HIGH ORDER VOLUMETRIC DIRECTIONAL PATTERN FOR ROBUST FACE RECOGNITION

Dissertation

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Doctor of Philosophy in Engineering

By

Almabrok Essa Essa

UNIVERSITY OF DAYTON

Dayton, Ohio

August, 2017

HIGH ORDER VOLUMETRIC DIRECTIONAL PATTERN FOR ROBUST FACE RECOGNITION

Name: Essa, Almabrok Essa

APPROVED BY:

Vijayan K. Asari, Ph.D.
Advisor Committee Chairman
Professor, Department of Electrical and Computer Engineering

Russell Hardie, Ph.D.
Committee Member
Professor, Department of Electrical and Computer Engineering

Eric J. Balster, Ph.D.
Committee Member
Associate Professor, Department of Electrical and Computer Engineering

Youssef Raffoul, Ph.D.
Committee Member
Professor, Department of Mathematics

Robert J. Wilkens, Ph.D., P.E.
Associate Dean for Research and Innovation
Professor, School of Engineering

Eddy Rojas, Ph.D., M.A., P. E.
Dean
School of Engineering

ABSTRACT

HIGH ORDER VOLUMETRIC DIRECTIONAL PATTERN FOR ROBUST FACE RECOGNITION

Name: Essa, Almabrok Essa
University of Dayton

Advisor: Dr. Vijayan K. Asari

The texture of objects in digital images is an important property that has been utilized in many computer vision and image analysis applications, such as pattern recognition, object classification, and region segmentation. Despite its frequent usage and many attempts to describe it in general terms, the texture lacks a precise definition. This makes the development of new texture descriptors a big challenge. In addition, researchers' interest has recently spread into the dynamic texture (video domain), where the problem becomes more challenging.

The main goal of feature description and representation techniques is to extract features from the image that are distinct and stable under different conditions during the image acquisition process. Texture descriptors can be generally classified into structural and statistical approaches. The structural methods consider the texture as a repetition of some primitives, with a specific rule of placement, while the statistical techniques characterize the stochastic properties of the spatial distribution of gray levels in an image using the gray tone co-occurrence matrix. In this work, we propose a combination of the structural and statistical approaches that can be utilized to recognize a variety of different textures, named High Order Local Directional Pattern (HOLDP) for still image based

feature extraction (static texture) as well as High Order Volumetric Directional Pattern (HOVDP) for video based feature extraction (dynamic texture).

Recently, the conventional Local Directional Pattern (LDP) has received a great deal of attention in face recognition applications. However, it only describes the micro structures of the texture images, because it considers only a small neighborhood size. In fact, our proposed HOLDP descriptor can capture more detailed discriminative information by not only extracting the micro structures but also the macro structures of the texture images, which can be done by the help of a pyramidal multi-structure approach. The pyramid based multi-structure presented in this dissertation research can be created by encoding the directional information from different neighborhood layers of the image for each pixel position, and then concatenating the feature vectors of each neighborhood layer to form the final HOLDP feature map.

Identifying human faces in video is a challenging problem due to the presence of large variations in facial pose and expression, as well as poor video resolution. To address this, Volumetric Directional Pattern (VDP) is proposed [1]. VDP is an oriented volumetric descriptor that is able to extract and fuse the information of multiple frames, temporal (dynamic) information, and multiple poses and expressions of faces in input videos to produce strong feature vectors. Meanwhile, to demonstrate the generality and capability of the HOLDP method, we develop another novel video based feature extraction technique, namely High Order Volumetric Directional Pattern (HOVDP) as an extension of VDP. HOVDP combines the movement and appearance features together by considering the n^{th} order directional variation patterns of all neighboring pixel layers from three consecutive frames. From extensive experiments on still image based and video based face recognition benchmarks, we demonstrate the excellent performance of our proposed techniques compared to the state-of-the-art approaches.

To my respectiveful parents − Essa Essa and Aisha Hammad

(*I hope that this achievement will fulfill the dream you had for me when I was a kid*)

To my supportive brothers and sisters − Mohamed, Ali, Fatima,

Rabia, Najib, Mustafa, Adel, Walid, and Yasin

(*I hope that this success will reward your support through all circumstances*)

To my beloved wife − Houyam Elhrari

(*I hope that this accomplishment will be our ever-standing glory*)

To my lovely daughters − Raghad Essa and Robeen Essa

(*I hope that this gift will inspire you along the way of your success*)

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude and warmest affection to Professor Vijayan K. Asari who has been a fountain of knowledge and inspiration.

I would like to thank my Ph.D. advisory committee members, Dr. Russell Hardie, Dr. Eric J. Balster and Dr. Youssef Raffoul for their valuable suggestions. My grateful appreciation is extended to UD Vision Lab colleagues, all the faculty, staff and graduate students in this department, from whom I learned so much.

I also would like to extend my sincere appreciation for the generous financial support from my country represented by Libyan Ministry of Higher Education and Scientific Research. I look forward to serve the country and to share the knowledge with the next generations.

Finally, I would like to express my deep love and appreciation to my parents, my brothers, my sisters, my wife, my daughters, and my dearest friends for their love, concern and encouragement throughout my study. I am truly thankful for having all of you in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

(P, R) P Number of Neighbor Pixels, R The Radius

AHOLDP Adaptive High Order Local Directional Pattern

AHOVDP Adaptive High Order Volumetric Directional Pattern

DT Dynamic Texture

ELDP Enhanced Local Directional Pattern

FR Face Recognition

GGDA Grassmannian Graph-Embedding Discriminant Analysis

HOLDP High Order Local Directional Pattern

HOVDP High Order Volumetric Directional Pattern

ICA Independent Component Analysis

IED Image Encoding & Decoding

k-NN k-Nearest Neighbors

KSANP Kernel Sparse Approximated Nearest Points

LBF Local Boosted Features

LBP Local Binary Pattern

LDP Local Directional Pattern

LDPv Local Directional Pattern Variance

LIBSVM Library for Support Vector Machines

PCA Principal Component Analysis

PML Projection Metric Learning

RKHS Reproducing Kernel Hilbert Space

RNP Regularized Nearest Points

SANP Sparse Approximated Nearest Points

SVM Support Vector Machine

VDP Volumetric Directional Pattern

CHAPTER I

INTRODUCTION

1.1 Image Textures

The texture of objects in digital images can be generally categorized into two main types, static texture and dynamic texture which is an extension of texture to the temporal domain. Local feature detection and description has gained much attention in recent years since photometric descriptors computed for regions of interest have proven to be very successful in many computer vision applications. In the context of texture (feature) analysis methods, there are two common types of techniques: 1) The structural approaches, where the image texture is considered as a repetition of some primitives with a specific rule of placement; and 2) The statistical methods. The stochastic properties of the spatial distribution of gray levels in an image are characterized by the gray tone co-occurrence matrix. A set of textural features derived from the co-occurrence matrix is widely used to extract textural information from digital images [2].

1.2 Face Recognition

Face recognition (FR) is one of the most suitable technologies that has been spread in several applications such as biometric systems, access control and information security systems, surveillance systems, content-based video retrieval systems, credit-card verification systems, and more generally image understanding. FR is a biometric approach that employs automated methods to verify

or recognize the identity of a living person based on their physiological characteristics. The key of each face recognition system is the feature extractors, which should be distinct and stable under different conditions, such as illumination variation, random noise, and alignment error. FR system can generally be categorized into one of the two main scenarios based on the characteristics of the images to be matched, such as still image based (still-to-still) FR or video based (video-to-video) FR. Also it could be a video-to-still image based face recognition system [3].

1.2.1 Still Image Based Face Recognition

In the context of face image description and representation, there are two common types of techniques that have been applied to still image based face recognition: 1) Subspace based holistic features; and 2) Local appearance based features. One of the initial subspace based methods is principal component analysis (PCA), which is particularly known as eigenface [4]. In addition, some other global features like independent component analysis (ICA) [5], gradientface [6], etc. have showed promising results in image representation for face detection and recognition. However, all of these representations suffer from illumination variation and alignment error.

Successful face appearance representations include Gabor features [7], elastic bunch graph matching [8], and local binary pattern (LBP) [9]. LBP has been proposed to represent visual objects, and successfully applied for different applications in facial image analysis like human detection, face recognition or expression recognition. LBP is basically a fine-scale descriptor that captures small texture details, in contrast to Gabor features which encode the facial shape and appearance over a range of scales [10, 11]. Nevertheless, LBP considers only first order intensity pattern changes in a local neighborhood, which may fail to extract detailed information, especially during changes in face image due to the noise and illumination variation problems. One of the newest local appearance methods that tries to avoid the shortcomings of LBP is local directional pattern (LDP) [12]. LDP encodes the directional information in the neighborhood instead of the

intensity as LBP does. However, LDP still suffers in non-monotonic illumination variation, random noise, and changes in pose, age, and expression conditions. Recently, a new extension of LBP is proposed to extract more detailed information named high order local derivative pattern [3]. This technique has been investigated for texture analysis and shown robust performance to classify the image textures.

1.2.2 Video Based Face Recognition

Dynamic texture (DT) or temporal texture is a texture with motion that includes the class of video sequences, which offers some stationary properties in time. Recently researchers start to investigate the domain of video, where the problem of face recognition becomes more challenging due to pose variations, different facial expressions, illumination changes, occlusions and so on. However, DT provides many samples of the same person, thus providing the opportunity to convert many weak examples into a strong prediction of the identity. In [13], Zhao and Pietikinen introduced an extended version of LBP named volume local binary patterns (VLBP) for video based facial expression recognition. They claim that the features extracted in a small local neighborhood of the volume can be boosted by combining the motion and appearance. These features are insensitive to translation and rotation. However, there are some illumination limitations since this method deals with small local neighborhood of a pixel as well as it utilizes the image intensity directly.

Nowadays, manifold features (linear subspaces), if the features lie in Euclidean spaces, have proven a powerful representation for video based face recognition. Huang et al. [14] recently introduced a new method called projection metric learning on Grassmann manifold (PML), which is combined with Grassmannian graph-embedding discriminant analysis (GGDA) [15]. In this technique, each video sequence can be treated as a set of face images without considering the temporal information. It serves as both of metric learning and dimensionality reduction method for

the Grassmann manifold to map the manifold to a reproducing kernel Hilbert space (RKHS). Although kernel-based methods have been successfully used in many computer vision applications, poor choice of kernels can often result in degraded classification performance [16], especially when the data lies in non-Euclidean spaces.

1.3 Scientific Contributions

The novel scientific contributions in this dissertation are summarized as follows:

- Development of a new feature extraction technique to calculate the n^{th}-order direction variation patterns, named **High Order Local Directional Pattern (HOLDP)**, which extracts n^{th}-order local information by encoding various distinctive spatial relationships from each neighborhood layer of a pixel in the pyramidal multi-structure way. Then the feature vectors of all neighborhood layers can be concatenated to form the final HOLDP feature vector. Derived from a general definition of texture in a local neighborhood, the conventional LDP encodes the directional information in the small 3×3 local neighborhood of a pixel, which may fail to extract detailed information due to sensor noise and illumination variation problems, etc. Therefore, HOLDP is introduced to capture more detailed and discriminative information than the traditional LDP.

- Development of a novel video based feature extraction technique named **Volumetric Directional Pattern (VDP)** [1], which addresses a difficult problem of identifying human faces in videos due to the presence of large variations in facial pose and expression, and poor video resolution. VDP has been successfully applied to video based face recognition as well as hyperspectral imagery (HSI) classification, and it has shown a promising classification accuracy among the other competitors [17].

- Development of a new video based facial feature extractor, named **High Order Volumetric Directional Pattern (HOVDP)**, based on the HOLDP. HOVDP can be viewed as an extension of VDP [1, 17]. It combines the movement and appearance features together by considering the relation of all neighboring pixels from three consecutive frames layer by layer. Unlike the conventional VDP operator that encodes the directional information in the small 3×3 local neighborhood of a pixel of each three consecutive frames, HOVDP extracts n^{th}-order volumetric information by encoding various distinctive spatial relationships from each neighborhood layer of a pixel in the pyramidal multi-structure way, and then concatenating the feature vector of each neighborhood layer to form the final HOVDP feature vector.

1.4 Dissertation Outline

The dissertation is structured as the following. Chapter II describes recent advances in FR application where several state-of-the-art techniques based on the concept of local pattern descriptors for face recognition are introduced. In Chapter III, we present our high order local directional pattern descriptor (i.e., HOLDP) theoretically and experimentally. Chapter IV proposes a new video-based face recognition descriptor (i.e., VDP), and then introduces the high order volumetric directional pattern scheme (i.e., HOVDP). Finally, Chapter V summarizes this dissertation and suggests future directions of this research.

CHAPTER II

RELATED WORK

In recent years, many research work have been done on extracting image features. The main goal of an image descriptor is to extract features from the image that are distinct and stable under different conditions during the image acquisition process. Many computer vision applications employ the texture analysis algorithms. One of the high performing texture algorithms group is based on the concept of local pattern descriptors (e.g., local binary pattern (LBP) and local directional pattern (LDP)) which describe the relationship of pixels to their local neighborhood. They only detect the important local textures by labeling each pixel with the code of texture primitive that best matches the local neighborhood. Fig. 2.1 shows some of these texture primitives that can be detected by the local pattern descriptors that include spots, line ends, flat area, edges, corners and so on. In the figure, ones are represented as white circles and zeros are black [18].

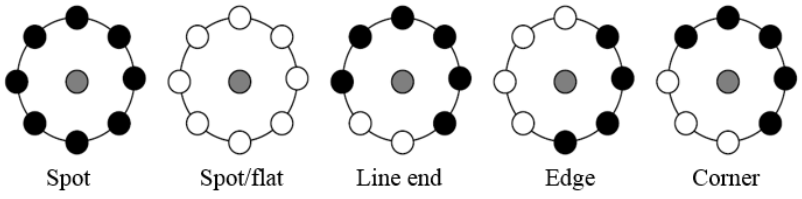

Figure 2.1: Different texture primitives detected by local pattern descriptors

2.1 Local Binary Pattern

During the past few years, local binary pattern (LBP) has aroused increasing interest in image processing and computer vision. LBP is a nonparametric method which extracts local structures of images efficiently by comparing each pixel with its neighboring pixels. If a neighbor pixel has an equal or higher gray level than the center pixel, then a one is assigned to that pixel, which is otherwise a zero. Finally, the LBP binary code for the center pixel is produced by concatenating the eight ones or zeros which will be converted to decimal number to produce the new value of that central pixel.

