# FACIAL EXPRESSION RECOGNITION USING 2D GAUSS-HERMITE MOMENTS OF IMAGES

**By**

**Saif Muhammad Imran**

**M.Sc. Engg. (EEE)**

**Department of**

ELECTRICAL AND ELECTRONIC ENGINEERING

**BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

**August 2014**

The thesis titled '**Facial Expression Recognition using 2D Gauss-Hermite Moments of Images**' submitted by **Saif Muhammad Imran** Roll No:**P0412062238** Session **April/2012** has been accepted as satisfactory in partial fulfilment of the requirement for the degree of **Master of Science in Electrical and Electronic Engineering** on 11th August 2014.

## Board of Examiners

1._____

Dr. S.M. Mahbubur Rahman
Associate Professor
Department of Electrical and Electronic Engineering — Chairman
Bangladesh University of Engineering and Technology — (Supervisor)
Dhaka-1205, Bangladesh.

2._____

Dr.Taifur Ahmed Chowdhury
Professor and Head
Department of Electrical and Electronic Engineering — Member
Bangladesh University of Engineering and Technology — (Ex-Officio)
Dhaka-1205, Bangladesh

3._____

Dr. Md. Kamrul Hasan
Professor
Department of Electrical and Electronic Engineering — Member
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh.

4._____

Dr. Md. Haider Ali
Professor
Department of Computer Science and Engineering — Member
University of Dhaka — (External)
Dhaka-1000, Bangladesh

## Candidate's Declaration

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

_____

Saif Muhammad Imran

Roll: P0412062238

## ACKNOWLEDGEMENTS

I would like to convey my heartfelt gratitude to my thesis supervisor, Dr. S.M. Mahbubur Rahman for his all-out help and able guidance throughout the tenure of his supervision. His plentiful resources and pragmatic ideas in the field of computer vision have allowed me to work with exhaustive datasets and numerous untapped and untested theories in this field. He has always been a great mentor for a novice like me, and his appreciation and constructive criticism in every stage of this thesis, even for whatever insignificant reasons it may be, have only bolstered and instilled the confidence in me to successfully complete the dissertation. I would like to thank Dr. Kamrul Hasan for taking the Advanced Digital Signal Processing course and Dr. Sheikh Anawarul Fattah for taking the Biomedical Signal Processing course respectively, two of the most enjoyable courses that I encountered during my MSc degree. These courses opened the horizon for me to ultimately take signal processing as my core area of research. I am indebted to my parents without whose financial and mental support it would have not been possible to complete the thesis within such a short period of time. And finally, I would like to thank the Almighty the Lord of the Worlds for blessing me with so many respected personalities around me and this auspicious achievement, in particular.

## ABSTRACT

Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; in behavioural sciences and in clinical practices. Expression recognition can be considered to consist of deformations of facial parts and their spatial relations or changes in the pigmentation of the face. The challenge of such recognition lies in classifying expressions both in the case of posed and spontaneous forms; where the former is an intentional expression and the latter is natural. Two major approaches of facial expression recognition include the holistic and landmark. This thesis deals with holistic based feature extraction since such approach considers the entire face images instead of selecting a few interested areas using computationally expensive algorithms for locating landmarks. Commonly used feature extraction techniques in holistic expression recognition methods include the principal component analysis (PCA), the linear discriminant analysis (LDA) and their variants, independent component analysis and even using the orthogonal moments. However, most of the existing approaches fail to consider either the local inherent spatial changes of the facial expression e.g., PCA or LDA. Although orthogonal moments carry local information of facial regions, the previously proposed methods select higher order moments heuristically without any justification.

In this thesis, the Gauss-Hermite Moments (GHMs) are used for developing a holistic facial expression recognition algorithm since the GHMs are widely used in visual signal processing. Based on a novel concept of scattering ratio, moments are selected having higher discrimination power of expression in the GHM subspace. Further, due to the existence of significant correlations among certain expressions in the case of spontaneous form as compared to posed form, the GHM features are projected to a new expression subspace where the information are de-correlated using the PCA. Finally, these feature vectors are used to recognize the expressions using the well known support vector machine classifier. Experiments are carried out using two exhaustive databases, namely, the Cohn Kanade and Facial Recognition Grand Challenge, the former representing posed expressions while the latter spontaneous expressions. Experimental results on mutually exclusive subjects reveal that the proposed method can provide the recognition accuracies of at least 7 % and 4 % higher than the existing methods for posed and spontaneous expressions, respectively.

# TABLE OF CONTENTS

## LIST OF ACRONYMS

| | |
|---|---|
| CK | Cohn-Kanade |
| EGHM | Expressive Gauss-Hermite Moments |
| FACs | Facial Action Coded System |
| FRGC | Face Recognition Grand Challenge |
| GHBF | Gauss-Hermite Basis Functions |
| GHM | Gauss-Hermite Moments |
| GHP | Gauss-Hermite Polynomials |
| LDA | Linear Discriminant Analysis |
| MCM | Mixture Covariance Matrices |
| PCA | Principal Component Analysis |
| RBF | Radial Basis Function |
| ROC | Receiver Operating Characteristics |
| RTIGHM | Rotation and Translation Invariant Gauss-Hermite Moments |
| SFFS | Sequential Forward Feature Selection |
| SVM | Support Vector Machines |
| ZM | Zernike Moments |

## LIST OF SYMBOLS

| | |
|---|---|
| $H_p(x)$ | Hermite polynomials of order *p* |
| $\delta_{pq}$ | Kronecker Delta function of order p,q |
| $\widetilde{H_p}(x)$ | Normalized Hermite polynomials |
| $\sigma$ | Spread factor/Scale parameter for controlling textures and ridges in image |
| $I(x, y)$ | 2D continuous image |
| $I(i, j)$ | 2D digital image |
| $M_{pq}$ | GHM of order p,q |
| $K$ | Total number of pixels in an image |
| $N$ | Maximum order of GHM |
| $R$ | Length of RTIGHMs used to combat rotation and translation effects in patterns of image |
| $\tilde{I}(i, j)$ | Approximate image formed after image reconstruction |
| $(\bar{x},\bar{y})$ | Centroid coordinates normalized over the interval [-1,1] |
| $(i_0, j_0)$ | Centroid coordinates of the digital image |
| $T_{pq}$ | Translation invariant GHM |
| $\psi_i$ | Rotation invariant GHM |
| $\mathbf{M}^v$ | Moment vector formed by all the Gauss-Hermite moments upto order N |
| $\mathbf{\Psi}^v$ | RTIGHM vector formed by all the moments up to order *R*. |
| $\mathbf{X}$ | Feature vector formed by combining the GHM vector $\mathbf{M}^v$ and RTIGHM vector $\mathbf{\Psi}^v$ |

$\lambda^i$ Number of members in each class i

$C$ Total number of classes

$\mu_r^i$ Class mean of the $r$ -th moment of class $i$

$\mu_r$ Grand mean of the $r$ th moment of all classes

$(\sigma_r^i)^2$ variance of r-th moment for that particular class *i*

$L$ Length of X and is equal to $(N+1)^2 + R$

$l$ Length of vector in the neutral space

$S_r^B$ Interclass scatter for the r-th moment in X

$S_r^W$ Intraclass scatter for the r-th moment in X

$SR_r$ Scatter ratio of the r-th moment

$\mathbf{X}_S$ Sorted moment vector of a single expression of a particular class according to the scatter ratio

$\widetilde{\mathbf{X}}_s$ Input moment vector from forward selection method

$\Gamma$ Length of vector $\widetilde{\mathbf{X}}_s$

$\mathbf{F}s$ Feature vector obtained from scatter ratio and forward selection method

$\mathbf{X}_{Nt}$ Neutral expression training matrix of dimension ($n \times L$), where n is the number of neutral expression observations in the training set and *L* is the length of each feature vector

$\mathbf{C}$ Empirical covariance matrix obtained from $\mathbf{X}_{Nt}$

$\mathbf{u}$ Mean vector obtained from $\mathbf{X}_{Nt}$

$\mathbf{W}_{Nt}$ Orthonormal matrix of eigenvectors which diagonalize the covariance

matrix $\mathbf{C}$

$\mathbf{F}_{Nt}$ Projected moments of training set in neutral space. The matrix has a dimension of $(n \times l)$ where *n* is the number of observation and $l$ is the number of components selected in the neutral space

$\mathbf{X}_{Ex}$ All expressions of testing set matrix in the neutral space. The matrix has a dimension of$(m \times l)$

$\mu_{F_{Nt}}$ Mean vector obtained from $\mathbf{F}_{Nt}$ in the neutral space. The vector has a dimension of $(1 \times l)$

$\mathbf{F}_{Exp}^{\Delta}$ Feature vector obtained from differentially expressive moments

$\mathbf{F}_{S}^{\Delta}$ Proposed feature vector formed by combining two individual feature vectors $\mathbf{F}s$ and $\mathbf{F}_{Exp}^{\Delta}$

## LIST OF FIGURES

## LIST OF TABLES

# 1 : INTRODUCTION

## 1.1 Introduction

In the 21st century, deploying artificial intelligence in social and interpersonal communications has turned out into an intriguing and sizzling area of research. What just started as a monologue 'dancing and speaking robot' for entertainment in the late twentieth century, the concept of human computer interfaces has now made its way to interactive lie-detecting and person-recognizing apps in smart-phones and even went on to building mood-meters and recognizing emotions by judging every now-and-then activities of its 'master'. Today's researchers are exploring and developing new ways and algorithms to build more robust human computer interfaces from other characteristics of human; face, voices, pulse rate or even brain signals. The endeavor of this thesis is to build one such interactive system from facial expressions of humans.

