4.

[155] H. Stark, "An extension of the Hilbert transform product theorem," *Proc. of the IEEE*, vol. 59, no. 9, pp. 1359–1360, Sep. 1971.

[156] S. L. Hahn, "Multidimensional complex signals with single-orthant spectra," *Proc. of the IEEE*, vol. 80, no. 8, pp. 1287–1300, Aug. 1992.

[157] K. G. Larkin, "Natural demodulation of two-dimensional fringe patterns. II-stationary phase analysis of the spiral phase quadrature transform," *J. Opt. Soc. Am. A*, vol. 18, no. 8, pp. 1871–1881, Aug. 2001.

[158] M. Craig, "Analytic signals for multivariate data," *Math. Geol*, vol. 28, no. 3, pp. 315–329, Apr. 1996.

[159] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, no. 4, pp. 713–720, 1988.

[160] G. Fornaro, G. Franceschetti, and R. Lanari, "Interferometric SAR phase unwrapping using green's formulation," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 3, pp. 720–727, May 1996.

[161] D. Massonet and T. Rabaute, "Radar interferometry: Limits and potential," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, no. 2, pp. 454–464, Mar. 1993.

[162] S. Pandit, N. Jordache, and G. Joshi, "Data-dependent systems methodology for noise-insensitive phase wrapping in laser interferometric surface characterization," *J. Opt. Soc. Am. A*, vol. 11, no. 10, pp. 2584–2592, Oct. 1994.

[163] S. M.-H. Song, S. Napel, N. J. Pelc, and G. H. Glover, "Phase unwrapping of MR phase images using Poisson equation," *IEEE Trans. on Image Processing*, vol. 4, no. 5, pp. 667–676, Apr. 1995.

[164] J. Y. Wang, "Optical resolution through a turbulent medium with adaptive phase compensations," *J. Opt. Soc. Am. A*, vol. 67, no. 3, pp. 383–390, Mar. 1977.

[165] M. H. Lee, J. F. Holmes, and J. R. Kerr, "Statistics of speckle propagation through turbulent atmosphere," *J. Opt. Soc. Am. A*, vol. 66, no. 11, pp. 1164–1172, Nov. 1976.

[166] J. M. Tribolet, "A new phase unwrapping algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 2, pp. 170–177, Feb. 1977.

[167] M. Hedley and D. Rosenfield, "A new two-dimensional phase unwrapping algorithm for MRI images," *Magnet. Res. Med.*, vol. 24, no. 1, pp. 177–181, Mar. 1992.

[168] D. J. Bone, "Fourier fringe analysis: the two-dimensional phase unwrapping problem," *Appl. Opt.*, vol. 30, no. 25, pp. 3627–3632, Sep. 1991.

[169] J. M. Huntley, "Noise-immune phase unwrapping algorithm," *Appl. Opt.*, vol. 28, no. 16, pp. 3268–3270, Aug. 1989.

[170] D. L. Fried, "Least square fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Am. A*, vol. 67, no. 3, pp. 370–374, Mar. 1977.

[171] B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *J. Opt. Soc. Am. A*, vol. 69, no. 3, pp. 393–399, Mar. 1979.

[172] J. L. Marroquin and M. Rivera, "Quadratic regularization functionals for phase unwrapping," *J. Opt. Soc. Am. A*, vol. 12, no. 11, pp. 2393–2400, Nov. 1995.

[173] M. D. Pritt and J. S. Shipman, "Least-squares two-dimensional phase unwrapping using FFT," *IEEE. Trans. Geosci. Remote Sensing*, vol. 32, no. 3, pp. 706–708, May 1994.

[174] B. Friedlander and J. M. Francos, "Model based unwrapping of 2-D signals," *IEEE Trans. on Signal Processing*, vol. 44, no. 1, pp. 2099–3007, Jan. 1996.

[175] Z. P. Liang, "A model based method for phase unwrapping," *IEEE Trans. Med. Imaging*, vol. 15, no. 6, pp. 893–897, Dec. 1996.

[176] B. Yegnanarayana, D. Saikia, and T. Krishnan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 3, pp. 610–623, Jun. 1984.

[177] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Eigenphases vs. eigenfaces," in *Proc. Int. Conf. on Pattern Recognition*, Cambridge, UK, 2004, pp. 810–813.

[178] A. Lumini and L. Nanni, "Detector of image orientation based on borda count," *Pattern Recognition Letters*, vol. 27, no. 3, pp. 180–186, Feb. 2006.

[179] Q. Fang, P. M. Meaney, and K. D. Paulsen, "The multidimensional phase unwrapping integral and applications to microwave topographical image reconstruction," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3311–3324, Nov. 2006.

[180] B. Yegnanarayana and A. Raghunathan, "Representation of images through group delay functions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 2, pp. 237–240, Feb. 1987.

[181] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, July 2004.

[182] J. Kim, K. Seo, and K. Chung, "A systematic approach to classifier selection on combining multiple classifiers for handwritten digit recognition," in *Proc. Int. Conf. on Document Analysis and Recognition*, Washington, DC, USA, 1997, pp. 459–462.

[183] E. M. Renda and U. Straccia, "Web meta search: Rank vs. score based rank aggregation methods," in *Proc. of 2003 ACM symposium on Applied computing*, New York, USA, 2003, pp. 841–846.

[184] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

List of Publications

Papers in refereed Journals

1. A. K. Sao and B. Yegnanarayana, “Face verification using template matching”, *IEEE Trans. Information Forensics and Security*, vol. 2, no. 3, pp. 636-641, Mar. 2007.

2. A. K. Sao, B. Yegnanarayana and B. V. K. Vijay Kumar , “Significance of image representation for face recognition”, *Signal, Image and Video Processing* , vol. 1, no. 3, pp. 225-237, 2007.

3. A. K. Sao and B. Yegnanarayana, “On the use of phase of the Fourier transform for face recognition under variations in illumination ”, accepted for publication in *Signal, Image and Video Processing*.

4. A. K. Sao and B. Yegnanarayana, “Edge extraction using zero-frequency resonator”, communicated to *Signal, Image and Video Processing*, Oct. 2009.

5. A. K. Sao and B. Yegnanarayana, “Laplacian of smoothed image as representation for face recognition ”, communicated to *Journal of Visual Communication and Image Representation*, Nov. 2009.

Papers in Conferences

1. A. K. Sao and B. Yegnanarayana, “Template matching approach for pose problem in face verification”, in Proc. *International Workshop on Multimedia Content Representation Classification and Security (MRCS)*, Istanbul, Turkey, 2006, pp. 191-198.

2. A. K. Sao and B. Yegnanarayana, “Analytic phase-based representation for face recognition”, in Proc. *International Conference on Advances in Pattern Recognition (ICAPR 2009)* , Kolkata, India, 2009, pp. 453-456.