The original LBP operator was introduced by Ojala et al. [9] for texture analysis, and has proved a simple yet powerful approach to describe local structures. LBP was originally defined for 3×3 neighborhood pixels, which gives an 8 bit binary code that is derived from comparing each pixel with its central pixel. If a neighbor pixel has a higher intensity value than the center pixel (or the same intensity value) then a 1 is assigned to that pixel, which is otherwise a 0. Formally, it is expressed as

$$LBP = \sum_{p=0}^{7} f(d_p - d_c) \times 2^p \tag{2.1}$$

and

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.2}$$

where d_c and d_p denote the intensity values of the central pixel and its surrounding pixels respectively.

The LBP operator has a number of extensions that have been extensively used in many applications such as, face image analysis [19, 20], image and video retrieval [21, 22], environment modeling [23, 24], visual inspection [25, 26], motion analysis [27, 28], biomedical and aerial image analysis [29, 30]. LBP has been exploited for facial representation in different tasks, which include

face detection [31, 32, 33], face recognition [10, 11, 34, 35, 36, 37, 38], facial expression analysis, and demographic (gender, race, age, etc.) classification [39, 40, 41, 42, 43, 44, 45]. The development of LBP methodology can be well illustrated in facial image analysis, and most of its recent variations have been proposed in this area [46].

Some of the extended versions of LBP were using neighborhoods of different sizes and rotation invariant, which allow to deal with large scale structures [47, 48]. Moreover, a single LBP code can be generated for each pixel in an image using all the previously mentioned extensions. This causes two fundamental limitations, trade-off between resolution and the presence of noise. To overcome with that, a Fuzzy Local Binary Pattern (FLBP) was introduced [49, 50]. FLBP extends the LBP by incorporating fuzzy logic in the representation of local patterns of texture. Fuzzification allows FLBP to contribute to more than a single bin in the distribution of the LBP values used as a feature vector.

Another limitation of LBP is that it does not consider the texture features from the magnitude component of the image local differences, as well as the local features from multi-resolution of the image, so that a completed LBP (CLBP) has been introduced [51]. CLBP represents a local region of an image by its local central information and a local difference sign-magnitude transform (LDSMT). The most generalized version of LBP and more discriminant and less sensitive to noise in uniform regions than LBP is local ternary pattern (LTP) [52]. Unlike the LBP operator, LTP extends the LBP to 3 codes, in which intensity value in a tolerance interval zone of width $\pm\tau$ around the center pixel is assigned to 0. If a neighbor pixel has a higher intensity value than this zone then a 1 is assigned to that pixel, which is -1 if it is below this.

2.2 Local Directional Pattern

One of the newest local appearance methods that tries to avoid the shortcoming of LBP is local directional pattern (LDP) [12, 53, 54, 55]. LDP encodes the directional information in the neighborhood instead of the intensity which LBP does. LDP is a gray-scale pattern that characterizes the spatial structure of a local image texture. It computes the edge response values in eight different directions at each pixel position by convolving the image with the Kirsch masks in eight different orientations, centered on its own position. Then it uses the relative strength magnitude to encode the image texture. The Kirsch operator is a derivate filter which is used to find edges in all eight directions of a compass. It has been proven to be a strong edge detection technique for face recognition tasks [56, 57]. Specifically, it takes a single mask, denoted as $M_i(x, y)$ for $i = 0, 1, ..., 7$, and rotates it in 45 degree increments through all 8 compass directions as shown in Fig. 2.2. An example of Kirsch kernel filtered images for one sample face image can be seen in Fig. 2.3, where all eight directional features are extracted with their corresponding masks. Since the edge responses are more noisy and illumination insensitive than intensity values, the resultant LDP feature maintains more information and describes the local primitives stably, including different types of curves, corners, and junctions.

Mathematically, given an input image $I(x, y)$, the eight different directional edge response m_i can be computed by

$$m_i = I(x, y) * M_i \qquad for\ i = 0, 1, ..., 7. \tag{2.3}$$

where $*$ represents a convolution operation. Fig. 2.4 shows these edge responses and the corresponding bit positions [58]. All eight directional features of objects are extracted with its corresponding masks. The response values are not equally important in all directions. The presence of a corner or an edge shows high response values in some particular directions. Therefore, in order to

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

East (M_0) North East (M_1) North (M_2) North West (M_3)

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

West (M_4) South West (M_5) South (M_6) South East (M_7)

Figure 2.2: Kirsch edge masks in all eight directions.

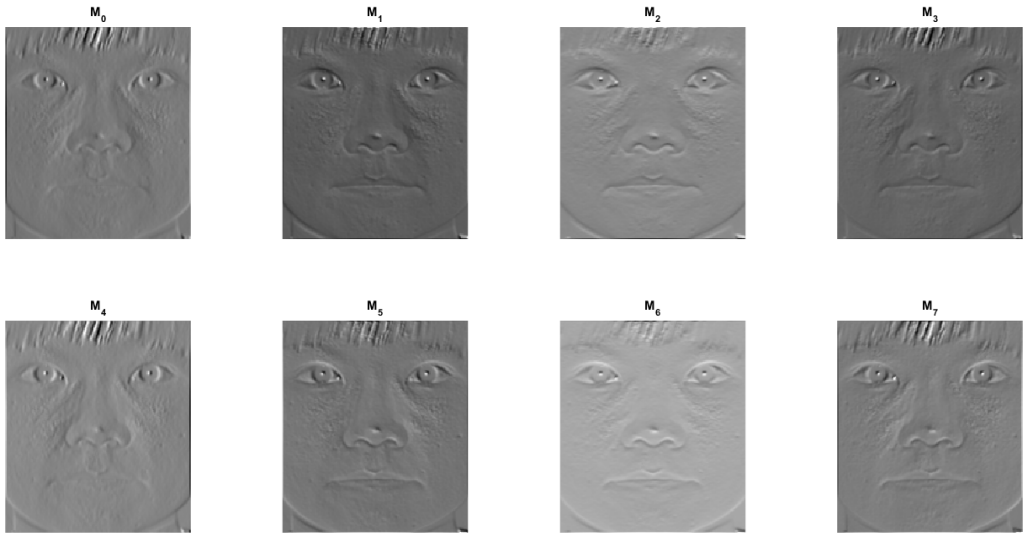

Figure 2.3: Kirsch kernel filtered output images.

generate the LDP, we need to know the t most prominent directions. Then, the top t directional bit responses are set to 1 and the rest (8-t) bits of 8-bit LDP pattern are set to 0.

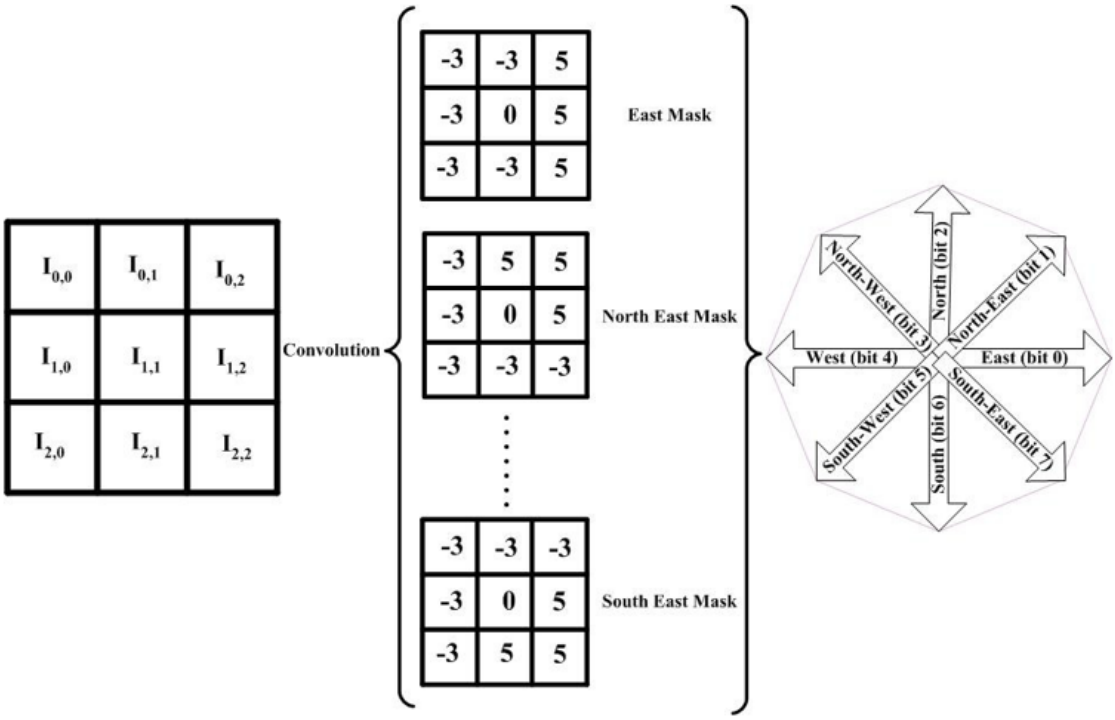

Figure 2.4: Eight directional edge response and LDP binary bit positions.

To encode the directional information in the neighborhood, a binary coding strategy is applied by exploiting the center pixel value in each 3×3 neighborhood regions. Finally, to retrieve the LDP feature map, we change that binary codes into the corresponding decimal codes, which can be calculated by

$$LDP_t = \sum_{i=0}^{i=7} f(m_i - m_t) \times 2^i \tag{2.4}$$

where m_t represents the t^{th} most significant directional response and the thresholding function $f(x)$ can be defined as in Eq. (2.2). An example of the LDP code calculation with $t = 3$ can be seen in Fig. 2.5 [58].

There are several extended versions of LDP such as local directional pattern variance (LDPv) [58, 59] and enhanced local directional pattern (ELDP) [60]. LDPv code is generated from the

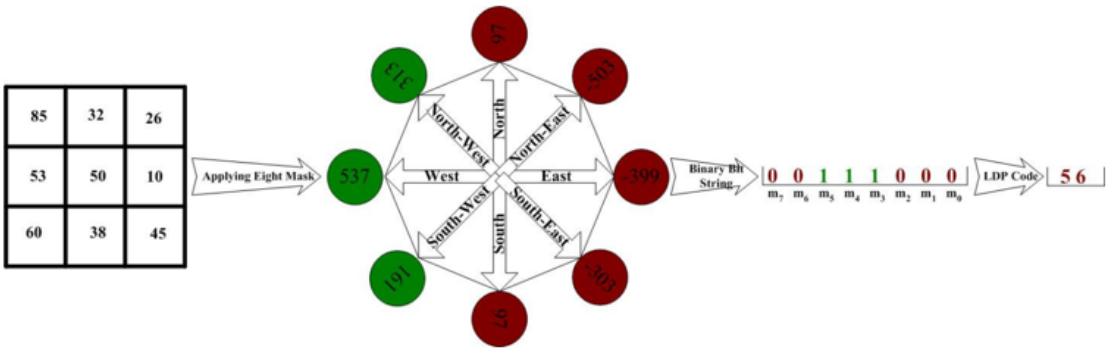

Figure 2.5: LDP code calculation with $t = 3$.

integral projection of each LDP code weighted by its corresponding variance, while ELDP encodes the image using the directions of the most prominent edge response value and the second most prominent one from the eight different directional edge response values, which are created by the original LDP. And then using their index i as the representation of the direction of m_i to create a double-digit octal number $(ab)_8$. Finally, the ELDP code is formed by converting that octal code into a decimal number. However, these extensions may increase the difficulty of recognition by extracting more redundant texture features as in LDPv or by losing some important features as in ELDP when using just two directional edge response values as the representation of each pixel of the image.

Nevertheless, both the above-mentioned descriptors LBP and LDP with their extensions consider only the first order intensity pattern changes in a local neighborhood which may fail to extract detailed information, especially during changes in face image due to the noise and illumination variation problems.

2.3 Local Boosted Features

To overcome some of LBP and LDP limitations that deal with small local neighborhood changes, local boosted features (LBF) is proposed [61]. LBF capture sufficient detailed discriminative information by encoding various distinctive spatial relationships from two neighborhood sizes of a pixel. It includes two main stages: 1) edge detection, where the input image magnitude is extracted by convolving the image with kernels; 2) image encoding and decoding, where the relative strength is used to encode the image texture.

For the edge detection, the input image is convolved with Kirsch masks in eight different orientations. The direction of the edge is determined by the mask that produces the maximum output value. Then the image encoding and decoding (IED) strategy is applied [56, 57]. To encode the local boosted features in the neighborhood, a binary coding strategy is applied to two neighborhood layers of each pixel in the image. Given a central pixel g_c in the image and its P circularly and evenly spaced R-radius neighbors g_p, $p = 0, 1, ..., P - 1$ that can be seen in Fig. 2.6 [51]. The first neighbor layer is considered as $P = 8$ and $R = 1$, and the second layer as $P = 8$ and $R = 2$. Therefore, there are 8 neighboring pixels for each surrounding layer excluding the pixel under consideration (central pixel), the values of neighboring pixels that are not in the center of grids can be estimated by interpolation. Then for each neighborhood layer, the 8 neighboring pixels (excluding the central pixel) are compared with their median. If a neighbor pixel has a higher edge value than the median value (or the same value) then a 1 is assigned to that pixel, which is otherwise a 0. Then a histogram is built for each layer. Finally, the boosted feature vector is formed by concatenating these two histograms.

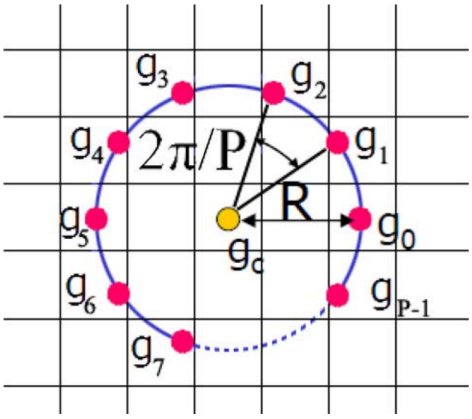

Figure 2.6: Central pixel and its P circularly and evenly spaced neighbors with radius R.

Given an input image $I(x, y)$ and Kirsch masks in eight different orientations M_i for $i = 0, 1, ..., 7$, the eight different directional edge response values $d_{i,j}$ can be computed by

$$d_{i,j} = I(x, y) * M_i, \qquad i = 0, 1, ..., 7 \qquad and \qquad j = 1, 2. \tag{2.5}$$

where '$*$' represents a convolution operation.