## 1.2 Expression Recognition: A Review

### 1.2.1 Background

Facial expression is one of the most powerful means for humans to communicate their emotions, intentions, and opinions to each other. Psychological study reveals that while overall impact of the text content of a conversation is limited to only 7% and the intonation of the voice contributes by 38%, the facial expressions carry the most part of the conveyed information, i.e., 55% [1].Thus, automated facial expression recognition has emerged as a highly active field of research in cognitive or behavioural science, with important applications in lie detection [2], intelligent communication in social media [3], and multimodal human-computer interface [4,5].

Facial expressions are outcomes of complex inter-relations of emotions of a person. A person may have a single expression for his/her mixed feelings such as a combination of happiness and surprise or that of disgust and contempt. Also each of the basic emotions can have multiple sub-classes. For examples, joy has several sub-class emotions including the cheerfulness, zest, contentment, pride, optimism,

enthrallment and relief [6]. However, psychological studies reveal that basic emotions have corresponding universal facial expressions for all cultures. Most automatic facial expression analysis approaches found in the literature attempt to directly map facial expressions into one of the six basic emotion classes introduced by Ekman and Friesen [7]. The set of prototypic emotional expressions includes disgust, sad, fear, anger, surprise and happy. As a consequence, recent advances in image analysis and pattern recognition open up the possibility of automatic detection and classification of these six expressions of facial images [8].

### 1.2.2 Challenges

While the human visual mechanisms for face detection are very robust, the same is not the case for interpretation of facial expressions. According to Bassili [9], a trained observer can correctly classify faces showing six basic emotions with an average of 87 percent. The result varies depending on several factors such as the familiarity with the face, that with the personality of the observed person, the general experience with different types of expressions, the attention given to the face or even the non visual cues including the context in which the expression appears.

Although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge. Most of the research efforts for automatic recognition are given in achieving optimal pre-processing, feature extraction or selection, and classification, particularly under conditions of input data variability. To attain a good classification performance for expressions, a control over the imaging conditions has been tried. The controlling of imaging conditions is detrimental to the widespread deployment of expression recognition systems, because many real-world applications require operational flexibility. The challenging issues in real-world applications typically cover the following aspects:

- **Variations due to view or pose of the head:** The process of capturing of face images can have an important consequence on expression recognition. Orientation and pose of subjects' head relative to the camera, in-plane and out of plane rotation of the head and misalignment of facial features with respect to each other and inter subjects' variation of facial dimension can have

adverse impact on feature extraction. Some of the temporary remedial steps to counter these effects include several registration algorithms, centroid adjustment, geometric transformation etc. However, research is still needed for transformation invariant expression recognition.

- **Variations due to environment clutter and illumination**: Complex and uneven image background pattern, occlusion, and uncontrolled lighting variation (illumination and shadow) have a potentially negative effect on recognition. They may potentially cause the contamination of feature extraction by information not related to facial expression. Normally these variations are addressed by several image segmentation algorithms, illumination correction, enhancement algorithms (e.g. histogram equalization) etc. Uncluttered backgrounds and controlled illumination are often used during image capturing to counter the above mentioned effects.
- **Miscellaneous sources of facial variability**: Facial characteristics display a high degree of variability due to a number of factors, such as differences across people (arising from age, illness, gender, or race, for example), growth or shaving of beards or facial hair, make-up, etc. These variations can also have negative impact on expression recognition. Efforts to minimize these unnecessary artifacts still remain under investigation.

### 1.2.3 General Approach

Researchers follow a common approach of steps for research in facial expression analysis. The three major steps in face expression analysis include the face detection, face expression data extraction and expression classification. There are several robust ways to detect face regions from an image or video sequence. After face detection, a mechanism must be devised for extracting the deformed content of facial images due to expression from the observations, either in static image or image sequence. There are two kinds of approaches to capture the deformation: appearance based and geometric based approach. While geometric based approach relies on the feature extraction from facial dimension (several facial characteristic points are sought in faces and stringent geometric normalization are carried out to measure deformation in expression [10,11]); appearance based approach relies more heavily on face models of appearance to capture deformation [12,13].Appearance based methods do

have to rely on neutral face images or face models in order to extract facial features that are relevant to facial actions and not caused by non-expression related factors, e.g., intransient wrinkles due to aging. In this thesis, discussion is limited to appearance based approach since it is expected to be a more pragmatic and less cumbersome way of feature extraction classification.

Facial features in appearance based method can be classified as being permanent/intransient or transient. Permanent/intransient features are the features like eyes, lips, brows and cheeks which remain permanently but may be deformed due to facial expressions. Tissue texture, facial hair as well as permanent furrows constitute other types of permanent facial features. The transient features include facial lines, brow, wrinkles and deepened furrows that appear with changes in expression and disappear on a neutral face. The forefront and regions surrounding mouth and eyes are also prone to contain transient facial features. The *facial* features for expression recognition can be represented in various ways; e.g. face as a whole unit (holistic representation) or from certain fiducial points of the face (landmark representation). In landmark-based approach, fiducial points are localized at strategic blocks (cluster of neighboring pixels) of neutral face called Facial Action Units (FAUs) which display muscle movement with facial expression. The activities of FAUs are then measured in face image that occur with change in expression [14,15]. On the other hand, holistic spatial analysis is an approach that employs image dimensional gray-level texture filters on the entire portion of the face. Hybrid approach has also been adopted for facial expression recognition by applying local holistic approach on blocks of facial images defined by a given set of fiducial points [16].

Most of the researches are focused on 2D face expression recognition. However a few number of research investigation has been done on developing 3D face expression recognition. Normally in 3D based systems, the facial structure is described with the aid of 3D face models by trying to recover the 3D geometric information of the scene or facial surface from the 3D shape of the face, making it possible for accommodating high level description of facial activities [17,18]. There are two types of 3D models, namely muscle and motion models [19,20]. The latter

can significantly improve the precision of motion estimations, since only physically possible motion is considered. However, the 3D models often require complex mapping procedures that generate heavy computational requirements. Many 3D models are often generated by laser beam; but it must be noted that laser resolution and pixel resolution are not the same; it is not possible to acquire depth information for all pixels. In addition, widespread usage of laser beam for public experiments can pose serious health hazard issues. To avoid laser beams, accurate head and face models often are constructed manually, which is a tedious process.

It must be noted that there are two major types of expression categories: *posed* and *spontaneous*. Posed expressions are the artificial expressions that a subject produces on request through a guided mechanism. On the other hand, in spontaneous expressions, the subject shows their day-to-day on the spot expressions without any guidance. Many psychologists have proved that the posed expressions are different from spontaneous expressions, in terms of their appearance, timing, temporal characteristics or the extent of exaggeration [6]. Genuine expressions are usually subtle and the differences between these two categories necessitate the need for developing expression recognition systems that performs well both in the spontaneous and posed expressions.

### 1.2.4 Related Works

In this section, several methodologies found in the literature for landmark based approach and holistic based approach for face expression recognition are discussed. Reasons for selecting the latter approach over the former in this thesis are also explained in this section.

The main objective of the land-mark based approach is to measure the activities/deformation of the AUs by locating fiducial points in the facial image. The next step is to create quantifiable description of emotions by Facial Action Coded System (FACS), developed by Ekman and Friesen [21]. FACs was designed based on human observations to detect subtle changes in facial features. Classification is done based on recognizing the AUs first and then using FACs to classify the expressions. Based on this concept, several other variants have also developed. Essa

et al. developed an optical flow based system to recognize action units from facial motions [19]. Lanitis et al. [22] for example, have developed a Point Distribution Model (PDM) from which the shape of the main features and their spatial relationships can be extracted. A parameterized description of shape for any instance of face is found to locate the features in new images. A wire frame model with 12 facial motion measurements based on a 2D or 3D model proposed by Tao and Huang [23] is also used for facial expression recognition. However, in all these approaches, the data extraction process is computationally expensive. At the initial stage, for every subject, fiducial points have to be located manually in its neutral image which hinders automation and real time classification. Capturing localized patterns of facial expression brings sensitivity to noise and depend highly on pre-processed images since this kind of approaches require precise alignment of multiple internal features [24]. Also in constrained environments, several edge detectors, edge projection analysis and facial geometry distance measure have to be used prudently to extract the fiducial points accurately.

