MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

Effective Facial Feature Representation Based on Directional Micro-Patterns

By

Faisal Ahmed

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT)
Organisation of Islamic Cooperation (OIC)
Gazipur 1704, Bangladesh

September, 2012

Effective Facial Feature Representation Based on Directional Micro-Patterns

By

Faisal Ahmed

Student No: 104604

Supervised By

Md. Hasanul Kabir, PhD

A Thesis Submitted to the Department of Computer Science and Engineering (CSE) in Partial Fulfillment of the Requirements for the Award of the Degree of Masters of Science in Computer Science and Engineering (M.Sc. Engg. (CSE))

Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT)

Gazipur 1704, Bangladesh

September, 2012

Abstract

Automatic recognition and analysis of human face allow many interesting applications in biometrics, human-computer interaction, and security industry, such as person identification, age estimation, gender classification, and facial expression analysis. Deriving an efficient and effective feature representation is the fundamental component for any successful facial recognition system. However, the inherent variability of facial images caused by different factors like variations in illumination, pose, facial expression, alignment, and occlusions make classification a challenging task. Therefore, the aim of the ongoing research in facial recognition is to increase the robustness of the underlying feature representation against these factors.

In this thesis, we work toward developing a simple, yet effective appearance-based feature descriptor for representing human facial image. A new local texture pattern, namely the directional ternary pattern (DTP) has been introduced that can effectively capture the texture properties of a local neighborhood and exhibits robustness against illumination variations and random noise. The proposed DTP operator encodes the texture information of a local neighborhood by labeling the edge response values in all eight directions around the center point using three different levels. Our encoding scheme employs a threshold in order to differentiate between uniform and high-textured face regions, and thus, ensures the generation of ternary micro-patterns consistent with the local texture property. The location and occurrence information of the DTP micro-patterns within the facial image is then represented with a spatial histogram, which functions as the facial feature descriptor.

The performance of the proposed method has been evaluated in three different applications, which are i) facial expression analysis, ii) face recognition, and iii) gender classification from facial image. Different publicly-available benchmark image datasets were used for the experiments. We have also compared the performance of the proposed method with some widely-used local pattern-based feature descriptors. Experimental results demonstrate that, the DTP descriptor is more robust in extracting facial information and provides higher classification rate compared to some existing feature representation techniques.

Acknowledgment

I would like to first thank the Almighty Allah for all the blessings he has bestowed upon me throughout my entire life and throughout my masters program. Without the mercy of Allah, I wouldn't be where I am right now. All thanks and praises be to Allah.

Secondly, I would like to thank my thesis supervisor, Dr. Md. Hasanul Kabir, for his support and guidance on this thesis. Dr. Kabir has been an instrumental to this work and my career. He taught me how to do research, think critically, be a graduate student, and teach effectively. His all-time guidance, encouragement and continuous observation made the whole matter as a successful one.

It was my pleasure to get the cooperation and coordination from the Head of the Deaprtment, Professor Dr. M.A. Mottalib during various phases of the work. I am grateful to him for his constant and energetic guidance, constructive criticism and valuable advice. The faculty members of the CSE department helped make my working environment a pleasant one, by providing a helpful set of eyes and ears when problems arose.

I would like to thank the jury members of my thesis committee for the many interesting comments and criticism that helped improve this manuscript. Lastly, I am deeply grateful to my friends and family for their unconditional suport. This work would have never been completed without the consistent support and encouragement from them throughout my masters programme.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we first present an overview of our thesis that includes the significance of the problem and the problem statement in detail. After that, we present our thesis objectives and contributions. The chapter ends with a short description on the organization of this thesis.

1.1 Background

With the increasing development of hardware and software technologies, the demand for personalized interaction with consumer products and applications is also increasing day by day. The most popular approach for achieving this is the automated recognition and analysis of human face, which has attracted much attention over the last two decades. One of the main reasons is that, facial recognition provides a more natural and non-intrusive approach than the other biometrics, such as fingerprint, voice, or retina, where the cooperation of the subject is a must.

An automated facial recognition system usually consists of a sequential arrangements of processing blocks, which is similar to a generic pattern recognition model. The fundamental components embedded in a facial recognition system are: image acquisition, pre-processing, feature extraction, classification, and post-processing, as shown in Figure 1.1. A key issue in successful face analysis is to find an efficient feature extraction method, which generates the facial feature vector for classifier training and testing. Although much work has been done, the inherent variability of facial images caused by different factors like variations in

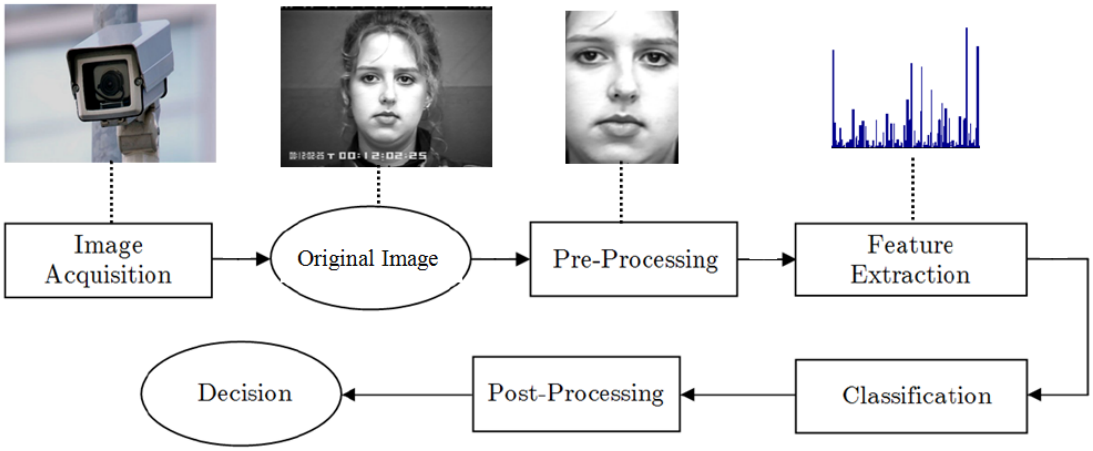


Figure 1.1: Components of a generic facial recognition system.

illumination, pose, alignment, occlusion, and aging makes robust feature extraction a difficult and challenging task [5]. Therefore, extraction of facial feature is an active research topic in computer vision, where the aim is to make the features robust against these factors.

1.2 The Significance of the Problem

The ability to identify and analyse facial images could lead to many useful applications in biometrics, human-computer interaction, data-driven animation, surveillance, and social robotics [6]. Nowadays, availability of inexpensive video cameras has fueled the widespread development of facial recognition-based customized applications for devices like laptops, cell-phones etc. The reason is that, augmenting person identification, gender classification, or facial expression recognition with applications specific to a person or gender can provide a more user-friendly environment and human-like interaction. There are various services which can be enabled through this kind of recognition system, such as personalized TV program [7, 8], intelligent digital photography [9], smart home [10] and many more.

In the last few years, social robotics has attained an increasing interest of scientific community. In this field of research, the objective is to design robots with the ability to move in populated environments and to interact with people.

Person identification is one of the primary tasks for achieving this. Vision-based face recognition provides a natural, intuitive, and nonintrusive approach for recognizing people. While for humans this is a trivial task, for robots it is still a very challenging problem. Despite some recent advances, vision-based face recognition is far from reaching human capability. For this reasons, new and more robust computer vision and pattern recognition approaches need to be investigated.

1.3 Problem Statement

Deriving an efficient, robust, and effective feature representation is a critical issue for any successful facial recognition system [11]. Although it receives considerable attention, the inherent variability of facial appearances makes recognition in unconstrained environment a difficult and challenging task. Human faces are non-rigid, dynamic objects with a large diversity in shape, color and texture, due to multiple factors such as head pose, lighting conditions (contrast, shadows), facial expressions, occlusions (glasses), ageing, and other facial features (make-up, beard) [12].Therefore, the aim of the ongoing research in automated facial analysis is to increase the robustness of the underlying facial feature descriptor against these factors.

An extracted facial feature can be considered an efficient representation if it can fulfill three criteria: first, it minimizes within-class variations of expressions while maximizes between-class variations; second, it provides robustness in uncontrolled environment; and third, it can be described in a low-dimensional feature space to ensure computational speed during the classification step [13,14]. Therefore, it is both a challenge and our motivation to find a facial feature descriptor satisfying the following three criteria: i) distinctiveness; ii) robustness; and iii) computationally inexpensive cost.

1.4 Thesis Objectives

Our thesis aimed at designing an effective appearance-based facial feature descriptor based on a robust local texture encoding scheme, which overcomes the limitations of the existing texture patterns. The designed facial feature representation should satisfy the following criteria:

- The feature descriptor should provide robustness against illumination variations and random noise in order to be used in real-life scenarios.
- It should be applicable in different face related problem domains, such as face recognition, facial expression analysis, and gender classification.
- The feature descriptor should achieve superior performance in unconstrained environment than the existing state-of-the-art techniques.
- The feature extraction method should be computationally efficient and effective.

1.5 Thesis Contributions

The main contributions of this thesis are summarized as follow:

- In this thesis, we present a simple, yet effective and robust facial feature descriptor constructed with a new local texture pattern, the directional ternary pattern (DTP) for representing human facial image. The proposed DTP operator encodes the texture information of a local neighborhood by quantizing the edge response values in all eight directions around the center point using three different levels. The proposed encoding scheme employs a threshold in order to differentiate between uniform and high-textured face regions, and thus, ensures the generation of ternary micro-patterns consistent with the local texture property. The location and occurrence information of the DTP micro-patterns within the facial image is then used as the feature descriptor.

- The recognition performance of the proposed method is evaluated in three different face related problems: i. facial expression recognition, ii. face recognition, and iii. gender classification. Different machine learning methods are exploited to classify images from several benchmark databases. Extensive experiments show the superiority of the DTP feature descriptor against some well-known appearance-based feature representation methods.

- We investigate DTP features for low-resolution facial recognition, a critical problem but seldom addressed in the existing work. Compared to the previous work, DTP features provide better performance, which is very promising for real-world applications.

- In real-world scenarios, images can easily be corrupted with random noise. Therefore, in this thesis, we empirically evaluate the performance of our proposed method under the presence of Gaussian noise. Experimental results show that, DTP performs more robustly under the presence of noise than the other existing methods.

1.6 Organization of the Thesis

The rest of the thesis will be organized as follows: in Chapter 2 we present the background and related work on different facial feature representation methods. We also present a description on the pattern recognition algorithms used in our experiments. We state the detailed description of the proposed facial feature representation scheme in Chapter 3. The experimental setup and the performance evaluation of our proposed method goes on Chapter 4. And we draw conclusion in Chapter 5 with a note of the future scope of research in this area.

Chapter 2

Literature Review

In this chapter, we first present a discussion on different geometric and appearance-based facial feature representation, which is followed by a review on different local appearance-based methods. Finally we end the literature review with a description of the pattern recognition methods used in this study.

2.1 Facial Feature Representation Methods

During the last two decades, many methods have been proposed for different face-related problems, where different facial feature extraction techniques have been introduced. Based on the types of features used, facial feature extraction approaches can be roughly divided into two different categories: geometric feature-based methods and appearance-based methods [6].

