Facial Expression Classification using Optimized KNN and Classifier Fusion System (CFS)

THESIS SUBMITTED FOR PARTIAL REQUIREMENT OF MASTER OF SCINECES IN COMPUTER SCIENCE

BY

MUHAMMAD SHAHID KHAN

405-FBAS/MSCS/S08

SUPERVISED BY

ASSISTANT PROF. DR. AYYAZ HUSSAIN

DEPARTMENT OF COMPUTER SCIENCE AND

SOFTWARE ENGINEERING,

FACULTY OF BASIC AND APPLIED SCIENCES

INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD

PAKISTAN

2011

1 - Computer systems

2. Data processing

Dated: ....

Final Approval

It is certified that we have examined the thesis titled "Facial Expression Classification using Optimized KNN and Classifier Fusion System" submitted by Muhammad Shahid Khan, Registration No. 405-FBAS/MSCS/S08, and found as per standard. In our judgment, this research project is sufficient to warrant it as acceptance by the International Islamic University, Islamabad for the award of MS Degree in Computer Science.

Committee

External Examiner

Khalid Saleem

Dr. Khalid Saleem

Assistant Professor

Department of Computer Science

Quaid-e-Azam University, Islamabad

Internal Examiner

Muhammad Nadeem

Mr. Muhammad Nadeem

Assistant Professor

Department of Computer Science and Software Engineering

International Islamic University, Islamabad

Supervisor

Dr. Ayyaz Hussain

Assistant Professor

Department of Computer Science and Software Engineering

International Islamic University, Islamabad

ABSTRACT

The purpose of this study is to develop an optimized classifier which should be able to precisely classify human facial expressions. There are six basic types of facial expressions i.e. happiness, sadness, fear, disgust, surprise, anger and one neutral expression. Our work starts with extraction of facial features usimg Local Binary Pattern (LBP), Discrete Cosine Transform (DCT) and Statistical Methods (Discrete Wavelet Transform) for facial features extraction. LBP & DWT features set are reduced using Principal Component Analysis (PCA) and then these features are fused together to get more appropriate features set. These features are used to finally differentiate between different facial expressions using proposed method (Optimized KNN) and Classifier Fusion System. Optimized KNN uses the concepts of distance weightings of neighbors and shared nearest neighbors. It assigns weight to each sample and then classifies new samples on the basis of top weighted sample. We have used JAFFE database for experiments and testing of O-KNN. JAFFE contains 213 images of 7 facial expressions posed by 10 Japanese female models. A precise and careful extraction of facial features helped a lot in precise classification using Optimized KNN & CFS. The calculations for each set of features and also for fused set are done separately of each quantitative measure Accuracy, ROC, Sensitivity, Specificity and AUCH.

DECLERATION

I hereby declare that this work, neither as a whole nor as a part has been copied out from any source. It is further declared that I have conducted this research and have accomplished this thesis entirely on the basis of our personal efforts and under the sincere guidance of my supervisor Assistant Prof. Dr. Ayyaz Hussain. If any part of this project is proved to be copied out from any source or found to be reproduction of some other project, I shall stand by the consequences. No portion of the work presented in this dissertation has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning.

Muhammad Shahid Khan

405-FBAS/MSCS/S08

A Dissertation submitted to the

Department of Computer Science and Software Engineering

International Islamic University Islamabad

As a partial fulfillment of requirements for the award of

The degree of

MS in Computer Science

Dedicated To

My beloved parents

And

My beloved wife

ACKNOWLEDGEMENTS

First of all, I am very thankful to Allah for His blessings which have brought me up to this level of education & respect. Then I would like to thank my supervisor, Assistant Prof. Dr. Ayyaz Hussain, for his guidance to me on every step and specially for encouraging me by advising through the sentences "Nothing is difficult", "Be confident" and "Keep trying". Sir, thanks for your tremendous support on my day to day activity and progress. I would like to thank Sir Nazir for his efforts and guidance in my practical working and teaching the basic concepts of image processing and data mining. I would also like to thank other faculty members of my university for their sincere appreciation, comments and suggestions.

I would also like to say thanks to my employer, Army Welfare Trust, and especially my manager, Mr. Imran Siddiqui, for providing me time for studying whenever I needed, and for tolerating me when I was tired after a sleepless night of study. Thanks are also for my dearest friends and cousins to support me directly and indirectly.

Finally I am thankful to my parents for their love, care and support in my life which has been directly encouraging me for my study. My parents' prayers have always been a big support in solving my problems. Allah may bless my parents with long life, health and happiness.

Muhammad Shahid Khan

405-FBAS/MSCS/S08

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

1. Introduction

1.1. Background and Rationale of the Study

Whenever a person comes across a person, it is very interesting and important to us to know and understand one's facial expressions so that a good conversation can be started for effective message exchange and successful interaction. A person's face, when captured with the help of camera, contains lots of information in the small image and cameras usually capture very large and high resolution photos. This image not only stores information about facial expressions but also portray the body language, surroundings information and meta data known as EXIF profile describing time, scale, zoom, camera, orientation, aperture, focal length, exposure length and flash settings. Simply it can be said that an image may contain the hell of information which is useful in variety of applications.

Today as number of images has increased on World Wide Web particularly and this count is increasing in huge amount daily, similarly interest and need to get image information and understanding this information has also increased. To find some specific images or carrying some useful information in this huge bundle and variety of images is also the need of time and many different techniques have been applied to know image information in general like searching images with respect to its size, color, black-and-white, type (face, photo, clip art, line drawing etc.), brightness variations, texture and shape. Searching images by color is actually generating its color histogram for different pixel values and actually shows human understandable colors. Similarly if retrieval of image information by its shape is required then user at least provide some sketch or data sample to match it with resting data samples and find out same shapes.

1.1.1. Facial Features

Facial features are different parts of the face which are important to categorize a face and its properties, these features can be categorized as ear, eye, facial-hair, mouth, nose and physiognomy. Each of the above features has its importance but eyes and mouth always play main and important role in facial features list. A brief overview of these features is given below.

- **Ear**
  Ear is a distinct facial feature but it is less important in facial features applications because particularly in expressions classification, it has no participation. One

reason is that it has actually many variations in its shape and also in its positions because its shape and position depends on shape and position of the mastoid process of the temporal bone.

- **Eye**
  Eye is also very distinct facial feature and important as well in any type of application because it is main face organ which performs apparently in any action-reaction and communication. Shape of eye ball depends on the shape of orbit and orbital edge. Main orbital shapes are angular, round, low and high.

- **Facial Hair**
  Facial hairs are developed mostly in men between 14-20 age starting from upper lips sides and later spreading to make complete beard, similarly in woman it starts developing after menopause but in very less amount as compared to a man.

- **Mouth**
  Mouth is one the important feature of face and lips shape depends on shape and size of teeth, jaw structure and projection of jaws.

- **Nose**
  Nose is also apparent facial feature which lies in the middle of human face and it has important role in facial applications.

1.1.2. Facial Features Extraction

Automatic Facial Recognition is an artificial intelligence process that starts with the steps of image normalization, noise removal, facial features extraction, facial features selection and classification of facial expressions based on these facial features. During this recognition process, features extracted from the image are actually matched and it is tried to find the close feature set so that it can be easily classified in that category. There have been done a lot of working and research in face recognition field which has resulted different outstanding classification techniques but still it is very challenging and complex research area due to variant conditions of facial images i.e. large number of variations in facial expressions, facial details, face poses and illumination conditions and especially people in different regions or different cultures can interpret facial expressions in different ways.

1.1.3. Classification

An algorithmic function which performs classification and has discretely implemented classification and results in assigning class to unknown object is known as classifier. To get maximum results in facial expression classification, facial features should be very precise. Extraction of precise and important facial features is also very helping in increasing the efficiency of classifier where as collection of unimportant and poorly extracted features leads to poor classification and results in a situation where even a very good classifier cannot perform efficient classification. Different machine learning techniques are performing efficiently in different fields but these methods still tradeoff for its complexity, performance, and time and space consumption.

For facial expressions classification problem, use of non-parametric techniques to classify the data is preferred when probability density functions are not known or not easy to estimate. A non-parametric mostly has techniques with no rely on data distribution and has ability to accommodate the complex data but this works very well with small data set sizes and is not very good with very large data set sizes. K-Nearest Neighbor is the simplest and non-parametric technique where it is type of instance-based learning or lazy learning because all computation is deferred until classification and commonly used in facial expression classification and has better results. The principle idea of KNN technique is to assign the query sample with the class label which is most commonly found among k nearest neighbor samples.

To improve the KNN classification accuracy, different extended forms of KNN have been developed so far but still these extended forms need to get improved and the performance of these techniques is compromising between accuracy & time efficiency. Our focus is to develop a technique named as Optimized KNN classifier using the concept of distance weightings and shared neighborhood. The proposed technique has been tested one by one on different facial feature sets which are extracted using Local Binary Pattern (LBP), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and alternate fusions of these individual feature sets.

1.2. Contributions

Basic idea in Optimized KNN technique is a thought from concepts of weighting data samples and the importance of neighbors. Firstly it is observed that weights assignment to data samples bring ease in sorting for importance and secondly neighbors in a specified neighborhood circle are not only important for testing data sample and also for its neighbors. It can be used to classify facial expressions with a very good accuracy results. To improve classification results, facial features are extracted using LBP, DCT and DWT then an optimized set of features is chosen using PCA. After this, selected features from different techniques is fused together to make a fine collection of important features. These features are used by the proposed classifier and classifier fusion system to calculate accuracy as well as its performance. In order to develop a robust classifier which can work in both noisy and non-noisy environment, the developed technique is also tested for noisy facial expression images.

The developed technique first detects the noise in the images and then an appropriate noise removal process is applied to remove noise and these pre-processed images are used to extract features. After appropriate pre-processing and feature extraction processes, the extracted features are fused together. Fused set of features are used by the optimized KNN which uses weights and shared neighborhood concept to improve classification accuracy. Furthermore, classification fusion system are used which uses a linear combination of KNN algorithm to give better performance in terms of all the quantitative measures. The developed technique is tested on JAFFE database which contains images of ten different Japanese females; each female has 7 expressions which constitute six basic expressions and one neutral expression. The performance of Optimized KNN is verified by comparing the parameters results where it is also compared with existing techniques results. In parallel with O-KNN, Classifier Fusion System is also tested for the same scenarios and same set of facial features and finally results are discussed.

- Optimized Local Binary Pattern
- Optimization in KNN
- Multiple combinations of features from LBP, DCT and DWT
- Classifier Fusion System to overcome failure of one classifier over other

1.3. Layout of the Thesis

Chapter No.2 describes in detail about the work history of feature extraction and selection techniques. Chapter No.3 discusses about the related work of classification need in today era and its different algorithmic methods to classify data. Chapter No.4 discusses about our proposed technique in detail and also shows how differently this technique has approved classification process starting from features extraction to classification. Chapter No.5 discusses in detail about experimental results in the form of detail graphs and data tables. It gives comparison between different techniques and finally suggested technique. Chapter No. 6 concludes all working and experimental results in a summary and finally it discusses the limitations of the method and also future work and improvements opportunities.