Bartlett. et. al. [24] found that the holistic spatial analysis and motion analysis, outperformed the landmark-based method for facial action recognition in several other criterions including accuracy. Many of the approaches in holistic based approach employ data-driven kernels learned from the statistics of the face image ensemble [25]. An alternative to landmark based image analysis, holistic analysis, emphasizes preserving the original images as much as possible and allowing the classifier to discover the relevant features in the images [13,25]. This is because deformations of facial features are usually characterized by shape and texture changes of face images that lead to high spatial gradients. In holistic spatial representation, ‘eigenfaces’ is based on the principal components of the image pixels [26,27]. The PCA finds an orthogonal set of dimensions that account for the principal directions of variability in the dataset. A low-dimensional representation of the face images with minimum reconstruction error is obtained by projecting the images onto the first few principal component axes. The PCA has been applied successfully to recognizing both facial identity [26,27], and facial expressions [28,29]. Another holistic spatial representation is obtained by a class-specific linear projection of the image pixels [30,31]. In such analysis, pre-processed images are subdivided into

several classes, and a set of images from each class is trained to seek a projection matrix that would maximally separate the between class while minimizing the within class variance. Gabor wavelet based filters [16,32,33], Local Binary Patterns [34], Local Directional Number Pattern [35] are some of the other local holistic based feature extraction methods where block based texture analysis of facial images are carried out. Feature extraction from holistic analysis has also been carried out by means of rotation and scale invariant orthogonal moments; Zernike [36], Tbechichef [37], Affine Moment invariants [38] and Krawtchouk Moments [39]. In all these approaches, the region of interest (face) is masked, the images are projected into the moment subspace, and then a set of moments are used as feature vectors which are finally used to train the classifiers.

A good reference point to start the study of facial expression analysis is how human visual system reacts to facial expression. A good characteristic of human visual system is that face is perceived as a whole, not as a collection of facial features [40]. The presence of the features and their spatial relationship with each other appears more important than the details of the features [41]. Moreover, holistic gray level information appears to play an important role in human face processing [42], and may therefore contain important information for face image analysis by computer as well. Holistic spatial analysis is also used in systems when configuration and shape can be often difficult to parameterize. Since marking and selecting fiducial points can often be cumbersome, instead of considering fiducial points individually, a good ploy can be to utilize the holistic approach benefits by taking algorithm specific blocks of facial images to extract features; often known as hybrid approach. However, it is expected that the feature set that performs well for the holistic approach will also show a similar performance for the feature sets extracted from FAUs of landmark based approach or localized blocks of the hybrid approach. It can be argued that the importance of face expression lies in choosing the feature set rather than the approach whether it is holistic or landmark-based or hybrid.

The final stage of any face expression recognition system is the classification stage. Several classifiers have been tested on extracted features. Some of the most effective and widespread classifiers to name are minimum distance classifiers [14,43], support

vector machine (SVM) [35,44], the naïve-Bayes classifier [45-46], artificial neural network [47]-[48] amongst the few.

## 1.3 Scope of Works

Existing holistic facial expression algorithms that extract features from the PCA, independent component analysis, LDA, mixture covariance matrixes (MCM) fail to address local facial deformation in the face image. Orthogonal moments such as Zernike, Tbechichef have also been tried as features with a view to capture the local spatial dynamics of the images. But these methods fail to recognize the discriminant information specifically due to the fact that the moments are chosen heuristically without any mathematical justification. To overcome these drawbacks a mathematical and experimental thorough study is necessary to develop an efficient orthogonal moment based facial expression algorithm. Among the existing various orthogonal moments, the two dimensional Gauss-Hermite Moments (GHM) can be chosen as initial face templates.In addition to its capacity to preserve information in compressed form, it is now being used widely for its efficient image reconstruction properties due to its orthogonality and geometrically invariant moment features. Also, it has the capacity to capture non-linear hidden structure of the data. It is also used popularly to model human visual signal processing, since the width of the Gaussian weight function of the Hermite polynomial expansion provides a flexibility in isolating the visual features just as human visual system does. After modelling the expression image in 2D GHMs, a feature selection strategy may be applied to extract relevant and effective feature for efficient expression classification.

## 1.4 Objective

With a view to developing a novel 2D GHM based face expression recognition algorithm, this thesis presents a detailed research study by maintaining the following specific objectives:

- To build a robust feature extraction technique by exploiting the orthogonal moments of static holistic expression images using the 2D Gauss-Hermite basis functions (GHBF) which are popular in visual signal processing.

- To improve the classification accuracy by developing a feature selection strategy that can select orthogonal moments having high discrimination capacity with a specific target for posed type expressions.
- To obtain expressive moment components from the prescribed neutral subspace to improve recognition accuracy in case of expression face images having highly correlated features (spontaneous expression).
- To investigate the proposed methodology in detail on several established datasets so that drawbacks and future modes of improvement can be highlighted.

In order to fulfil the above mentioned objectives, the 2D GHMs of static images are first estimated. After the moment estimation, a feature selection strategy is developed that can filter the more discriminant moments most representative of the expression subspace. Effectively, for posed expressions or images showing exaggerated expressions, using these selected moments in the feature space give a very good recognition rate. However, for spontaneous expressions of each class, which is correlated highly with the neutral expression, in addition to the developed feature vector, a neutral expression template is sought by projecting a set of neutral 2D GHMs of mutually exclusive subjects into a new feature space (called the neutral space) using PCA. It is expected that the interpersonal variation which is irrevelant and unnecessary in expression classification can also be filtered out in this feature space. Taking the virtual neutral expression subspace as reference, a difference component between neutral and expressive 2D GHMs is calculated. These difference components along with the discriminant moments (selected by scatter ratio) are then combined to form the feature vector of expressions of any class, which is referred to as the expressive GHM (EGHM). Finally these feature vectors are fed to a very reputed and online classifier, support vector machine (SVM), which is a widespread classifier for expression recognition.

The methodology is tested on exhaustive and well known posed expression database namely; the Cohn Kanade (CK) and the spontaneous database namely; the Facial Recognition Grand Challenge (FRGC). In each of the expression class, subjects are all mutually exclusive so that no two subjects fall in the same class. Since, personal

identity information conveyed by a face is an unwanted source of variability in expression recognition, this ensures person independency of the methodology and gives it a universal application. Performance of the proposed GHM based expression recognition algorithm is compared with the existing PCA, LDA, Zernike based methods by evaluating the recognition accuracy, type I and type II errors as well as Receiver Operating Characteristics (ROC) curves.

## 1.5 Outline

The thesis is organized as follows.

In Chapter 2, a brief review of 2D Gauss-Hermite moments, their reconstruction ability of images and the formation of rotation and translation invariance moments are given.

In Chapter 3, the proposed recognition scheme both for the posed and spontaneous expression is elaborated by the help of illustrations, block diagrams and mathematical expressions.

Chapter 4 presents the experimental results comparing the proposed method with the existing recognition methods.

Finally, Chapter 5 provides the conclusion along with the scopes for future work.

# 2 :GAUSS-HERMITE MOMENTS--A REVIEW

## 2.1 Introduction

Moments and functions of moments have been widely used in pattern recognition [49,50], edge detection [51,52], image segmentation [53], texture analysis [54] and other domains of image analysis [55,56] and computer vision [57,58].Among all kinds of moments, the geometric one is firstly introduced and has been used due to its simplicity and explicit geometric meaning. In 1961, Hu first introduced moment invariants by using nonlinear combinations of geometric moments [59].However, the basis functions of geometric moments are not orthogonal and the recovery of the image from these moments is very difficult [35]. In general, for image analysis, the bases of the space, onto which the image function are projected, would be desirable to be orthogonal. The advantage of using orthogonal basis is that the orthogonality between the bases helps in reducing the calculation. Moreover, the error is easier to estimate when a limited number of projection is used and reconstruction could also be more simple [55].In the early 1980s of the last century, Teague introduced the orthogonal Legendre and Zernike moments using the corresponding Legendre polynomials and Zernike polynomials as kernel functions for image analysis [35]. Since then, many authors in literature have proposed to use orthogonal moments [60], based on the theory of orthogonal polynomials such as Legendre [61] , Zernike [62] and Tchebichef moments [63] for image reconstruction.