After obtaining the eight different directional edge response values of the first and second neighborhood layers $d_{i,1}$ and $d_{i,2}$ for $i = 0, 1, ..., 7$ respectively, the initial step of IED computation is to compute the median of $d_{i,1}$ and $d_{i,2}$. After that, we compare each pixel from each layer with its corresponding median value to form the binary code. Then to retrieve the edge feature map, we change that binary codes into the corresponding decimal codes D_1 and D_2, which can be computed by

$$D_1 = \sum_{i=0}^{7} f(d_{i,1} - m_1) \times 2^i \tag{2.6}$$

and

$$D_2 = \sum_{i=0}^{7} f(d_{i,2} - m_2) \times 2^i \tag{2.7}$$

where m_1 and m_2 are the medians of each 8 neighboring pixels for the first and second neighborhood layers respectively. The thresholding function $f(x)$ can be defined as in Eq. (2.2). Figure 2.7 illustrates the whole procedure of LBF technique. Although LBF was evidenced by yielding promising recognition rates than the above-mentioned descriptors LBP and LDP, it still has limitations that it considers just the first and second neighborhood layers of a pixel of interest and it does not consider all the pixels from the second neighborhood layer, which may fail to capture premiere features with large scale instructions.

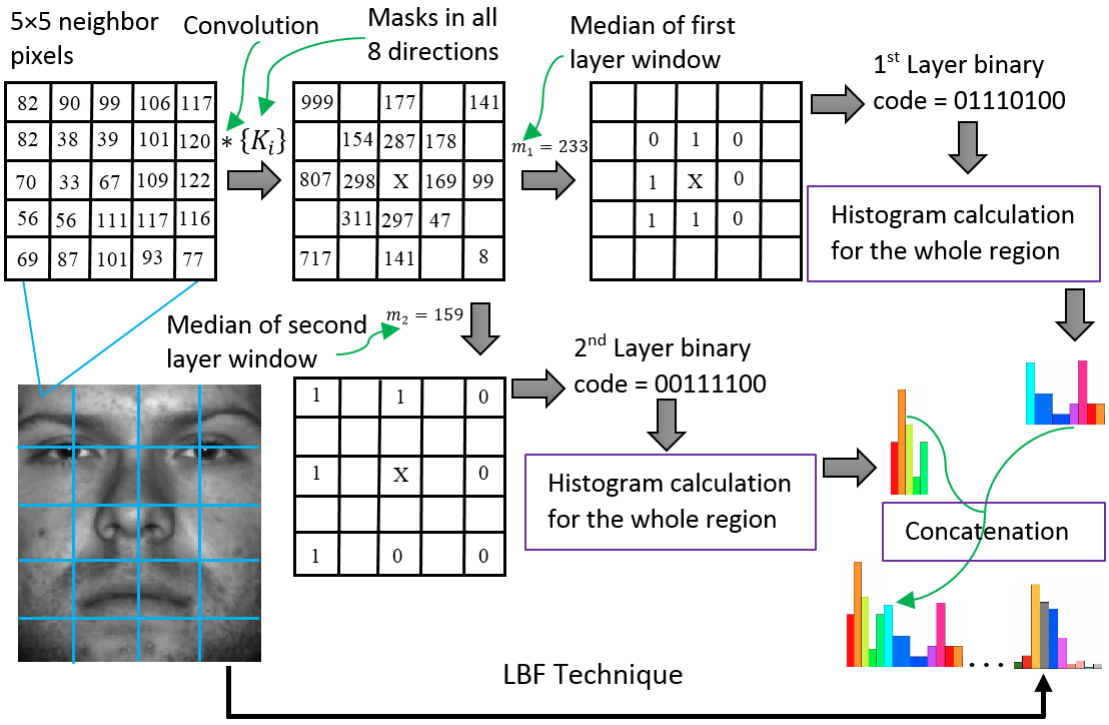

Figure 2.7: An example of calculating LBF code.

CHAPTER III

HIGH ORDER LOCAL DIRECTIONAL PATTERN WITH PYRAMIDAL MULTI-STRUCTURE TECHNIQUE

In this chapter, we introduce a new feature extraction technique to calculate the n^{th}-order directional variation patterns, named high order local directional pattern (HOLDP). After that, the spatial histogram is built for modeling the distribution information of a face image. Lastly, several experimental results will be given to show the strength of HOLDP for face recognition task.

3.1 Methodology

Derived from a general definition of texture in a local neighborhood, the conventional LDP encodes the directional information in the small 3×3 local neighborhood of a pixel, which may fail to extract detailed information especially during changes in the input image due to several factors such as random noise and illumination variation problems, etc. Therefore, in this research work we introduce a novel feature extraction technique that calculates the n^{th} order directional variation patterns, named high order local directional pattern (HOLDP). The proposed HOLDP can capture more detailed discriminative information than the traditional LDP. Unlike the LDP operator, our proposed technique extracts n^{th}-order local information by encoding various distinctive spatial relationships from each neighborhood layer of a pixel in the pyramidal multi-structure way. Then

we concatenate the feature vector of each neighborhood layer to form the final HOLDP feature vector. Several observations can be made for HOLDP:

- Under the proposed framework, the well known LDP is a special case of HOLDP, which simply calculate the 1^{st} order pattern information in the local neighborhood of a pixel.
- The relation between the neighbor layers and the pixel under consideration could be easily weighted in HOLDP based on the distance between each layer and the central pixel. Because of that, the pixels within the closest layer to the central pixel has more weight than the others.
- Due to the same format and feature length of different order HOLDPs, they can be readily fused, and the accuracy of the face recognition can be significantly improved after the fusion.

3.2 The Proposed Method

Given a central pixel g_c in the image and its P evenly spaced l neighbors $g_{p,l}$ where $p = 0, 1, ..., P-1$ and $l = 1, 2, ..., n$ (number of neighboring layers) as can be seen in Fig. 3.1. The traditional LDP simply calculate the 1^{st} order edge directional value along p direction, with limitations of the total number of involved neighbors $P = 8$ and $l = 1$. To provide a stronger discriminative capability in describing detailed texture information than the 1^{st} order as used in LDP, we propose to use different layers of neighborhood configuration.

The directional edge extraction process can be done by convolving the input image $I(x, y)$ with Kirsch masks in eight different directions $M_i(x, y)$ for $i = 0, 1, ..., 7$, which can be seen in Fig. 2.2. An example of Kirsch kernels filtered images for one sample face image can be seen in Fig. 2.3, where all eight directional features are extracted with their corresponding masks.

To make our calculation simple and easy to compute the high order relevant edge values along p direction, let us assume $P = 8$ and $l = 1$ to calculate the 1^{st} order which means (3×3) neighborhood

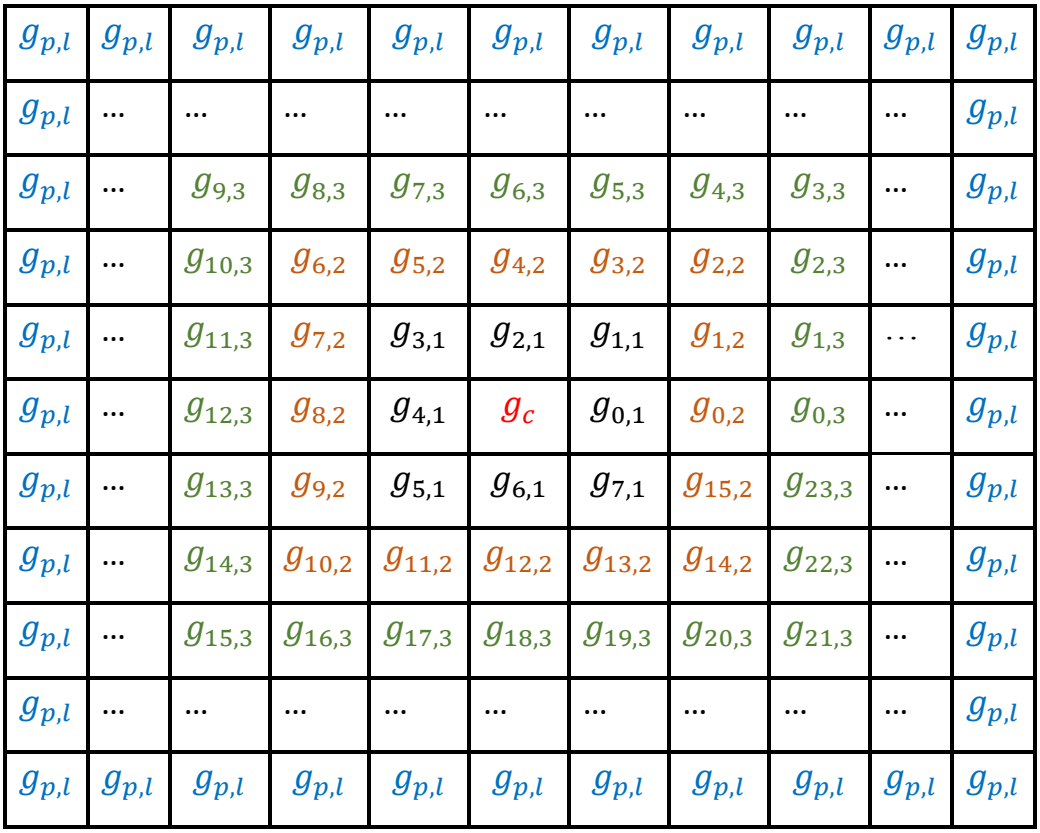

Figure 3.1: Central pixel and its P evenly spaced neighbors within l neighboring layers.

pixels, $P = 16$ and $l = 2$ to calculate the 2^{nd} order for (5×5) neighborhood that means twenty four neighbor pixels will be under consideration, as can be seen in Fig. 3.2 top, and $P = 8 \times n$ and $l = n$ to calculate the n^{th} order.

For example, the 1^{th} and 2^{nd} order local directional patterns $(HOLDP_1)$ and $(HOLDP_2)$ could be computed as follows; Firstly, we need to find the eight different directional edge response values of the first and second neighborhood layers (1^{th} and 2^{nd} LDP order), which can be done by

$p_{6,2}$	$p_{5,2}$	$p_{4,2}$	$p_{3,2}$	$p_{2,2}$
$p_{7,2}$	$p_{3,1}$	$p_{2,1}$	$p_{1,1}$	$p_{1,2}$
$p_{8,2}$	$p_{4,1}$	X	$p_{0,1}$	$p_{0,2}$
$p_{9,2}$	$p_{5,1}$	$p_{6,1}$	$p_{7,1}$	$p_{15,2}$
$p_{10,2}$	$p_{11,2}$	$p_{12,2}$	$p_{13,2}$	$p_{14,2}$

Central Pixel with its 1st & 2nd Neighborhood Layers Edge Values

$m_{3,2}$		$m_{2,2}$		$m_{1,2}$
	$m_{3,1}$	$m_{2,1}$	$m_{1,1}$	
$m_{4,2}$	$m_{4,1}$	X	$m_{0,1}$	$m_{0,2}$
	$m_{5,1}$	$m_{6,1}$	$m_{7,1}$	
$m_{5,2}$		$m_{6,2}$		$m_{7,2}$

Sixteen Edge Responses

$b_{3,2}$		$b_{2,2}$		$b_{1,2}$
	$b_{3,1}$	$b_{2,1}$	$b_{1,1}$	
$b_{4,2}$	$b_{4,1}$	X	$b_{0,1}$	$b_{0,2}$
	$b_{5,1}$	$b_{6,1}$	$b_{7,1}$	
$b_{5,2}$		$b_{6,2}$		$b_{7,2}$

Sixteen Binary Bit Positions

Figure 3.2: Two neighborhood layers and sixteen edge responses with their binary bit positions.

$$m_{i,1} = p_{i,1}, \tag{3.1}$$

$$m_{i,2} = \frac{1}{3} \sum_{j=-1}^{j=1} p_{g,2}, \tag{3.2}$$

and

$$g = \mod (2i + j, 16), \qquad for\ i = 0, 1, ..., 7 \tag{3.3}$$

where $m_{i,1}$ and $m_{i,2}$ are the eight different directional edge response values of the first and second neighborhood layers respectively. $p_{g,2}$ is the edge response value after convolving the input image with Kirsch kernels, the subscripts g and 2 are the number of surrounding pixels of each direction i and the second neighborhood layer (second order) respectively, and (mod) is the modulo operation.

From equations 4.10 and 4.11, two observations can be made. The first one is that by using the modulo operation (mod), we maintain the circularly neighbors configuration. The second one is that by tacking the mean of each three pixels in this case (2^{nd} order), we give less weight to the pixels in the second layer than the pixels in the first layer, which reduces the total number of second layer's pixels from sixteen pixels to eight. For example, the edge response value at 0° direction can be calculated as

$$m_{0,2} = \frac{p_{15,2} + p_{0,2} + p_{1,2}}{3} \tag{3.4}$$

Secondly, based on the observation that every corner or edge has high response values in particular directions, we are interested to know t the most prominent directions after convolving the input image with all 8 masks. Then the local directional pattern of each pixel position in each neighbor layer for the 1^{th} and 2^{nd} order can be formed as

$$HOLDP_1 = \sum_{i=0}^{7} f(m_{i,1} - m_{t,1}) \times 2^i \tag{3.5}$$

and

$$HOLDP_2 = \sum_{i=0}^{7} f(m_{i,2} - m_{t,2}) \times 2^i \tag{3.6}$$

where $m_{t,1}$ and $m_{t,2}$ are the t^{th} most significant directional responses of the first and second neighboring layers respectively. The thresholding function $f(x)$ can be defined as in Eq. (2.2).

After identifying the local directional pattern of each pixel in each neighborhood layer ($HOLDP_1$ for the first layer and $HOLDP_2$ for the second layer), a histogram is built to represent the whole a distinguishing features of an image from each neighbor layer separately. Then we concatenate these two histograms to finalize the feature vector of our proposed technique, which statistically describes the face image characteristics. This way, it is easy to combine multiple HOLDP with more than two orders, by concatenating different histograms one by one.

