Facial Expression Recognition Using Shearlet Transform and Kirsch Masking

Kaniz Fatema-Tuz-Zohora  
dept. of Computer Science and Engineering  
East West University  
Dhaka, Bangladesh  
kanizfatema5200@gmail.com

Adnan Ferdousi  
dept. of Computer Science and Engineering  
East West University  
Dhaka, Bangladesh  
adnanferdousi@gmail.com

Taskeed Jabid  
dept. of Computer Science and Engineering  
East West University  
Dhaka, Bangladesh  
taskeed@ewubd.edu

Abstract—In this paper, a new feature extractor is proposed that is combined of Discrete Shearlet Transform (DST) system and Kirsch Compass Kernel (KCK) with Local Binary Pattern (LBP) for extracting features from images. Discrete Shearlet Transform gives multi-resolution images of an image by highlighting the major edges along with detail features. On the other hand, Kirsch Compass Kernel (KCK) provides 8 different directional images with prominent edges which sharpen the edges in each direction. These characteristics of DST and KCK help LBP to extract the significant features of an image. The extracted features by LBP help to classify the images by using Support Vector Machine (SVM).

Index Terms—feature extractor, facial expression recognition, local binary pattern, discrete shearlet transform, kirsch compass kernel

I. INTRODUCTION

At the present time, the automatic facial expressions recognition (FER) in computer vision field has been played the crucial role in industrial and research sectors [1]. Facial expression (such as disgust, anger, fear, happy, sad and surprise) is a non-verbal communication process, which notify the mood or mental state of a person [2]. It is a very natural process for human beings by which they disclose their personal emotions, feelings and intentions. Automated facial expression recognition is a challenging task in computer vision and image processing. Facial expression recognition had been developed since nineties and since then people are still working on it for improving [3]. There are many applications of automatic facial expression recognition such as understanding human attitude or mode, detection of mental disorders, fatigue detection of operators as well as drivers, playing music depending on one's mood, security, lie detection, syntheses of humans expressions and automatic counseling systems. There are numerous methods and techniques to extract features from images for facial expression recognition. Local Binary Pattern (LBP) is one of the most popular Feature Extractor (FE) in computer vision field. It is a gray scale invariant texture primitive which has a significant popularity for the texture describing of an image [4]. Shearlet is a multi-scale framework for the efficient representation of multidimensional data. Shearlet is an instinctive extension of Wavelet and Curvelet. Shearlet has multivariate functions those are generally governed by anisotropic features like detecting edges in images [5]. Kirsch Compass Kernel (KCK) is a non-liner edge detector, in some predetermined directions KCK finds the maximum strength of edge from a two dimensional image [6]. The advantages of those anisotropic features of Shearlet and maximum edge strength feature of KCK are taken to extract features as well as those features are used for detecting facial Expressions in biometric recognition sectors. In this paper, a Shearlet-Kirsch based feature extractor system is proposed which extracts features from the images for both genders of female and male, to recognize facial expressions. Two types of databases is used in this paper, Cohn-Kanade (CK) database [15] which contains images of both gender and Japanese Female Facial Expression (JAFFE) database [16] for the proposed system.

II. RELATED WORK

Discrete shearlet transform (DST) was developed by W.-Q Lim [5], an efficient multi-scale directional Representation. Q Lim showed the implementation of the shearlet transform by multi-resolution analysis (MRA). J. He used Shearlet for texture classification [7], where different filters and thresholds of shearlet transform system are used for rotation invariant texture. Kirsch compass Kernel (KCK), a non-liner edge detector mask, is used for finding maximum strength of edges in some prefixed directions by Russell A. Kirsch [6]. These predetermined directions are 8 compass directions, that are North, North-West, West, South-West, South, South-East, East, North-East. KCK is used to extract initial edges of micro calcification clusters [8]. To the certain edge direction each masking has great response. This derivative mask scale can be changed according to requirements. To obtain most possible edge point of images, threshold values also can be adjusted. The LBP operator was proposed by Ojala T. et al [9]. Ojala constructed a novel descriptor that characterized by the differences between the central pixel value and its neighboring pixel values and form a binary pattern of a local image patch. As shearlet has multi-resolute features, both shearlet and LBP could provide significant image features together for face recognition [10]. A novel and discriminative texture descriptor was proposed by Xianbiao Qi et al [11] which is based on LBP. MRELBP is the extension of ELBP and LBP, which can classify texture.
In the proposed method, Shearlet and Kirsch Compass Kernel is used together with LBP for getting uniquely generated different type of prominent features of image. Hence the objective is to do the comparison of a variety of directions and analyze the performance.