As another kind of continuous orthogonal moments, the GHMs were firstly introduced by Shen [64]. The study of these moments is far from complete and mainly limited to their countable applications. Shen and Wu detected the moving objects using the moments [65,66,67]. Similarly, Wang and Dai[68]and Wang et al.[69]introduced the moments to the fingerprint classification in biometrics. Besides, other applications such as iris identification [70], SAR image segmentation [71] and stereo matching based on GHMs have recently been reported [72].Other applications include the character recognition [73], singular point detection of fingerprint images [74], recognition of 1D signals from iris images [75] and even in license plate analysis [76].

## 2.2 Image Representation by 2D Gauss-Hermite Moments

In this section, a brief review of Hermite polynomials and their orthogonality relations with the Gaussian weighting function is given. The method of obtaining the orthogonal Gauss-Hermite moments of the face images from the polynomials and the reconstruction of images from these moments are presented next.

### 2.2.1 Hermite Polynomials

The Hermite polynomial of order $p \in \mathbb{Z}^1$ on the real line $x \in \mathbb{R}^1$ is given by [77]

$$H_p(x) = (-1)^p \exp(x^2) \frac{d^p}{dx^p} \exp(-x^2) \tag{2.1}$$

These polynomials may be computed efficiently using the following recursive relations

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_{p+1}(x) &= 2xH_p(x) - 2pH_{p-1}(x) \qquad p \geq 1 \end{aligned} \tag{2.2}$$

The Hermite polynomials satisfy the orthogonality property with respect to the weight function $w(x) = \exp(-x^2)$ such that

$$\int_{-\infty}^{\infty} \exp(-x^2) H_p(x) H_q(x) dx = 2^p \, p! \sqrt{\pi} \delta_{pq} \tag{2.3}$$

where $\delta_{pq}$ is the Kronecker delta function. An orthonormal relation may be obtained by using a normalized version of the Hermite polynomials given by

$$\widetilde{H_p}(x) = \sqrt{2^p \, p! \sqrt{\pi}} \exp(-x^2/2) H_p(x) \tag{2.4}$$

A generalized version of (4) may be obtained by using a spread factor $\sigma$ $(\sigma > 0)$ on the real line $x \in \mathbb{R}^1$. In such a case, the so called generalized Gauss-Hermite (GH) polynomials may be written as

$$\widetilde{H}_p(x;\sigma) = \sqrt{2^p \, p! \sqrt{\pi}} \exp(-x^2/2\sigma^2) H_p(x/\sigma) \tag{2.5}$$

for which the orthonormal relation is maintained as

$$\int_{-\infty}^{\infty} \widetilde{H}_p(x;\sigma) \widetilde{H}_q(x;\sigma) dx = \delta_{pq} \tag{2.6}$$

### 2.2.2 Gauss-HermiteMoments

Let $I(x,y) \in L_2(\mathbb{R}^2)$ be a continuous square integrable 2D signal. The set of 2D GHMs of order $(p,q)(p \in \mathbb{Z}^1)(q \in \mathbb{Z}^1)$ denoted as $M_{pq}$ may be obtained from 2D GH basis functions (GHBF) expressed in terms of two independent *p-th* and q-*th* order of GH polynomials (GHP) using the following relations [77]

$$M_{pq} = \iint_{\mathbb{R}^2} I(x,y)\widetilde{H}_p(x;\sigma)\widetilde{H}_q(y;\sigma)dxdy \tag{2.7}$$

**Fig.2.1**showsfew shapes of several orders of 1D GHBFs. Each of the basis functions are orthogonal to each other.**Fig.2.2** shows2D GHBF obtained from the tensor product of two separate set of 1D GHP along *x* and *y* axes. The GHMs of the 2D signal may be considered as the projections of the signal onto these 2D basis functions. Thus, these moments characterize the signal at different spatial modes those are defined by certain linear combinations of the derivatives of the Gaussian functions.

#### 2.2.2.1 Capacity to capture local information in an image

GHM's base functions of different orders have different number of zero crossings and different shapes, so it is believed that they have the capacity to separate image features based on different modes, which is already a highly intriguing research in pattern analysis, shape classification and detection of moving objects. Moreover, the base functions of GHMs are much more smoothed; are less sensitive to noise and avoid the artifacts introduced by window function's discontinuity [78].

**Fig.2.1**clearly shows that the 1D modes have several zero crossings and with increase in order of modes, the density of the zero-crossings increase. **Fig.2.2** shows how the tensor product of the modes behaves in 2D form, which is the region of our interest now. The figure reveals that due to local patches of brightness at several areas of the 2D frame, it has the ability to extract local information from several regions and positions of the image. So GHMs can work as local feature extractor. It must be noted that modes and order convey the same meaning.

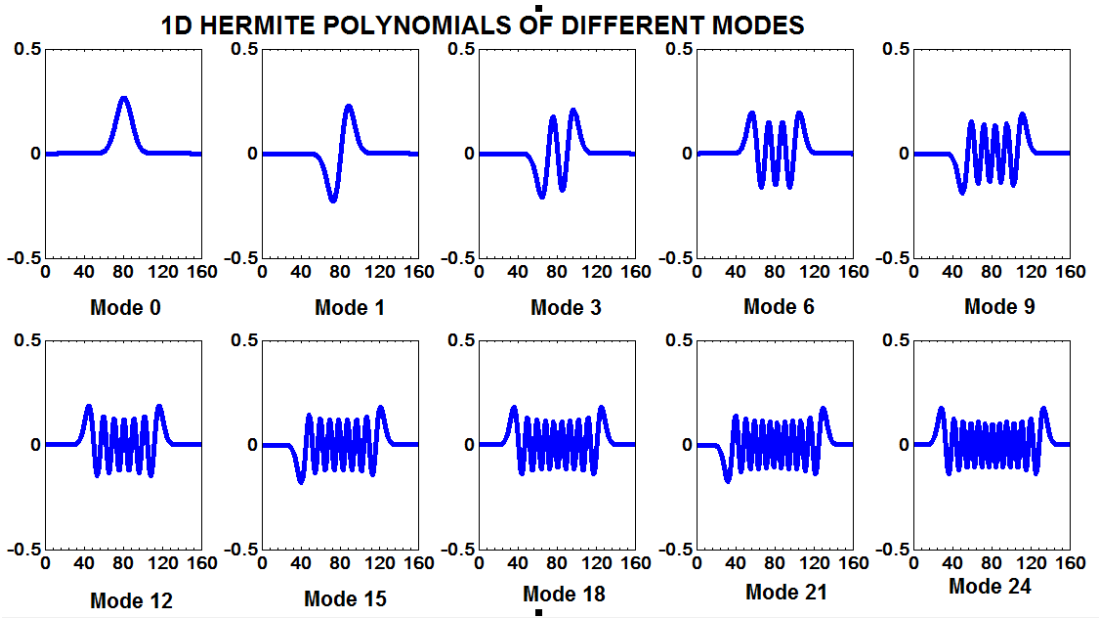
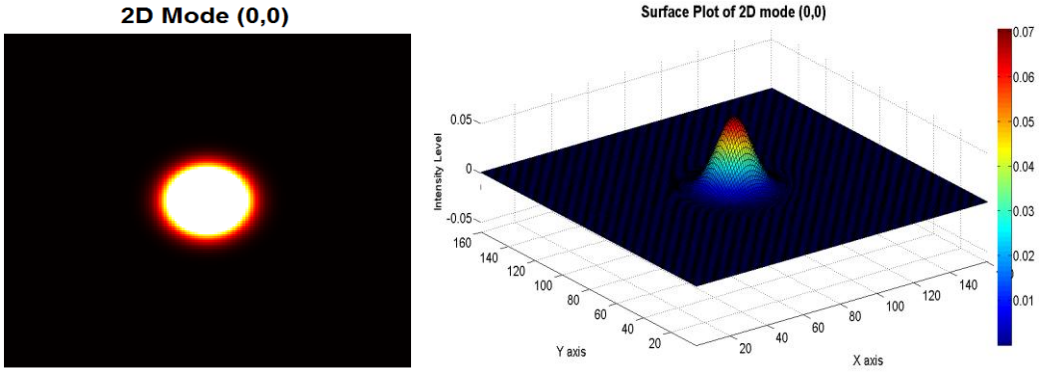
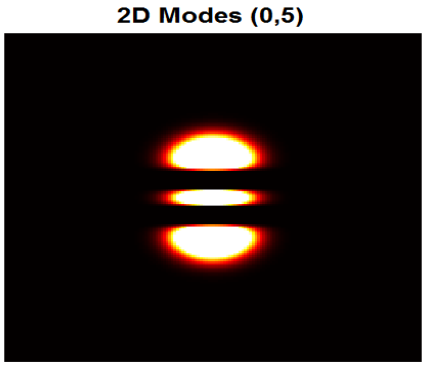