The 3^{rd} order local directional pattern $(HOLDP_3)$ could be computed by concatenating the 2^{nd} order local directional pattern histogram that has been obtained above with the local directional pattern histogram of the third neighborhood layer, which can be found by

$$HOLDP_3 = \sum_{i=0}^{7} f(m_{i,3} - m_{t,3}) \times 2^i \tag{3.7}$$

where the thresholding function $f(x)$ can be defined as in Eq. (2.2). $m_{t,3}$ is the t^{th} most significant directional responses and $m_{i,3}$ is the eight different directional edge response values of the third neighborhood layer, which is computed by

$$m_{i,3} = \frac{1}{5} \sum_{j=-2}^{j=2} p_{g,3} \tag{3.8}$$

and

$$g = \mod (3i + j, 24), \qquad for\ i = 0, 1, ..., 7 \tag{3.9}$$

From equation 3.8, we can say that using the pyramidal multi-structure approach, which can be seen in Fig. 3.3 (the red cells), we give less weight to the pixels in the third layer than the pixels in the second layer which is less than the ones in the first layer. In other words, to get the edge

response value for each direction, we consider five pixels as one pixel from the third layer, three pixels as one pixel from the second layer, and one pixel from the first layer. For example, the edge response values at 270° direction using the pyramidal multi-structure approach can be calculated as

$$m_{6,1} = p_{6,1} \ , \tag{3.10}$$

$$m_{6,2} = \frac{p_{11,2} + p_{12,2} + p_{13,2}}{3}, \tag{3.11}$$

and

$$m_{6,3} = \frac{p_{16,3} + p_{17,3} + p_{18,3} + p_{19,3} + p_{20,3}}{5} \tag{3.12}$$

Based to the analysis above, the n^{th} order local directional pattern $(HOLDP_n)$ of each pixel position in each neighbor layer from the input image can be defined as

$$HOLDP_n = \sum_{i=0}^{7} f(m_{i,n} - m_{t,n}) \times 2^i \tag{3.13}$$

where $n = 1, 2, ...$ is the local directional pattern order (the number of neighborhood layers), $m_{t,n}$ is the t^{th} most significant directional responses of each neighboring layer n, the thresholding function $f(x)$ can be defined as in Eq. (2.2), and $m_{i,n}$ is the eight different directional edge response values of each neighborhood layer n, which can be computed as

$$m_{i,n} = \frac{1}{2n-1} \sum_{j=-n+1}^{n-1} p_{g,n} \tag{3.14}$$

and

$$g = \mod (ni + j, 8n) \qquad for\ i = 0, 1, ..., 7 \tag{3.15}$$

$p_{9,3}$	$p_{8,3}$	$p_{7,3}$	$p_{6,3}$	$p_{5,3}$	$p_{4,3}$	$p_{3,3}$
$p_{10,3}$	$p_{6,2}$	$p_{5,2}$	$p_{4,2}$	$p_{3,2}$	$p_{2,2}$	$p_{2,3}$
$p_{11,3}$	$p_{7,2}$	$p_{3,1}$	$p_{2,1}$	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$
$p_{12,3}$	$p_{8,2}$	$p_{4,1}$	X	$p_{0,1}$	$p_{0,2}$	$p_{0,3}$
$p_{13,3}$	$p_{9,2}$	$p_{5,1}$	$p_{6,1}$	$p_{7,1}$	$p_{15,2}$	$p_{23,3}$
$p_{14,3}$	$p_{10,2}$	$p_{11,2}$	$p_{12,2}$	$p_{13,2}$	$p_{14,2}$	$p_{22,3}$
$p_{15,3}$	$p_{16,3}$	$p_{17,3}$	$p_{18,3}$	$p_{19,3}$	$p_{20,3}$	$p_{21,3}$

Central Pixel with its 1st, 2nd & 3rd Neighborhood Layers Edge Values

$m_{3,3}$			$m_{2,3}$			$m_{1,3}$
	$m_{3,2}$		$m_{2,2}$		$m_{1,2}$	
		$m_{3,1}$	$m_{2,1}$	$m_{1,1}$		
$m_{4,3}$	$m_{4,2}$	$m_{4,1}$	X	$m_{0,1}$	$m_{0,2}$	$m_{0,3}$
		$m_{5,1}$	$m_{6,1}$	$m_{7,1}$		
	$m_{5,2}$		$m_{6,2}$		$m_{7,2}$	
$m_{5,3}$			$m_{6,3}$			$m_{7,3}$

Twenty Four Edge Responses

$b_{3,3}$			$b_{2,3}$			$b_{1,3}$
	$b_{3,2}$		$b_{2,2}$		$b_{1,2}$	
		$b_{3,1}$	$b_{2,1}$	$b_{1,1}$		
$b_{4,3}$	$b_{4,2}$	$b_{4,1}$	X	$b_{0,1}$	$b_{0,2}$	$b_{0,3}$
		$b_{5,1}$	$b_{6,1}$	$b_{7,1}$		
	$b_{5,2}$		$b_{6,2}$		$b_{7,2}$	
$b_{5,3}$			$b_{6,3}$			$b_{7,3}$

Twenty Four Binary Bit Positions

Figure 3.3: Three neighborhood layers and twenty four edge responses with their binary bit positions.

3.2.1 The Adaptiveness of the Proposed Technique

After the eight different directional edge response values of each neighboring layer are computed, the image encoding and decoding (IED) strategy is applied [56, 57]. Unlike the conventional LDP which needs to know the most prominent edge values to set them to 1 and the rest to 0, the local directional patterns in this section can be formed adaptively by comparing the 8 neighboring pixels

(excluding the central pixel) with their median of each neighborhood layer. If a neighboring pixel has a higher edge value than the median value (or the same value) then a 1 is assigned to that pixel, which is otherwise a 0. We tried to use other different measures to adaptively find the relationship of the neighboring pixels of each layer such as thresholding the pixels with their mean, standard deviation, and variance, but we found that the median provides better accuracy rates. Therefore, the n^{th} order adaptive local directional pattern $(AHOLDP_n)$ of each pixel position in each neighbor layer from the input image can be defined as

$$AHOLDP_n = \sum_{i=0}^{7} f(m_{i,n} - m_n) \times 2^i \tag{3.16}$$

where $n = 1, 2, ...$ is the local directional pattern order (the number of neighborhood layers), m_n is the median of each neighboring layer n, and $m_{i,n}$ for $i = 0, 1, ..., 7$ is the eight different directional edge response values of each neighborhood layer n, which can be computed using equations 3.14 and 3.15. The thresholding function $f(x)$ can be defined as in Eq. (2.2).

3.3 Experimental Results

For evaluation, we use four face recognition benchmarks that include diversity of illumination and lighting conditions, namely extended Yale B database [62, 63], AT&T (ORL) dataset [64], Georgia Tech (GT) face database [65], and AR database [66, 67]. All the images are resized to 64×64. After that, we extract the information from each image using our proposed technique HOLDP and represent it as one histogram vector. The length of this feature vector (histogram) depends on the order of the local directional pattern descriptor, which means it is $n \times 256$. For example, the first order is 256 bins, the second order is 512 bins, and so on. In addition, the proposed HOLDP and AHOLDP techniques are compared with five spatial feature extraction methods that have common characteristics including LBP, CLBP, FLBP, and LTP, which their codes are publicly available. As well as the conventional LDP, which considers as a special case of the proposed

HOLDP technique (the 1^{st} order). When it comes to the face recognition process, the objective is to compare the encoded feature vector from one image with all other candidate feature vectors using a library for support vector machines classifier (LIBSVM) [68].

Two different experiments are conducted to verify the effectiveness and efficiency of the proposed HOLDP framework. The first one is exploring the effectiveness of different local directional order (different neighborhood layers for each pixel of the image) as changing the number of the most prominent response values $\{t = 2, 3, \cdots, 6\}$. The second one is the effectiveness of the proposed HOLDP and AHOLDP are evaluated by comparing them with five popular texture feature extractors including LBP, CLBP, FLBP, LTP and LDP. To avoid any bias, we randomly select the data for training and testing, then the experiments were repeated 10 times, after that the average results is calculated for comparison. Note that, we coded LDP technique since there is no source code publicly available and it is a special case of our proposed HOLDP technique. The source codes for LBP, CLBP, FLBP, and LTP were available respectively at:

- http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab
- http://www.comp.polyu.edu.hk/ cslzhang/code/CLBPMatlab
- http://www.cb.uu.se/ gustaf/textureDescriptors/FLBPMatlab
- http://parnec.nuaa.edu.cn/xtan/Publication.htm/LTPMatlb

3.3.1 Extended Yale B Database

The extended Yale B database has a total of 2280 face images for 38 subjects representing 60 illumination conditions per subject under the frontal pose, Fig. 3.4 shows some sample faces of one subject of this dataset. In the figure, it is clear that how the illumination problems extremely affect the input images. To show the effectiveness of the proposed HOLDP technique, we randomly select half of the data for training (30 images/subject) and the other half for testing. Then summarize the highest recognition rates as changing the number of neighborhood layers (the order of the proposed

approach) in the range $(1-4)$ along with changing the threshold t (the most prominent edge response values) in the range $\{t = 2, 3, \cdots, 6\}$ as well as with the adaptive HOLDP for each pixel of the input image in Table 3.1.

The performance results of well known local appearance based feature algorithms for face recognition like LBP, CLBP, FLBP, LTP, and LDP with the proposed methods HOLDP and AHOLDP on extended Yale B dataset are presented in Table 3.2. Note that, we ran the same sets of training and testing for all methods, since LBP, CLBP, and FLBP are nonparametric methods, we extract their histograms directly. While for LTP approach, there is one main parameter τ that splits LTP into positive and negative parts, then the histogram is built for each part to form the final LTP feature description of the original image. Therefore, a set of different thresholding numbers of $\{\tau = 0.1, 0.2, 0.5, 1, 2, \cdots, 7\}$ are experimented, and it is found that $\tau = 0.5$ yields optimal performance for this dataset.

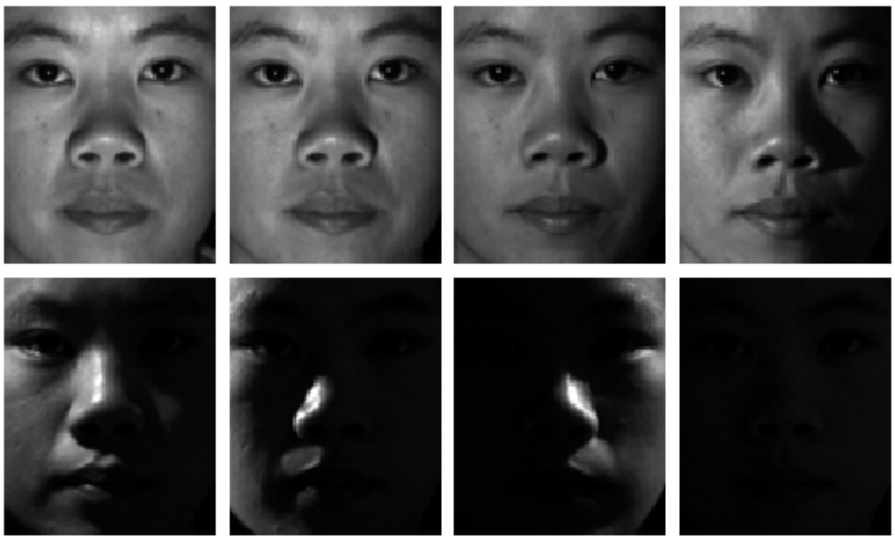

Figure 3.4: Samples of one subject from the extended Yale B database.

Table 3.1: Recognition rates as changing the threshold t and the proposed approach order on extended Yale B dataset.

Descriptor	**Recognition Accuracy (%)**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOLDP
1^{st} Order	98.38 %	96.80 %	98.73 %	98.52 %	97.36 %	**98.92 %**
2^{nd} Order	98.92 %	97.82 %	99.14 %	99.10 %	97.89 %	**99.31 %**
3^{rd} Order	99.19 %	98.38 %	99.29 %	99.43 %	98.45 %	**99.45 %**
4^{th} Order	99.47 %	98.59 %	99.40 %	99.54 %	98.68 %	**99.56 %**

Table 3.2: Performance comparison of the proposed methods with well-known face recognition algorithms on extended Yale B dataset.

Recognition Accuracy (%) for Each Method						
LBP	CLBP	LTP	FLBP	LDP	HOLDP	AHOLDP
97.91 %	62.92 %	88.36 %	92.21 %	98.73 %	99.54 %	**99.56 %**

3.3.2 AT&T Dataset (ORL)

The ORL database contains 400 face images corresponding to 40 distinct subjects each has 10 different images. Some sample faces are shown in Fig. 3.5. The images are taken at different times with different specifications, including varying slightly in illumination and pose, different facial expressions such as open and closed eyes, smiling and not smiling, and facial details like wearing glasses and not wearing glasses. The same procedure as the previous section is applied, so we randomly select half of the data for training (5 images/subject) and the other half for testing. Then summarize the highest recognition rates as changing the order of the proposed descriptor in the range $(1-4)$ along with changing the threshold t in the range $\{t = 2, 3, \cdots, 6\}$ as well as with the adaptive

HOLDP for each pixel of the input image in Table 3.3. While Table 3.4 presents the comparison results of LBP, CLBP, FLBP, LTP, and LDP with the proposed methods HOLDP and AHOLDP on ORL database. Note that, we ran the same sets of training and testing for all methods, LBP, CLBP, FLBP, and LTP with a set of different thresholding numbers of $\{\tau = 0.1, 0.2, 0.5, 1, 2, \cdots, 7\}$ and it is found that $\tau = 2$ yields optimal performance for this database.

Figure 3.5: Samples of one subject from the ORL database.

3.3.3 Georgia Tech (GT) Face Database

The Georgia Tech (GT) face database consists 750 color images corresponding to 15 different images of 50 distinct subjects. These images have large variations in both pose and expression and some illumination changes. Images are converted to gray scale and cropped into the size of 64×64. Some sample faces of one subject are shown in Fig. 3.6. The same procedure as the previous sections is applied, so we randomly select half of the data for training (8 images/subject)

Table 3.3: Recognition rates as changing the threshold t and the proposed approach order on ORL database.

Descriptor	**Recognition Accuracy (%)**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOLDP
1^{st} Order	95.80 %	95.90 %	96.50 %	96.65 %	96.40 %	**97.70 %**
2^{nd} Order	96.60 %	96.95 %	97.40 %	97.25 %	97.10 %	**98.20 %**
3^{rd} Order	96.85 %	97.15 %	97.90 %	97.80 %	97.55 %	**98.60 %**
4^{th} Order	96.85 %	97.05 %	98.25 %	97.85 %	97.95 %	**98.60 %**

Table 3.4: Performance Comparison of the Proposed methods with-well known face recognition algorithms on ORL database.

Recognition Accuracy (%) for Each Method						
LBP	CLBP	LTP	FLBP	LDP	HOLDP	AHOLDP
96.00 %	81.45 %	96.00 %	96.00 %	96.65 %	98.25 %	**98.60 %**

and the other half for testing. Then summarize the highest recognition rates in Table 3.5 to show the effectiveness of each neighborhood layer. While Table 3.6 presents the comparison results of LBP, CLBP, FLBP, LTP, and LDP with the proposed methods HOLDP and AHOLDP on GT database. Note that, we ran the same sets of training and testing for all methods, LBP, CLBP, FLBP, and LTP with LTP with a set of different thresholding numbers of $\{\tau = 0.1, 0.2, 0.5, 1, 2, \cdots, 7\}$ and it is found that $\tau = 3$ yields optimal performance for this dataset.