III. METHODS

All the methods of proposed system will be described in this section. The technique of proposed system is described after a basic introduction of individual methods.

A. Shearlet

Shearlet is a multiscale framework which is prolongation of wavelet transform. It originates from the affine group and is efficiently developed to detect one dimensional (1-D) and two dimensional (2-D) directional features in images. Shearlet was introduced in 2005 [13] for the sparse approximation analysis of multivariate functions \( f \in L^2(\mathbb{R}^2) \), which are typically governed by anisotropic features such as edges in images, parabolic scaling for angular and directional ridges, shearing and translation. There are two types of shearlet system, continuous shearlet system and discrete shearlet system. In this proposed system, discrete shearlet system is used, the affine systems with composite dilations are the collections of

\[
\psi_{j,l,k} = \left| \text{det} A \right|^{\frac{1}{2}} \psi(B^l A^j (x - k)) : j, l \in \mathbb{Z}, \ k \in \mathbb{Z}^2
\]

where \( \psi \in L^2(\mathbb{R}^2) \) discretizing the parameter set \( \mathbb{R}^+ \ast \mathbb{R} \ast \mathbb{R}^2 \) [13], [14]. \( A \) and \( B \) are 2x2 invertible matrices and \( |\text{det} B| = 1 \). Here, the dilations matrices \( A^j \) are known as scaling matrix (A) and matrices \( B^l \) is known as shearing matrix (S).

\[
A = \begin{bmatrix} a & 0 \\ 0 & a^2 \end{bmatrix}, (a > 0) \quad \text{and} \quad S = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, (s \in \mathbb{R})
\]

where \( A \) matrix is associated with scale transformations and \( S \) matrix is associated to area-preserving geometrical transformations, such as rotations and shear. A set of discrete shearlet systems that produce the frequency tiling shown in Fig:1(a)
can be described by combining shearlets supported on the horizontal cone

\[
D_0 = \{ (\omega_1, \omega_2) \in \mathbb{R}^2 : |\omega_1| \geq \frac{1}{8}, \ |\omega_2| \leq 1 \}
\]

and the vertical cone

\[
D_1 = \{ (\omega_1, \omega_2) \in \mathbb{R}^2 : |\omega_2| \geq \frac{1}{8}, \ |\omega_1| \leq 1 \}
\]

Each element \( \psi_{j,l,k} \) is supported on a pair of trapezoids, of approximate size \( 2^{2j} \ast 2^l \), oriented along lines of slope \( l2^{-j} \) in Fig:1(b). This transformation is based on manipulating the data to obtain the desired spatial-frequency tiling determined by the theory.