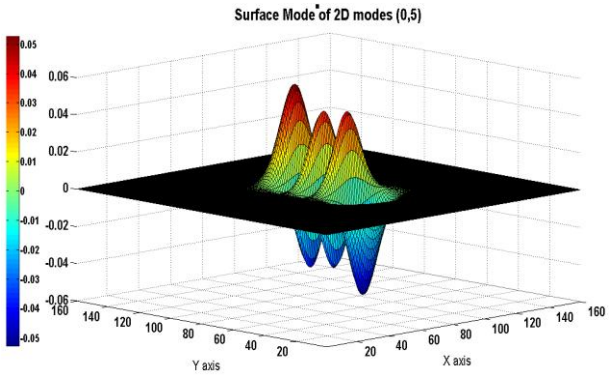

**Fig.2.1**: 1-D Hermite Polynomial of different modes.



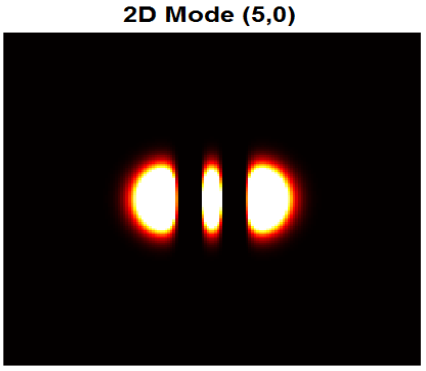

(a) (b)

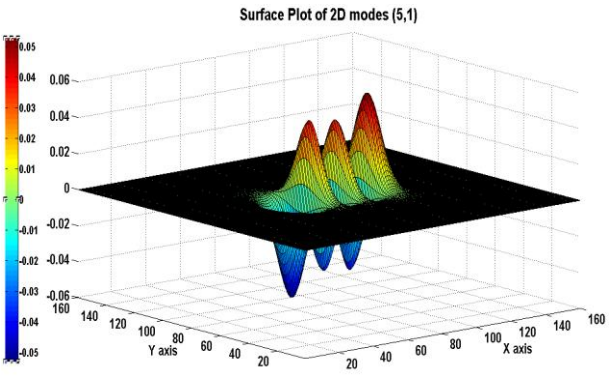
Figure continued on next page...



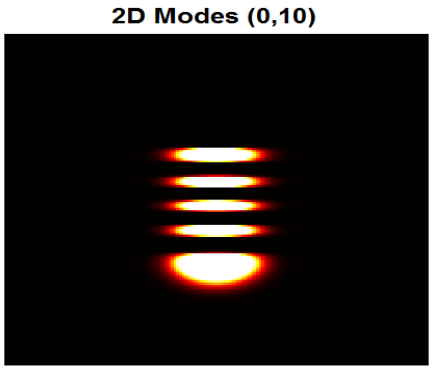



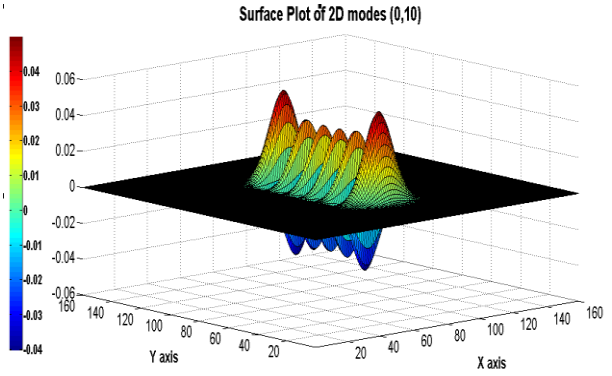

(c) (d)



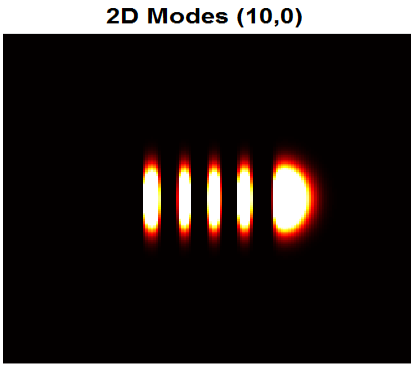



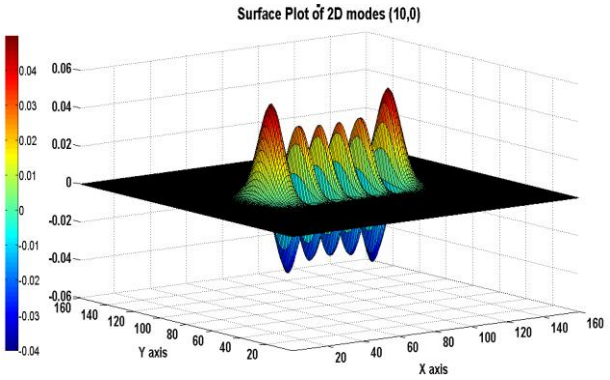

(e) (f)



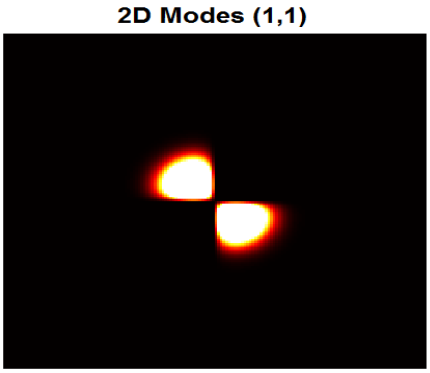

(g) (h)

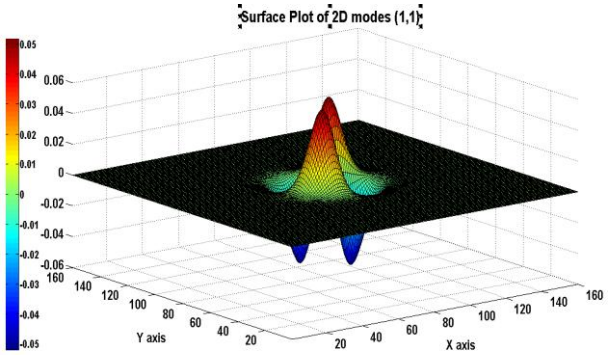


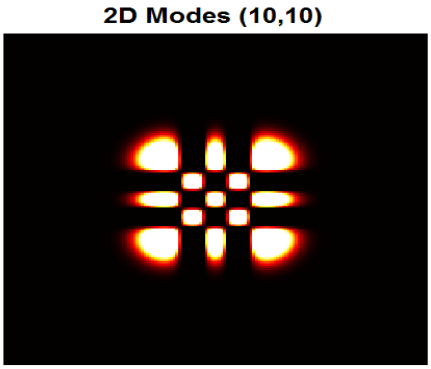

(i) (j)

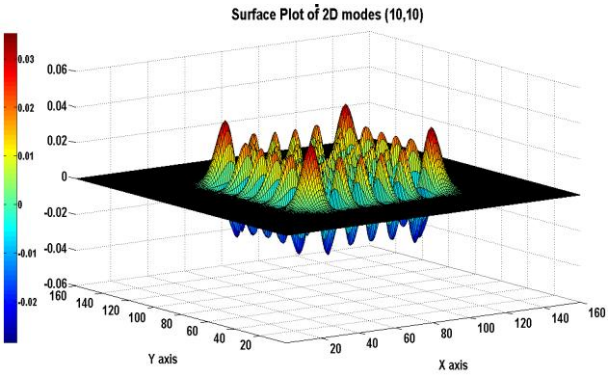
Figure Continued On Next Page...



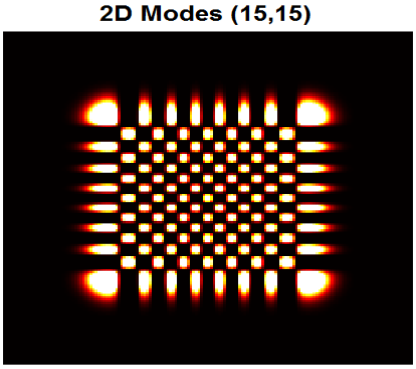



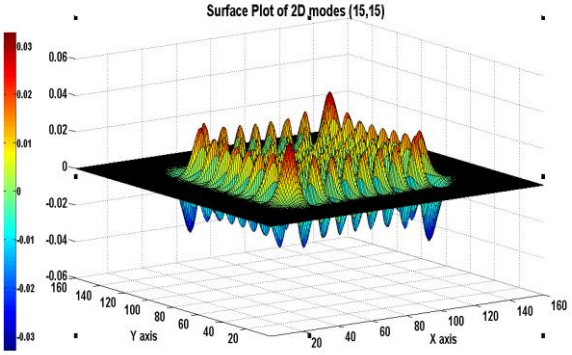

(k) (l)