2.1.1 Geometric Feature-Based Methods

In geometric feature-based methods, the feature vector is formed based on the geometric relationships, such as positions, angles or distances between different facial components (eyes, ears, nose etc.). Earlier methods for facial recognition are mostly based on these geometric feature representations.

For facial expression recognition, facial action coding system (FACS) [15] is a popular geometric feature-based method which represents facial expression with the help of a set of action units (AU). Each action unit represents the physical behavior of a specific facial muscle. Later, Zhang [1] proposed a feature extraction method based on the geometric positions of 34 manually selected fiducial

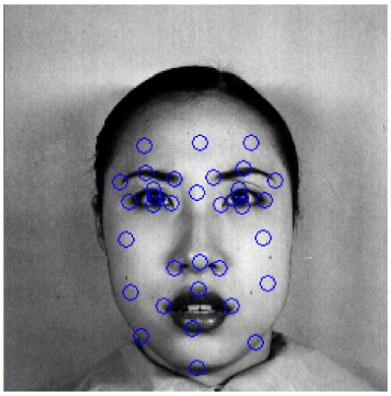


Figure 2.1: Locations of 34 fiducial points (Courtesy of Z. Zhang [1]).

points. A similar representation was adopted by Guo and Dyer [16], where they employed linear programming in order to perform simultaneous feature selection and classifier training. Recently, Valstar et al. [17,18] have studied facial expression analysis based on tracked fiducial point data and reported that, geometric features provide similar or better performance than appearance-based methods in action unit recognition.

For face recognition, elastic bunch graph match (EBGM) [19] algorithm is the most commonly-used geometric feature-based method. In this method, human faces are represented as graphs, where nodes are positioned on fiducial points and edges are labeled using distance vectors. A set of Gabor wavelets co-efficients is assigned to each node, which is called jet. While the geometrity of the faces is encoded by the edges, the jets represents the grey value distribution. Brunelli and Poggio [20] investigated a set of 16 geometric features for designing a gender classification model. The features were extracted with HyperBF networks. Later, Abdi et al. [21] experimented with a radial basis function (RBF) network and a perceptron, and achieved good classification result for both pixel-based inputs and measurement-based inputs (geometric features).

However, the effectiveness of geometric methods is heavily dependent on the accurate detection of facial components, which is a difficult task in changing and unconstrained environment, thus making geometric methods difficult to accommodate in many scenarios [6].

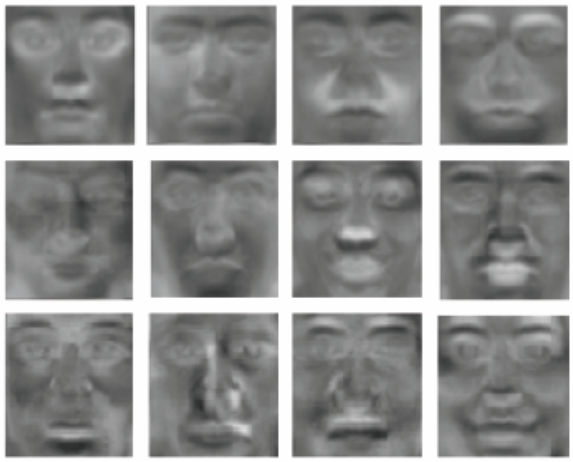


Figure 2.2: Facial feature representation based on PCA for expression recognition (Courtesy of Uddin et al. [2]).

2.1.2 Appearance-Based Methods

Appearance-based methods extract the facial appearance by applying image filter or filter bank on the whole face image or some specific facial regions. For facial expression recognition, Principal component analysis (PCA) [22], independent component analysis (ICA) [23, 24], Gabor wavelets [25, 26] and more recent enhanced ICA (EICA) [2] are the commonly-used appearance-based methods. Among these techniques, PCA is a global feature extraction method, where the whole facial image is taken into account during feature vector generation. This method provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in the sense of least mean squared reconstruction error [27]. On the other hand, ICA and Gabor wavelets extract the local features of an image, therefore called local feature descriptors.

In face recognition, PCA is usually called the 'eigenface' method, which was introduced by Turk and Pentland [28]. Later, linear discriminant analysis (LDA) [29], 2D PCA [30], local features analysis (LFA) [31], and dynamic link architecture (DLA) [32] methods were also investigated for feature representation in face recognition systems. Although PCA and ICA feature descriptors can effectively capture the variability of the training images, their performances deteri-

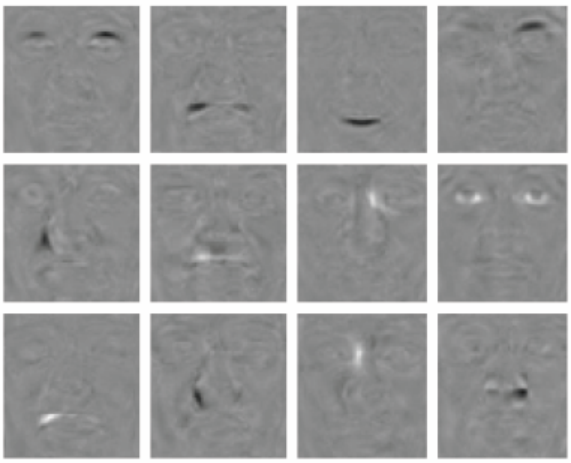


Figure 2.3: Facial feature representation based on ICA for expression recognition (Courtesy of Uddin et al. [2]).

orate in changing environment [33]. Donato et al. [34] presented a comprehensive analysis of different techniques for facial action recognition, which included PCA, ICA, local feature analysis (LFA), Gabor wavelets, and local principal components (PCs). Among these techniques, ICA and Gabor wavelets provided the best recognition rate. However, Gabor wavelets based facial image representation is time and memory intensive, since it requires convoluting facial images with multiple Gabor filters of many scales and orientations.

Recently, local appearance face descriptors based on local binary pattern (LBP) [11, 35] and its variants [36] have attained much attention due to their superior performances in uncontrolled environment. Local binary pattern is a simple, yet effective local texture description technique, which is computationally efficient and robust against non-monotonic illumination variation. The LBP operator encodes the local texture by quantizing the neighbor gray levels of a neighborhood with respect to center value and thus forms a binary pattern that acts as a template for micro-level information such as edge, spot, or corner. However, the LBP method performs weakly under the presence of large illumination variation and random noise [5], since a little variation in the gray level can easily change the LBP code. Later, local ternary pattern (LTP) [5] was introduced to increase the robustness of LBP in uniform and near-uniform regions by adding

an extra intensity discrimination level and extending the binary LBP value to a ternary code. More recently, Sobel-LBP [33] has been proposed to improve the performance of LBP by applying Sobel operator to enhance the edge information prior to applying LBP for feature extraction. However, in uniform and near-uniform regions, Sobel-LBP generates inconsistent patterns as it uses only two discrimination levels just like LBP. Another approach using derivative-based local texture pattern [37] takes the advantage of detailed high-order derivative descriptions and keeps the spatial relationships in local regions. Another method named local directional pattern (LDP) [6] employed a different texture encoding approach, where directional edge response values around a position is used instead of gray levels. The motivation was to exploit more stable edge response values instead of gray levels in order to encode the local texture, which increases the robustness of the underlying feature descriptor. Although this approach achieves better recognition performance than local binary pattern, LDP tends to produce inconsistent codes in uniform and near-uniform facial regions and is heavily dependent on the selection of the number of prominent edge directions.

2.2 Local Texture Pattern Operators

In this section, we present a review on some widely-used local pattern-based facial feature descriptors.

2.2.1 Local Binary Pattern (LBP)

Local binary pattern (LBP) is a gray-scale and rotation invariant texture primitive that describes the spatial structure of the local texture of an image. LBP was first introduced for texture classification [38] and later this method has been successfully applied in face-related problems, such as face recognition [35], gender classification [39, 40], and facial expression recognition [11]. The LBP operator selects a local neighborhood around each pixel of an image, thresholds the P neighbor gray values with respect to the center pixel and concatenates the re-

sult binomially. The resulting binary value is then assigned to the center pixel. Formal definition of the LBP operator takes the following form:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c)2^p \tag{2.1}$$

$$s(v) = \begin{cases} 1, & v \geq 0 \\ 0, & v < 0 \end{cases} \tag{2.2}$$

Here, i_c is the gray value of the center pixel (x_c, y_c), i_p is the gray value of its neighbors, P is the number of neighbors and R is the radius of the neighborhood. In practice, the LBP operator considers the signs of the differences of the gray values of P equally spaced neighbors with respect to the central pixel in a local neighborhood, which is then represented using a P-bit binary number. If any neighbor does not fall exactly on a pixel position, then the value of that neighbor is estimated using bilinear interpolation. The LBP histogram of the encoded image block is then used as a texture descriptor for that block. The basic LBP encoding process is illustrated in Figure 2.4.

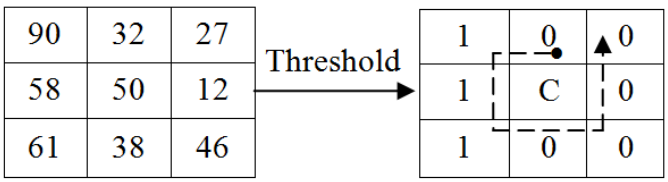


Figure 2.4: Illustration of the LBP encoding process. Here, the LBP binary code for pixel C is 01110000.

One extension to the original LBP operator, known as the uniform LBP (ULBP), exploits certain LBP patterns, which appear more frequently in a significant area of the image. These patterns are known as the uniform patterns as they contain very few spatial transitions (bitwise 0/1 changes) in a circular sequence of bits, which is represented by a uniformity measure U. The U value of an LBP pattern is defined as the number of bitwise transitions from 0 to 1 or vice versa in that pattern. One example of a uniform pattern is 00011111. It has a U value of 1 as there is only one transition from 0 to 1. Ojala et al. [38] observed

that, LBP patterns with $U \leq 2$ are the fundamental properties of texture, which provide a vast majority of all the 8-bit binary patterns present in any texture image. Therefore, uniform patterns are able to describe significant local texture information, such as bright spot, flat area or dark spot, and edges of varying positive and negative curvature [38]. All the other patterns ($U > 2$) are grouped under a miscellaneous label. The mapping of $LBP_{P,R}$ to $LBP^{U2}_{P,R}$ (LBP patterns with U value ≤ 2) is implemented with a lookup table of $2P$ entries.