1.4. Summary

In this chapter, all basic concepts regarding image types, its different attributes and processing are discussed. It details about main facial expressions and how these expressions are important in our daily life conversations. Facial expression varies from one person to the other person so it is tough to categorize these expressions in some easy way. Different attributes of a human face are listed and discussed. Then a brief description is added regarding how this working and study will help out the existing knowledge and literature to proceed in some direction. Each study and work always gives a clue to the future works sooner or later. At the end of the chapter layout of thesis document is described for a quick chapter summary.

Chapter 2

2. Feature Extraction and Selection Techniques

This chapter gives a brief overview of some existing techniques for facial feature extractions. An image when read through computer called digital image; is actually a two dimensional array of pixels where each pixel is the smallest unit of image. Images are categorized in two types usually; one is gray scale and other is color. Pixel of a gray scale image is denoting only the intensity information which varies from 0 to 255 and "0" represents black pixel of very weak intensity and "255" represents white pixel which is of very high intensity value. This intensity scale actually shows gray scale values. Gray scale image is different from black and white image in a sense that it has only two values black "0" and white "255" where gray scale image contains values varying from 0 to 255. This image uses 8 bits to save the value of each gray scale image pixel. A color image, as compared to gray scale image, has each pixel containing three different components for each color red, green and blue which are also known as color planes. To draw one pixel of color image, these three components are required to join together.

Digital image processing is a scientific area where digital images are processed and this process takes image as input and performs some required functionality, usually algorithms, and the result output might be an image or a set of features or some extracted information. Facial expression is a visible symptom of the intentions, personality and psychology of a person. It expresses our emotions and important communicative indications. In words of psychologists [10], facial expressions contribute 55% weight during communication when someone delivers a message where as language has 7% weight and voice has 38% weights contribution to one's message conveying. In 1971, Ekman and Friesen posited six basic emotions named as happiness, sadness, fear, disgust, surprise and anger [1]. Face parts, particularly eyes, show the most important indications about the motions and moods. Eye contact actually performs conversional turns; give the communications involvements and interests. A person's eye has much information about he is feeling and thinking, even blinking eyes means a person is nervous or easy feeling. A professor Joe Tecce from Boston College suggests that blink level also reveals stress level.

Human facial expressions have much diversity due to varying facial features from one person to other, and from one area people to another area people which has made facial expression classification a very challenging task. Features are generally *categorical or nominal* (set of unordered items such as gender of "male" or "female" or a blood type of "A", "B", "AB" or "O"), *ordinal* (set of ordered items such as "large", "medium" or "small"), *integer-valued* (a count of a number of occurrences of a particular word in email) or *real-valued* (measurement of blood pressure). Mostly categorical and ordinal features are grouped together.

As an image contains lot of information in very small units (pixels) so processing these variables in a large quantity is computation expensive so a feature set is developed which is reduced set of data. Facial features are mainly categorized in two areas [11]; firstly geometric-based features (which includes shapes and locations of facial micro organs i.e. nose, eyebrows, eyes, & lips) and secondly surface-based features (which includes face appearances changing and skin textures i.e. face wrinkles and furrows). There are different available features extracting methods which gets different feature sets depending on the logic of algorithm, here these methods are being discussed one by one in this literature review.

2.1. Local Binary Pattern:

Local Binary Pattern (LBP) descriptor [2] is one of the simplest and efficient feature extraction methods; it uses pixel level intensities distribution in the image and generates a binary pattern known as LBP code from which histograms of these codes are generated either for complete image or for some specified facial regions. LBP is a non-parametric operator and describes the local spatial structure of an image. It calculates a bit-code from binary derivates of pixels; calculates the difference of central pixel value with its surrounding pixels values, arrange these differences in an ordered form of clock-wise move, and finally this bit-pattern code is converted into decimal value which is the new LBP code for the central pixel. Basic LBP operator works for 3×3 pixels which are nine in total number and the central pixel gets LBP code for surrounding eight pixels. It is described by the following diagram;

129	125	199
120	110	90
139	119	87

If $i_n > i_c$ then 1 otherwise 0. $\rightarrow$

1	1	1
1		0
1	1	0

Binary: 11100111 $\rightarrow$ Decimal: 230

Fig. 2.1 Basic LBP Operator

This diagram shows grey scale values of 3×3 pixels which is using following formula 1 and formula 2 for calculating the LBP code.

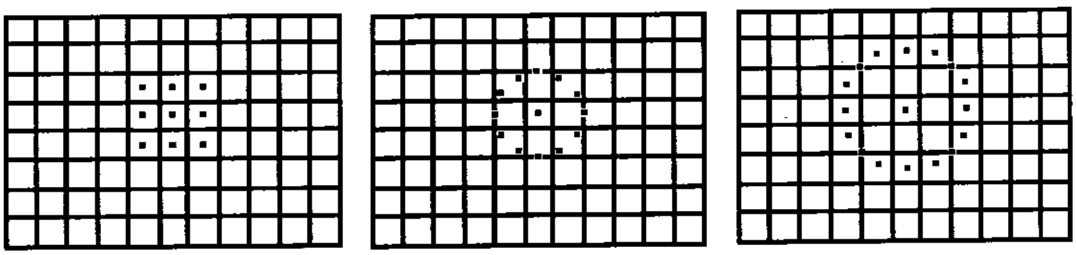
$$LBP(x_c, y_c) = \sum_{n=0}^{7} S(i_n - i_c)2^n$$

$$S(x) = \begin{cases} 1 \text{ if } (x => 0) \\ 0 \text{ if } (x < 0) \end{cases}$$ (1) Basic LBP Operator

Here x_c *and* y_c shows the position of center pixel, i_n and i_c is gray scale value of surrounding pixel and central pixel respectively. All pixels in image is labeled with LBP code and a histogram is generated from LBP coded image which is used as features set for further classification.

2.2. Local Binary Pattern Extension:

Local Binary Pattern has one its very successful extension form [30] with the concept of circular neighborhood of varying radius and with the relaxation of any number of neighborhood pixels and any radius value. This figure shows LBP extension where varying radius values like R=1.0, R=L5 and R=2.0 are used and similarly it shows the selection of different values for pixels like P=8, P=12 and P=16. This form can be represented by the following formula;



(P=8, R= 1.0) (P=12, R= 1.5) (P=16, R= 2.0)

Fig. 2.2 LBP Operator Extension [30]

In this formula, P is showing selected number of pixels and R is showing radius value. Here g_j and g_c represents the neighborhood pixels and central pixel respectively. After generating LBP code image, histograms of image is created which give information about different micro patterns in the image and this histogram bins are treated as features set.

$$LBP_{P,R}^{u^2} = \sum_{j=0}^{p-1} S(g_j - g_c)2^j$$

$$S(x) = \begin{cases} 1 \text{ if}(x => 0) \\ 0 \text{ if}(x < 0) \end{cases}$$

(2) LBP Operator Extension

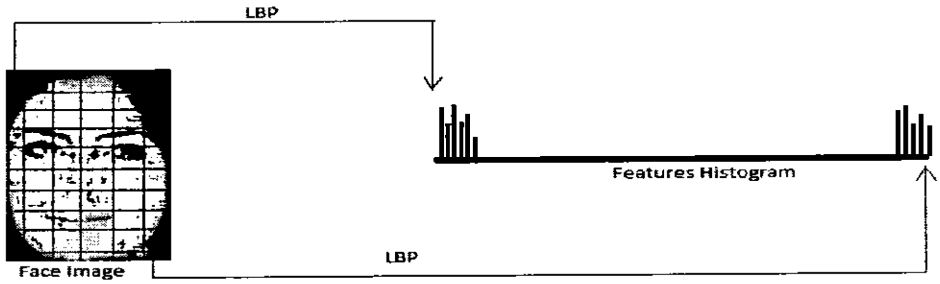


Fig. 2.3 LBP Histogram

Another way is to divide complete image into small regions, LBP coded is done for each region, then histograms of each region is generated and these histograms are combined together to show global histogram of the face. These regions are more accurate in helping image information reading and interpreting. The following circular results show different micro patterns of facial image.

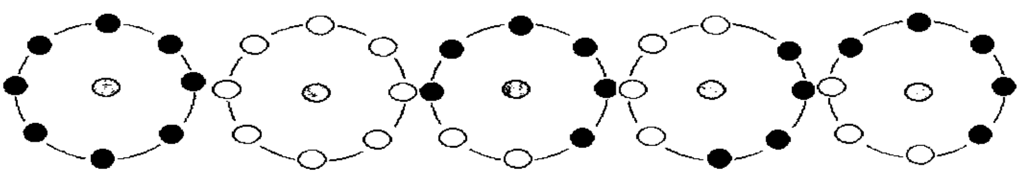


Fig. 2.4 LBP Detectable Texture Primitives

A circular LBP operator gives us different micro textures like spot, line end, edge and corner. It is shown in Fig.4, here black circles show “Zero” value and white circles show “Ones”. Features set are the collection of histogram bins.

2.3. Local Direction Pattern:

Local Directional Pattern (LDP) is an approach for facial feature extraction; it uses edge responses of each pixel in all eight directions, and generates a code named as LDP. Kirsch masks are used in eight directions ($M_0 \sim M_7$) with the central pixel position as shown in Fig. 2.5.

$$M_0\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\quad M_1\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}\quad M_2\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\quad M_3\begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}$$

$$M_4\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}\quad M_5\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}\quad M_6\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\quad M_7\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}$$

Fig. 2.5 Kirsch Edge Response Masks in Eight Directions

Each direction mask is represented by M which shows edge response of pixel in that particular direction but every direction response is not equally important because only edge or corner which shows high response are important for calculating LDP code. It finds out K prominent responding directions and set values as one while remaining 8-k bit values are set zeros. This complete method is described in the following Fig. 6.

20	90	83
95	12	11
89	29	56

Mask Index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask Value	-651	-27	285	213	221	-435	-127	-399
Rank	2	6	1	8	7	5	3	4
Code Bit	1	0	1	0	0	0	1	1
LDP Code	197							

Fig. 2.6 Local Directional Pattern Code Generation

First row shows eight masks index $m_0 \sim m_7$ and second row is showing its calculated values. If it is carefully observed then value (-651) has high variance and then (-285), (-127), and (-399) are showing high responses so this method assigns these values as 1 and remaining as 0 so the LDP code is 10100011 which means 163. In this way complete image is processed for LDP code and finally face is described by LDP histograms as shown below.