3.3.4 AR Database

The AR face database contains over 4000 color face images of 126 people, including frontal views of faces with different facial expressions, illumination conditions and occlusions. In our

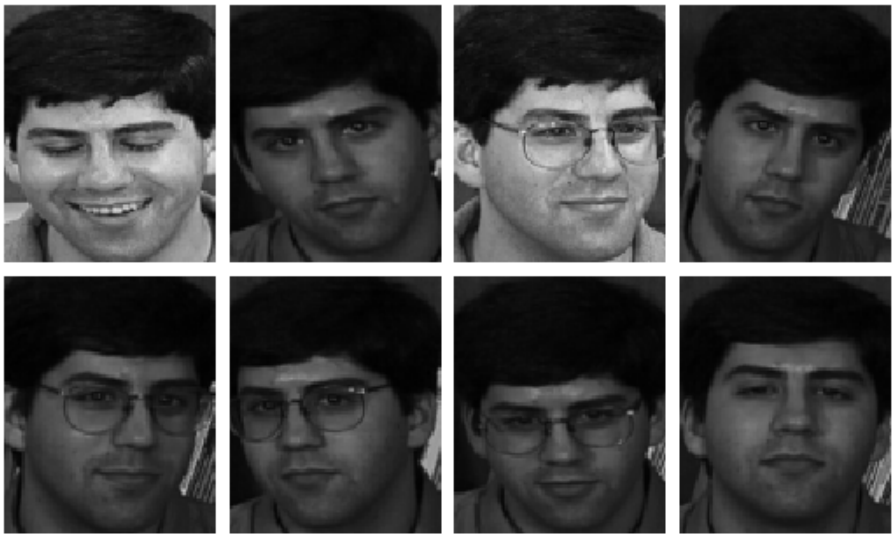

Figure 3.6: Samples of one subject from the GT database.

Table 3.5: Recognition rates as changing the threshold t and the proposed approach order on GT database.

Descriptor	**Recognition Accuracy (%)**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOLDP
1^{st} Order	87.22 %	87.40 %	89.28 %	87.94 %	87.62 %	**91.57 %**
2^{nd} Order	87.80 %	88.94 %	89.80 %	88.94 %	88.40 %	**93.37 %**
3^{rd} Order	88.05 %	89.11 %	89.82 %	89.51 %	88.57 %	**93.00 %**
4^{th} Order	87.60 %	88.97 %	89.22 %	89.02 %	88.34 %	**92.68 %**

experiments, a subset with large variations in both illumination and expression was chosen, which corresponds to 50 male subjects and 50 female subjects. For each subject, there are two sections one for training and the other for testing. Each section contains 7 images per subject. To show the effectiveness of each neighborhood layer, we summarize the highest recognition rates in Table

Table 3.6: Performance comparison of the proposed methods with-well known face recognition algorithms on GT database.

Recognition Accuracy (%) for Each Method						
LBP	CLBP	LTP	FLBP	LDP	HOLDP	AHOLDP
89.00 %	80.48 %	90.40 %	90.85 %	89.28 %	89.82 %	**93.37 %**

3.7. While Table 3.8 presents the comparison results of LBP, CLBP, FLBP, LTP, and LDP with the proposed methods HOLDP and AHOLDP on AR database. Note that, we ran the same sets of training and testing for all methods, LBP, CLBP, FLBP, and LTP with LTP with a set of different thresholding numbers of $\{\tau = 0.1, 0.2, 0.5, 1, 2, \cdots, 7\}$ and it is found that $\tau = 5$ yields optimal performance for this database.

Figure 3.7: Samples of one subject from the AR database.

Table 3.7: Recognition rates as changing the threshold t and the proposed approach order on AR database.

Descriptor	**Recognition Accuracy (%)**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOLDP
1^{st} Order	88.57 %	88.00 %	90.14 %	90.42 %	88.57 %	**90.00 %**
2^{nd} Order	90.57 %	91.00 %	92.14 %	92.14 %	90.14 %	**92.85 %**
3^{rd} Order	90.42 %	91.85 %	93.00 %	93.28 %	91.85 %	**94.28 %**
4^{th} Order	91.00 %	92.00 %	93.42 %	93.85 %	93.00 %	**95.57 %**

Table 3.8: Performance comparison of the proposed methods with-well known face recognition algorithms on AR database.

Recognition Accuracy (%) for Each Method						
LBP	CLBP	LTP	FLBP	LDP	HOLDP	AHOLDP
87.57 %	78.28 %	87.28 %	85.42 %	90.42 %	93.85 %	**95.57 %** %

3.4 Discussion

In this chapter, we have introduced a new feature descriptor named HOLDP. Throughout the performance evaluations, we found that HOLDP provides better performance for face recognition regardless of extreme variations of illumination environments and slightly differences in pose and expression conditions. In addition, compared to the other state-of-the-art methods, we conclude that our method provides better accuracy in all test cases. Furthermore, it is observed that a number of neighborhood layers and the threshold will affect recognition accuracy. From the results above it is clear that the high-order local patterns provide a stronger discriminative capability in describing detailed texture information than the first-order local pattern as used in the traditional LDP technique.

When it comes to the adaptiveness of the proposed HOLDP, AHOLDP does provide the highest recognition rates than the HOLDP in all test cases. In general, considering all comparison results, we can assess that HOLDP can be a promising candidate for face recognition application.

It is a true challenge to build an automated system which equals human ability to recognize faces. While traditional face recognition is typically based on still images, face recognition from video sequences has become popular recently due to more abundant information than still images and more adopted in real applications.

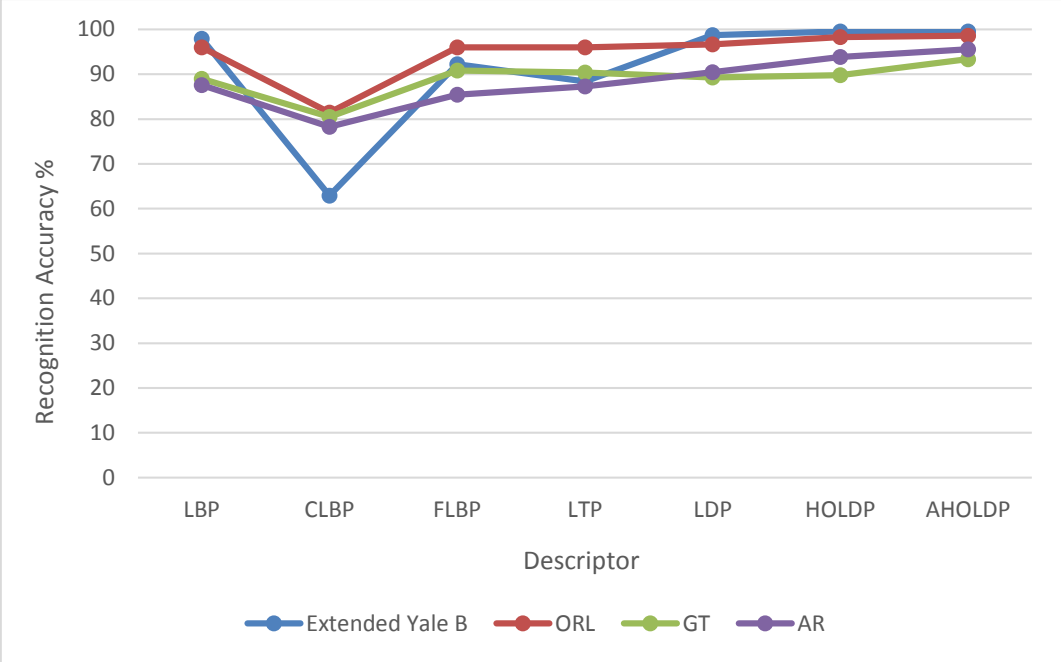

Figure 3.8: Recognition rates comparison on extended Yale B, ORL, GT, and AR databases.

CHAPTER IV

HIGH ORDER VOLUMETRIC DIRECTIONAL PATTERN

In this chapter, we introduce two new video-based descriptors, volumetric directional pattern (VDP) [1] and high order volumetric directional pattern (HOVDP), for robust face recognition, addressing a difficult problem of identifying human faces in video due to the presence of large variations in facial pose and expression, as well as poor video resolution. VDP is an oriented volumetric descriptor that is able to extract and fuse the information of multi frames, temporal (dynamic) information, and multi poses and expressions of faces in input video to produce strong feature vectors, which are used to match with all the videos in the database. Then based on our novel HOLDP technique that we presented in Chapter 3, we develop the HOVDP, which can be viewed as an extension of VDP. HOVDP combines the movement and appearance features together considering the n^{th} order volumetric directional variation patterns of all neighboring pixel layers from three consecutive frames.

4.1 Methodology

The main goal of the volumetric directional pattern is extracting and fusing the temporal information (dynamic features) from three consecutive frames which are distinct under multi poses and facial expressions variations. The overall end-to-end video based face recognition framework can be explained in Fig. 4.1.

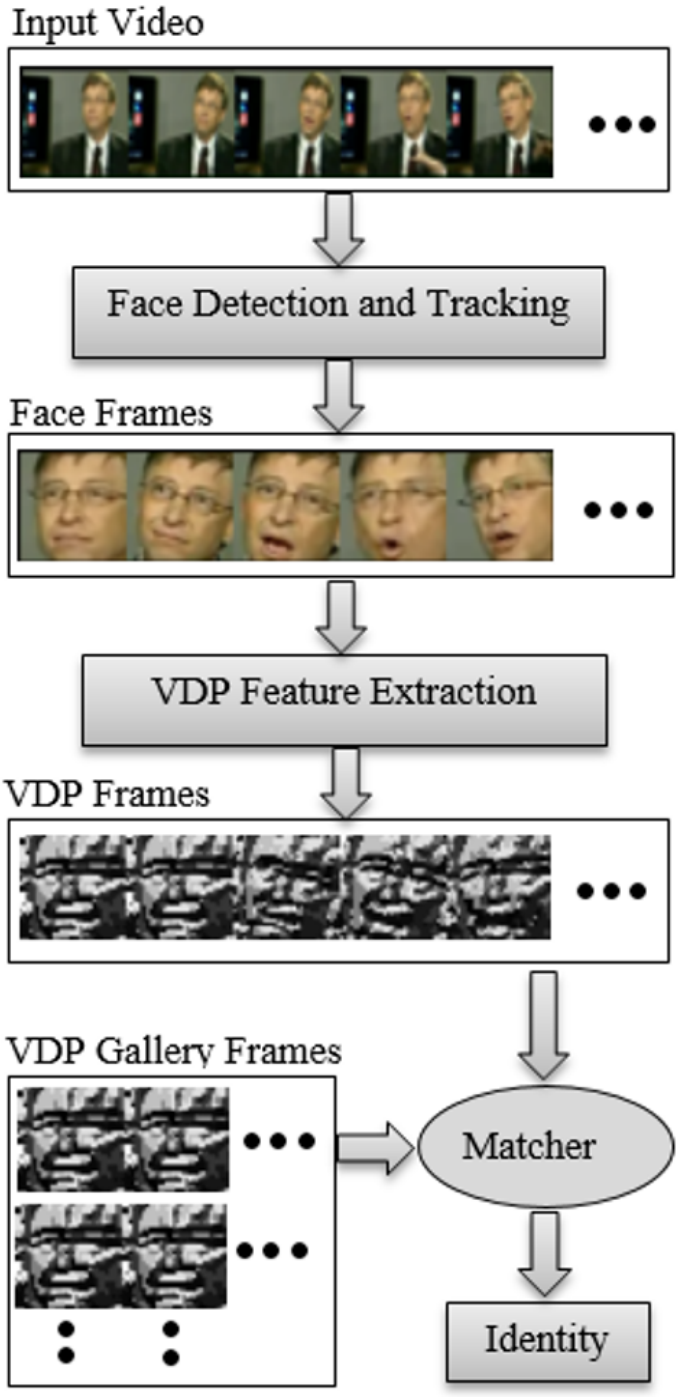

Figure 4.1: Video-to-video face recognition pipeline.

Given a video as input and a gallery of videos, we perform face recognition process throughout the whole video clip. Firstly, we detect and track faces using Viola and Jones face detector [69]. Then for each frame we extract and combine the dynamic features of its two neighborhood frames using our novel algorithm, similar procedure to the gallery videos. Then a histogram is built for each frame. These histograms are concatenated to form the final VDP-feature vector.

4.2 Volumetric Directional Pattern

Volumetric Directional Pattern (VDP) is a gray-scale pattern that characterizes and fuses the temporal structure (dynamic information) of three consecutive frames. The proposed VDP has been developed to merge the movement and appearance features together. VDP is a twenty-four bit binary code assigned to each pixel of an input frame, which can be calculated by comparing the relative edge response value of a particular pixel from three consecutive frames in different directions by using Kirsch masks in eight different orientations centered on its own position for one frame and the corresponding positions of the other two frames.

Given a central pixel in the middle (center) frame of three consecutive frames, the eight different directional edge response values (first order) c_i for $i = 8, 9, ..., 15$ are used to create an eight bit binary number which can describe the edge response pattern of each pixel in the center frame (frame of interest). Meanwhile, the eight different edge response values p_i for $i = 16, 17, ..., 23$ and n_i for $i = 0, 1, ..., 7$ are used to create an eight bit binary number each, which can describe the edge response pattern of each pixel in the previous frame and next frame respectively. Fig. 4.2 shows the twenty four edge responses and their corresponding bit binary positions, as well as the fusing strategy of this 24-bit code. The twenty four different directional edge response values for each pixel location can be computed by

$$n_i = \sum_{0}^{7} dot(I_{3\times3}, M_i), \tag{4.1}$$

$$c_i = \sum_{8}^{15} dot(I_{3\times3}, M_{i-8}), \tag{4.2}$$

and

$$p_i = \sum_{16}^{23} dot(I_{3\times3}, M_{i-16}) \tag{4.3}$$

where $dot(\cdot)$ represents the dot product operation, M_i is the mask, and $I_{3\times3}$ is 3×3 neighbors of the center pixel of each frame. n_i, c_i and p_i are the spatiotemporal directional response values of the first layer for the next, center, and previous frames respectively.