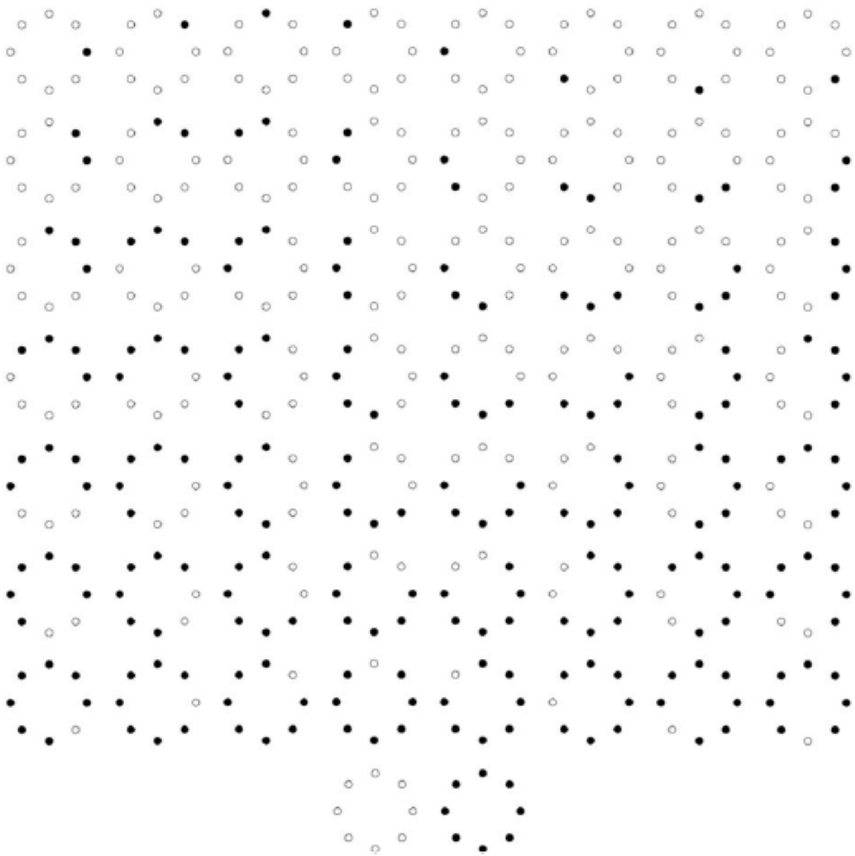


Figure 2.5: Uniform LBP Patterns for $P = 8$. The black and white dots represent bit value of 1 and 0 in the LBP code, respectively (Courtesy of Guo et al. [3]).

2.2.2 Local Ternary Pattern (LTP)

LBP codes are resistant against monotonic illumination variations. However, the LBP operator thresholds at exactly the value of the center pixel, and therefore,

are sensitive to noise in uniform and smooth regions since a little variation can cause its value to alter with respect to the center value. Given that many facial regions are relatively uniform, it is potentially useful to improve the robustness of the underlying descriptors in these areas [5]. To address this issue, local ternary pattern (LTP) [5] was proposed, which extends LBP to a 3-valued code in order to provide more consistency in both smooth and high-textured regions under the presence of noise. In the LTP encoding process, gray values in a zone of width $\pm t$ about the center pixel are quantized to 0, and those above $+t$ and below $-t$ are quantized to $+1$ and -1, respectively. Hence, the indicator $s(v)$ in equation (2.2) is replaced by a 3-valued function:

$$s'(i_p, i_c) = \begin{cases} 1, & i_p \geq i_c + t \\ 0, & |i_p - i_c| < t \\ -1, & i_p \leq i_c - t \end{cases} \tag{2.3}$$

Here, t is a user-defined threshold. The combination of these 3 levels gives the final LTP code, as illustrated in Figure 2.6.

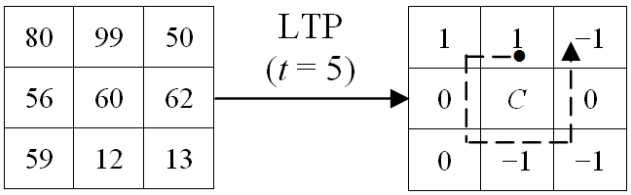


Figure 2.6: Illustration of the LTP encoding process. Here, the LTP code is 1100(-1)(-1)0(-1) for $t = 5$.

2.2.3 Local Directional Pattern (LDP)

LDP [4,6] is a gray-scale texture pattern which characterizes the spatial structure of a local image texture. A LDP operator computes the edge response values in all eight directions at each pixel position and generates a code from the relative strength magnitude. Since the edge responses are more illumination and noise insensitive than intensity values [6, 41], the resultant LDP feature describes the

local primitives including different types of curves, corners, and junctions, more stably and retains more information. Given a central pixel in the image, the eight directional edge response values m_i, i = 0,1,...,7 are computed by Kirsch masks M_i in eight different orientations centered on its position, as shown in Figure 2.7.

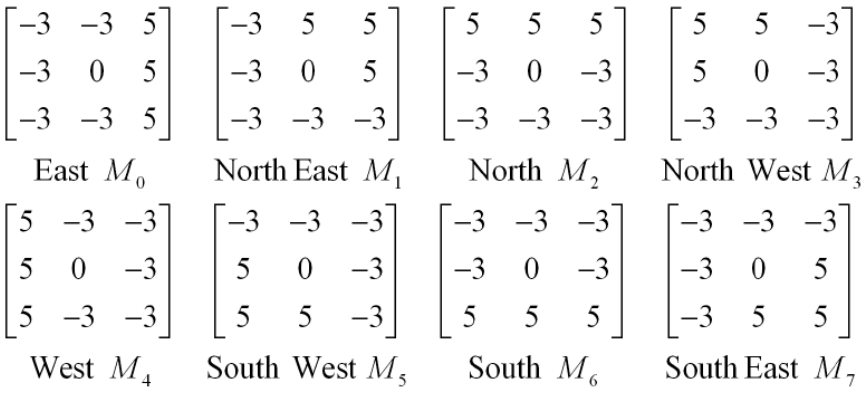


Figure 2.7: Kirsch edge response masks in all eight directions.

Presence of edge or corner will cause high edge-response values in their respective directions. Likewise, uniform or smooth regions will provide edge response values of same or similar magnitudes in different directions. Therefore, The LDP operator sets the most prominent k directions to 1 and others to 0 in order to obtain a 8-bit binary pattern based on the relative strength of the edge response values in different directions. Formally, the LDP code is derived by

$$LDP_k = \sum_{i=0}^{7} b_i(m_i - m_k)2^i \tag{2.4}$$

$$b_i(a) = \begin{cases} 1, & a \geq 0 \\ 0, & a < 0 \end{cases} \tag{2.5}$$

Here, m_k is the magnitude of the k-th most significant directional response. Since the edge responses are less sensitive to illumination and noise than intensity values, the resultant LDP feature retains more information and characterizes the texture primitives in a more robust manner. The LDP encoding process is illustrated in Figure 2.8.

After computing the LDP code for each pixel (r, c), the input image I of size $M \times N$ is represented by a LDP histogram H using equation 2.6. The resultant

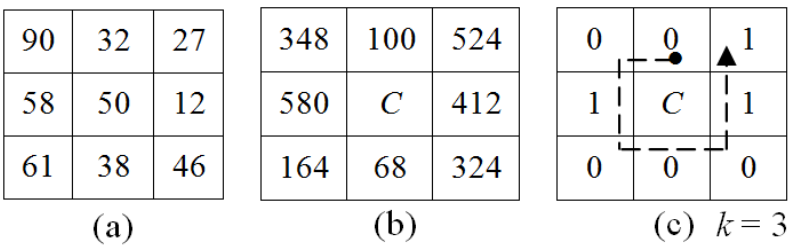


Figure 2.8: Illustration of the LDP encoding process, (a) original image, (b) magnitude of eight directional edge responses, (c) LDP binary code = 00100011 for center C.

histogram H is the LDP descriptor of that image.

$$H(t) = \sum_{r=1}^{M} \sum_{c=1}^{N} f(LDP_k(r,c), t) \tag{2.6}$$

$$f(a,t) = \begin{cases} 1, & a = t \\ 0, & otherwise \end{cases} \tag{2.7}$$

Here, t is the LDP code value. For a particular value of k, the LDP descriptor is a $C_k^8 = 8!/(8! \times (k-8)!)$ element feature vector.

Figure 2.9: An example of the LDP encoding, (a) Original image, (b) LDP encoded image (Courtesy of Jabid et al. [4]).

2.3 Machine Learning Techniques

For a facial recognition system, different machine learning approaches such as linear programming, template matching, linear discriminant analysis, or support vector machine (SVM) can be used. In this thesis, we have used template matching and support vector machine for recognizing facial images.

2.3.1 Template Matching (TM)

Template matching is used to verify the effectiveness of the proposed feature representation for face recognition. In template matching, the feature vector generated from the facial image of one person is compared with all the candidates' template feature vectors. The dissimilarity between the testing sample and the template feature vector is a test of goodness-of-fit that can be measured using a non-parametric statistic test. We used the chi-square statistic in our experiments. The chi-square measure is defined as:

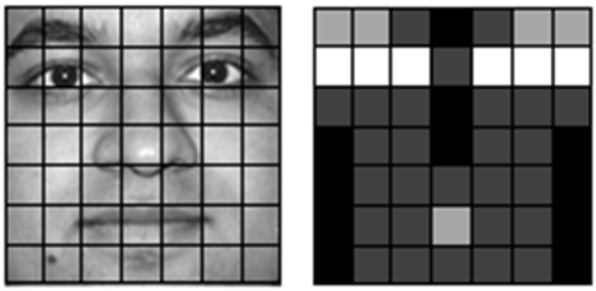
$$D(S, M) = \sum_{j=1}^{J} \frac{(S(j) - M(j)) \times (S(j) - M(j))}{S(j) + M(j)} \tag{2.8}$$

Here, S is the histogram of the testing sample, M is the model histogram of a category, and J is the number of bins in the histogram. After calculating the dissimilarity value for each of the candidates, the testing sample is assigned to the candidate with the smallest dissimilarity value.

Based on the face physiology knowledge, it can be said that, facial features extracted from some specific face regions (such as eye and mouth areas) contribute more than features extracted from other regions [6]. In this context, if the face regions can be weighted based on the importance of the contained information of the corresponding regions, it could potentially increase the discrimination ability of the underlying classifier. Therefore, weighted chi-square statistic is used to give more or less importance to particular face regions while measuring the dissimilarity value. The weighted chi-square measure is defined as:

$$D_w(S, M) = \sum_{i,j} w_i \frac{(S_i(j) - M_i(j)) \times (S_i(j) - M_i(j))}{S_i(j) + M_i(j)} \tag{2.9}$$

Here w_i is the weight assigned with the region R_i. Adopted weights for different regions of a partitioned (7×7) sample face image are illustrated in Figure



(a) (b)

Figure 2.10: (a) Sample face image divided into 7×7 sub-regions, (b) Adopted weights for different regions are shown in color. Here, black indicates weight 0, dark gray indicates 1, light gray indicates 2, and white indicates 4.

2.3.2 Support Vector Machine (SVM)

Support vector machine (SVM) is a state-of-the-art machine learning approach based on the modern statistical learning theory. It has been successfully applied in different classification problems. SVM performs the classification by constructing a hyper plane in such a way that the separating margin between positive and negative examples is optimal. This separating hyper plane then works as the decision surface.

Given a set of labeled training samples $T = \{(x_i, l_i),\ i = 1,\ 2,\ ...,\ L\}$, where $x_i \in R^P$ and $l_i \in \{-1, 1\}$, a new test data x is classified by

$$f(x) = sign(\sum_{i=1}^{L} \alpha_i l_i K(x_i, x) + b) \tag{2.10}$$

Here, α_i are Lagrange multipliers of dual optimization problem, b is a threshold parameter, and K is a kernel function. The hyper plane maximizes the separating margin with respect to the training samples with $\alpha_i > 0$, which are called the support vectors.