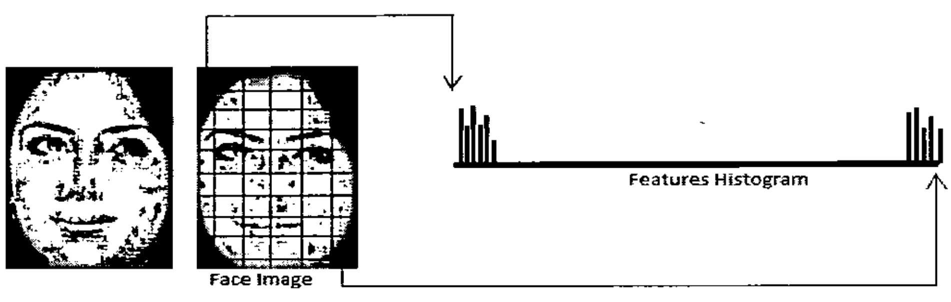


Fig. 2.7 Facial representation using LDP histogram

Histogram generated through LDP code shows very fine results of image; local texture features, corners, spots and edges. LDP technique can work better than LBP technique even if the image is passed under white noise using Gaussian method. This result is shown in the following diagram;

20	90	83
95	12	11
89	29	56

→ LBP=00111000
LDP=00010011

After white noise by Gaussian Method. →

20	90	83
95	12	11
89	29	56

→ LBP=00101000
LDP=00010011

Fig. 2.8 LDP and LBP under Gaussian Method

After noise as shown in Fig. 8, LBP code gets affected at 5^{th} bit where as LDP code shows unchanged code.

2.4. Local Ternary Pattern:

LTP is also an extension of LBP operator and the difference is in basic LBP operator it has 59 bins of histogram while in LTP bins are further reduced to 24 only which give very good results. As in LBP, it also uses the concept of uniform patterns; which allows maximum two bit wise transitions in code. It uses the following formula to calculate LTP code.

$$y(x) = \begin{cases} 0 & (x < -0) \\ 1 & (-\theta <= x >= 0) \\ 2 & (x > 0) \end{cases}$$

$$LTP(x_c, y_c) = \sum_{n=0}^{7} S(g_n - g_c) \qquad \text{(3) LTP Operator}$$

Here the central pixel is evaluated against the 8 neighbors using a threshold θ. LTP sows grey scale values invariance because, similar to LBP, it actually uses the difference between grey scale values rather than original grey scale values. Let's take an example to calculate LTP.

129	125	199
120	110	90
139	119	87

→

2	2	2
2		1
2	2	0

→ LTP: 22102222

Fig. 2.9 Calculation of LTP Code

2.5. Discrete Cosine Transform:

Discrete Cosine Transform (DCT) uses sine functions to transform images into frequency domain. It expresses a function or a signal in terms of a sum of sinusoids with different frequencies and amplitudes and is a famous technique to reduce dimensionality and used for image compression generally. DCT method selects only sine waves from transformed image in frequency domain. The following equation is used to calculate DCT coefficients *C (u, v)* of image *f (x, y)*.

$$C(u,v)=\frac{2}{\sqrt{mn}}(a(u)a(v))\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f(x,y)\cos[\frac{(2x+1)u\pi}{2m}]\cos[\frac{(2y+1)v\pi}{2n}] \quad \text{(4) DCT}$$

DCT has very good quality to pack high information which makes it very robust for image compressions. DCT arranges its coefficients in priority order and high significant coefficients are chosen first which show high variance, this variance decreases with high frequency, so for dimension reduction, low frequency coefficients are usually selected. It is almost equivalent to Discrete-Fourier-Transform but it uses only real numbers and roughly twice the length operating on real data with even symmetry. Due to its strong energy compaction property, it is usually preferred for lossy data compression. Discrete Fourier Transform is different from DCT in a sense that DCT uses only cosines function while DFT uses both cosines and Sines functions.

2.6. Discrete Wavelet Transform:

Discrete wavelet transform is in which small wavelets are discretely sampled. It has a large number of practical applications n science, engineering, and mathematics and computer science. A wavelet transform has one advantage over Fourier transform which is it captures both frequency and location information. It was actually invented by Hungarian mathematician n 1988. The first DWT is HAAR Transform which simply pairs up input values, storing the difference and passing the sum. It demonstrates the localization and provides a time-frequency representation of the signal. It was developed for further development in Fourier Transform because it notice both time and frequency as compared to Fourier transform in which only frequency is noted. A wavelet is small piece of wave which exists for a finite domain only whereas a sinusoidal wave which is used by Fourier transform has its infinity repetition.

2.7. Summary

In this chapter, a literature survey is presented about facial features extraction techniques. For an accurate classification of facial expressions, a precise and small set of features is required and a good technique always brings small set of facial features without loss of important information from the image. Local Binary Pattern, its one extension, DCT and DWT are main facial features extraction techniques discussed in this chapter and the simple and accurate results are shown by Optimized LBP technique. Some other extensions of LBP are also discussed in summarized form along with its procedures and calculating algorithms.

Chapter 3

3. Classification Techniques

This chapter describes basic concepts of classification and discusses different classification techniques in general and K Nearest Neighbor in particular. In recent some decades, electronic data has tremendously increased due to storing, editing and processing purposes of historical data but this huge data is of no use unless meaningful information is extracted from this hell of data. For this purpose, machine learning techniques have been very popular and useful to extract information, to mine data, to consolidate data, to predict future trends and to picture data. Classification is helpful at large level for organizations in retrieval of legal data from large data in a very short time span and type and size of data always varies from one organization to other organization but most of organizations do not correlate information with their processes and result is poor utilization of storage resources.

An algorithmic function that has power to classify data samples and is discretely coded for this job is known as “Classifier”. Input data samples to classifier are known generally as instances and categories used by classifier are known as classes. Classification generally thought as supervised or unsupervised; supervised learning means classification of new instances based on learning from training data set which is already classified manually or by system but it is hundred percent correct classified training data sets where unsupervised classification is known as clustering like grouping data samples into classes which based on similarity measures by inheritance.

The above both techniques classification and clustering are very general methods of pattern recognition where others are regression (assigning real-value), sequence labeling (assigning class to each member of sequence) and parsing (assigning parse tree to input sentence). Different classification techniques are used for different purposes and each technique has its varying properties, different strengths and different weaknesses. Here some of these techniques are discussed one by one. Probabilistic or statistical classification is one of the classification sub class and it gives us the probability of the input sample being the member of each of the possible output classes and on the other hand, non-probabilistic techniques simple output the best class for output.

3.1. Support Vector Machine:

SVM is a discriminative classification approach, more accurate and based on Structural Risk Minimization (SRM) principle. It is an inductive approach to learn from finite training data and also controls the generalization ability of learning machines. It also guarantees lowest true error. It uses both negative and positive training datasets which is not very common in other classification techniques and it creates a decision surface known as hyper plane which separates both negative and positive training sets. The closest points to the decision surface are called support vectors and others non-supporting vectors if removed have no effect on decision. SRM maximizes classification accuracy as it entails to find optimal hyper plane which is helpful in reducing error rate but this thing has some serious drawbacks like complexity of training dataset and categorizing algorithm and also high consumption of memory and time.

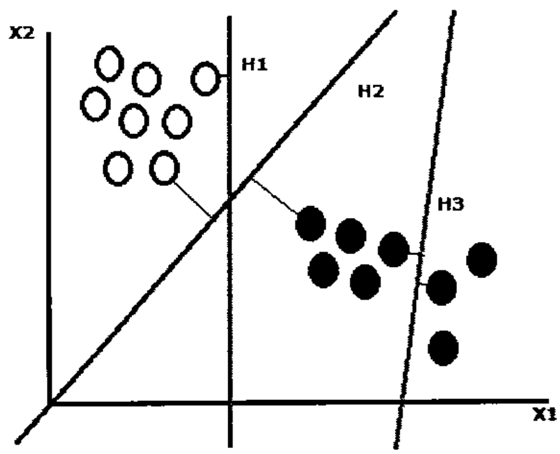


Fig. 3.1 SVM Hyper Planes

This diagram represents that two different classes are shown by filled and blank circles. The maximum-margin hyper plane is need to found out which has maximum distance from data points so three different hyper planes are drawn to best separate data samples from each other. Green (H1) and blue (3) hyper planes are not suited as they are not highly separating while red hyper plane H2 is best suited because it separates maximum and known as optimal hyper terminal.

3.2. K-Nearest Neighbors:

KNN performs classification based on the instances found in nearest neighbors that's why it is also called instance based classification. On the basis of categories, the feature space is partitioned into regions and a testing sample is assigned the category which is closest and frequent to testing sample in the k nearest neighbors of feature space. KNN is outclass classifier because of its simplicity so it requires only a small training set, an integer to specify the value of K and a metric to find the closeness between testing samples and training samples. Euclidean distance formula is generally preferred to measure closeness during testing phase while in training phase only feature space is built and training set is categorized. During testing phase, K closest samples are selected and the more categories are selected as predicted category for input sample.

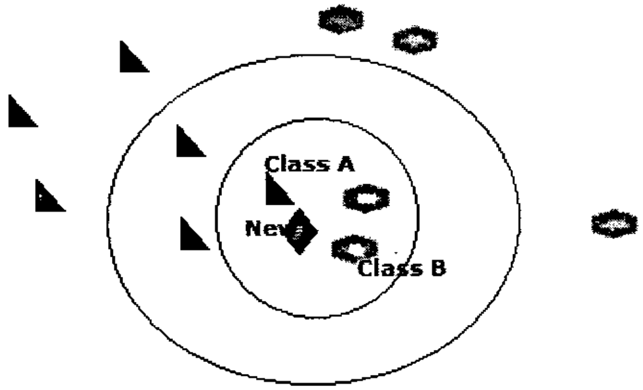


Fig. 3.2 KNN Classification

KNN classification is shortly presented using this diagram here. Suppose diamond shape red color object as a new sample is entered which needs to be classified as in class "A" (right triangle shaped blue color) or class "B" (hexagon shaped green color). In the first circle where "K" is 3, classifier founds class "B" samples more in count than class "A" so it will classify "New" sample as class "B". If the value of "K" is increased to 5 then class "A" has more count than class "B" so classifier will classify "New" sample as class "A". In this case, uneven distribution of samples can lead to misclassifications.

3.3. Bucketing Algorithm:

It is an enhancement for conventional KNN to increase its efficiency by decreasing time consumption, the processor and the physical memory usage. In this algorithm first the feature space is divided into identical cells according to categories. Data points in each cell are stored in a list and during the classifying stage, the distance between the internal data points of each cell and the input data point is computed. The process terminates when the distance between input data point and the cell currently examined is larger than the distance between input data point to the cell which has been examined. The cell with shortest distance to the input data point gives us classification result.

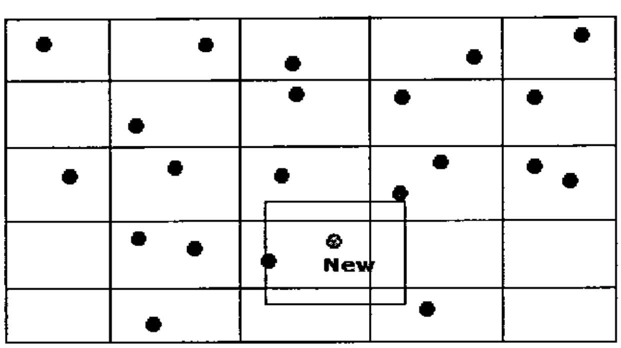


Fig. 3.3 Bucketing Algorithm

This diagram shows simple working of KNN bucketing algorithm. Here green circle represents "New" data sample and red points represents existing class data members.