In order to generate the VDP-feature vector, we need to know the t most prominent directional bits for all three consecutive frames. These t bits are set to 1 and the rest of 8-bit VDP pattern of each frame are set to 0. Then a binary code is formed to each pixel from each frame, which will be mapped to its own bin to build a histogram. Finally, we concatenate these three histograms of these three consecutive frames to obtain the final VDP-feature vector, which is the descriptor for each center frame (frame of interest) that we used to recognize the face image by the help of a classifier. The final VDP code can be derived by

$$VDP = \sum_{i=0}^{7} f(n_i - n_t) \times 2^i \| \sum_{i=8}^{15} f(c_i - c_t) \times 2^{i-8} \| \sum_{i=16}^{23} f(p_i - p_t) \times 2^{i-16} \tag{4.4}$$

where the thresholding function $f(x)$ can be defined as in Eq. (2.2).

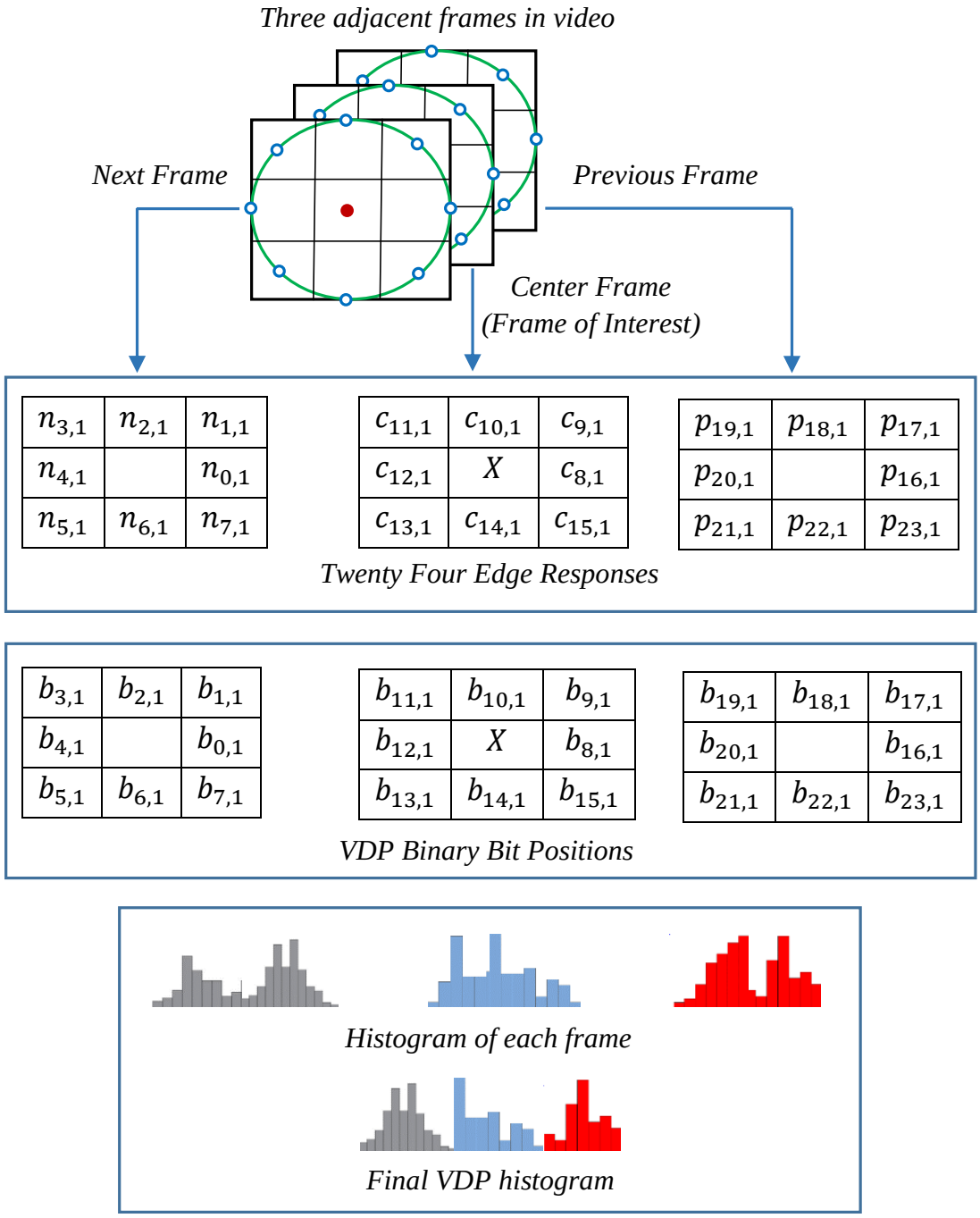

Figure 4.2: The twenty four edge responses with their VDP binary bit positions and concatenating them to form the final VDP feature vector.

4.3 High Order Volumetric Directional Pattern

Derived from a general definition of texture in a local neighborhood, the conventional VDP or the first order volumetric directional pattern encodes the directional information in the small 3×3 local neighborhood of a pixel of each three consecutive frames, which may fail to extract detailed information especially during changes in data collection environments due to several factors such as differences in the pose, random noise, illumination variation problems, etc. Therefore, we improve the proposed VDP to tackle this problem by calculating the n^{th} order volumetric directional variation patterns. The improved version of VDP is named high order volumetric directional pattern (HOVDP). The proposed HOVDP can capture more detailed discriminative information than the traditional VDP. Unlike the VDP operator, the extension of our proposed VDP technique extracts n^{th}-order volumetric information by encoding various distinctive spatial relationships from each neighborhood layer of a pixel in the pyramidal multi-structure way, and then concatenating the feature vector of each neighborhood layer to form the final HOVDP feature vector. Several observations can be made for HOVDP:

- Under the proposed framework, our proposed VDP is a special case of HOVDP, which simply calculate the 1^{st} order volumetric directional information in the local neighborhood of a pixel.
- The relation between the neighbor layers and the pixel under consideration could be easily weighted in HOVDP based on the distance between each layer and the central pixel. Because of that, the pixels within the closest layer to the central pixel has more weight than the others.
- Due to the same format and feature length of different order HOVDP, they can be readily fused, and the accuracy of the face recognition can be significantly improved after the fusion.

4.4 The Proposed Method

The proposed high order volumetric directional pattern (HOVDP) technique is an oriented and multi-scale volumetric directional descriptor that is able to extract and fuse the information of multi frames, temporal (dynamic) information, and multi poses and expressions of faces in input video to produce strong feature vectors. Given a central pixel in the middle (center) frame of three consecutive frames, to calculate the second order volumetric directional pattern (VDP_2) we first compute the first order (VDP_1), which is exaltedly the same as the original VDP by using the equations (4.1) - (4.4). Then,

$$n_{i,1} = n_i, \tag{4.5}$$

$$c_{i,1} = c_i, \tag{4.6}$$

$$p_{i,1} = p_i, \tag{4.7}$$

and

$$VDP_1 = VDP \tag{4.8}$$

where $p_{i,1}$, $c_{i,1}$, and $n_{i,1}$ are an eight bit binary number that describes the edge response pattern of each pixel of the first layer in the previous frame for $i = 16, 17, ..., 23$, center frame for $i = 8, 9, ..., 15$, and next frames $i = 0, 1, ..., 7$ respectively.

To make our calculation simple and easy to compute the high order relevant edge values, let us assume v_i the magnitude values for each layer separately after convolving the input image $I(x, y)$ with Kirsch masks in eight different directions $M_i(x, y)$ for $i = 0, 1, ..., 7$, which can be seen in Fig. 4.3. Then the directional edge values for the particular layer can be found as

$$c_{i,1} = v_{i,1}, \tag{4.9}$$

$$c_{i,2} = \frac{1}{3} \sum_{j=-1}^{j=1} v_{g,2}, \tag{4.10}$$

and

$$g = \mod (2i + j, 16), \qquad for\ i = 0, 1, ..., 7 \tag{4.11}$$

where $c_{i,1}$ and $c_{i,2}$ are the eight different directional edge response values of the first and second neighborhood layers respectively. $v_{g,2}$ is the magnitude value after convolving the input image with Kirsch kernels, the subscripts g and 2 are the number of surrounding pixels of each direction i and the second neighborhood layer (second order) respectively, and (mod) is the modulo operation which is used to maintain the circularly neighbors configuration.

$v_{6,2}$	$v_{5,2}$	$v_{4,2}$	$v_{3,2}$	$v_{2,2}$
$v_{7,2}$	$v_{3,1}$	$p_{2,1}$	$v_{1,1}$	$v_{1,2}$
$v_{8,2}$	$v_{4,1}$	X	$v_{0,1}$	$v_{0,2}$
$v_{9,2}$	$v_{5,1}$	$v_{6,1}$	$v_{7,1}$	$v_{15,2}$
$v_{10,2}$	$v_{11,2}$	$v_{12,2}$	$v_{13,2}$	$v_{14,2}$

Central Pixel with its 1st & 2nd Neighborhood Layers Edge Values

Figure 4.3: Magnitude values of two neighborhood layers.

Based on the observation that every corner or edge has high response values in particular directions, we are interested to know t the most prominent directional bits for all three consecutive frames in order to generate the VDP-feature vector of each neighborhood layer. These t bits are set to 1 and the rest of 8-bit pattern in each layer from each frame are set to 0. Then a binary code is

formed to each pixel in each layer from each frame, which will be mapped to its own bin to build a histogram for that particular layer of each frame. The volumetric directional pattern of each pixel position in the second neighbor layer can be formed as

$$VDP_2 = \sum_{i=0}^{7} f(n_{i,2} - n_{t,2}) \times 2^i \| \sum_{i=8}^{15} f(c_{i,2} - c_{t,2}) \times 2^{i-8} \| \sum_{i=16}^{23} f(p_{i,2} - p_{t,2}) \times 2^{i-16} \quad (4.12)$$

where the thresholding function $f(x)$ can be defined as in Eq. (2.2).

After identifying the volumetric directional pattern of each pixel in each neighborhood layer from each frame (VDP_1 for the first layer and VDP_2 for the second layer), a histogram is built to represent the whole distinguishing features of each neighbor layer from each frame separately, and then to obtain the final 2^{nd} order VDP-feature vector, which is the descriptor for each center frame (frame of interest), concatenate these histograms starting from the same layer order one by one which can be seen in Fig. 4.4.

In a general formulation, the n^{th} order volumetric directional pattern ($HOVDP_n$) of each pixel position in each neighbor layer from each frame can be defined as

$$HOVDP_n = \sum_{i=0}^{7} f(n_{i,n} - n_{t,n}) \times 2^i \| \sum_{i=8}^{15} f(c_{i,n} - c_{t,n}) \times 2^{i-8} \| \sum_{i=16}^{23} f(p_{i,n} - p_{t,n}) \times 2^{i-16} \quad (4.13)$$

where the subscript $n = 1, 2, ...$ is the volumetric directional pattern order (the number of neighborhood layers), $n_{t,n}$, $c_{t,n}$, and $p_{t,n}$ are the t^{th} most significant directional responses of each neighboring layer n from next frame, center frame, and previous frame respectively, and $f(x)$ is the thresholding function that can be defined as in Eq. (2.2). $n_{i,n}$, $c_{i,n}$, and $p_{i,n}$ are the eight different directional edge response values of each neighborhood layer n from next frame, center frame, and previous frame respectively which can be computed as

$$c_{i,n} = \frac{1}{2n-1} \sum_{j=-n+1}^{n-1} v_{g,n} \quad (4.14)$$

and

$$g = \mod (ni + j, 8n) \qquad for\ i = 0, 1, ..., 7 \tag{4.15}$$

The same procedure is applied to find the eight different directional edge response values of next frame $n_{i,n}$ and previous $p_{i,n}$.

4.4.1 The Adaptiveness of HOVDP

To encode the local structures information in the neighborhood adaptively, the image encoding and decoding (IED) strategy is applied [56, 57]. Because of that there is no need to any pre-experiments to find the most prominent edge values to set them to 1 and the rest to 0, the local directional patterns in this section can be formed adaptively by comparing the 8 neighboring pixels (excluding the central pixel) with their median of each neighborhood layer for each frame. If a neighboring pixel has a higher edge value than the median value (or the same value) then a 1 is assigned to that pixel, which is otherwise a 0. Therefore, the n^{th} order adaptive volumetric directional pattern $(AHOVDP_n)$ of each pixel position in each neighbor layer from each frame can be defined as

$$AHOVDP_n = \sum_{i=0}^{7} f(n_{i,n} - n_{m,n}) \times 2^i \| \sum_{i=8}^{15} f(c_{i,n} - c_{m,n}) \times 2^{i-8} \| \sum_{i=16}^{23} f(p_{i,n} - p_{m,n}) \times 2^{i-16} \tag{4.16}$$

where the subscript $n = 1, 2, ...$ is the volumetric directional pattern order (the number of neighborhood layers), $n_{m,n}$, $c_{m,n}$, and $p_{m,n}$ are the medians of each neighboring layer n from the next frame, center frame, and previous frame respectively. The thresholding function $f(x)$ can be defined as in Eq. (2.2).