SVM makes binary decisions. To achieve multi-class classification, the common approach is to adopt the one-against-rest or several two-class problems. In

our study, we used the one-against-rest approach. Radial basis function (RBF) kernel was used for the classification problem. The radial basis function K can be defined as

$$K(x_i, x) = exp(-\gamma||x_i - x||^2), \gamma > 0 \tag{2.11}$$

$$||x_i - x||^2 = (x_i - x)^t(x_i - x) \tag{2.12}$$

Here, γ is a kernel parameter. A grid-search was carried out for selecting appropriate kernel parameter values, as suggested by [42].

Chapter 3

Proposed Feature Descriptor Based on Directional Ternary Pattern (DTP)

In this chapter, we explain the proposed directional ternary pattern (DTP), a new local texture pattern for effective facial feature representation, which includes the description of the basic DTP operator and how to construct facial feature descriptor based on DTP. After that, we present a modification of the DTP encoding scheme that can effectively reduce the DTP feature vector length.

3.1 Directional Ternary Pattern (DTP)

Both LBP and LTP methods use merely the gray level intensity values to encode the local texture of an image, which makes these methods unstable under the presence of large illumination variations. LBP is also susceptible to random noise. Therefore, some recent local pattern operators exploit more stable gradient information or edge response values instead of gray levels. One such method is the Sobel-LBP [33] that combines the LBP with the Sobel operator in order to enhance the local features and thus facilitates the extraction of more detailed information. Another approach is the local directinal pattern (LDP) [4, 6] that employs eight directional edge response values to encode the local texture. However, since both Sobel-LBP and LDP methods use two discrimination levels (0 and 1) to generate binary texture patterns, these methods tend to produce inconsistent codes in uniform and smooth regions, where the intensity variations among the neighbors are negligible. Considering these limitations, we present the

directional ternary pattern (DTP) - a new local texture primitive for representing facial feature. Instead of quantizing the gray level intensities like LBP or LTP, the DTP operator encodes the more stable edge response values of a local neighborhood, which retains more information of the local image content. In addition, while the binary LDP and Sobel-LBP coding fails to differentiate between smooth and high-textured facial regions, the proposed method employs a ternary coding scheme that discriminates between smooth and high-textured face regions with a threshold, and thus ensures the generation of consistent texture micro-patterns.

3.1.1 The Basic DTP Encoding Scheme

Edge responses are more stable than intensity values [6]. Therefore, an encoding scheme that exploits the edge responses in different directions can retain more information of the local region. The proposed directional ternary pattern (DTP) operator encodes the local texture by assigning a 3-valued code to each pixel based on the edge response values in different directions about the center pixel. Here, eight directional edge response values are computed by Robinson masks centered on a pixel oriented in eight different directions as shown in Figure 3.1.

−1	−2	−1
0	0	0
1	2	1

N, E, S, W
(rotated 90°)

−2	−1	0
−1	0	1
0	1	2

NW, NE, SE, SW
(rotated 90°)

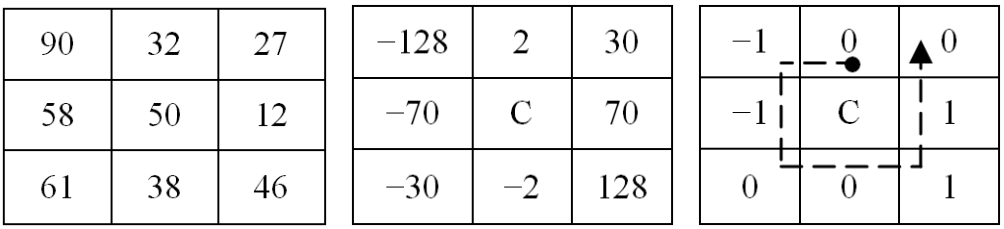
Figure 3.1: Robinson eight directional edge response masks. Different orientations are obtained by rotating these masks by 90°. Here, N, S, E, and W correspond to North, South, East, and West, respectively.

By applying these eight masks, we obtain eight edge response values, each of which represents the edge significance in its corresponding direction. Presence of edge or corner will produce high edge-response values in their respective direc-

tions. On the other hand, uniform or smooth regions will produce edge response values of similar magnitudes in different directions. Therefore, unlike the LDP operator [6] that always sets the most prominent k directions to 1 and others to 0 for forming a binary code regardless of the local region being uniform or not, the DTP operator employs a ternary coding scheme that differentiates between the smooth and high textured regions using a threshold. After computing the average μ of all eight directional edge response values, those responses within a $\pm t$ margin about the mean μ are quantized to 0, those above $\mu + t$ and those below $\mu - t$ are quantized to $+1$ and -1, respectively, as shown in equation 3.1.

$$S_{DTP}(r_i) = \begin{cases} 1, & r_i > \mu + t \\ 0, & \mu - t \leq r_i \leq \mu + t \\ -1, & r_i < \mu - t \end{cases} \tag{3.1}$$

Here, r_i is the edge response value in the i-th direction, μ is the average edge response value and t is a user-defined threshold. Figure 3.2 illustrates the basic DTP encoding method.



(a) (b) (c)

Figure 3.2: Illustration of the DTP operator, (a) Original image, (b) Eight directional edge responses, (c) DTP code = 0(-1)(-1)00110 for t=40.

One of the main advantages of the DTP operator is that, it exploits edge response values instead of gray levels for forming the texture patterns, which is more stable under the presence of noise and illumination variations. In addition, it ensures consistent pattern generation through the more robust ternary encoding scheme that exploits a threshold in order to differentiate between smooth and high-textured face regions. In order to reduce the length of the feature vector,

each DTP code is further split into its corresponding positive and negative parts, and treated as two separate binary patterns, namely P_{DTP} and N_{DTP}. Thus, the number of features reduces from 3^8 (= 6561) to 2×2^8 (= 512). Here, P_{DTP} and N_{DTP} take the form:

$$P_{DTP} = \sum_{i=0}^{7} S_P(S_{DTP}(r_i))2^i, \ where \ S_P(v) = \begin{cases} 1, & v = 1 \\ 0, & otherwise \end{cases} \tag{3.2}$$

$$N_{DTP} = \sum_{i=0}^{7} S_N(S_{DTP}(r_i))2^i, \ where \ S_N(v) = \begin{cases} 1, & v = -1 \\ 0, & otherwise \end{cases} \tag{3.3}$$

Figure 3.3 illustrates the generation of P_{DTP} and N_{DTP} codes from the original DTP code.

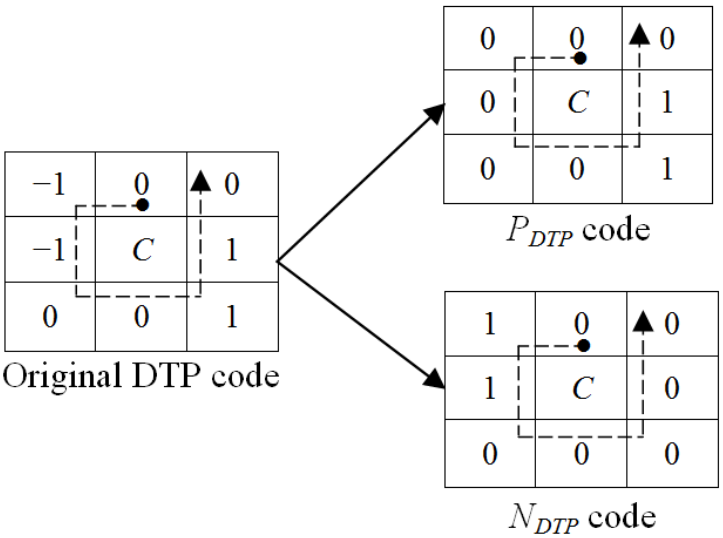


Figure 3.3: Generation of P_{DTP} and N_{DTP} codes from the original DTP code. Here, the original DTP code is 0(-1)(-1)00110 and the corresponding P_{DTP} and N_{DTP} codes are 00000110 and 01100000, respectively.

3.1.2 Facial Feature Representation with DTP

After applying the DTP operator on all the pixels of an image we get two encoded images, one for P_{DTP} and the other for N_{DTP}. Histograms generated from these two encoded images are spatially concatenated to form a combined histogram, the

DTP histogram, which functions as a feature representation for the face image. The DTP histogram generation process is shown in Figure 3.4.

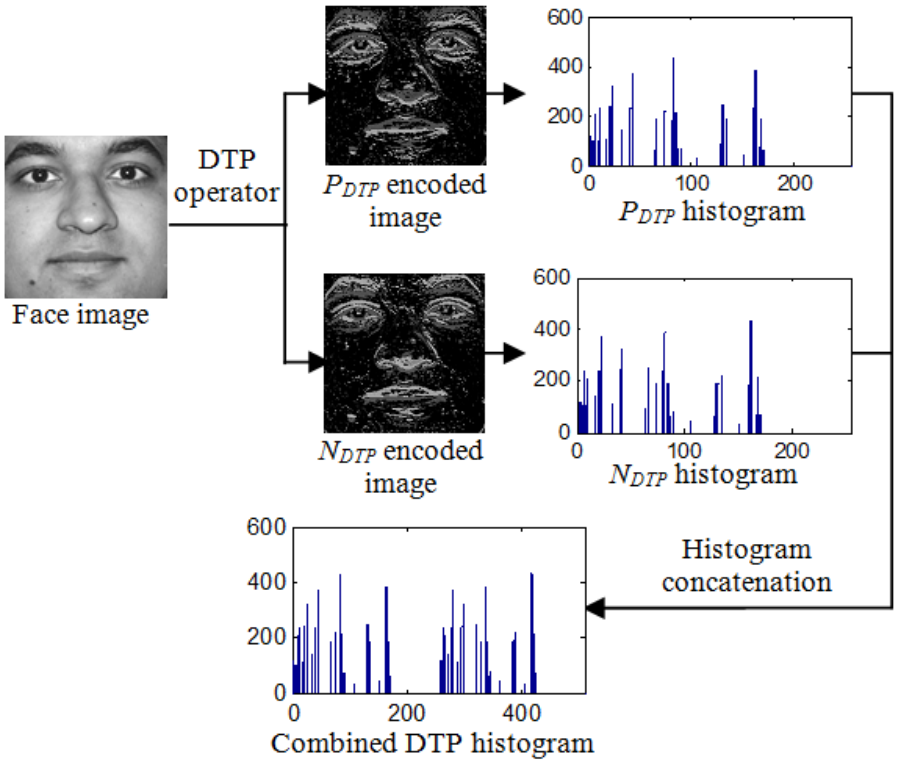


Figure 3.4: Illustration of DTP histogram generation process.

Histograms generated from the whole encoded image contain no location information of the micro-patterns, but merely their occurrences are expressed. However, presence of location information and spatial relationships provides a better facial feature representation and describes the image content more accurately [6, 35, 43]. Therefore, the DTP histogram is modified to an extended histogram in order to incorporate some degree of location information. First, each image is partitioned into a number of regions and individual DTP histograms are generated from each of those regions. Finally, the histograms of all the regions are concatenated to obtain an extended DTP histogram. For the facial expression recognition process, this histogram collection is used as the facial feature vector. The process is illustrated in Figure 3.5.