3.4. K-Dimensional Trees:

This algorithm is another enhancement for conventional KNN to increase its efficiency. It is a binary search tree and generalized in high dimensions and organizes data points into k-dimensional feature space. During training stage, feature space is partitioned into K-dimensional tree and each node is associated with a hyper-rectangle and a hyper-plane which is orthogonal to one of the coordinate axis and this hyper-rectangle is divided into two parts by the hyper-plane and these two parts are associated with the successor of node. It ends the phase of training as the number of training data points in the hyper-rectangle is less than a given threshold. During the classifying stage, an input data point is mapped into one of the nodes in k-dimensional tree. This algorithm searches the tree in descending order to find all the training data points. The classification result is obtained based on the label of the training data point which has the shortest distance to the input data point.

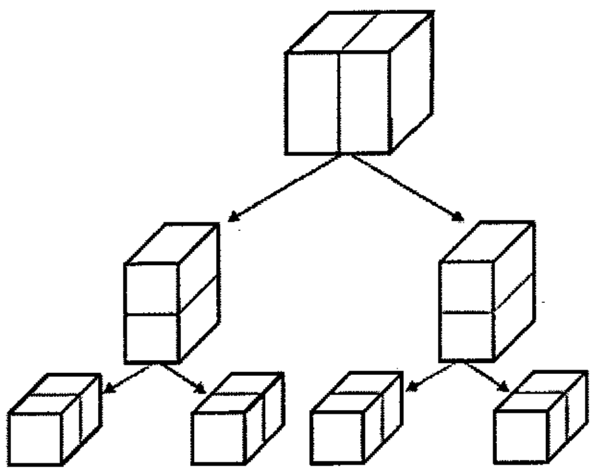


Fig.3.4 K-Dimension Tree

This diagram shows K-dimensional tree where top hyper rectangle is shown and divided by a hyper plane and these two divided parts are associated with its successor nodes which are again divided by hyper plane and this process continues as explained above for classification.

3.5. Hybridized KNN and SVM:

HKNNSVM is also one of the improved methods to deal with classification in an effective way which uses both KNN & then KNN with SVM as hybridized. In this technology, KNN is first used to prune training samples and then second KNN hybridized with SVM is used to classify the testing samples and it uses Euclidean formula to calculate distance. In the first KNN, if the class label of training sample is same as the label of the majority of its k nearest neighbors, the training sample is reserved, whereas others are pruned. For the pruned samples set, the second KNN and SVM are applied to classify. If K nearest neighbors has all the same labels, the sample is labeled.

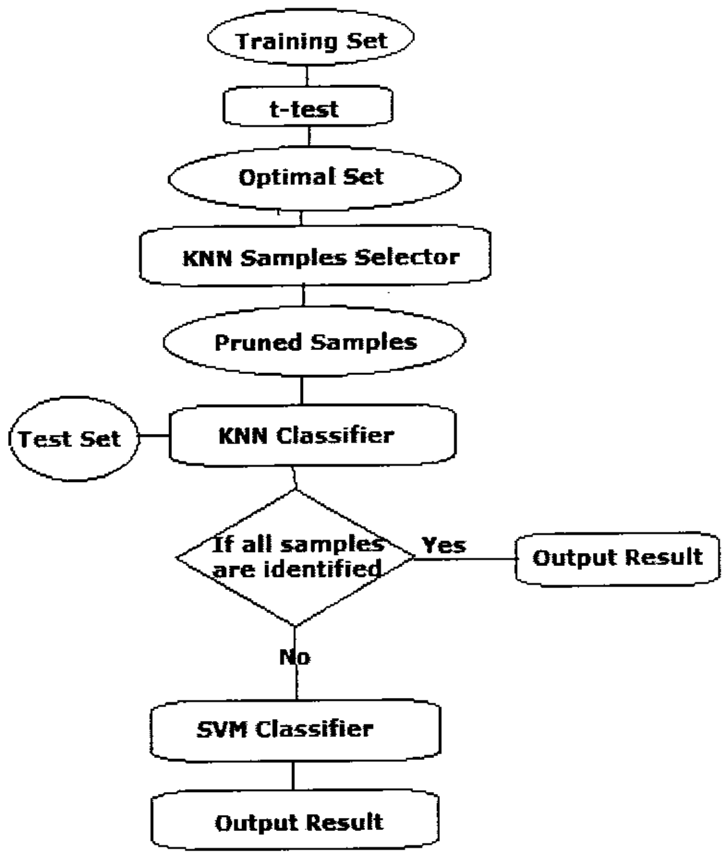


Fig. 3.5 Hybridized KNN – SVM

This process is shown in this figure that how first classifier "KNN" works its best to classify test dataset at first if all samples are classified then its output is accepted as result but if it is unable to classify test data sample then second classifier "SVM" comes into action and performs classification. It is wonderful that KNN classifies this test data set with little complexity and computation time at first stage.

3.6. Modified KNN:

It is a modified version of conventional KNN approach. It performs classification in the same way as number of class labels with larger count in K neighbors but it uses the concept of weighted terminology. It calculates weights in a different way like fraction of the same labeled neighbors to the total number of neighbors. In this method, validity of the data samples in training set is computed at first and then a weighted KNN is performed on any test samples. Validity of training sample x is the number of points with the same label to the label of x. Weighted KNN now uses weighted vote rather than majority vote from neighbors which based on Euclidean distance formula. Now validity and distance formula combine to assign the weight of data sample and data sample with highest weight is selected as predicted label.

3.7. Nearest Feature Line Method:

In this approach, any two features points of the same class are generalized by a line known as Feature Line FL. It is shown by the following diagram.

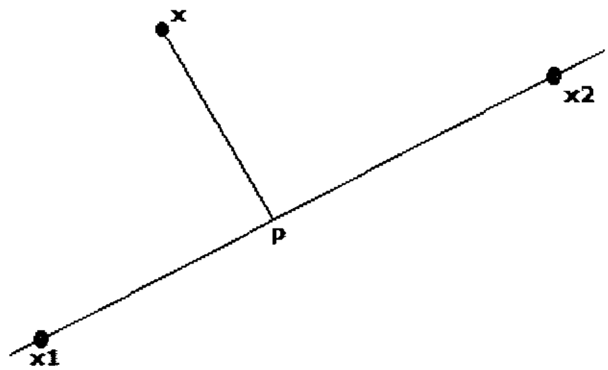


Fig. 3.6 Nearest Feature Line

The classification is based on the nearest distance from the query feature point to each FL. The query feature point x is projected onto FL as p. The distance between x *and* $\overline{x_1x_2}$ to is defined as;

$$d\left(x.\overline{x_1x_2}\right) = \|x - p\|$$ (5) Nearest Feature Line Distance

3.8. Nearest Neighbor Line Method:

During this approach, it calculates the feature lines of only nearest neighbor class samples rather than feature lines of all available classes NFL method [20]. It is shown in the following diagram.

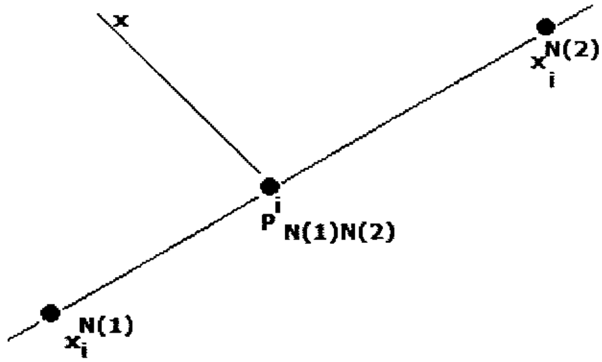


Fig. 3.7 Nearest Neighbor Line

$x_i^{N(1)}$ and $x_i^{N(2)}$ are two nearest neighbors of same class i to the query item x. The straight line passing through these two "i" class neighbors is called Nearest Neighbor Line. The distance between "x" and NL can be calculated as;

$$d\left(x, \overline{x_i^{N(1)} x_i^{N(2)}}\right) = \left\| x - p^i_{N(1)N(2)} \right\| \qquad \text{(6) Nearest Neighbor Line Distance}$$

NNL is the NL with the lowest distance over all the "c" classes.

3.9. Center-based Nearest Neighbor:

It defines a line called center-based line (CL). This CL connects a known labeled sample point and the center of sample class i.e. Let x_i^c is the training sample and O^c is the center of class c which can be calculated as;

$$O^c = \frac{\sum_{i=1}^{N_c} x_i^c}{N_c} \qquad \text{(7) Center-Based Nearest Neighbor Center of Class}$$

The above method is shown in below diagram.

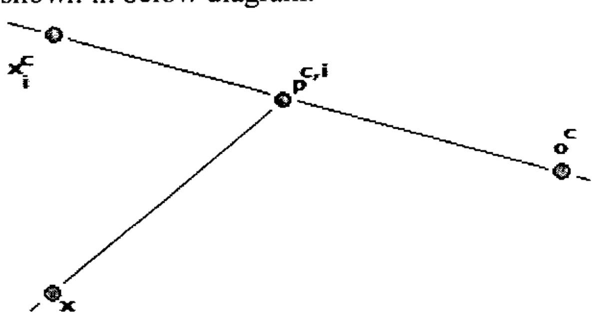


Fig. 3.8 Center-based Nearest Neighbor

For an unknown sample x, the distance between x and CL can be defined as following which is used as basic measure for classification. $p^{c,i}$ is the projection of x on CL. Distance formula for this algorithm is shown in the following formula;

$$d\left(x, \overline{x_i^c o^c}\right) = \left\| x - p^{c,i} \right\| \qquad (8)\ \text{Center-Based Nearest Neighbor Distance}$$

3.10. Problem Statement:

Facial expression classification is one of the very challenging problems due to subtlety, complexity and variability of facial expressions. Feature extraction, dimensionality reduction and classification are key steps to solve this challenging classification problem. An accurate and efficient classification of facial expression plays an important role because of its wide range of real world applications. Major problems related to facial expression classification includes

- Feature Extraction and Dimensionality reduction to obtain more compact representation of data with limited loss of information or more suitable for classification.
- Type of features more appropriate for facial expression recognition.
- An appropriate classification system which can give better efficiency and accuracy on noisy facial expressions than state-of-the-art methods by using minimum number of features.

3.11. Summary

In this chapter, basic concepts of classification are discussed in details. Why a classifier is required to handle today large amount of data and how it actually works. A classifier is an algorithm that is specific to classify unknown data samples into proper categories or even for future prediction from the huge dataset. Classifier may be complex or simple and computational high or low and more accurate or less but all this can be suffered depending on the type of its application and availability of resources. KNN is the first to be simple and important discussed here along with its variations while SVM is more accurate but it is computationally high. Different extensions and variations of KNN are discussed one by one and each extension has different way to perform classification. At the end, problem statement is addressed which explains how multi facial feature techniques and their fusion is really helpful for a precise features and finally classification.