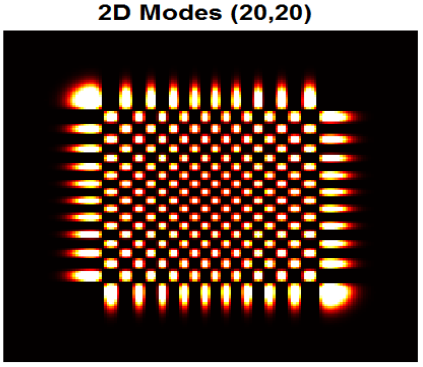



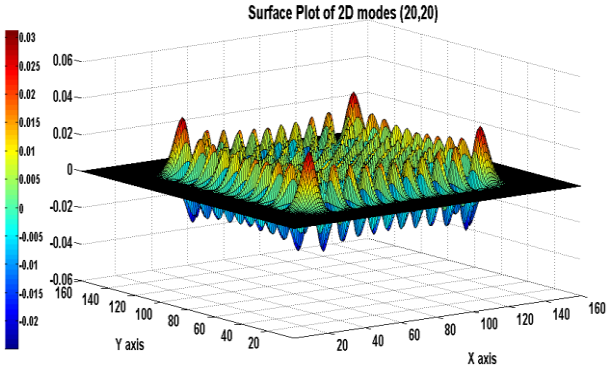

(m) (n)



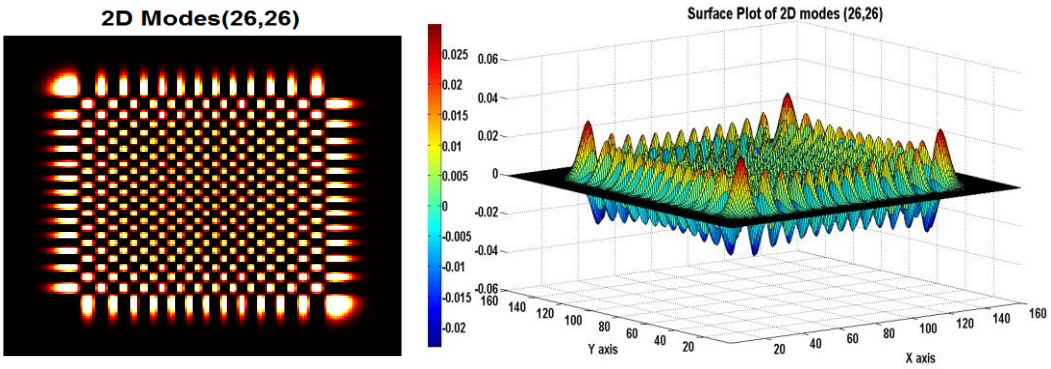

(o) (p)




(q) (r)

Figure continued on next page...




(s) (t)

**Fig.2.2**: 2D Gauss-Hermite Polynomials of different modes. The intensity plot is shown in the left (a,c,e,g,i,k,m,o,q,s) and corresponding surface plot is shown in the right (b,d,f,h,j,l,n,p,r,t)).

### 2.2.2.2 Reconstruction Capacity of GHMs

An important aspect of GHM is the ability to reconstruct images based on its moment descriptors. GHMs have a high compression capacity which allows storing the image efficiently in reduced dimension. This helps to avoid the 'curse of dimensionality' problem often faced by storing higher resolution gray level images. **Fig.2.3**helps describe how the moments can be again used to reconstruct the image in its gray level intensity.

Ideally, from all possible moments, the 2D signal *I(x,y)* may be reconstructed without any error as

$$I(x,y)=\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}M_{pq}\widetilde{H}_p(x;\sigma)\widetilde{H}_q(y;\sigma) \tag{2.8}$$

It is to be noted that the GHMs are obtained from two real lines $x\in\mathbb{R}^1$ and $y\in\mathbb{R}^1$, and hence a modification is required for obtaining moments from the discrete coordinates of face images.

In order to compute the moments for a digital image *I(i,j)* whose size is $K\times K$ ,the normalized coordinates over the square image $[-1\le x,y\le 1]$ is recommended for a comparable evaluation of $\sigma$ selection:

$$x=\frac{2i-K+1}{K-1}\text{ and }y=\frac{2j-K+1}{K-1} \tag{2.9}$$

The discrete version of Gauss-Hermite polynomial $\widetilde{H}_p(i,K;\sigma)$ is computed on the interval *[-1,1]*.It is in fact the equidistant sampling as a substitute for the continuous Gauss-Hermite polynomial:

$$\widetilde{H}_p(i,K;\sigma) = [2^p\, p!\sqrt{\pi}\sigma]^{-1/2} \exp(-x^2/2\sigma^2) H_p(x/\sigma)$$
$$\widetilde{H}_q(j,K;\sigma) = [2^q\, q!\sqrt{p}\sigma]^{-1/2} \exp(-y^2/2\sigma^2) H_q(y/\sigma) \quad (2.10)$$

In terms of discrete implementation, the 2D moments for the face images may be obtained as

$$M_{pq} = \frac{4}{(K-1)^2} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} I(i,j) \widetilde{H}_p(i;\sigma) \widetilde{H}_q(j;\sigma) \quad (2.11)$$

And the image can be restored from its Gauss Hermite moments of order *(0,0)*upto*(N,N)*by

$$\tilde{I}(i,j) = \sum_{p=0}^{N} \sum_{q=0}^{N} M_{pq} \widetilde{H}_p(i,K;\sigma) \widetilde{H}_q(j,K;\sigma) \quad (2.12)$$

Note that the (a) in **Fig.2.3**represents the actual image and all the other images show the effect of reconstructing the images from low order (15, 15) GHMs to high order (50, 50) GHMs.

The biggest motivation obtained from this quality of GHM is that PCA and LDA can also be used to reconstruct the gray level image, and since both the algorithms have been used in a wide scale in holistic image analysis, GHMs might also have the ability to extract features from holistic image analysis.

A crucial choice for obtaining GHMs is the scale parameter $\sigma$ ($\sigma > 0$) that has to be selected before moment computation and image reconstruction. Generally speaking, given the same moments for image reconstruction, a large $\sigma$ can entertain more smooth regions and texture in the image, while a smaller $\sigma$ can capture very minute details of an image. Therefore, in order to have a better image reconstruction, we should take the influence of $\sigma$ into account and make a reasonable trade-off between smooth and detailed regions. Since the expression image may contain wrinkles or creases in forehead, we have kept $\sigma$ to fall within 0.1.

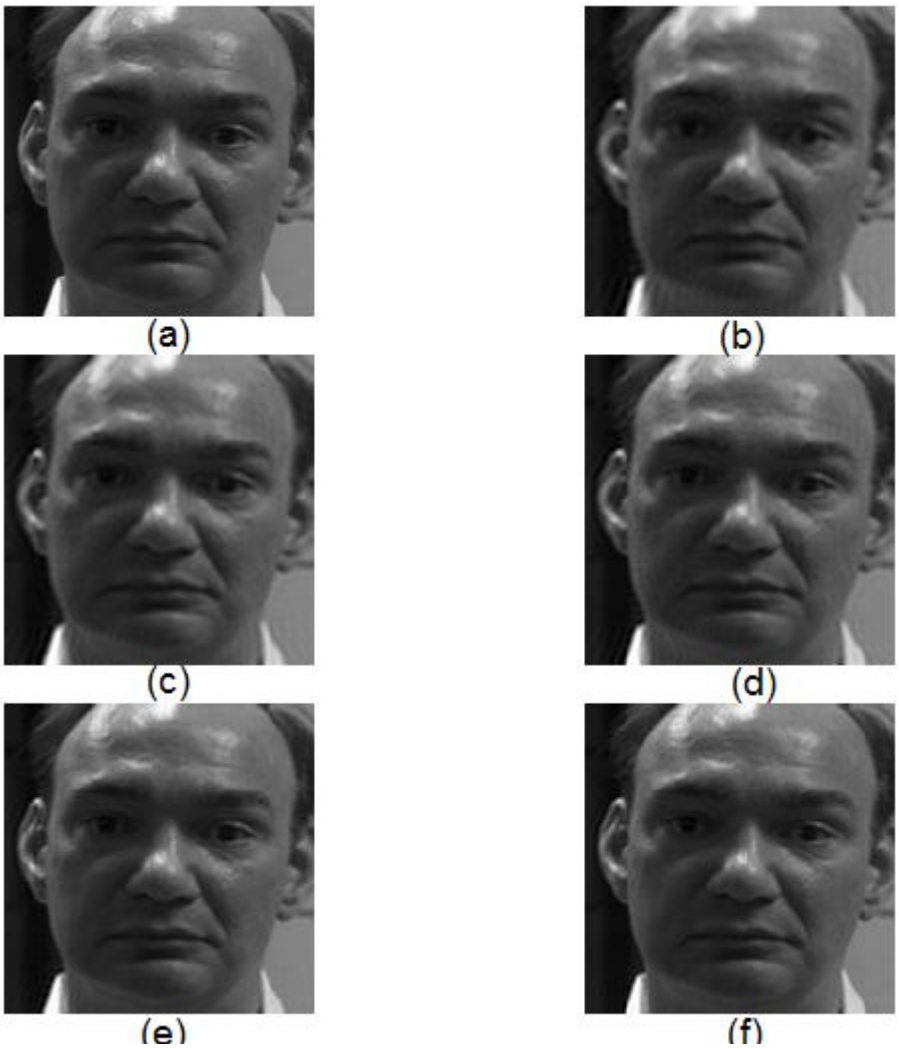



**Fig.2.3**: Image reconstruction capacity with GHM descriptors (a) represents the actual image of the subject, while (b) (c) (d) (e) and (f) represent the image reconstructed by lower to higher order of moments 15, 20, 25, 40, 50 in that order respectively.