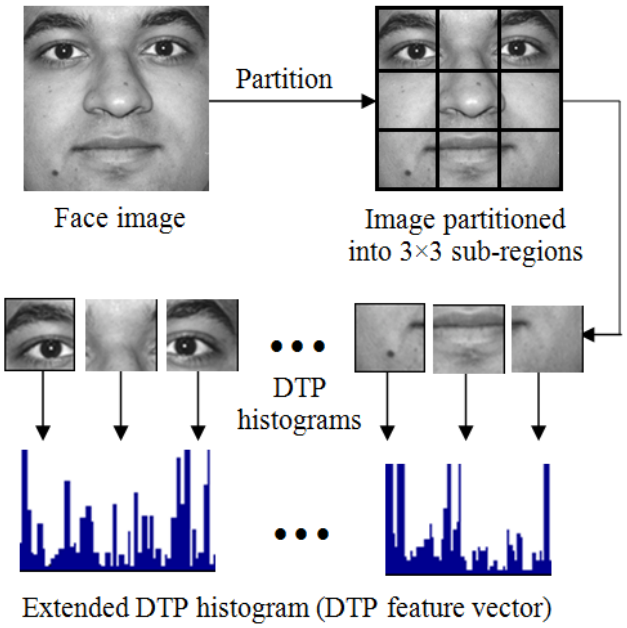


Figure 3.5: Each face image is partitioned into a number of regions and individual DTP histograms generated from each of the regions are concatenated to form the feature vector.

3.2 Compressing the DTP features

This section presents a modified encoding scheme for the DTP operator that effectively reduces the number of possible DTP features from 2×2^8 to only 3^4, without any loss of information.

3.2.1 Compressed DTP encoding

The strength of the DTP features lies in the quantization of stable edge response values and the discrimination between smooth and high-textured facial regions. To compute the eight directional responses, the DTP operator uses the Robinson compass masks. The reason is that, Robinson compass masks are easier to implement than other compass masks (such as Kirsch masks) as they rely only on the co-efficients of 0,1, and 2 and are symmetrical about their directional axis [44]. Therefore, only the results on four of the masks are required to be computed. The results from the other four masks can be obtained by negating the results

from the first four.

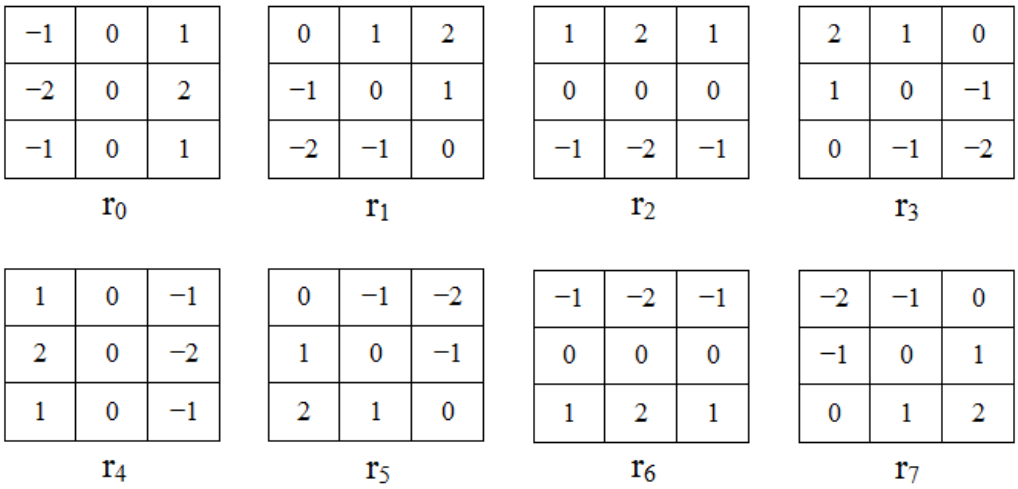


Figure 3.6: Robinson edge response masks. It can be observed that, the masks r_0, r_1, r_2, and r_3 are symmetrical to r_4, r_5, r_6, and r_7, respectively.

Now, let us assume that, N and S are the edge response values from two opposite directions (such as north and south). Due to the symmetric property of the Robinson masks, the magnitude of the edge responses N and S will always be the same, only the sign will be the opposite. As a result, the average edge response value around a point will alwasy be 0. Therefore, while quantizing the edge response values with a threshold, the edge responses from any two opposite directions N and S will always follow any of the three properties listed below for any $t > 0$:

- If $N > (\mu + t)$, then $S < (\mu - t)$.
- If $N < (\mu - t)$, then $S > (\mu + t)$.
- If $(\mu - t) \leq N \leq (\mu + t)$, then $(\mu - t) \leq S \leq (\mu + t)$.

From Figure 3.7, it can be observed that, based on the position of the one edge response, we can tell where the opposite edge response will lie for all $t > 0$. Therefore, instead of quantizing both of the edge response values in opposite directions, we can quantize one, without any loss of information, which effectively

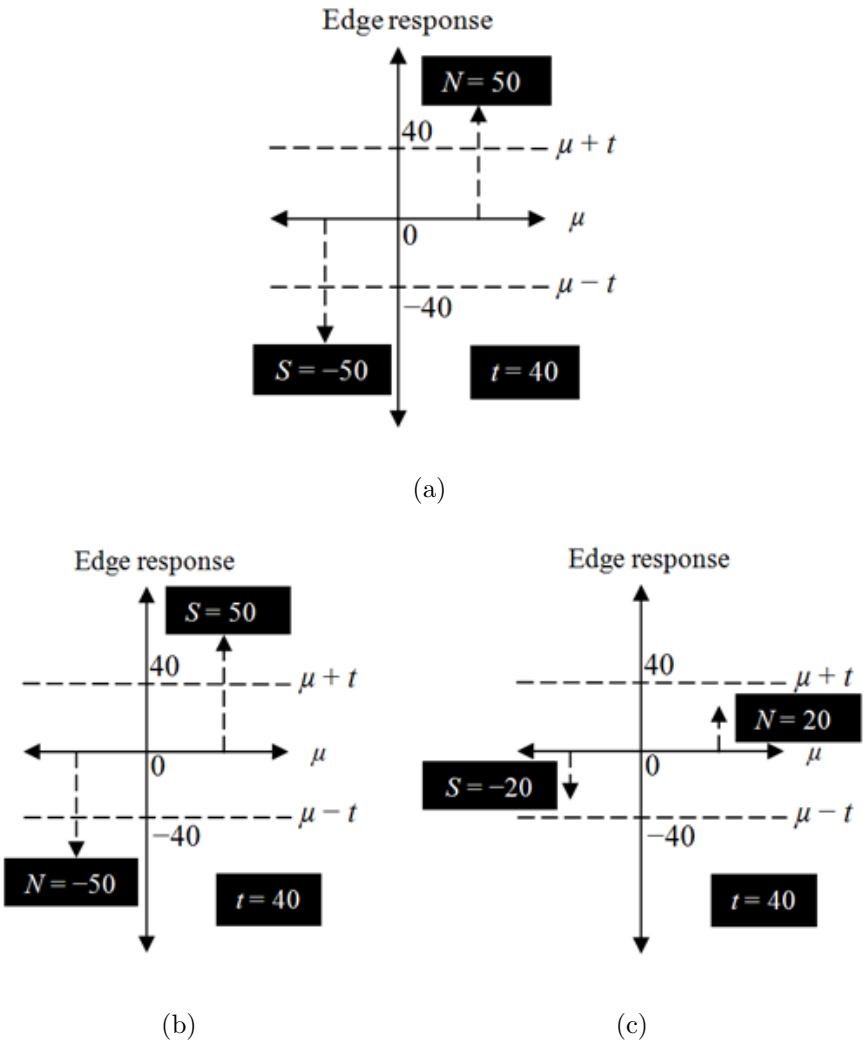


Figure 3.7: An illustration of the possible 3 conditions for $t = 40$ and different values of N and S.

decrease the number of possible DTP patterns from 3^8 to 3^4. In this new encoding scheme, each time we select two opposite edge response values and use one single level to represent both. Thus, we need only a 4-digit base-3 number to represent a local neighborhood. The encoding scheme can be represented formally by the following equation:

$$cDTP = S_{DTP}(R_N) \times 3^3 + S_{DTP}(R_E) \times 3^2 + S_{DTP}(R_{NE}) \times 3^1 + S_{DTP}(R_{NW}) \times 3^0 \tag{3.4}$$

$$S_{DTP}(v) = \begin{cases} 0, & v > \mu + t \\ 1, & \mu - t \leq v \leq \mu + t \\ 2, & v < \mu - t \end{cases} \tag{3.5}$$

Here, R_N, R_E, R_{NE} and R_{NW} are the edge response values from the north, east, north-east, and north-west directions, respectively. The difference between the basic DTP encoding and this compressed DTP encoding scheme is that, here we are using one base-3 digit for labeling two edge responses, while in the basic DTP encoding, each edge response is labeled with one base-3 digit. Therefore, the total number of possible DTP patterns in the compressed encoding scheme is 3^4 (=81) instead of 3^8 (=6561). The compressed DTP encoding is illustrated in Figure 3.8. In this figure, for the given 3×3 neighborhood, the corresponding compressed DTP code is 1012.

3.2.2 Facial Feature Representation with Compressed DTP

Applying the compressed DTP operator on all the pixels of an image will result in an encoded DTP image, where the value of each pixel will range between 0 and 80. The distribution information of these compressed DTP micro-patterns are then represented as a spatial histogram, namely the CDTP histogram. In order to incorporate some degree of location information of the micro-patterns, the CDTP histogram is modified to an extended histogram just like the DTP histogram. This is done by partitioning the compressed DTP encoded image into a number of regions and generating individual CDTP histograms from each of the regions. Finally, the histograms of all the regions are spatially concatenated to obtain the extended CDTP histogram, which functions as the facial feature vector. The process is illustrated in Figure 3.9.

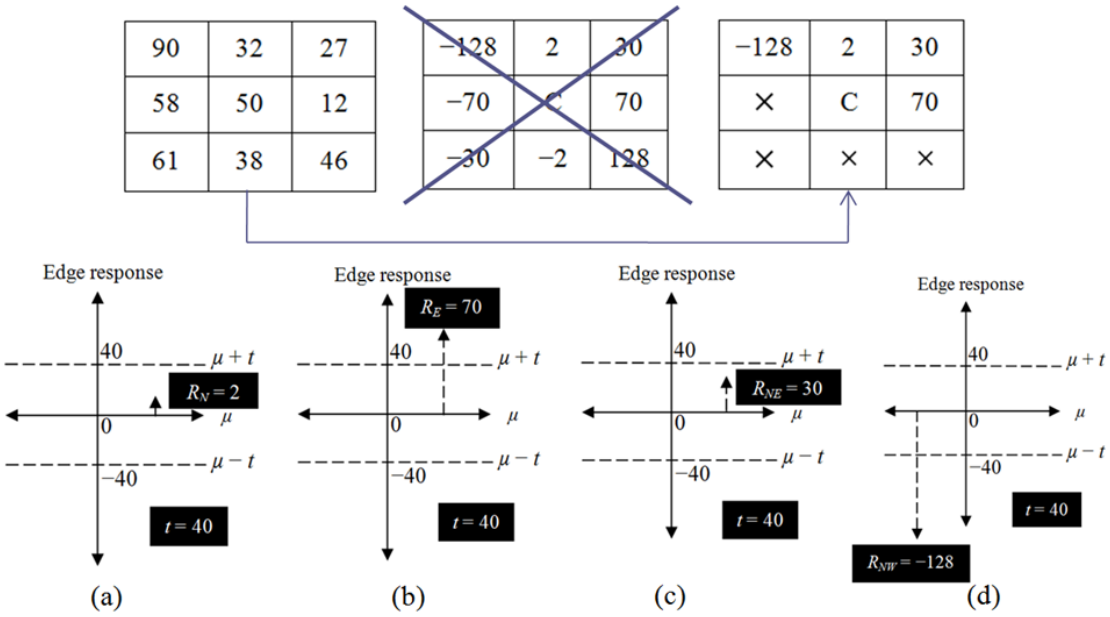


Figure 3.8: Illustration of the compressed DTP encoding ($t = 40$) using four directional edge response, (a) Quantization of edge response from North (N) (Compressed DTP code value = 1), (b) Quantization of edge responses from East (E) (Compressed DTP code value = 0), (c) Quantization of edge response from North-East (NE) (Compressed DTP code value = 1), (d) Quantization of edge response from North-West (NW) (Compressed DTP code value = 1). Here, the compressed DTP code is obtained by concatenating the quantization results of a, b, c, and d, which is 1012.