Chapter 4

4. Facial Expression Classification using Optimized KNN and Classifier Fusion System (CFS)

In this chapter, all basic concepts of our proposed technique Optimized KNN and classifier fusion system are being discussed in details and their applications with respect to the existing methods. Facial expressions can be more accurately categorized if facial features are extracted by some efficient technique and process. This complete process should be carefully started, run and completed because at each stage any minor ambiguity or misinformation can lead to wrong and weak inputs to the next stage of process so it has been carefully done at each step. The process has some more important phases i.e. features extraction, features fusion and classification, all these are very important for next stage to give a fine output. CFS is one which is using both KNN and O-KNN alternatively as needed when data samples are not categorized by O-KNN.

4.1. Facial Expression Classification Process

This complete working is done in steps at each stage but finally to reach classification step, it has to be organized in a process form to show the steps and way it works out. This complete process can be describes in the following way.

4.1.1. Image Standardization

This process starts working on JAFFE images database by cropping the images by removing unimportant regions and calculating its standard deviation.

4.1.2. Facial Features Extraction

Local Binary Pattern and its optimized form are found to be best feature extraction techniques in both ways of its simplicity and complexity. It has been started by extracting facial features in JAFFE database using LBP, its optimized form, DCT and DWT step by step to see the power of LBP operator. This is our first priority to actually extract a very fine set of features from images using these techniques very carefully.

4.1.2.1. Optimized LBP

Optimized LBP is achieved through hit and trail base because finding its appropriate number of radius and appropriate number of data points is a very tough task if one's does it manually and its automated process a lot of time and computational capacity. This technique has many variables to check performance and still many unanswered questions in it.

4.1.2.2. Discrete Cosine Transform

DCT is our next feature extraction technique used in our process. It is prewritten in MATLAB and is used as it. It has strong power of features extraction with minimum loss of information from the image.

4.1.2.3. Discrete Wavelet Transform

DWT is third features extraction technique used in our practical thesis and the purpose of this technique is to bring a different aspect in our process so that Wavelet and Fourier transform can also be compared on this image dataset.

4.1.3. Facial Features Fusion

The above three different datasets are now fused together to test the power of multi techniques at one place. This process of fusion is very simple and follow the below method.

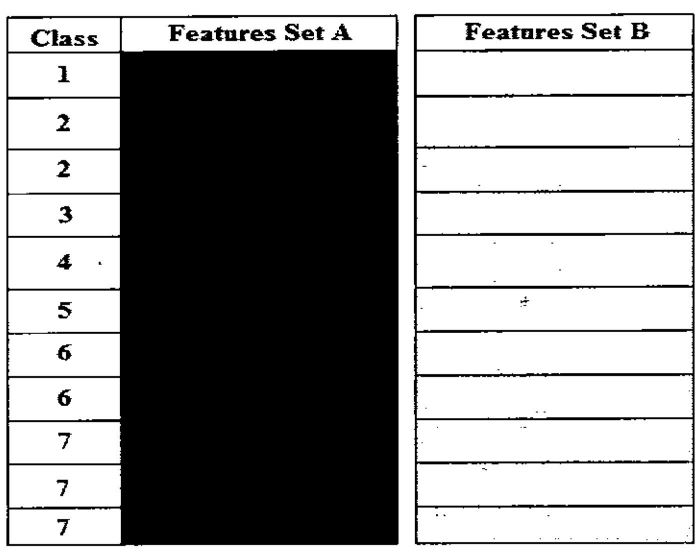


Fig. 4.1 Fusion of Features Datasets

Here F represents feature and class codes represents number of classes. Fusion of these three datasets is tested one by one but it has not actually improved results as expected so it still required some close look to more optimization. LBP and DWT features set is also further processed through Principle Component Analysis (PCA) to further get the awesome blend of features.

4.1.4. Classification

After getting features set and its different fusion sets, it can be processed by classifier and now the selected KNN and its modified form is used which is optimized KNN to classify facial expressions using these features. Simple KNN is not very effective in multi classification and in our case for seven classes its performance has been very unsatisfactory for all the above three features sets and its fusion as well.

4.1.4.1. Optimized KNN

Optimized KNN is an advanced and modified version of basic KNN technique. A simple KNN is an instance-based classification technique to classify data samples which gets majority vote from neighbors and assign the selected label to the testing sample. This technique has been optimized with the concept of weighting data samples in accordance to the neighborhood importance.

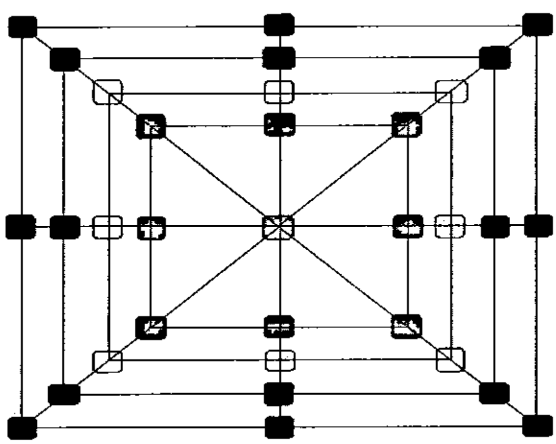


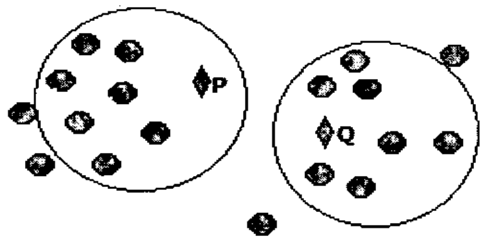
Fig. 4.2 Optimized KNN

4.1.4.1.1. Neighbors

In this proposed method, this algorithm first draws a neighborhood radius and selects data samples under this radius which are known as neighbors. Neighbors can be shown in the following diagram. There are two data samples here "P" and

"Q", each data sample has a neighborhood radius and all data samples in that circle are called neighbors.

Neighborhood Radius = 7

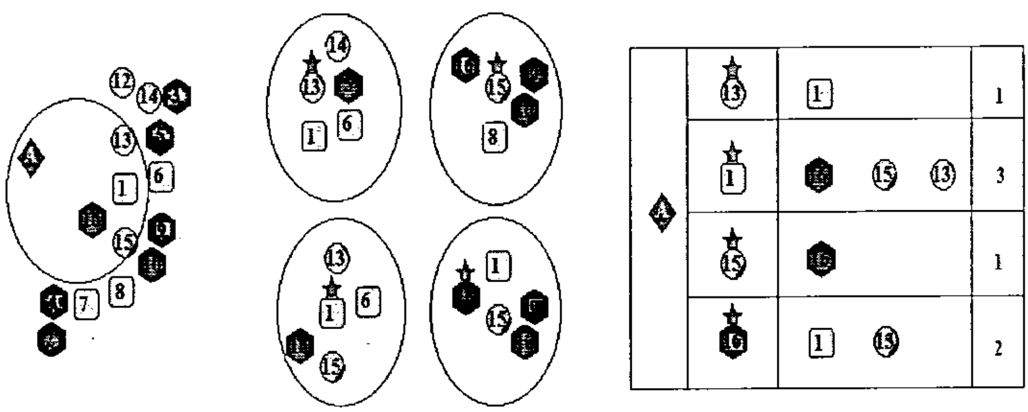


Neighbors

Fig. 4.3

4.1.4.1.2. Shared Neighbors

If two data samples share neighbors which means neighborhood circles overlap and some common neighbors found then this neighborhood is very useful in finding similarity between data samples. This concept can be shown below.



Shared Neighbors

Fig. 4.4

This diagram shows complete picture of shared neighbor concept in which three different neighborhood circles of three different data samples p1, p2 and p3 overlap and has common neighbors which gives a strong clue in classifying new objects.

4.1.4.1.3. Procedure to Weighting Data Samples

Now one of these neighbors is selected to find its neighbors in the same radius value and total number of shared neighbors between testing sample and selected neighbor is counted and this count is attached as a weight to the neighbor data sample and this process is repeated for each neighbor of testing data sample and then these weighted neighbors are sorted in descending order. Now top weighted neighbor is selected as label for current testing data sample.

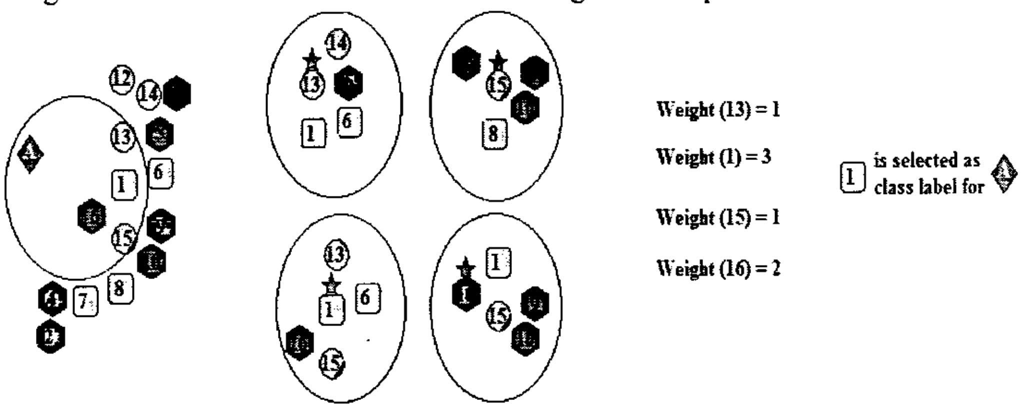


Fig. 4.5 Weighting shared neighbors

4.1.4.2. Classifier Fusion System

In our next step, O-KNN is applied to assign some label to the testing sample and if O-KNN is unable to predict any label then simple KNN method uses its majority vote and assign label to the testing data sample. In this way, two classifiers KNN and O-KNN are fused together to enhance each other output when one technique is failed to label the testing sample. This CFS has a little addition in our O-KNN approach because almost all data samples are labeled in some at least category.

This process proceeds step by step with different datasets and classifier implementations so a close look of this complete process can shown in a diagram form as below.

4.2. Process Diagram

The complete process can be presented in a diagram form as below.