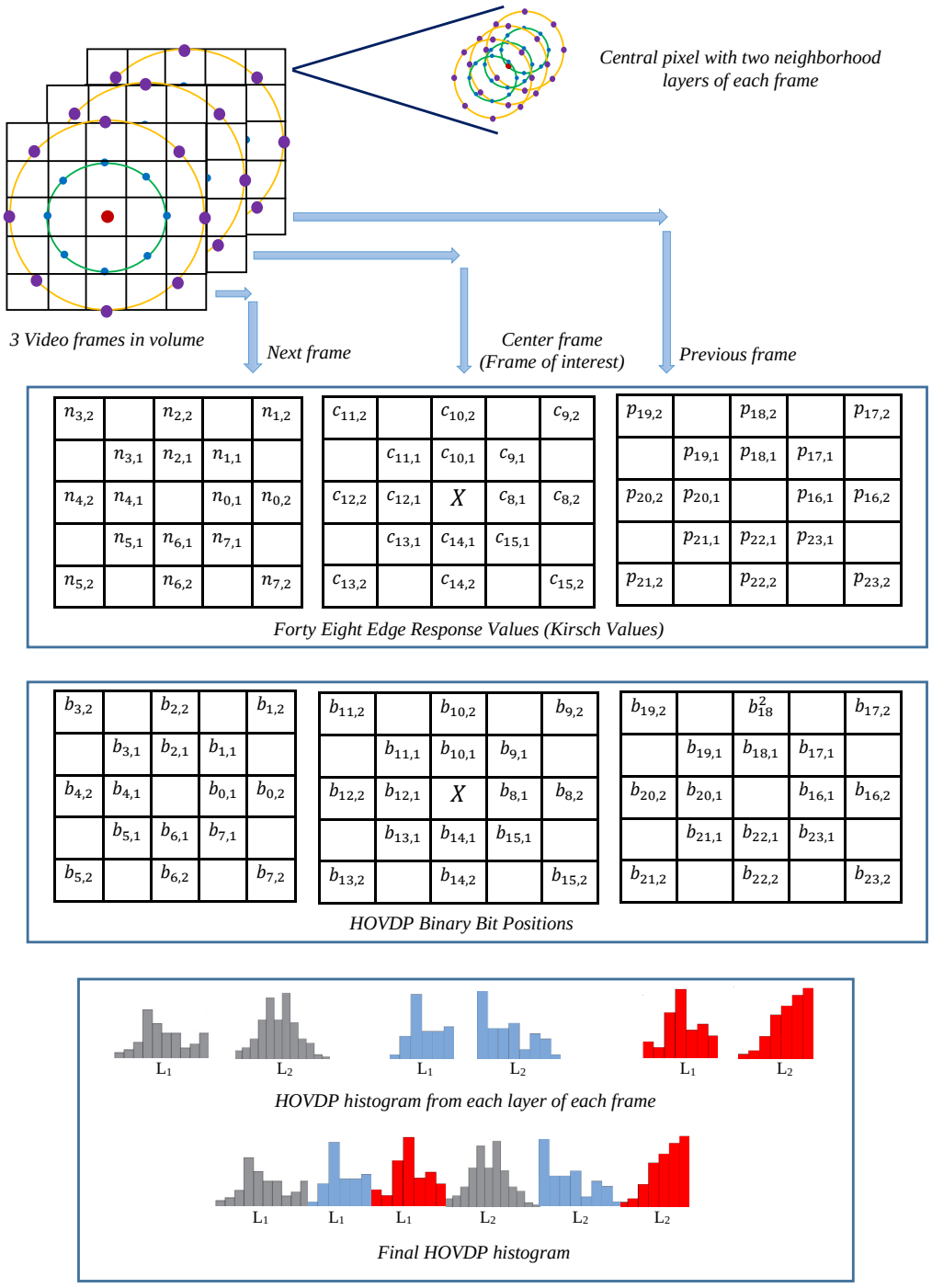

Figure 4.4: Procedure of 2^{nd} order VDP. L_1 and L_2 are the histograms of first and second layers respectively for each frame.

4.5 Experimental Results

To evaluate the illumination, pose, and expression variations robustness of the introduced method, we tested it on two publicly available datasets named, YouTube celebrities dataset [70] and Honda/UCSD database [71, 72]. All the face images in this work were detected by using the Viola and Jones face detector [69]. After manually removing the false detections, all the detected face images were resized to 64×64 and then the spatiotemporal information was extracted using the proposed VDP and HOVDP techniques. When it comes to the face recognition process, we represent the face using a VDP-feature histogram and HOVDP-feature histogram. The objective is to compare the encoded feature vector from one frame with all other candidates feature vector using two well-known different classifiers. The first one is support vector machine (SVM) classifier (we used LIBSVM), and the second one is a k-nearest neighbors classifier (k-NN). The corresponding face of the VDP/HOVDP feature vector with the lowest measured value indicates the match found.

Two different experiments are conducted for each database to verify the effectiveness and efficiency of the proposed HOVDP framework. The first one explores the effectiveness of different volumetric directional order (different neighborhood layers for each pixel of the image) as changing the number of the most prominent response values $\{t = 2, 3, \cdots, 6\}$. The second one evaluates the effectiveness of the proposed VDP, HOVDP, and AHOVDP by comparing them with four popular video based face recognition techniques. To avoid any bias, we randomly select the data for training and testing.

4.5.1 YouTube Celebrities Dataset

YouTube celebrities database is a large-scale video dataset which contains 1910 video sequences of 47 different celebrities (actors and politicians) that are collected from YouTube. The dataset is considered as one of the most challenging video databases due to the large illumination, pose,

and expression variations as well as low resolution and motion blur. In this part, we evaluated the proposed VDP and HOVDP on all 47 celebrities, while some of the state-of-the-art compared methods were evaluated on some of the subjects (e.g. in [73] they use only the first 29 celebrities). Following the prior works [74, 14], for each subject three video sequences are randomly selected as the training data, with the other six video clips are randomly selected for testing. We conduct one experiment by randomly selection of training/testing data. The clips contain different numbers of frames (from 8 to 400) which are mostly low resolution and highly compressed. Figure 4.5 shows some examples of cropped faces in this dataset.

Figure 4.5: YouTube celebrities database. Each row represents different samples of one subject.

To show the effectiveness of the proposed VDP (the special case of HOVDP) and HOVDP techniques, we summarize the recognition rates via changing the number of neighborhood layers (the order of the proposed approach) in the range $(1-4)$, varying the threshold t (the most prominent edge response values) in the range $\{t = 2, 3, \cdots, 6\}$, as well as comparing with the adaptive HOVDP in Table 4.1. From Table 4.1, it is found that $t = 4$ yields optimal performance for this dataset. Additionally, It is clear that the second order VDP improves the face recognition accuracy of the first order VDP in all test cases. While the third order and fourth order decrease the accuracy rates due to the fact that increasing the scale (the neighbors pixels) causes to extract and fuse the information of different poses and different locations of the face components, which produces confused feature vectors. Therefore, there is no need to increase the descriptor order which slightly differs on the case of still images based face recognition task as we found in the Chapter 3.

The performance results of well-known face recognition algorithms like regularized nearest points (RNP) [73], sparse approximated nearest points between image sets (SANP) and its kernel extension (KSANP) [74], and projection metric learning on Grassmann manifold (PML) [14] combined with grassmannian graph-embedding discriminant analysis (GGDA) [15] denoted as (PML-GGDA), with the proposed methods VDP and HOVDP on this dataset are presented in Table 4.2. Notice that the results we compared with are as we got from their original references which are mentioned in the table. Meanwhile, a part of this dataset was used in RNP [73] and three video sequences were randomly selected as the training data, with the other three sequences randomly selected as the testing data.

4.5.2 Honda/UCSD Dataset

Honda/UCSD database consists of 59 videos sequences of 20 different subjects. There are pose, illumination and expression variations across the sequences for each subject. Each video consists of about $12-645$ frames. Figure. 4.6 has shown some examples [72]. Each row corresponds

Table 4.1: Recognition rates using all video frames as changing the threshold t and the proposed approach order on YouTube celebrities database.

Descriptor	**Recognition Accuracy (%) Using libsvm Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	86.50 %	86.46 %	86.75 %	86.18 %	86.28 %	76.67 %
2^{nd} Order	**87.33 %**	**87.28 %**	**87.74 %**	**87.14 %**	**87.04 %**	77.38 %
3^{rd} Order	83.62 %	85.36 %	85.51 %	85.65 %	84.76 %	86.28 %
4^{th} Order	84.41 %	85.69 %	85.71 %	85.98 %	85.45 %	**86.56 %**
Descriptor	**Recognition Accuracy (%) Using knn Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	85.08 %	85.45 %	85.96 %	86.03 %	85.88 %	76.67 %
2^{nd} Order	**85.34 %**	**85.75 %**	**86.62 %**	**86.44 %**	**85.95 %**	76.35 %
3^{rd} Order	82.23 %	84.47 %	84.94 %	84.86 %	84.52 %	85.02 %
4^{th} Order	82.13 %	84.48 %	84.64 %	84.71 %	84.42 %	**85.41 %**

Table 4.2: Performance comparison of the proposed methods with well-known race recognition algorithms on YouTube celebrities database.

Recognition Accuracy $\pm$ Standard Deviation for Each Method				
SANP	KSANP	PML-GGDA	VDP (Proposed)	HOVDP (Proposed)
55.64 $\pm$ 5.74 %	65.46 $\pm$ 5.53 %	70.32 $\pm$ 3.69 %	86.75 $\pm$ 0 %	**87.74 $\pm$ 0 %**

to an image set of a subject. In our experiment, we use the standard training/testing configuration provided in [71], which means 20 sequences are used for training and the remaining 39 sequences for testing. We report results using all frames as well as with limited number of frames. Specifically, we conduct the experiments following the prior works [73, 74] by executing three parts of experiments, 1) using only the first 50 frames/video clip, 2) using only the first 100 frames/video clip, 3) using all video frames. In case a set contains fewer than the selected frames, all frames

are used for classification. The performance results of the proposed techniques VDP and HOVDP as changing the number of neighborhood layers (the order of the proposed approach) in the range $(1-4)$ along with changing the threshold t (the most prominent edge response values) in the range $\{t = 2, 3, \cdots, 6\}$ using set lengths 50 frames/clip, 100 frames/clip, and all frames/clip respectively are presented in Tables 4.3, 4.4, and 4.5.

The performance results of well known video based face recognition algorithms like SANP, KSANP, and RNP with the proposed methods VDP and HOVDP on this dataset are presented in Table 4.6. Notice that, the results we compared with are as we got from their original references which are mentioned in the table. For our proposed methods, we select the t that yields optimal performance for the comparison.

Table 4.3: Recognition rates using only 50 frames of each Video as changing the threshold t and the proposed approach order on Honda/UCSD database.

Descriptor	**Recognition Accuracy (%) Using libsvm Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	86.85 %	85.45 %	86.10 %	86.10 %	81.40 %	87.50 %
2^{nd} Order	**87.75 %**	**86.85 %**	**87.00 %**	**87.45 %**	**84.30 %**	**88.20 %**
3^{rd} Order	85.70 %	84.65 %	86.45 %	85.75 %	82.75 %	84.50 %
4^{th} Order	86.70 %	84.45 %	86.75 %	85.55 %	82.85 %	86.65 %
Descriptor	**Recognition Accuracy (%) Using knn Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	88.75 %	90.11 %	89.65 %	88.65 %	88.55 %	89.60 %
2^{nd} Order	**89.25 %**	**90.30 %**	**90.10 %**	**89.65 %**	**88.80 %**	**90.45 %**
3^{rd} Order	86.25 %	87.05 %	88.70 %	87.60 %	87.85 %	90.15 %
4^{th} Order	87.00 %	88.05 %	88.40 %	87.55 %	87.50 %	89.90 %

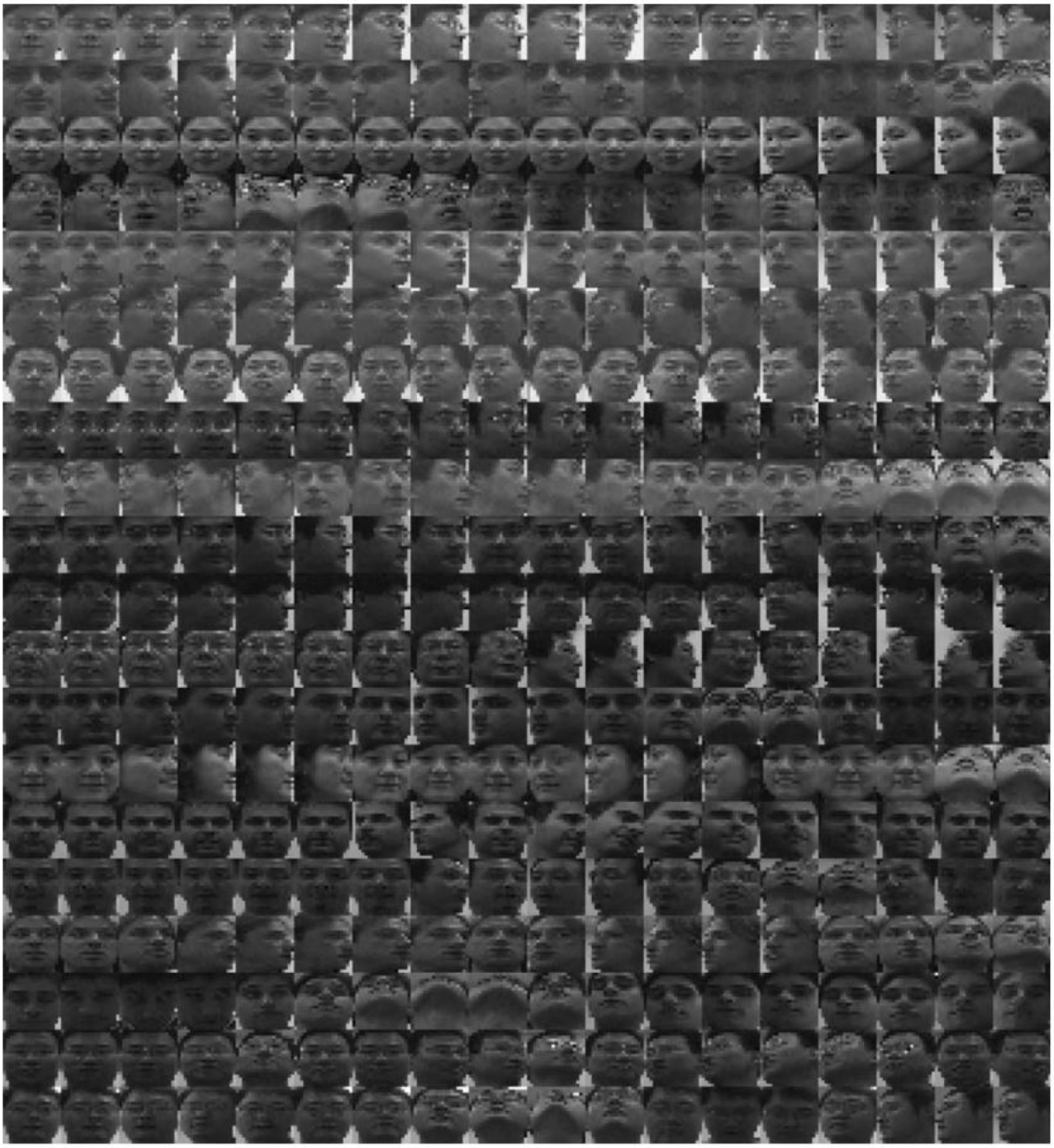

Figure 4.6: Honda/UCSD dataset. Each row represents different samples of one subject.

4.6 Discussion

In this chapter, we introduced two new feature descriptors, namely VDP and HOVDP. Throughout the performance evaluation in terms of face recognition accuracy, we found that VDP and its

Table 4.4: Recognition rates using only 100 frames of each video as changing the threshold t and the proposed approach order on Honda/UCSD database.