B. Kirsch Compass Kernel

Kirsch Compass Kernel (KCK) is an edge detector which is used for finding edges with the maximum strength in a few predetermined directions [6], [8]. A 3x3 single kernel mask is taken by KCK and it is rotated via all 8 compass directions by 45° anti-clock wise as Fig.2. The maximum edge magnitude of the Kirsch operator across all the 8 directions is calculated as,

\[
h_{n,m} = \max_{z=1,2,3,\ldots,8} \sum_{i=-1}^{1} \sum_{j=-1}^{1} (g_{ij}^{(z)}) \ast f_{n+1,m+j}
\]

where \( z \) computes the 8 compass direction kernels. The 3x3 kernel mask across 8 directions are

\[
\begin{align*}
\text{North} & = \begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \quad \text{Northwest} = \begin{bmatrix} +5 & +5 & -3 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \\
\text{West} & = \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix}, \quad \text{Southwest} = \begin{bmatrix} +5 & 0 & -3 \\ +5 & -3 & -3 \\ +5 & -3 & -3 \end{bmatrix} \\
\text{South} & = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ +5 & +5 & +5 \end{bmatrix}, \quad \text{Southeast} = \begin{bmatrix} -3 & 0 & -3 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{bmatrix} \\
\text{East} & = \begin{bmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{bmatrix}, \quad \text{Northeast} = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{bmatrix}
\end{align*}
\]
C. Local Binary Pattern

Local Binary Pattern (LBP) is an operator which is proposed by Ojala et al [9]. This operator takes a local patch of image as input and finds the differences between the central pixel and its neighboring pixel values. If the neighboring pixel has equal or greater value than central pixel then it set 1 otherwise it set 0 (zero) for that neighboring pixel and generate an 8 bit binary pattern. Update the central pixel value of that input patch by the decimal value of that corresponding binary value. Let us consider, $X_c$ is the central pixel and $X_i$ is the neighborhood pixel then the mathematical representation is,

$$LBP_{r,p}(X_c) = \sum_{i=0}^{p-1} s(X_i - X_c)2^i,$$

where, $r$ is the redial distance from the central pixel, $p$ is the number of neighborhood pixels \{X_i\}_{i=0}^{p-1} for each central pixel and Function $s(x)$ is the sign function that generates a P-bit binary numbers and resulting into $2^p$ distinct LBP patterns.

D. Proposed Method

The proposed method of this paper, can extract local pattern of features based on the discrete Shearlet system and Kirsch Compass Kernel to detect facial expressions. Images are taken as input and then perform discrete Shearlet System and Kirsch Compass Kernel with LBP for extracting strong features from the images.

Here, discrete shearlet system and kirsch compass kernel are separately applied on input image which give different directional images as an output. Laplacian-pyramid scheme is used to decompose the image $f_{j-1}^d$ into low-pass $f_j^a$ and high-pass $f_j^d$ images, where $f_0^a$ is the given MxN input image. Computing $f_j^d$ on a pseudo-polar grid and applying filtering along the angular direction, which provides the matrix $Pf_j^d$. A band-pass filtering is applied to the matrix $Pf_j^d$ and then the inverse two-dimensional FFT is applied. The imaginary part of this transformation is discarded, just only the real part is considered from this calculation. All these output images in Fig:4(a) are derived from input image but in different representation, as it is a multiscale framework with anisotropic features. It provides 6 output images where first image represents the high-pass image and the rest 5 images represent the low-pass image. This high-pass image needs some adjustment for visualizing the image in Fig:4(b).

Similarly, for kirsch compass kernel (KCK), it provides 8 images as output in Fig:5 for each input image. KCK highlights the edges in fixed directions with the maximum strength. This mask is rotated from one to another positive direction by 45°.

After getting the multidimensional and multidirectional images, each output image is divided into four blocks and LBP is applied to each block to find every detailed features from 14 output images (6 images from the shearlet system and 8 images from the KCK). The block wise histogram of each image is taken and all block wise histogram outputs are concatenated as shown in Fig:6. Finally combining the
histogram together of all 14 output images, which is the extracted final feature by the proposed system as output. From these extracted features, accuracy is checked by SVM classifier that supports multi-class properties.