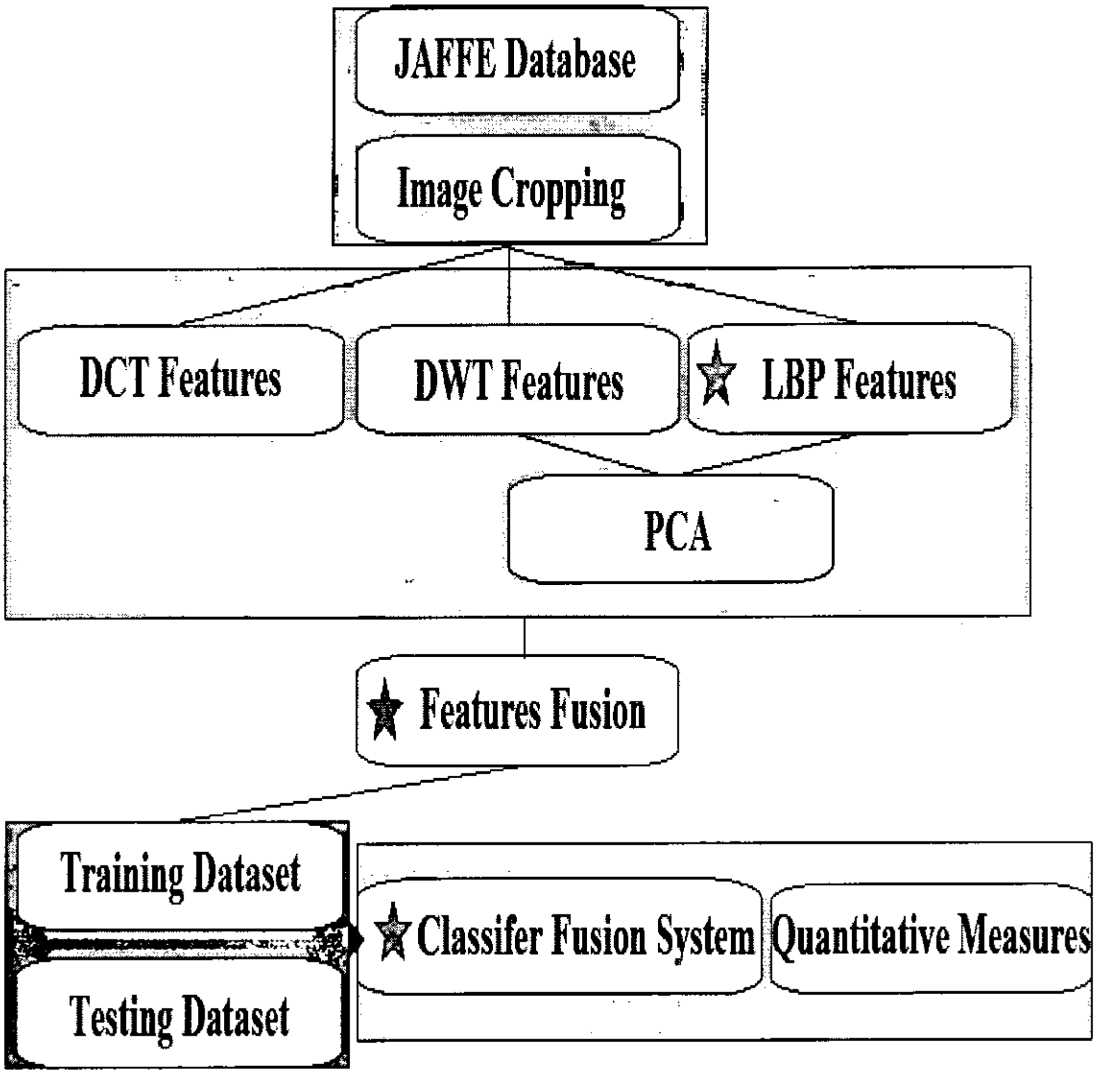


Fig. 4.6 Process Diagram

4.3. Algorithm : Classifier Fusion System

```
Inputs:     TsD = Testing Dataset, TrD = Training Dataset, DF = Distance-Formula,
            NR = Neighborhood-Radius and K
    Begin
            For Me = TsD-Sample In TsD

                Nr-D = Find-Neighbors (DF, NR, Me, TrD)
                WtD = Weight-Shared-Neighbors (DF, NR, Nr-D, TrD)
                SrD = Sort-DS (WtD)
                Predicted-Label = Select-Top-One (SrD)

                If (Predicted-Label <> Class-Labels)
                    K-Samples = Find-K-Samples (NrD, K)
                    Predicted-Label = Top-Counted-Class-Label (K-Samples)
                End If

                Me = Predicted-Label

            End For
    End

Output: Classified Testing Dataset
```

Summary

In this chapter, our proposed method is explained in a form of process step by step and explains how it proceeds further. A diagram is also shown to give a quick view of complete process starting from images normalization and then features extraction using LBP, DCT and DWT techniques and finally fusion of these features datasets to get a big dataset of excellent features. It also explains that how each dataset is tested by partitioning into training and testing databases and applying our proposed technique on these datasets one by one. At the last stage, different quantitative measures are calculated to see accuracy results. An algorithm is also included in this chapter to describe the concept of classifier fusion system.

Chapter 5

5. Experimentation and Results

In this chapter, experiments are performed and mention results are calculated and shown. At first, all images are loaded and processed for removing unimportant regions from the image like hair and dress or shoulders side because these areas are unimportant for facial expression classification so each image is cropped for its face part only.

5.1. Performance Measures:

The following performance measures have been used to evaluate these techniques.

Actual Values

		Happy	**Not Happy**
Prediction Outcomes	**Happy**	**True Positive**	**False Positive**
	Not Happy	**False Negative**	**True Negative**

Fig. 5.1 Performance Measures Matrix

5.1.1. Accuracy

Accuracy is calculated using the formula by dividing total number of true positives plus total number of true negatives with total number of true positives plus false positives plus true negatives plus false negatives and it can be shown as below.

$$\text{Accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number oftrue positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

It is calculated for all seven classes, in this case if one class is treated as positive while others remaining six classes are treated as negatives.

5.1.2. Receiver Operating Characteristics (ROC)

Receiving Operating Characteristics is actually a graphical plot between sensitivity and 1- specificity.

5.1.3. Sensitivity

Sensitivity is termed as true positive rate which is actually obtained by dividing number of true positives with number of true positives plus number of false negatives.

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of true\ positives + number\ of\ false\ negatives}$$

5.1.4. Specificity

Specificity refers to the ability of the test to identify negative results, and it is obtained by dividing number of true negatives divided by number of true negatives plus number of false positives.

$$Specificity = \frac{number\ of\ true\ negatives}{number\ of false\ positives + number\ of\ true\ negatives}$$

5.1.5. Precision

Precision is calculated by dividing true positive with true positives plus false positives.

$$\Pr ecision = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ postives}$$

5.1.6. Recall

Recall is calculated by dividing true positive with true positives plus false negatives.

$$\operatorname{Re} call = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}$$

5.2. Tool Used

MATLAB is used as automation tool for implementing features extraction technique and also classification techniques and all results are cross verified with WEKA tool classifiers as well.

5.3. Data Sets

JAFFE database is being used for testing purpose of this technique. JAFFE images database contains 213 images of 7 different facial expressions posed by 10 Japanese female models. Each image is of 256 by 256 sizes.

5.4. Scenarios

The following three basic features extraction techniques and datasets are used separately and in combined form. There are following scenarios tested in our experiments.

Scenario No.	Features Extraction Methods			Classifier
1.	DCT	--	--	Classifier Fusion System (CFS)
2.	--	Basic LBP	--	CFS
3.	--	Basic LBP + PCA	--	CFS
4.	--	Optimized LBP	--	CFS
5.	--	Optimized LBP + PCA	--	CFS
6.	--	--	DWT	CFS
7.	--	--	DWT + PCA	CFS
8.	DCT	Optimized LBP	--	CFS
9.	--	Optimized LBP	DWT	CFS
10.	DCT	--	DWT	CFS
11.	DCT	Optimized LBP + PCA	--	CFS

Fig. 5.2 Features Extraction Scenarios

As first step of image normalization is very important in removal of unimportant regions from the image like hair and shoulders etc and crop only face from the image. Now LBP technique is applied to extract features and then apply PCA on LBP features to further reduce the dimensionality of feature space and get two features sets LBP and LBP-PCA where LBP-PCA is a LBP features set further reduced using PCA. The following parameters are being used in applying LBP technique which has somehow improved LBP features set.

- Histogram Bins
- Data Points
- Radius
- Total LBP Features

Histograms bins varyingly give different performances as observed 200 up to 400 bins give good results range. Similarly, data points also can be selected differently its good performing range is from 4 to 16 as multiples of 4. Next variable for LBP improved version is its radius and according to our experiments as the value of this variable is increased starting from 1 up to 20 it keeps improving the performance step by step. The last variable is total number of LBP features selected for using in classification. The next features set extracted is LBP-PCA which is achieved by applying PCA on LBP coded image and this feature set has not actually improved the LBP features performance rather it has decreased classification performance.

Next feature extraction technique used to extract wonderful features is Discrete Cosine Transform (DCT). It is Fourier Transform technique but it uses Cosine function rather than Sine function. This technique has wonderful power to shorten the number of features so that only 6 to 20 features count can be wonderful results containing in classification. Discrete Wavelet Transform (DWT) is our third technique used to extract features from images. It is also very effective technique and it has given us good results. After extracting three features sets fusion is performed as simple fusion or combining of these features sets in different ways. The following terms are being used for different feature sets.

Features Set	Description
LBP	It contains only Local Binary Pattern features
LBPPCA	It contains LBP features applied with Principle Component Analysis
DWT	It contains only Discrete Wavelet Transform features
DWTPCA	It contains DWT features applied with PCA
DCT	It contains only Discrete Cosine Transform features
LBP-DCT	It contains LBP fused with DCT features
LBP-DWT	It contains LBP fused with DWT features
DWT-DCT	It contains DWT fused with DCT features

Fig. 5.3 Features Datasets

The above feature sets are our first achievement to improve the process and try to achieve maximum useful features and see which is better than other. The same process can be used for any other image databases.

Now after feature sets, work on our proposed technique of Optimized KNN is started. The concept of weighted data samples is being used in this method. As explained above in our technique it has improved KNN for multi-classification for more than expected. A comparison has been performed between simple KNN results and optimized KNN and results are amazingly more accurate up to 100% in some cases. Here is detail analysis of our experiments.

5.4.1. Scenario No. 1 (Basic LBP)

In our first scenario, the process is using features set from basic LBP operator which uses 3×3 pixels from its neighbors. It uses following parameters to extract features.

- Features Set = Basic LBP
- Histogram Bins = 650
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier		True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	1.00	0.11	0.44	0.60	1.00	
	Sad	0.33	0.06	0.56	0.50	0.33	
	Surprised	1.00	0.06	0.44	0.75	1.00	
	Angry	0.67	0.11	0.50	0.50	0.67	
	Disappointed	0.33	0.22	0.56	0.20	0.33	
	Fear	0.33	0.00	0.56	1.00	0.33	
	Neutral	0.00	0.00	0.61		0.00	52.38
Classifier Fusion System	Happy	0.67	0.06	0.61	0.67	0.67	
	Sad	0.67	0.06	0.61	0.67	0.67	
	Surprised	1.00	0.00	0.56	1.00	1.00	
	Angry	0.33	0.11	0.67	0.33	0.33	
	Disappointed	1.00	0.06	0.56	0.75	1.00	
	Fear	0.00	0.11	0.72	0.00	0.00	
	Neutral	0.67	0.06	0.61	0.67	0.67	61.90

Fig. 5.4 Basic LBP Results

5.4.2. Scenario No. 2 (LBPPCA)

In this scenario, the process is using features set from basic LBP operator which uses 3*3 pixels from its neighbors and this features set is further reduced using PCA method. It uses following parameters to extract features.