## 2.3 Translation and Rotation Invariants of Gauss-Hermite Moments

Successful feature extraction from facial images may suffer due to misalignment and several postures of the subjects. In this section, our main objective is to find several combined translation and rotation invariant GHMs (RTIGHMs) of facial images that could be used to provide some robustness to the overall system of feature extraction.

### 2.3.1 Translation Invariant Moments

The chief property of translation invariance is eliminating the effect due to the subtle change of location of an image on moment computation. The chief step is the computation of Gauss-Hermite polynomials with translation of origin to the centroids in discrete case, which can be expressed by

$$\begin{aligned}\overline{H}_p(i,K;\sigma) &= [2^p\, p!\sqrt{\pi}\sigma]^{-1/2} \exp(-(x-\bar{x})^2 / 2\sigma^2) H_p((x-\bar{x})/\sigma) \\ \overline{H}_q(j,K;\sigma) &= [2^p\, p!\sqrt{\pi}\sigma]^{-1/2} \exp(-(y-\bar{y})^2 / 2\sigma^2) H_q((y-\bar{y})/\sigma)\end{aligned} \tag{2.13}$$

where $(\bar{x},\bar{y})$ are centroid coordinates normalized over the interval [-1,1] and defined in the digital image by

$$\begin{aligned}\bar{x} &= (2i_0 - K + 1)/(K-1) \\ \bar{y} &= (2j_0 - K + 1)/(K-1)\end{aligned} \tag{2.14}$$

Here $(i_0, j_0)$ is the centroid coordinates of the digital image *I(i,j):*

$$\begin{aligned}i_0 &= \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} i \bullet I(i,j) \Big/ \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} I(i,j) \\ j_0 &= \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} j \bullet I(i,j) \Big/ \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} I(i,j)\end{aligned} \tag{2.15}$$

Gauss-Hermite translation invariant moments can then be computed with the subsequent equation:

$$T_{pq} = \frac{4}{(K-1)^2} \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} I(i,j)\overline{H}_p(i,K;\sigma)\overline{H}_q(j,K;\sigma) \tag{2.16}$$

### 2.3.2 Rotation Invariant Moments

The main reason for using the rotation invariance forms of the Gauss-Hermite Moments is because during face acquisition, many subjects could have different poses and the subject photos could have been taken at different camera angles. We note that the extent to which rotation invariance moments could work for different poses is still not yet known and not under the purview of this thesis.

The rotation invariant moments upto fifth order is calculated in this thesis. The higher order moments could also be derived, but it is found that they are more vulnerable to noises and variations of image intensity function [60,79]; also the basis of lower order is the most used in practice. So, as recommended by [77], in total 18

moments (moments upto 5th order)are found to have significant rotation invariance properties.

The following equations are the formula to calculate the rotation invariant moments as given by [77].

Second and third order Moments:

$$\psi_1 = M_{20} + M_{02} \tag{2.19}$$

$$\psi_2 = (M_{30} + M_{12})^2 + (M_{21} + M_{03})^2 \tag{2.20}$$

$$\psi_3 = (M_{20} - M_{02})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] + 4M_{11}(M_{30} + M_{12})(M_{21} + M_{03}) \tag{2.21}$$

$$\psi_4 = M_{11}[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] - (M_{20} - M_{02})(M_{30} + M_{12})(M_{21} + M_{03}) \tag{2.22}$$

$$\psi_5 = (M_{30} - 3M_{12})(M_{30} + M_{12})[(M_{30} + M_{12})^2 - 3(M_{21} + M_{03})^2] + (M_{03} - 3M_{21})(M_{21} + M_{03})[(M_{21} + M_{03})^2 - 3(M_{30} + M_{12})^2] \tag{2.23}$$

$$\psi_6 = (M_{30} - 3M_{12})(M_{21} + M_{03})[(M_{21} + M_{03})^2 - 3(M_{30} + M_{12})^2] + (3M_{21} - M_{03})(M_{30} + M_{12})[(M_{30} + M_{12})^2 - 3(M_{21} + M_{03})^2] \tag{2.24}$$

Fourth Order Moments:

$$\psi_7 = M_{40} + 2M_{22} + M_{04} \tag{2.25}$$

$$\psi_8 = (M_{40} - M_{04})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] + 4(M_{31} + M_{13})(M_{30} + M_{12})(M_{21} + M_{03}) \tag{2.26}$$

$$\psi_9 = (M_{31} + M_{13})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] - (M_{40} - M_{04})(M_{30} + M_{12})(M_{21} + M_{03}) \tag{2.27}$$

$$\psi_{10} = (M_{40} - 6M_{22} + M_{04})[(M_{30} + M_{12})^4 - 6(M_{30} + M_{12})^2(M_{21} + M_{03})^2 + (M_{21} + M_{03})^4] + 16(M_{31} - M_{13})(M_{30} + M_{12})(M_{21} + M_{03})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] \tag{2.28}$$

$$\psi_{11} = (M_{40} - 6M_{22} + M_{04})(M_{30} + M_{12})(M_{21} + M_{03})[(M_{21} + M_{03})^2 - (M_{30} + M_{12})^2] + (M_{31} - M_{13})[(M_{30} + M_{12})^4 - 6(M_{30} + M_{12})^2(M_{21} + M_{03})^2 + (M_{21} + M_{03})^4] \tag{2.29}$$

Fifth Order Moments:

$$\psi_{12} = (M_{50} + 2M_{32} + M_{14})^2 + (M_{41} + 2M_{23} + M_{05})^2 \tag{2.30}$$

$$\psi_{13} = (M_{50} + 2M_{32} + M_{14})(M_{30} + M_{12}) + (M_{41} + 2M_{23} + M_{05})(M_{21} + M_{03}) \tag{2.31}$$

$$\psi_{14} = (M_{41} + 2M_{23} + M_{05})(M_{30} + M_{12}) - (M_{50} + 2M_{32} + M_{14})(M_{21} + M_{03}) \quad (2.32)$$

$$\psi_{15} = (M_{50} - 2M_{32} - 3M_{14})[(M_{30} + M_{12})^3 - 3(M_{30} + M_{12})(M_{21} + M_{03})^2] - (3M_{41} + 2M_{23} - M_{05})[(M_{21} + M_{03})^3 - 3(M_{21} + M_{03})(M_{30} + M_{12})^2] \quad (2.33)$$

$$\psi_{16} = (M_{50} - 2M_{32} - 3M_{14})[(M_{21} + M_{03})^3 - 3(M_{21} + M_{03})(M_{30} + M_{12})^2] + (3M_{41} + 2M_{23} - M_{05})[(M_{30} + M_{12})^3 - 3(M_{30} + M_{12})(M_{21} + M_{03})^2] \quad (2.34)$$

$$\psi_{17} = (M_{50} - 10M_{32} + 5M_{14})[(M_{30} + M_{12})^5 - 10(M_{30} + M_{12})^3(M_{21} + M_{03})^2 + 5(M_{30} + M_{12})(M_{21} + M_{03})^4] + (5M_{41} - 10M_{23} + M_{05})[(M_{21} + M_{03})^5 - 10(M_{30} + M_{12})^2(M_{21} + M_{03})^3 + 5(M_{21} + M_{03})(M_{30} + M_{12})^4] \quad (2.35)$$

$$\psi_{18} = (M_{05} - 10M_{23} + 5M_{41})[(M_{30} + M_{12})^5 - 10(M_{30} + M_{12})^3(M_{21} + M_{03})^2 + 5(M_{30} + M_{12})(M_{21} + M_{03})^4] - (5M_{14} - 10M_{32} + M_{50})[(M_{21} + M_{03})^5 - 10(M_{30} + M_{12})^2(M_{21} + M_{03})^3 + 5(M_{21} + M_{03})(M_{30} + M_{12})^4] \quad (2.36)$$

### 2.3.3 Combination of rotation and translation invariants

The combination of rotation and translation invariants yields a kind of GHM invariants which is independent of transformations involving both rotation and translation. Replacing $M_{pq}$ in all the Eqs [2.19-2.36] by $T_{pq}$ in Eq [2.16] we obtain the combined rotation and translation invariants $\psi_i$, upto the fifth order. The rotation and translation invariant moments (RTIGHM) are expected to withstand the displacement of the images, if it does exist. The calculated moments are then concatenated with the original projected moment vectors to obtain the final Gauss-Hermite feature vectors.