Descriptor	**Recognition Accuracy (%) Using libsvm Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	87.86 %	87.09 %	87.99 %	87.99 %	84.96 %	86.22 %
2^{nd} Order	**89.71 %**	**89.71 %**	90.21 %	88.89 %	86.25 %	89.60 %
3^{rd} Order	87.89 %	87.83 %	90.86 %	88.99 %	89.34 %	88.07 %
4^{th} Order	88.34 %	89.57 %	**91.68 %**	**89.36 %**	**90.20 %**	**90.84 %**
Descriptor	**Recognition Accuracy (%) Using knn Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	90.47 %	92.13 %	92.16 %	91.29 %	90.47 %	92.27 %
2^{nd} Order	**91.79 %**	**92.56 %**	**93.08 %**	**92.58 %**	**91.66 %**	**94.92 %**
3^{rd} Order	88.89 %	91.10 %	91.58 %	91.63 %	91.13 %	92.11 %
4^{th} Order	88.99 %	91.50 %	91.84 %	91.74 %	91.61 %	92.53 %

extension HOVDP are robust for video-based face recognition application. With a video as input and a gallery of videos, we performed face recognition process throughout the whole video clip frames. For each input frame, we extracted and combined the dynamic features of its two neighbors frames using our novel HOVDP algorithm, similar procedure was done to the gallery videos. Finally, we compared the encoded HOVDP-feature histogram from each frame with all other candidate HOVDP-feature vectors from all gallery video frames. By randomly selecting three video sequences per subject for training and other randomly six videos for testing as in the YouTube celebrities database and by using the stander sets for training and testing as in Honda/UCSD dataset, we finally showed that our method could achieve high recognition accuracy in most test cases.

From the evaluation results, it is found that the proposed HOVDP algorithm can successfully improve the accuracy rates compared to the original VDP in all test cases and exceed a set of state-of-the-art methods in most test cases. For the Honda/UCSD database, our proposed techniques

Table 4.5: Recognition rates using all video frames as changing the threshold t and the proposed approach order on Honda/UCSD database.

Descriptor	**Recognition Accuracy (%) Using libsvm Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	94.75 %	95.09 %	95.96 %	95.82 %	94.05 %	94.56 %
2^{nd} Order	**96.26 %**	**96.55 %**	**97.23 %**	96.57 %	95.74 %	96.53 %
3^{rd} Order	95.08 %	94.42 %	96.23 %	96.55 %	95.32 %	94.55 %
4^{th} Order	95.57 %	95.11 %	96.72 %	**97.13 %**	**95.84 %**	**96.88 %**
Descriptor	**Recognition Accuracy (%) Using knn Classifier**					
Order	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	AHOVDP
1^{st} Order	96.13 %	95.58 %	96.87 %	95.97 %	95.44 %	96.44 %
2^{nd} Order	**96.52 %**	**96.87 %**	**97.08 %**	**96.61 %**	**96.23 %**	**97.29 %**
3^{rd} Order	95.41 %	95.95 %	96.74 %	96.44 %	95.65 %	96.55 %
4^{th} Order	95.42 %	96.11 }	96.63 %	96.51 %	95.86 %	96.89 %

Table 4.6: Performance comparison of the proposed methods with well-known face recognition algorithms on Honda/UCSD database.

Number	**Recognition Accuracy (%) for Each Method**				
of Frames	SANP	KSANP	RNP	VDP (Proposed)	HOVDP (Proposed)
50	84.62 %	87.18 %	87.18 %	90.11 %	**90.45 %**
100	92.31 %	94.87 %	94.87 %	92.27 %	**94.92 %**
All frames	**100** %	**100** %	**100** %	96.87 %	97.29 %
Average	92.31 %	94.02 %	94.02 %	93.08 %	**94.22 %**

provide better recognition rates, although the other compared methods outperform ours in case of all frames are used. Meanwhile, our proposed HOVDP beats the others in case of smaller sets are used, which often occurs in real-world applications. For example, the tracking of a face may fail for a long sequence when only the first part of the video sequence is available for classification.

CHAPTER V

CONCLUSION AND FUTURE WORK

In this dissertation, three novel approaches were presented to extract features of face images: 1) high order local directional pattern (HOLDP) for calculating the n^{th}-order directional variation patterns for still images, 2) volumetric directional pattern (VDP) for videos, and 3) an extension of the VDP named high order volumetric directional pattern (HOVDP) for extracting and fusing the temporal information (dynamic features) in videos. The effectiveness of the methods was evaluated using different face recognition benchmarks, as well as compared with a set of state-of-the-art methods.

The first part of this research focused on the development of a local appearance feature extraction algorithm, which is capable of capturing discriminative information of still based face image. Derived from a general definition of texture in a local neighborhood, the conventional LDP encodes the directional information in the small 3×3 local neighborhood of a pixel, which may fail to extract detailed information, especially during changes in the input image due to random noise and illumination variation. To tackle this problem, a technique named HOLDP was developed. The key process of HOLDP is based on calculating the n^{th} order directional variation patterns by encoding various distinctive spatial relationships from each neighborhood layer of a pixel in the pyramidal multi-structure way. The output of HOLDP provides a spatial histogram for modeling the distribution information of a face image. From the evaluation results, it has been found that HOLDP

algorithm can successfully perform feature extraction tasks and exceed a set of state-of-the-art descriptors in all test cases.

The second part of this research concentrated on extracting and fusing various temporal information (dynamic features) from three consecutive frames for video based face images. The goal was to address a difficult problem of identifying human faces in video due to the presence of large variations in facial pose and expression, as well as poor video resolution. Based on this, two new descriptors were introduced namely VDP and HOVDP. VDP is a gray-scale pattern that characterizes and fuses the micro patterns of three consecutive frames to represent the appearance features deeply. Additionally, to provide a stronger discriminative capability in describing detailed texture information than the traditional VDP does, HOVDP was developed. HOVDP incorporates the concepts of HOLDP and VDP to encode various distinctive spatial relationships by considering not only the neighborhood layers of a pixel in the pyramidal multi-structure way but also the neighbor layers of a pixel in the adjacent frames. This combination significantly improved overall face recognition accuracy compared to the original VDP in all test cases and exceed a set of state-of-the-art methods in most test cases. From the evaluation results, it is observed that a fusion of static and dynamic features would provide a better representation for video based FR.

In future exploration, it would be a great interest to apply HOLDP, VDP, and HOVDP for various applications, such as hyperspectral image (HSI) classification which has been already done using VDP that was evidenced by yielding promising classification accuracy among the other competitors [17], shape localization, and automatic object detection, so that it will extensively show their robustness and usefulness, and further express significance of this research.

BIBLIOGRAPHY

[1] A. Essa and V. Asari, "Video-to-video pose and expression invariant face recognition using volumetric directional pattern," in *VISAPP 2015 - Proceedings of the 10th International Conference on Computer Vision Theory and Applications, Volume 2, Berlin, Germany, 11-14 March, 2015*. SciTePress, 2015, pp. 498–503.

[2] U. Raju, A. S. Kumar, B. Mahesh, and B. E. Reddy, "Texture classification with high order local pattern descriptor: local derivative pattern," *Global Journal of Computer Science and Technology*, vol. 10, no. 8, 2010.

[3] J. Suneetha, "A survey on video-based face recognition approaches," *International Journal of Application or Innovation in Engineering And Management (IJAIEM)*, vol. 3, no. 2, 2014.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[5] P. C. Yuen and J.-H. Lai, "Face representation using independent component analysis," *Pattern recognition*, vol. 35, no. 6, pp. 1247–1257, 2002.

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

[7] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[8] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[9] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[11] C.-H. Chan, J. Kittler, and K. Messer, *Multi-scale local binary pattern histograms for face recognition*. Springer, 2007.

[12] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (ldp) for face recognition," in *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, 2010.

[13] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, 2007.

[14] Z. Huang, R. Wang, S. Shan, and X. Chen, "Projection metric learning on grassmann manifold with application to video based face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 140–149.

[15] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, "Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2705–2712.

[16] R. Vemulapalli, J. K. Pillai, and R. Chellappa, "Kernel learning for extrinsic classification of manifold features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1782–1789.

[17] A. Essa, P. Sidike, and V. Asari, "Volumetric directional pattern for spatial feature extraction in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, accepted April 2017.

[18] T. Mäenpää and M. Pietikäinen, "Texture analysis with local binary patterns," *Handbook of Pattern Recognition and Computer Vision*, vol. 3, pp. 197–216, 2005.

[19] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer vision-eccv 2004*. Springer, 2004, pp. 469–481.

[20] A. Hadid, M. Pietikäinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–797.

[21] D. Huijsman and N. Sebe, "Content-based indexing performance: A class size normalized precision," in *Recall, Generality Evaluation, International Conference on Image Processing (ICIP'03)*, vol. 3, pp. 733–736.

[22] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 8, pp. 1371–1384, 2008.

[23] W. Ali, F. Georgsson, and T. Hellström, "Visual tree detection for autonomous navigation in forest environment," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 560–565.

[24] L. Nanni and A. Lumini, "Ensemble of multiple pedestrian representations," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 365–369, 2008.

[25] T. Mäenpää, J. Viertola, and M. Pietikäinen, "Optimising colour and texture features for real-time visual inspection," *Pattern Analysis & Applications*, vol. 6, no. 3, pp. 169–175, 2003.

[26] M. Turtinen, M. Pietikainen, and O. SilvÉn, "Visual characterization of paper using isomap and local binary patterns," *IEICE transactions on information and systems*, vol. 89, no. 7, pp. 2076–2083, 2006.

[27] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 657–662, 2006.

[28] V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Human activity recognition using a dynamic texture based method." in *BMVC*, vol. 1, 2008, p. 2.

[29] A. Oliver, X. Lladó, J. Freixenet, and J. Martí, "False positive reduction in mammographic mass detection using local binary patterns," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 286–293, 2007.

[30] S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3d teacher for car detection in aerial images," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[31] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved lbp under bayesian framework," in *Image and Graphics (ICIG'04), Third International Conference on*. IEEE, 2004, pp. 306–309.

[32] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block lbp representation," in *Advances in biometrics*. Springer, 2007, pp. 11–18.

[33] H. Zhang and D. Zhao, "Spatial histogram features for face detection in color images," in *Advances in Multimedia Information Processing-PCM 2004*. Springer, 2004, pp. 377–384.

[34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[35] S. Liao and A. C. Chung, "Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude," in *Computer Vision–ACCV 2007*. Springer, 2007, pp. 672–679.

[36] W. Zhang, S. Shan, H. Zhang, W. Gao, and X. Chen, "Multi-resolution histograms of local variation patterns (mhlvp) for robust face recognition," in *Audio-and Video-Based Biometric Person Authentication*. Springer, 2005, pp. 937–944.

[37] S. Z. Li, C. Zhao, M. Ao, and Z. Lei, "Learning to fuse 3d+ 2d based face recognition at both feature and decision levels," in *Analysis and Modelling of Faces and Gestures*. Springer, 2005, pp. 44–54.

[38] J. Zhao, H. Wang, H. Ren, and S.-C. Kee, "Lbp discriminant analysis for face verification," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 167–167.

[39] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[40] X. Feng, A. Hadid, and M. Pietikäinen, "A coarse-to-fine classification scheme for facial expression recognition," in *Image Analysis and Recognition*. Springer, 2004, pp. 668–675.

[41] S. Liao, W. Fan, A. Chung, and D.-Y. Yeung, "Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 665–668.

[42] G. Zhao and M. Pietikäinen, "Experiments with facial expression recognition using spatiotemporal local binary patterns," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 1091–1094.

[43] T. Gritti, C. Shan, V. Jeanne, and R. Braspenning, "Local features based facial expression recognition with face registration errors," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–8.

[44] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao, "Gender classification based on boosting local binary pattern," in *Advances in Neural Networks-ISNN 2006*. Springer, 2006, pp. 194–201.

[45] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *Advances in Biometrics*. Springer, 2007, pp. 464–473.

[46] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: a survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, no. 6, pp. 765–781, 2011.

[47] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[48] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *International Conference on Advances in Pattern Recognition*. Springer, 2001, pp. 399–408.

[49] D. K. Iakovidis, E. G. Keramidas, and D. Maroulis, "Fuzzy local binary patterns for ultrasound texture characterization," in *International Conference Image Analysis and Recognition*. Springer, 2008, pp. 750–759.

[50] T. Ahonen and M. Pietikäinen, "Soft histograms for local binary patterns," in *Proceedings of the Finnish signal processing symposium, FINSIG*, vol. 5, no. 9, 2007, p. 1.

[51] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[52] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE transactions on image processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[53] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[54] D.-J. Kim, S.-H. Lee, and M.-K. Sohn, "Face recognition via local directional pattern," *International Journal of Security and Its Applications*, vol. 7, no. 2, pp. 191–200, 2013.

[55] A. Essa and V. K. Asari, "Local directional pattern of phase congruency features for illumination invariant face recognition," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2014, pp. 90 940G–90 940G.

[56] A. Essa and V. Asari, "Face recognition based on modular histogram of oriented directional features," in *Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National*. IEEE, 2016, pp. 49–53.

[57] A. Essa and V. K. Asari, "Histogram of oriented directional features for robust face recognition," *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, vol. 4, no. 3, pp. 35–51, 2016.

[58] S. Z. Ishraque, T. Jabid, and O. Chae, "Face recognition based on local directional pattern variance (ldpv)," 2012.

[59] T. J. Hasanul Kabir and O. Chae, "Local directional pattern variance (ldpv): A robust feature descriptor for facial expression recognition," 2010.

[60] F. Zhong and J. Zhang, "Face recognition with enhanced local directional patterns," *Neurocomputing*, vol. 119, pp. 375–384, 2013.

[61] A. Essa and V. K. Asari, "Local boosted features for illumination invariant face recognition," *International Conference on Electronic Imaging*, 2017.

[62] A. S. Georghiades and P. N. Belhumeur, "Illumination cone models for faces recognition under variable lighting," in *Proceedings of CVPR98*, 1998.

[63] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 5, pp. 684–698, 2005.

[64] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*. IEEE, 1994, pp. 138–142.

[65] A. Nefian. Georgia tech face database. [Online]. Available: http://www.anefian.com/research/face_reco.htm

[66] A. M. Martinez, "The ar face database," *CVC technical report*, vol. 24, 1998.

[67] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[68] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[69] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[70] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[71] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–313.

[72] K. C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005.

[73] M. Yang, P. Zhu, L. Van Gool, and L. Zhang, "Face recognition based on regularized nearest points between image sets," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–7.

[74] Y. Hu, A. S. Mian, and R. Owens, "Face recognition using sparse approximated nearest points between image sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1992–2004, 2012.