- Features Set = LBPPCA
- Histogram Bins = 650
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.67	0.22	0.17	0.33	0.67	
	Sad	0.00	0.11	0.28	0.00	0.00	
	Surprised	0.33	0.11	0.22	0.33	0.33	
	Angry	0.33	0.11	0.22	0.33	0.33	
	Disappointed	0.00	0.11	0.28	0.00	0.00	
	Fear	0.00	0.00	0.28		0.00	
	Neutral	0.33	0.22	0.22	0.20	0.33	23.81
Classifier Fusion System	Happy	0.33	0.11	0.28	0.33	0.33	
	Sad	0.33	0.28	0.28	0.17	0.33	
	Surprised	0.00	0.22	0.33	0.00	0.00	
	Angry	0.33	0.06	0.28	0.50	0.33	
	Disappointed	0.33	0.11	0.28	0.33	0.33	
	Fear	0.00	0.00	0.33		0.00	
	Neutral	0.67	0.06	0.22	0.67	0.67	28.57

Fig. 5.5 Basic LBP with PCA Results

5.4.3. Scenario No. 3 (OLBP)

In this scenario, the process is using features set from optimized LBP operator which uses eight corners pixels from its neighbors of any specified radius. It uses following parameters to extract features.

- Features Set = OLBP
- Histogram Bins = 350
- Radius = 35
- Data Points = 8
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	1.00	0.17	0.61	0.50	1.00	
	Sad	1.00	0.11	0.61	0.60	1.00	
	Surprised	1.00	0.06	0.61	0.75	1.00	
	Angry	0.33	0.00	0.72	1.00	0.33	
	Disappointed	0.67	0.06	0.67	0.67	0.67	
	Fear	0.67	0.00	0.67	1.00	0.67	
	Neutral	0.00	0.00	0.78	0.00	0.00	66.67
Classifier Fusion System	Happy	1.00	0.00	0.94	1.00	1.00	
	Sad	1.00	0.00	0.94	1.00	1.00	
	Surprised	1.00	0.06	0.94	0.75	1.00	
	Angry	1.00	0.00	0.94	1.00	1.00	
	Disappointed	1.00	0.00	0.94	1.00	1.00	
	Fear	0.67	0.00	1.00	1.00	0.67	
	Neutral	1.00	0.00	0.94	1.00	1.00	95.24

Fig. 5.6 Optimized LBP Results

5.4.4. Scenario No. 4 (OLBPPCA)

In this scenario, the process is using features set from optimized LBP operator which uses eight corners pixels from its neighbors of any specified radius and this features set is further processed with PCA. It uses following parameters to extract features.

- Features Set = OLBPPCA
- Histogram Bins = 350
- Radius = 35
- Data Points = 8
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	1.00	0.11	0.50	0.60	1.00	
	Sad	0.33	0.06	0.61	0.50	0.33	
	Surprised	1.00	0.11	0.50	0.60	1.00	
	Angry	0.67	0.11	0.56	0.50	0.67	
	Disappointed	0.33	0.00	0.61	1.00	0.33	
	Fear	0.67	0.00	0.56	1.00	0.67	
	Neutral	0.00	0.11	0.67	0.00	0.00	57.14
Classifier Fusion System	Happy	0.67	0.06	0.72	0.67	0.67	
	Sad	0.67	0.06	0.72	0.67	0.67	
	Surprised	1.00	0.00	0.67	1.00	1.00	
	Angry	1.00	0.06	0.67	0.75	1.00	
	Disappointed	0.67	0.00	0.72	1.00	0.67	
	Fear	0.00	0.11	0.83	0.00	0.00	
	Neutral	1.00	0.06	0.67	0.75	1.00	71.43

Fig. 5.7 Optimized LBP with PCA Results

5.4.5. Scenario No. 5 (DCT)

In this scenario, the process is using features set from DCT operator. It uses following parameters to extract features.

- Features Set = DCT
- Features Selected = 350
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	1.00	0.11	0.56	0.60	1.00	
	Sad	0.33	0.22	0.67	0.20	0.33	
	Surprised	1.00	0.00	0.56	1:00	1.00	
	Angry	1.00	0.06	0.56	0.75	1.00	
	Disappointed	0.67	0.00	0.61	1.00	0.67	
	Fear	0.33	0.00	0.67	1.00	0.33	
	Neutral	0.00	0.06	0.72	0.00	0.00	61.90
Classifier Fusion System	Happy	1.00	0.00	0.89	1.00	1.00	
	Sad	0.67	0.06	0.94	0.67	0.67	
	Surprised	1.00	0.00	0.89	1.00	1.00	
	Angry	1.00	0.00	0.89	1.00	1.00	
	Disappointed	1.00	0.00	0.89	1.00	1.00	
	Fear	1.00	0.06	0.89	0.75	1.00	
	Neutral	0.67	0.00	0.94	1.00	0.67	90.48

Fig. 5.8 DCT Results

5.4.6. Scenario No. 6 (DWT)

In this scenario, the process is using features set from DWT operator. It uses following parameters to extract features.

- Features Set = DWT
- Features Count = 650
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.67	0.17	0.56	0.40	0.67	
	Sad	0.33	0.00	0.61	1.00	0.33	
	Surprised	1.00	0.06	0.50	0.75	1.00	
	Angry	1.00	0.11	0.50	0.60	1.00	
	Disappointed	0.33	0.11	0.61	0.33	0.33	
	Fear	0.33	0.06	0.61	0.50	0.33	
	Neutral	0.33	0.00	0.61	1.00	0.33	57.14
Classifier Fusion System	Happy	1.00	0.00	0.89	1.00	1.00	
	Sad	0.67	0.00	0.94	1.00	0.67	
	Surprised	1.00	0.00	0.89	1.00	1.00	
	Angry	1.00	0.06	0.89	0.75	1.00	
	Disappointed	0.67	0.06	0.94	0.67	0.67	
	Fear	1.00	0.00	0.89	1.00	1.00	
	Neutral	1.00	0.00	0.89	1.00	1.00	90.48

Fig. 5.9 DWT Results

5.4.7. Scenario No. 7 (DWTPCA)

In this scenario, the process is using features set from DWT operator and this feature set is further processed by PCA technique. It uses following parameters to extract features.

- Features Set = DWTPCA
- Features Count = 650
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.33	0.11	0.50	0.33	0.33	
	Sad	0.67	0.11	0.44	0.50	0.67	
	Surprised	0.33	0.06	0.50	0.50	0.33	
	Angry	0.33	0.06	0.50	0.50	0.33	
	Disappointed	0.33	0.06	0.50	0.50	0.33	
	Fear	1.00	0.22	0.39	0.43	1.00	
	Neutral	0.33	0.00	0.50	1.00	0.33	47.62
Classifier Fusion System	Happy	0.67	0.11	0.67	0.50	0.67	
	Sad	0.33	0.00	0.72	1.00	0.33	
	Surprised	0.33	0.00	0.72	1.00	0.33	
	Angry	0.67	0.00	0.67	1.00	0.67	
	Disappointed	0.67	0.06	0.67	0.67	0.67	
	Fear	1.00	0.00	0.61	1.00	1.00	
	Neutral	1.00	0.22	0.61	0.43	1.00	66.67

Fig. 5.9 DWTPCA Results

5.4.8. Scenario No. 8 (DCT – OLBP)

In this scenario, the process is using two different features set; one is DCT and second is optimized LBP and it have been simply fused these two sets by placing features side by side. It uses following parameters.

- First Features Set = DCT
- Second Features Set = OLBP
- O-LBP Histogram Bins = 350
- O-LBP Radius = 35
- O-LBP Data Points = 8
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	1.00	0.00	0.44	1.00	1.00	
	Sad	0.33	0.28	0.56	0.17	0.33	
	Surprised	1.00	0.06	0.44	0.75	1.00	
	Angry	0.33	0.00	0.56	1.00	0.33	
	Disappointed	0.67	0.00	0.50	1.00	0.67	
	Fear	0.33	0.06	0.56	0.50	0.33	
	Neutral	0.00	0.17	0.61	0.00	0.00	52.38
Classifier Fusion System	Happy	1.00	0.00	0.94	1.00	1.00	
	Sad	1.00	0.00	0.94	1.00	1.00	
	Surprised	1.00	0.06	0.94	0.75	1.00	
	Angry	1.00	0.00	0.94	1.00	1.00	
	Disappointed	1.00	0.00	0.94	1.00	1.00	
	Fear	1.00	0.00	0.94	1.00	1.00	
	Neutral	0.67	0.00	1.00	1.00	0.67	95.24

Fig. 5.11 DCT and OLBP Results

5.4.9. Scenario No. 9 (DCT – OLBPPCA)

In this scenario, the process is using two different features set; one is DCT and second is optimized LBPPCA and it have been simply fused these two sets by placing features side by side. It uses following parameters.

- First Features Set = DCT
- Second Features Set = OLBPPCA
- O-LBP Histogram Bins = 350
- O-LBP Radius = 35
- O-LBP Data Points = 8
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.33	0.06	0.56	0.50	0.33	
	Sad	0.67	0.28	0.50	0.29	0.67	
	Surprised	1.00	0.06	0.44	0.75	1.00	
	Angry	0.33	0.00	0.56	1.00	0.33	
	Disappointed	0.67	0.00	0.50	1.00	0.67	
	Fear	0.67	0.00	0.50	1.00	0.67	
	Neutral	0.00	0.17	0.61	0.00	0.00	52.38
Classifier Fusion System	Happy	1.00	0.06	0.94	0.75	1.00	
	Sad	1.00	0.00	0.94	1.00	1.00	
	Surprised	1.00	0.00	0.94	1.00	1.00	
	Angry	1.00	0.00	0.94	1.00	1.00	
	Disappointed	1.00	0.00	0.94	1.00	1.00	
	Fear	1.00	0.00	0.94	1.00	1.00	
	Neutral	0.67	0.00	1.00	1.00	0.67	95.24

Fig. 5.12 DCT and OLBPPCA Results

5.4.10.Scenario No. 10 (DCT - DWT)

In this scenario, the process is using two different features set; one is DCT and second is DWT and it have been simply fused these two sets by placing features side by side. It uses following parameters.

- First Features Set = DCT
- Second Features Set = DWT
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.67	0.00	0.67	1.00	0.67	
	Sad	0.67	0.17	0.67	0.40	0.67	
	Surprised	1.00	0.00	0.61	1.00	1.00	
	Angry	0.67	0.00	0.67	1.00	0.67	
	Disappointed	1.00	0.06	0.61	0.75	1.00	
	Fear	0.33	0.11	0.72	0.33	0.33	
	Neutral	0.33	0.06	0.72	0.50	0.33	66.67
Classifier Fusion System	Happy	1.00	0.00	0.94	1.00	1:00	
	Sad	0.67	0.00	1.00	1.00	0.67	
	Surprised	1.00	0.00	0.94	1.00	1.00	
	Angry	1.00	0.00	0.94	1.00	1.00	
	Disappointed	1.00	0.00	0.94	1:00	1.00	
	Fear	1.00	0.00	0.94	1.00	1.00	
	Neutral	1.00	0.06	0.94	0.75	1:00	95.24

Fig. 5.13 DCT and DWT Results

5.4.11. Scenario No. 11 (DWT – OLBP)

In this scenario, the process is using two different features set; one is OLBP and second is DWT and it have been simply fused these two sets by placing features side by side. It uses following parameters.