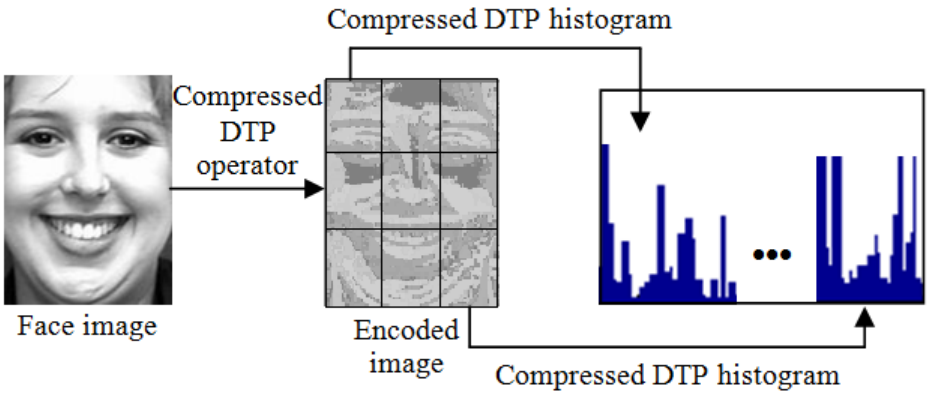


Figure 3.9: Generation of compressed DTP feature vector.

Chapter 4

Experiments and Results

In this chapter, we evaluate the performance of the proposed method for face recognition, facial expression recognition, and gender classification. We also present a comaprison of our method in terms of classification rate with some state-of-the-art local texture-based facial feature descriptors, namely local binary pattern (LBP) [11,35], local ternary pattern (LTP) [5], and local directional pattern (LDP) [4, 6]. We also assess the performance of the DTP descriptor for low-resolution and noisy images, which is presented at the end of this chapter.

4.1 Experimental Setup and Dataset Description

4.1.1 Face Recognition

The effectiveness of the DTP feature representation method for face recognition is evaluated in accordance to the Colorado State University (CSU) Face Identification Evaluation System that uses full-frontal face images from the FERET database [46]. The FERET database consists of a total of 14,051 gray-scale images representing 1,199 individuals. The images contain variations in lighting, facial expressions, pose angle, etc. In this work, only frontal faces from four different sets are considered. These sets are:

- fa set, used as a gallery set for training only, contains single frontal images of 1,196 people.

- fb set (1,195 images), the subjects were asked for an alternative facial expression than in the fa photograph.
- $dupI$ set (722 images), the photos were taken later in time.
- $dupII$ set (234 images), this is a subset of the $dupI$ set containing those images that were taken at least a year after the corresponding gallery image.

The face images were cropped from the original ones using the ground truth of the two eye positions and then normalized to 100×100 pixels. Figure 4.1 shows sample face images from the FERET database.

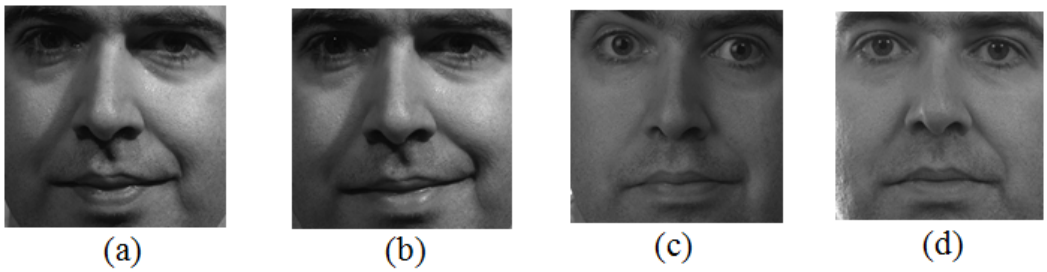


Figure 4.1: Sample face images of a person from different sets of the FERET database, (a) image from fa set, (b) image from fb set, (c) image from $dupI$ set, and (d) image from $dupII$ set.

4.1.2 Facial Expression Recognition

The recognition ability of the proposed method is evaluated based on a set of prototypic emotional expressions, which includes anger, disgust, fear, joy, sadness, and surprise. This 6-class expression can be further extended to a 7-class expression set by adding neutral face expression images. The performance evaluation is performed with two well-known image databases, namely the Cohn-Kanade (CK) facial expression database [45] and the Japanese female facial expression (JAFFE) database [25].

The CK database comprises 100 university students who were around 18 to 30 years old at the time of image acquisition. Among them, 65% were female,

15% were African-American, and 3% were Asian or Latino. A series of facial expression displays were performed by the subjects starting from neutral or near-neutral to one of the six prototypic emotional expressions stated before. The image sequences were digitized into 640×480 or 640×690 pixel resolution. In our setup, we first selected 1224 face image sequences from a total of 96 subjects, where each of the images was labeled as one of the six prototypic expressions. This 6-class expression dataset was then extended to a 7-class expression dataset by including additional 408 images of neutral expression face. Figure 4.2 shows the sample prototypic expression images from the CK database.

Figure 4.2: Sample expression images from the Cohn-Kanade database.

The JAFFE database comprises facial expression images of 10 Japanese female subjects. All the images were digitized into a resolution of 256×256 pixels. The images were taken from a frontal pose, and the subjects' hair was tied back in order to facilitate the exposure of all the expressive zones of the face. In the image scene, an even illumination was created using tungsten lights. Instead of revealing the actual names, the subjects are referred with their initials, which are KA, KL, KM, KR, MK, NA, NM, TM, UY, and YM. In our setup, the 6-class expression dataset comprises a total of 283 images, while the 7-class expression

set includes additional 50 neutral expression images. Figure 4.3 shows the sample prototypic expression images from the JAFFE database.

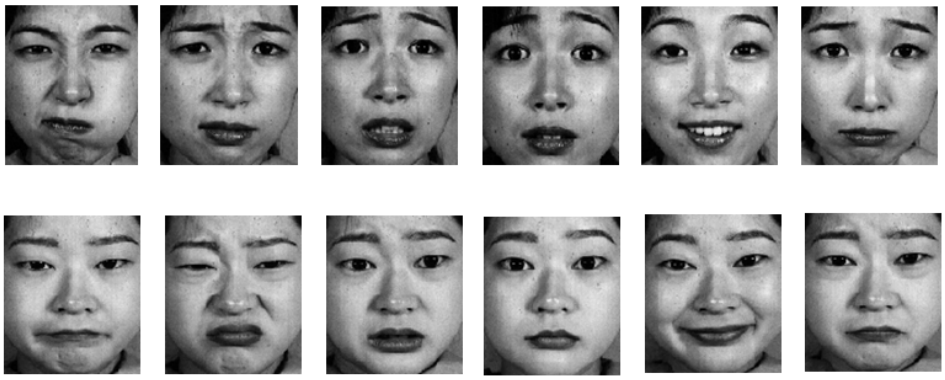


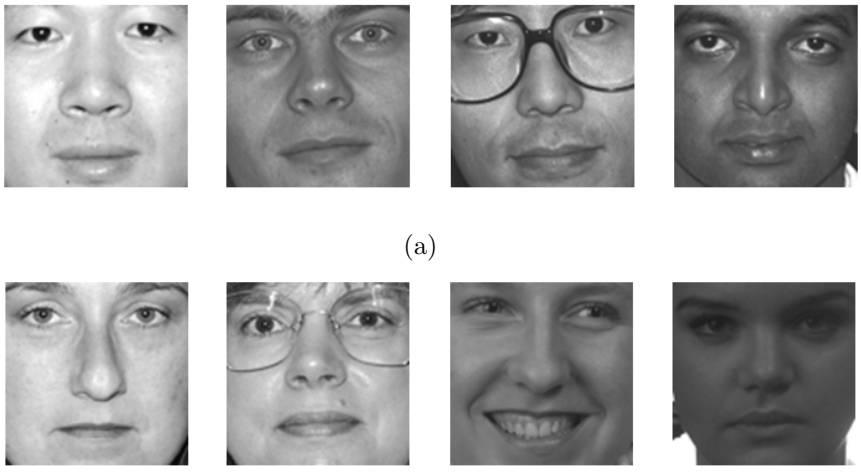
Figure 4.3: Sample expression images from the JAFFE database.

The selected images were cropped from the original ones based on the positions of the two eyes and normalized to 150 × 110 pixels. The ground-truth of eye position data was provided for cropping. No alignment of facial features (such as alignment of mouth) was performed in our setup. Figure 4.4 shows a sample cropped facial image from CK database.

Figure 4.4: Cropping of a sample face image from the original one.

4.1.3 Gender Classification

The effectiveness of the proposed method for gender classification is also evaluated with the face images from the FERET face database [46]. For our experiment,



(a)

(b)

Figure 4.5: Sample male and female face images from the FERET database, (a) male, (b) female.

we selected a total of 1800 images, among which 900 were male and the rest 900 were female face images. The selected images were cropped from the original 512×768 pixel images using the ground truth of the two eye positions and then normalized to 100×100 pixels. Figure 4.5 shows some sample male and female images from the FERET database.

4.2 Experimental Results

4.2.1 Performance Evaluation for Face Recognition

4.2.1.1 Optimal Parameter Selection

The classification rate of the proposed method can be influenced by adjusting two parameters: the threshold value t and the number of regions in which the facial image is to be partitioned. We first selected a fixed number of regions (7×7) and search for the optimal value of t. In our experiment, the DTP threshold value was set to 5, 10, 20, and 30 and the corresponding recognition rate was checked in order to find the optimal value. Next, we searched for the optimal number of regions with the determined t value. We considered three cases, where

Table 4.1: Recognition rate of DTP for different threshold values using non-weighted template matching.

Threshold (t) value	Recognition Rate		
	fb	dup I	dup II
5	0.92	0.67	0.65
10	0.94	0.68	0.66
20	0.93	0.66	0.62
30	0.92	0.62	0.59

Table 4.2: Recognition rate of DTP for different number of regions using non-weighted template matching.