IV. EXPERIMENTAL RESULTS

A. Dataset Description

In this paper, Cohn-Kanade and JAFFE databases are used for experiment. Cohn-Kanade database contains gray-scale images of 97 numbers of people with six and seven expressions [15]. In this dataset, there are only frontal viewed facial images of both male and female and most of them are Euro-American and few percent of other nations. Six expressions dataset containing disgust, angry, fear, happy, sad, surprise. While in the seven expressions dataset there is neutral expression along with other six expressions Fig:7(a). In JAFFE database there are 213 images posed by 10 Japanese female models [16]. JAFFE database contains 7 facial expressions, six are basic facial expressions and one is neutral expression shown in Fig:7(b).

B. Experiment

After applying the proposed extractor, the desired extracted patterns of facial expressions is gotten for both Cohn-Kanade and JAFFE databases. Then the accuracy of this extracted patterns are tested to check the performance and recognition rate of each facial expression using multi-class SVM classifier.

C. Result

The extracted patterns for facial expressions are tested in various ways with different values and different cases which is shown in below tables. The percentage of accuracy is checked for each facial image by n-fold cross validation with the value of n=5 and the kernel type is polynomial, which is shown in Table-I for both the CK and JAFFE databases. This result shows that the proposed method performed better than other three methods based on extracted features. For CK dataset, accuracy of proposed method is 98.9575% for six expressions and 98.1051% for seven expressions. Also for JAFFE database, it gives 94.0376% accurate result to recognizing facial expressions. The graphical representation of each method also shown in Fig:8 for both Cohn-Kanade (6 and 7-expressions) and JAFFE (7-expressions) databases.

The kernel wise accuracy of the extracted features of proposed method are classified into Table-II via SVM with different kernel types, such as linear, polynomial kernel and radial basis function (rbf). Fig:9 illustrates the kernel wise accuracy. For each kernel CK 6-expressions has the highest accurate level among all databases. Though JAFFE performs less than CK 6 and 7-expressions, still the rate of accuracy is 95.6997% for radial basis function (rbf) kernel. According to Table-II for each kernels, the accuracy rate of proposed method is more than 94% for JAFFE database while the rate of accuracy of CK database is more than 98%.

The accuracy of recognizing the each type of facial expression is shown by confusion matrix in Table-III and Table-IV for CK database and Table-V for JAFFE database. From the observation, six expressions dataset of CK gives more accurate result than the seven expressions dataset of CK. ‘Angry’ is the lowest recognized expression with the rate of 96.95% accuracy and ‘sad’ is the highest recognized expression.
with the rate of 98.75% accuracy in CK six expression database. On the other hand, Expression ‘happy’ is the lowest recognized expression with the rate of 91.78% accuracy and ‘angry’ is the highest recognized expression with the rate of 98.87% accuracy in CK seven expression database. The accuracy of seven expressions gets lower than six expressions because there are some images which is confused and getting mixed with other expressions, specially neutral expression. This is the reason for falling the accuracy rate to recognize facial expressions.

Similarly, the Confusion Matrix of JAFFE database is shown in Table-V. The observation states that the results of each expression is not so high as CK database but the accuracy of recognition of each expression is still in so good rate. Expression ‘sad’ is the lowest recognized expression with the rate of 85.67% accuracy and ‘happy’ is the highest recognized expression with the rate of 96.67% accuracy in JAFFE database. The accuracy rate is diminished due to incorrect labeling or some of confused images in the JAFFE database.

V. CONCLUSION
In this paper, a new descriptor is proposed for facial expression recognition based on shearlet and kirsch. Features of facial expression are extracted from image by applying Shearlet-Kirsch and computing LBP code from the transformed image. Experimental results using Cohn-Kanade database with six and seven expressions are compared with other methods to demonstrate the superiority of the proposed method, which successfully identifies more than 98% of facial expressions correctly. Also for JAFFE database with seven expressions, proposed method successfully identifies more than 94% of facial expressions correctly. Extensive experiments illustrate that the proposed approach is effective and efficient for facial expression recognition. After training, proposed system can be used in many different fields for human-computer interaction which requires recognition of facial expressions.

REFERENCES