- First Features Set = OLBP
- O-LBP Histogram Bins = 350
- O-LBP Radius = 35
- O-LBP Data Points = 8
- Second Features Set = DWT
- Training Set = 9
- Testing Set = 1
- Distance Formula = Euclidean

Classifier	Class	True Positive Rate (Sensitivity)	False Positive Rate	True Negative Rate (Specificity)	Precision	Recall	Accuracy
KNN	Happy	0.67	0.17	0.56	0.40	0.67	
	Sad	0.67	0.11	0.56	0.50	0.67	
	Surprised	1.00	0.00	0.50	1.00	1.00	
	Angry	0.67	0.06	0.56	0.67	0.67	
	Disappointed	1.00	0.00	0.50	1.00	1.00	
	Fear	0.00	0.06	0.67	0.00	0.00	
	Neutral	0.00	0.11	0.67	0.00	0.00	57.14
Classifier Fusion System	Happy	1.00	0.06	0.78	0.75	1.00	
	Sad	0.67	0.06	0.83	0.67	0.67	
	Surprised	0.67	0.00	0.83	1.00	0.67	
	Angry	1.00	0.00	0.78	1.00	1.00	
	Disappointed	1.00	0.06	0.78	0.75	1.00	
	Fear	0.33	0.00	0.89	1.00	0.33	
	Neutral	1.00	0.06	0.78	0.75	1.00	80.95

Fig. 5.14 DWT & OLBP Results

5.5. Summary

In this chapter, complete experimentations are performed on each database set and results are also calculated. Here it is explained that which quantitative measures are calculated in the process i.e. true positive rate, false positive rate, sensitivity, specificity, recall, precision and accuracy along with definitions and formulas as calculated by system. Each scenario is briefly explained with the number of variables that counts in the process with specific values used for the under testing dataset. Each result is also shown in the form of charts. Here it is observed that Optimized LBP has better features set as compared to basic LBP operator and similarly at each stage Optimized KNN has better results than KNN classifier.

Chapter 6

6. Conclusion & Future Work

6.1. Conclusion

After all experimental results, the conclusion is that optimized LBP is better approach for facial features extraction and next to this is DCT which has also amazingly accurate features set. When LBP code is generated from far circular diagonals points i.e. 35 in our case, it gave a best description for image features. This optimized LBP can be named as Diagonal Binary Pattern. Optimized KNN has best results with both optimized LBP and DCT technique, so best combination are O-LBP/ DCT as features extraction techniques and O-KNN as classifier. Optimized KNN has improved results from KNN due to its concept of shared neighbors weightings which mean same neighbors lie in same circle and share more resemblance within circle. Fusion of different datasets in our experiments has not yielded very satisfactory results as it was expected but this idea can be optimized to get better results. Combination of DCT with LBPPCA dataset has somehow improved results of both DCT and LBPPCA as examined separately.

6.2. Future Work

In this thesis, some different ways are being used to get maximum results for classification of facial expression. Staring from features extraction techniques using some basic LBP operator and then its optimized form to classification techniques like KNN and its more optimized form has been the main focus of our research work but some other techniques has also been used for comparison and classifier accuracy check and these features extraction techniques are DCT and DWT and this all has gone with a lot of new experiences and outcomes along with classification accuracy. A different technique is also tried to combine these different features sets at one place and then tested this classifier on these combinations which have somehow shown different results and more stunning outcomes.

In this chapter, this work is concluded and suggested some points of improvements in our complete working as future work. First of all, cropping of face area from image is done but it is simply done using manual cropping so there can be some formal way which is more precise to crop face area from image and to remove unnecessary regions from the image. Next, Simple LBP is very good features extraction technique which uses its 3×3 neighbors to calculate LBP operator whereas as it has been observed that Optimized LBP has tremendous results when its radius value is increased and also give amazing change in results if points are selected along the diagonals of circle. It would require a lot of heavy processing machines to check results for variations in number of points and its radius values. Selection of these points can be explored in some ways of sequence to generate LBP code which can give very good accuracy results. It can be also checked and explored that whether a spider natural web waving sequence can be helpful in this technique of LBP code sequence.

If the radius value is increased and selected points then its LBP code value also becomes very large which is somehow confusing in some ways; firstly how these values can be interpreted with gray scale values and second how to decide its histogram bins value. This LBP code actually is also needed to be more precise and more meaningful and lying in some reasonable decimal range for such larger radius. As the number of histogram bins concerned, it is here explored manually that which bins count is more suited but histogram generation should also be improved or some technique should be used to get total number of histogram bins with respect to radius value and number of points selected and LBP code.

After exploring LBP features extraction technique, two different features set like LBP and DCT are combined but it has not improved results as it was being hoped and it can be further explored some way to combine two different features sets and bring some good results. On classification side neighbor radius, value of K and importance of neighborhood for classifier KNN is deeply observed and analyzed. KNN is found as better approach and easy way for binary classification but multi classification was a challenge for KNN so have worked on simple KNN approach to optimize multi-class classification and our proposed technique has done this job very well because it has very good classification results as compared to simple KNN but there are some areas which need to be tested and improved like distance formula, data partitioning strategy and data permutation scheme because it varies results from 80% to 98 %. Accuracy results for facial expression classification are good but not very satisfactory for practical applications so it further required research to get good techniques at each level of this process.

6.3. Applications

This technique is very helpful in classifying persons and their behaviors in the field of security and psychology. It is very interesting application if we can understand the mood and behaviors of persons who are entering in our premises. It is also very helpful in recognizing the behavior and status of prisoners during intelligence investigation and intelligence queries.

References

[1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality Social Psychology, Vol. 17, No. 2, pp.124-129, (1971)

[2] W. Wang, F. Chang, J. Zhao and Z. Chen, "Automatic Facial Expression Recognition using Local Binary Pattern," Proceedings of the 8th World Congress on Intelligent Control and Automation, Jinan, China, (2010)

[3] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Pattern," Springer-Verlag Berlin Heidelberg, ECCV, LNCS 3021, pp. 469–481, (2004)

[4] T. Jabid, M. H. Kabir, O. Chae, "Facial Expression Recognition using Local Directional Pattern (LDP)," Proceedings of 2010 IEEE 17th International Conference on Image Processing, Hong Kong, (2010)

[5] Z. Zhong, Z. Peidong, L. Yong, D. Qun, Y. Hong'an, "Face recognition based on Wavelet-Curvelet-Fractal Technique," Journal of Electronics, China, Vol.27 No.2, (2010)

[6] S. M. Lajevardi, Z. M. Hussain, "Automatic Facial Expression Recognition: Feature Extraction and Selection," Springer-Verlag London Limited, (2010)

[7] I. Bucia, I. Nafornita, "Feature Extraction through Cross-Phase Congruency for Facial Expression Analysis," International Journal of Pattern Recognition and Artificial Intelligence Vol. 23, No. 3, 617–635, World Scientific Publishing Company, (2009)

[8] F. Dornaika, F. Davoine, "Simultaneous Facial Action Tracking and Expression Recognition in the presence of Head Motion," Springer Science Business Media, LLC, Int J Comput Vis, 76: 257–281 (2007)

[9] P. Chen, G. Wang, Y. Yang and J. Zhou, "Facial Expression Recognition based on Rough Set Theory and SVM," G. Wang et al. (Eds.): LNAI 4062, pp. 772–777, Springer-Verlag Berlin Heidelberg, (2005)

[10] A. Mehrabian, "Communication without words," Psychol. Today 2(4), 53–56, (1968)

[11] Tian, Y.L. Kanade, T. Cohn, J.F, "Facial Expression Analysis. Hand Book of Face Recognition," pp. 247–275 (2005)

[12] T. Yang, V. Kecman, "Face recognition with Adaptive Local Hyperplane Algorithm," Pattern Anal Applic,13:79–83, Springer-Verlag London Limited, (2010)

[13] P. Vincent, Y. Bengio, "K-local Hyperplane and Convex Distance Nearest Neighbor Algorithms," Advances in neural information processing systems (NIPS), vol 14. MIT Press, Cambridge, pp 985–992 (2002)

[14] L. SZ, L. J., "Face Recognition using the Nearest Feature Line Method," IEEE Trans Neural Netw, 10:439–443 (1999)

[15] W. Zheng, L. Zhao, C. Zou, "Locally Nearest Neighbor Classifiers for Pattern Classification," Pattern Recognition, 37:1307–1309 (2004)

[16] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification," John Wiley & Sons, Inc., New York, (2001)

[17] C. Shan, S. Gong, P. W. Mcowan, "Facial Expression Recognition based on Local Binary Patterns, A Comprehensive Study," Image and Vision Computing 27 803–816, Elsevier B.V, (2009)

[18] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7) 971–987 (2002)

[19] J. Shlens, "A Tutorial on Principal Component Analysis," (2005)

[20] K. Cabeen, P. Gent, "Image Compression and the Discrete Cosine Transform," College of the Red Woods, (2000)

[21] S. B. Kotsiantis, "Supervised Machine Learning: A Review of classification Technique," Informatica 31, 249-268 (2007)

[22] H. Parvin, H. Alizadeh and B. M. Bidgoli, "Modified K-Nearest Neighbor," Proceedings of the World Congress on Engineering and Computer Science, WCECS, San Francisco, USA, (2008)

[23] L. Hong, Lee, C. Heng, Wan, T. Fui, Yong and H. Meian, Kok, "A Review of Nearest Neighbor-Support Vector Machines Hybrid Classification Models," Journal of Applied Sciences, ISSN 1812-5654, Asian Network for Scientific Information, (2010)

[24] Y. Cai, D. Ji ,D. Cai, "A KNN Research Paper Classification Method Based on Shared Nearest Neighbor," Proceedings of NTCIR-8 Workshop Meeting, Tokyo, Japan, (2010)

[25] L. Hong, Lee, C. Heng, Wan, T. Fui, Yong and H. Meian, Kok, "Bucketing Algorithm," Journal of Applied Sciences, ISSN 1812-5654, Asian Network for Scientific Information, (2010)

[26] L. Hong, Lee, C. Heng, Wan, T. Fui, Yong and H. Meian, Kok, "K-Dimensional Tree Algorithm," Journal of Applied Sciences, ISSN 1812-5654, Asian Network for Scientific Information, (2010)

[27] M.Z. Jahromi, E Parvinnia, R. John, "A method of learning Weighted Similarity Function to improve the Performance of Nearest Neighbor," Information Sciences 179 2964–2973, Elsevier Inc, (2009)

[28] Q. Gao, Z.Z. Wang, "Center-based Nearest Neighbor Classifier," Pattern Recognition 40 346 – 349, Elsevier Ltd. (2007)

[29] H. C. Akakin, A. A. Salah, L. Akarun, B Sankur, "2D/ 3D Facial Feature Extraction," SPIE-IS&T, Vol. 6064 60641D-1, (2006)

[30] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) 971–987, (2002)