Number of regions	Recognition Rate		
	fb	dup I	dup II
3×3	0.86	0.59	0.57
5×5	0.91	0.64	0.62
7×7	0.94	0.68	0.66

images were divided into 3×3, 5×5, and 7×7 regions. Non-weighted template matching was used for the classification task. Table 4.1 shows the recognition rate of the DTP feature representation method for different threshold values. Here, the recognition rate can be defined formally by the following formula:

$$\text{Recognition rate} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} \tag{4.1}$$

It can be observed that, the best recognition rate is achieved for $t = 10$. Table 4.2 shows the recognition rate of DTP for different number of regions when $t = 10$. Here, the highest recognition rate is achieved for images partitioned into 7×7 sub-regions, which indicates that, increasing the number of sub-regions can enhance the performance. However, it also increases the length of the feature vector and thus imposes a higher level of computational complexity.

Table 4.3: Recognition rate of different feature representation methods using weighted template matching.

Operator	Recognition Rate		
	fb	dup I	dup II
LBP	0.95	0.65	0.65
LTP	0.96	0.65	0.66
LDP	0.95	0.63	0.66
DTP	0.97	0.71	0.69
Compressed DTP	0.96	0.72	0.66

4.2.1.2 Performance on FERET Database

The performance of the DTP feature representation method is compared with 3 widely-used local texture operators, namely local binary pattern (LBP) [35], local ternary pattern (LTP) [5], and local directional pattern (LDP) [6]. Experiments were carried out on images partitioned into 7×7 sub-regions using weighted template matching. The adopted weights were same as shown in Figure 3.1. Table 4.3 shows the recognition rate of different texture operators. It can be observed that, both DTP and compressed DTP achieves superior performances than the other mehtods. However, the recognition rates of the DTP and the compressed DTP methods are not identical as they employ different encoding schemes, and therefore, the generated patterns are different which results in slight differences in the classification rates.

4.2.2 Performance Evaluation for Facial Expression Recognition

In order to evaluate the proposed method in expression recognition, we carried out a ten-fold cross-validation to measure the classification rate. In a ten-fold cross-validation, the whole dataset is randomly partitioned into ten subsets, where each subset comprises an equal number of instances. One subset is used as the

testing set and the classifier is trained on the remaining nine subsets. The average classification rate is calculated after repeating the above process for ten times. A support vector machine with a radial-basis function (RBF) kernel was used as the classifier.

4.2.2.1 Optimal Parameter Selction

For facial expression recognition, we compute the classification rates for images divided into 3×3, 5×5, and 7×6 regions. The threshold t was set to 40 empirically using the same optimal parameter selection procedure described in the section 4.2.1.

4.2.2.2 Performance on CK Database

For expression recognition also, we have compared the performance of the proposed method with the three well-known local pattern operators, namely local binary pattern (LBP) [11], local ternary pattern (LTP) [5], and local directional pattern (LDP) [6]. Table 4.4 and Table 4.5 show the recognition rate of these local pattern-based feature descriptors against the Cohn-Kanade 6-class and the 7-class expression dataset, respectively. In both cases, DTP and compressed DTP exhibits superior performances in recognizing expression images. For the 6-class dataset, DTP achieves an excellent recognition accuracy of 97.5%, while the accuracy of compressed DTP is 97.3%. On the other hand, for the 7-class dataset, the recognition accuracy of DTP and compressed DTP are 95.8% and 95.2%, respectively. Here, inclusion of neutral expression images results in a decrease in the accuracy. For both the 6-class and the 7-class recognition problem, the highest classification rate is obtained for images partitioned into 7×6 regions. The confusion matrix of recognition using the DTP feature descriptor for the 6-class and the 7-class dataset are shown in Table 4.6 and Table 4.7, respectively, which provides a better picture of the recognition accuracy of individual expression types. It can be observed that, for the 6-class recognition, all the expressions can be recognized with high accuracy. On the other hand, for the 7-class dataset,

while anger, disgust, fear, and surprise can be recognized with high accuracy, the recognition rates of joy, sadness, and neutral expressions are lower than the average.

Table 4.4: Recognition rate (%) for the CK 6-class expression dataset.

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	79.3	89.7	90.1
LTP	91.3	92.3	94.6
LDP	80.2	91.9	93.7
DTP	94.5	97.1	97.5
Compressed DTP	94.2	96.8	97.3

Table 4.5: Recognition rate (%) for the CK 7-class expression dataset.

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	73.8	80.9	83.3
LTP	85.3	88.5	88.9
LDP	75.7	86.3	88.4
DTP	90.3	93.9	95.8
Compressed DTP	89.9	93.5	95.2

4.2.2.3 Performance on JAFFE Database

For the JAFFE database, the DTP feature descriptor achieves classification rates of 92.5% and 88.7% for the 6-class and the 7-class expression datasets, respectively. Table 4.8 and Table 4.9 show the recognition rates of different feature descriptors with the JAFFE 6-class and the 7-class expression datasets, respectively. It can be observed that, DTP achieves the highest classification rates in all cases. The recognition rate on the JAFFE database is relatively lower than the

Table 4.6: Confusion matrix for the CK 6-class recognition using DTP feature representation (for images partitioned into 7×6 regions). Rows represent true class and columns represent classification rate (%).

		1	2	3	4	5	6
1	Anger	**97.7**	0	0	0	0	2.3
2	Disgust	0	**97.9**	0	1.6	0	0.5
3	Fear	0	0.4	**96.0**	0	0	3.6
4	Joy	0.6	0.5	0	**98.9**	0	0
5	Sad	0	0	0	0	**100**	0
6	Surprise	0	0	0	3.4	0	**96.6**

CK database. The main reason is the incorrect labeling of some of the facial expression images in the JAFFE database. Figure 4.6 shows examples of incorrect labeling of expression images in the JAFFE database.

Figure 4.6: Some mislabeled sample expression images in the JAFFE Database.

4.2.3 Performance Evaluation for Gender Classification

To evaluate the performance of the proposed method for gender classification, we have used the dataset described in section 4.1.3. Experiments were carried out on images partitioned into 7×7 sub-regions. For gender classification, the optimal value of t was found to be 10. A support vector machine was used for the classification task. Table 4.10 shows the classification rate of the different feature descriptors. From the classification rate, it is evident that, the performance

Table 4.7: Confusion matrix for the CK 7-class recognition using DTP feature representation (for images partitioned into 7×6 regions). Rows represent true class and columns represent classification rate (%).

		1	2	3	4	5	6	7
1	Anger	**98.5**	0	0	0	0	1.5	0
2	Disgust	0	**98.0**	0	0	0	2.0	0
3	Fear	0	0.4	**98.5**	0	0	1.1	0
4	Joy	0.5	0.5	0	**93.2**	0	5.8	0
5	Sad	0	0	0	0	**93.4**	0	6.6
6	Surprise	0	0	0	1.9	0	**98.1**	0
7	Neutral	5.9	0	0	0	3.0	0	**91.1**

Table 4.8: Recognition rate (%) against the JAFFE 6-class expression dataset.

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	84.1	87.6	90.5
LTP	84.3	87.9	90.9
LDP	83.2	88.9	90.7
DTP	87.5	90.1	92.5
Compressed DTP	87.1	89.8	92.2

of the proposed method is superior with comparison to other existing feature descriptors.

4.2.4 Performance Evaluation for Low-Resolution Images

Automated facial analysis is useful in smart meeting, surveillance and many other applications, where often only low-resolution video data is available. Since geometric methods like detection of facial action units are difficult to accommodate in these scenarios, appearance-based methods seem to be a better solution. Therefore, the performance of the proposed method is also evaluated on low-resolution

Table 4.9: Recognition rate (%) against the JAFFE 7-class expression dataset.

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	81.5	82.3	85.3
LTP	84.6	85.0	86.7
LDP	83.3	85.3	85.9
DTP	85.3	86.9	88.7
Compressed DTP	85.9	86.5	88.2

Table 4.10: Male and female classification rate (%) of different feature descriptors from facial images.

Feature descriptor	Classification rate(%)		
	Male	Female	Overall
LBP	92.78	92.78	92.78
LTP	92.89	92.89	92.89
LDP	89.81	90.08	89.94
DTP	93.21	93.02	93.11
Compressed DTP	92.9	93.10	93.05

images. Experiments were conducted on images from the CK 6-class expression dataset. We considered 3 different image resolutions: 75×55, 48×36, and 37×27. The original images were down-sampled to obtain these low-resolution images. All the images were partitioned into 7×6 regions while forming the feature vector. Here also, the performance of the DTP feature descriptor is compared with LBP, LTP, and LDP. Table 4.11 shows the recognition rates of these methods for low resolution expression images. From the recognition accuracy, it is evident that, facial feature representation based on DTP is more robust than other existing local texture patterns over a useful range of low resolutions.

Table 4.11: Recognition rate (%) using low-resolution images from the CK 6-class dataset.

Feature descriptor	Image resolution		
	75×55	48×36	37×27
LBP	88.9	83.5	79.7
LTP	89.7	85.9	83.3
LDP	90.7	89.1	84.4
DTP	93.9	92.2	89.1
Compressed DTP	93.5	91.9	88.7

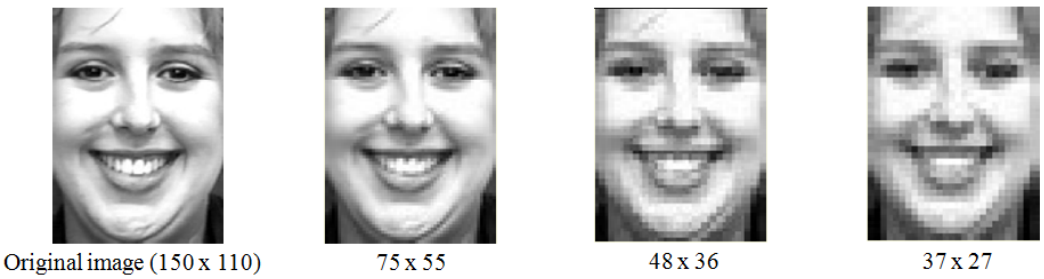


Figure 4.7: Sample low resolution images used in the experiments.

4.2.5 Performance Under the Presence of Noise

To investigate the robustness of the porposed DTP descriptor under the presence of noise, further experiments were conducted on the images from the CK 6-class expression dataset. In the experimental setup, the images in the testing set were contaminated with Gaussian white noise of different variances, while the training samples were kept unchanged. All the images were partitioned into 7×6 regions during feature vector generation. Table 4.12 shows the corresponding recognition rates of LBP, LTP, LDP, and DTP aganist images corrupted with Gaussian white noise of mean 0 and different variances (0.01, 0.02, 0.03, 0.04, and 0.05). It can be observed that, in all cases DTP achieves significantly higher recognition rates than the other micro-pattern operators. The superiority of DTP encoding is due to the utilization of stable edge responses and its discriminating capability of

Table 4.12: Recognition rate (%) on images from the CK 6-class dataset corrupted with Gaussian white noise with zero mean and different variances.

Feature descriptor	Noise variance				
	0.01	0.02	0.03	0.04	0.05
LBP	73.7	66.7	64.5	62.3	61.9
LTP	77.1	70.4	67.3	65.0	62.3
LDP	70.9	61.3	55.1	52.4	48.2
DTP	87.7	79.7	73.4	69.6	66.5
Compressed DTP	87.5	79.3	73.1	68.8	65.3

smooth and high textured areas from different face regions.

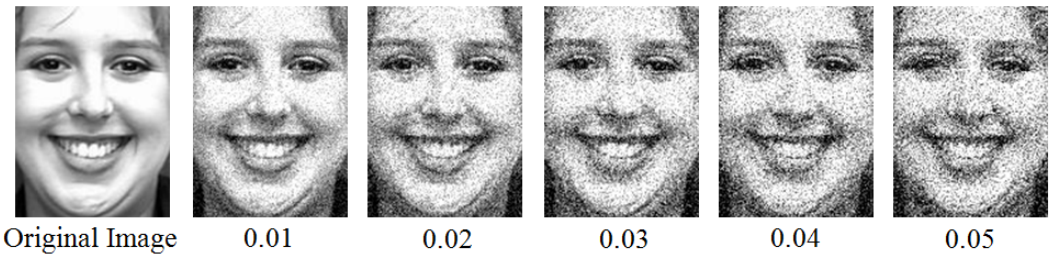


Figure 4.8: Sample expression images contaminated with Gaussian white noise with zero mean and different variances.

Our experimental results validate that, the proposed DTP feature representation performs better than the existing local pattern-based feature descriptors for different face-related problems. In addition, utilization of the symmetric property of the Robinson compass masks enables the computation of DTP features faster than other edge response based texture patterns such as LDP.

4.2.6 Computation Time Analysis

In order to determine the feasibility of the proposed method for real-time recognition systems, we have also examined the computation time of the DTP features. Table 4.13 shows the average computation time to calculate local pattern features from an image with a resolution of 100×100. It can be observed that, the

Table 4.13: Average computation time for different local pattern operators

Operator	Computation Time (in seconds)
LBP	0.39
LTP	0.53
LDP	0.77
DTP	0.66
Compressed DTP	0.41

average computation time for DTP and compressed DTP features from a single image is 0.66s and 0.41s, respectively, which is feasible for real-time scenarios. If we compare the computation time with the other existing texture operators, we can see that, both DTP and compressed DTP have lower computation time than the other existing edge response based texture operator, namely LDP. However, the computation time for DTP and compressed DTP is slightly higher than the LBP operator. This difference is due to time required for the computation of edge response values.

Chapter 5

Conclusions

5.1 Summary of Contributions

This thesis addresses the issue of designing an effective and robust facial feature representation that eliminates the limitations of the existing approaches and can be used in different face-related applications, such as face recognition, facial expression recognition, and gender classification.

We have introduced a new local texture pattern, the directional ternary pattern (DTP), and a robust facial feature descriptor constructed with the DTP codes for representing facial image. The DTP operator integrates the local edge responses for texture encoding, and also differentiates smooth and non-smooth areas with three different levels. We present a comprehensive study on the effectiveness of the proposed method, which has been evaluated for three face-related porblems: i) face recognition, ii) facial expression recognition and iii) gender classification. Extensive experiments using different benchmark databases show the superiority of the DTP descriptor against three well-known appearance-based feature representation methods, namely LBP, LTP, and LDP.

We have investigated DTP features for low-resolution facial recognition, a critical problem but seldom addressed in the existing work. Compared to the previous work, DTP features provide better performance, which is very promising for real-world applications. In addition, random noise in image is very common for real-world applications. Therefore, the feature descriptor should also be able to deal with random noise and produce stable features. In this work, the performance of the proposed method has also been evaluated against images corrupted

with Gaussian white noise. Experimental results reveal that, the proposed DTP feature descriptor can perform more robustly and accurately under the presence of noise. The effectiveness of the proposed method is due to its texture discriminating capability, and robustness under illumination variations and presence of noise as compared to existing representations, therefore, can be used in real-time applications for consumer products.

5.2 Limitations and Future Works

- In our work, we have only considered single static images for facial expression recognition. Psychological experiments by Bassili [47] have suggested that facial expressions can be recognized more accurately from sequence images than from a single image. In future, we plan to incorporate temporal information with the DTP method for facial expression recognition from video sequence images. This can be done by introducing Hidden Markov Model for classification.

- Our proposed method uses a global threshold t to generate DTP codes, which may not be ideal for every local neighborhood in the image. Therefore, we plan to incorporate an automated local threshold selection method with the DTP encoding. One possible way is to consider the variance of the edge response values in the local neighborhood in order to find the local threshold.

Bibliography

[1] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiment with a multi-layer perceptron," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, pp. 893–911, 1999.

[2] M. Z. Uddin, J. J. Lee, and T. S. Kim, "An enhanced independent component-based human expression recognition from video," *IEEE Transaction on Consumer Electronics*, vol. 55, no. 4, pp. 2216–2224, 2009.

[3] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using lbp variance(lbpv) with global matching," *Pattern Recognition*, vol. 43, pp. 706–719, 2010.

[4] T. Jabid, M. Kabir, and O. Chae, "Local directional pattern (ldp) for face recognition," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 4, pp. 2423–2437, 2012.

[5] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, LNCS 4778*, 2007, pp. 168–182.

[6] 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.

[7] M. C. Hwang, L. T. Ha, N. H. Kim, and C. S. Park, "Person identification system for future digital tv with intelligence," *IEEE Transaction on Consumer Electronics*, vol. 53, no. 1, pp. 218–226, 2007.

[8] P. Corcoran, C. Iancu, F. Callaly, and A. Cucos, "Biometric access control for digital media streams in home networks," *IEEE Transaction on Consumer Electronics*, vol. 53, no. 3, pp. 917–925, 2007.

[9] D. S. Kim, I. J. Jeon, S. Y. Lee, P. K. Rhee, and D. J. Chung, "Embedded face recognition based on fast genetic algorithm for intelligent digital photography," *IEEE Transaction on Consumer Electronics*, vol. 52, no. 3, pp. 726–734, 2006.

[10] F. Zuo and P. H. N. de With, "Real-time embedded face recognition for smart home," *IEEE Transaction on Consumer Electronics*, vol. 51, no. 1, pp. 183–190, 2005.

[11] 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.

[12] Y. Rodriguez, "Face detection and verification using local binary patterns," Ph.D. dissertation, Ecole Polytechnique Federale De Lausanne, 2006.

[13] K. K. et al., "Facial feature extraction based on private energy map in dct domain," *ETRI Journal*, vol. 29, no. 2, pp. 243–245, 2007.

[14] K. C. et al., "A probabilistic network for facial feature verification," *ETRI Journal*, vol. 25, no. 2, pp. 140–143, 2007.

[15] P. Ekman and W. Friesen, "Facial action coding system: A technique for measurement of facial movement," *Palo Alto: Consulting Psychologists Press*, 1978.

[16] G. D. Guo and C. R. Dyer, "Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 346–352.

[17] M. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *IEEE CVPR Workshop*, vol. 3, 2005, pp. 76–84.

[18] M. Valstar and M. Pantic, "Fully automatic facial action unit detection and temporal analysis," in *IEEE CVPR Workshop*, 2006, p. 149.

[19] L. Wiskott, J. Fellous, N. Kruger, 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.

[20] R. Brunelli and T. Poggio, "Hyperbf networks for gender classification," in *DARPA Image Understanding Workship*, 1992, pp. 311–314.

[21] H. Abdi, D. Valentin, B. Edelman, and A. O'Toole, "More about the difference between men and women: Evidence from linear neural network and principal component approach," *Neural Computing*, vol. 7, no. 6, pp. 1160–1164, 1995.

[22] C. Padgett and G. Cottrell, "Representation face images for emotion classification," *Advances in Neural Information Processing Systems*, vol. 9, pp. 894–900, 1997.

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

[24] C. C. Fa and F. Y. Shin, "Recognizing facial action units using independent component analysis and support vector machine," *Pattern Recognition*, vol. 39, no. 9, pp. 1795–1798, 2006.

[25] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transaction on Pattern Analysis and Machince Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[26] Y. Tian, "Evaluation of face resolution for expression analysis," in *IEEE Workshop on Face Processing in Video*, 2004.

[27] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary pattern," in *European Conference on Computer Vision*, vol. LNCS 3021, 2004, pp. 469–481.

[28] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[29] K. Etemad and R. Chellappa, "Dicriminant analysis for recognition of human face images," *Journal of the Optical Society of America*, vol. 14, pp. 1724–1733, 1997.

[30] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[31] P. Penev and J. Atick, "Local feature analysis: A general statistical theory for object representation," *Network: Computation in Neural Systems*, vol. 7, pp. 477–500, 1996.

[32] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300–311, 1993.

[33] S. Zhao, Y. Gao, and B. Zhang, "Sobel-lbp," in *IEEE International Conference on Image Processing*, 2008, pp. 2144–2147.

[34] G. Donato, M. S. Bartlett, J. C. Hagar, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Transaction on Pattern Analysis and Machince Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

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

[36] G. Zhao and M. Pietikainen, "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition," *Pattern Recognition Letters*, vol. 30, no. 12, pp. 1117–1127, 2009.

[37] A. Shobeirinejad and Y. Gao, "Gender classification using interlaced derivative patterns," in *International Conference on Pattern Recognition*, 2010, pp. 1509–1512.

[38] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transaction on Pattern Analysis and Machince Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[39] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao, "Gender classification based on boosting local binary pattern," in *International Symposium on Neural Networks*, no. 2, 2006, pp. 194–201.

[40] H. Lian and B. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *International Symposium on Neural Networks*, no. 2, 2006, pp. 202–209.

[41] H. Kabir, T. Jabid, and O. Chae, "Local directional pattern variance (ldpv): A robust feature descriptor for facial expression recognition," *International Arab Journal of Information Technology*, vol. 9, pp. 382–391, 2012.

[42] C. W. Hsu and C. J. Lin, "A comparison on methods for multiclass support vector machines," *IEEE Transaction on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

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

[44] S. E. Umbaugh, *Digital Image Processing and Analysis*. CRC Press, 2011.

[45] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *IEEE International Conference on Automated Face and Gesture Recognition*, 2000, pp. 46–53.

[46] P. Phillips, H. Wechler, J. Huang, and P. Rauss, "The feret database and evaluation procedure for face recognition algorithms," *Image and*, vol. 16, no. 10, pp. 295–306, 1998.

[47] J. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049–2058, 1979.

Appendix A

List of Publications

Journal Papers (SCI/SCIE Indexed):

[1] **Ahmed F.** et al., "Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)", *International Arab Journal of Information Technology* [SCIE, IF: 0.127], ISSN: 1683-3198, 2012 [Accepted, In Press].

Conference Papers:

[1] **Ahmed F.** and Kabir M.H., "Directional Ternary Pattern (DTP) for Facial Expression Recognition", In: *30th IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA, pp. 265-266, January 2012.

[2] **Ahmed F.** and Kabir M.H., "Facial Feature Representation with Directional Ternary Pattern (DTP): Application to Gender Classification", In: *13th IEEE International Conference on Information Reuse and Integration (IRI)*, Las Vegas, USA, pp. 159-164, August 2012.

[3] Kabir M.H. and **Ahmed F.**, "Face Recognition with Directional Ternary Pattern (DTP)", In: *4th International Conference on Graphic and Image Processing (ICGIP)*, Singapore, November 2012 [Accepted for Presentation].

[4] **Ahmed F.**, Hossain E., Bari A.S.M.H., and Shihavuddin ASM, "Compound Local Binary Pattern (CLBP) for Robust Facial Expression Recognition", In: *12th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, Hungary, pp. 391-395, 2011.