## 2.4 Conclusion

The above properties of GHMs indicate that these moments have some very interesting attributes to use them as features for facial expression recognition. Aside from the capacity to preserve local information in images, it also possesses orthogonal, translation, rotation and scale invariant properties that can be used to our advantage for feature extraction.

# 3 : EXPRESSION RECOGNITION USING GHMS

## 3.1 Introduction

In Chapter 1, a general description of facial expression recognition and some of the related works in this field are explained. Chapter 2 goes more specific by discussing the literature review of GHMs in pattern recognition and how it can be utilized for efficient feature extraction. This chapter delineates the proposed methodology and describes how it best represents the expression characteristics by utilizing both the localized spatial information and distinguishing moments of interest amongst the several expression classes at the same time. The chapter highlights some of the concept diagrams and mathematical expressions that motivate us to formulate this proposition. At the end of the chapter, we kept scatter plot of typical members of expression classes for both the databases to illustrate the effectiveness of our method in expression recognition.

## 3.2 Feature Extraction and Selection

For each expression, the initial moment vector is formed by all the GHMs upto order $N$ where $p,q = 0,1,2,\cdots,N$ and the rotation and translation invariant GHMs (RTIGHM) described in Chapter 2. We consider RTIGHMs of length $R$. If $\mathbf{x}$ becomes the initial moment vector, $M_{pq}$ be the GHM of order $p,q$ and $\Psi_i$ be the$i^{th}$RTIGHM, $\mathbf{X}$ can be written as follows:

$$\mathbf{x} = [\mathbf{M}^v \ \mathbf{\Psi}^v] \tag{3.1}$$

where

$$\mathbf{M}^v = [M_{00}, M_{01}, \cdots, M_{0N}, M_{10}, \cdots, M_{0N}, M_{11}, \cdots, M_{NN}] \tag{3.2}$$

and

$$\mathbf{\Psi}^v = [\Psi_1, \Psi_2, \cdots, \Psi_R] \tag{3.3}$$

Here $\mathbf{M}^v$ and $\mathbf{\Psi}^v$ denote the normal GHM vector and RTIGHM vectors respectively. The total length of moment vector in $\mathbf{x}$ is $L = (N+1)^2 + R$.

From this feature space in GHM domain, the work of the thesis is now to build effective feature space that can recognize expression with improved ability.

The major contribution of this thesis is two fold. First, we propose a feature space for expression recognition where the moments of higher discrimination amongst all the moments in $\mathbf{X}$ of the given expression classes are selected. Based on the discrimination power, the moments are sorted and around two hundred to two fifty sorted moments are then fed to Sequential Forward Feature Selection (SFFS), a wrapper approach of feature selection that select only few features based on improved classification accuracy. By this step, feature dimension is set to reduce by a wide margin.

Second, we propose to establish a neutral subspace within the GHM space which can be exploited by extracting the differential information amongst all the expression in the neutral space. It is believed that this information can be used as an added advantage to find the minute details of expression indistinguishable even by human visualization. In this section, the concept of feature extraction procedure is discussed in two separate subsections.

### 3.2.1 Discriminant Analysis Using Scattering Ratio

Discriminant analysis is a statistical technique to classify objects into mutually exclusive and exhaustive groups based on a set of measurable object's features. The term discriminant analysis comes with different names with different fields of study. It is often called pattern recognition, supervised learning or supervised classification.
In the proposed method, an off-line automatic learning method is adopted to select discriminant GHMs from the GHMs of the entire training data. Generally the concept of scattering is based on the interclass and intra class distance of separation amongst the member classes. We propose a customization of both the between class and within class scatter distances to make it fit to our method. Since GHBF are orthogonal and independent, we consider the mean and variances of each of the moments independently and utilize these statistics for finding out the distinguishing information. The following figure gives a concept diagram of the between class and within class distance of separation amongst the member classes.

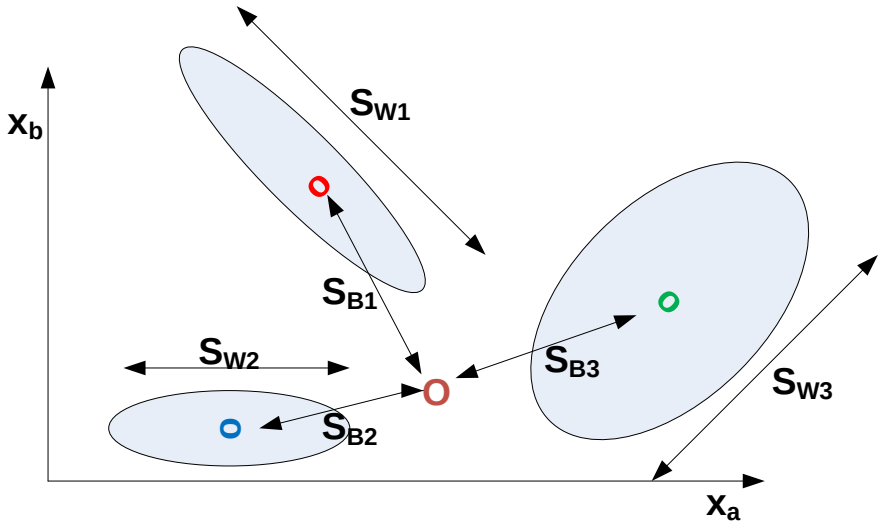



**Fig.3.1**: Aschematic diagram of within and between scatter distances.

In**Fig.3.1**, it is seen that there are three member classes in the feature space. Different sizes and shape of the ellipse indicate the degree of variability of the feature of its members in the respective classes. Each of the ellipses generates around the mean of each class. Here '**o**' marker indicates the mean of each class, while '**O**' marker represents the grand mean of all the classes. The distance between '**o**' and '**O**' is termed as the interclass distance, also referred to as between class scatter $\mathbf{S_B}$, while the degree/distance of separation of each member from its class is parameterized by intra class distance, or within-class scatter $\mathbf{S_W}$.

In the proposed method, the between class or interclass scatter for a moment of a particular order is obtained as:

$$S_r^B = \sum_{i=1}^{C} \lambda^i \mid \mu_r^i - \mu_r \mid r = 0,1,2,\cdots,L \tag{3.4}$$

where $\lambda^i$ is the number of members in each class, $L$ the length of features in the initial moment space **x**. $\mu_r^i$ and $\mu_r$ are the class mean and grand mean respectively. The grand mean is calculated by taking the weighted average of the class mean of each class, and is defined as follows:

$$\mu_r^i = \sum_{i=1}^{C} \lambda^i \mu_r \Big/ \sum_{i=1}^{C} \lambda^i \tag{3.5}$$

The within class or intra-class parameter $S_r^W$ of the order of *r* is defined as:

$$S_r^W = \sum_{i=1}^{C} \sum_{j=1}^{\lambda^i} | (x_r^i(j))^2 - (\sigma_r^i)^2 | \tag{3.6}$$

$x_r^i(j)$ is the moment of order *r* of class i, and jth member of ith class, and $(\sigma_r^i)^2$ is the variance of the moment of order *r* for that particular class *i*. The variance is calculated by

$$(\sigma_r^i)^2 = (x_r^i - \mu_r^i)^2 / (\lambda^i - 1) \tag{3.7}$$

which is an unbiased estimator of the variance of population from which $x_r^i$ is drawn.

There is justification for using the mean and variance in calculating the between class scatter and within class scatter respectively. It is apparent that the distance between variance of all the groups/classes (grand variance) and the variance of within groups/classes (group variance) does not provide a clear picture of the inter-distance boundary and the location of each class in the feature space while calculating the between class scatter. On the other hand, the major reason for using variance in calculating the within-class scatter is because due to large number of members in each class, it is expected that the discriminant information is in the variance of the members instead of the class mean. Variance measures the spread or variability of the distribution and calculating the distance of samples from this parameter might give a better idea of the stretchness of the data. So we use the variance for calculating the within class scatter.

The scatter ratio of all the moments for a given training expression GHM set is obtained as

$$SR_r = \frac{S_r^B}{S_r^W}, r = 1, 2, \cdots, L \tag{3.8}$$

Now, the objective of this scatter ratio is to seek for features that maximize this ratio. This means that the algorithm wants only those features that can maximize the between class scatter distance (signifying that the class mean should be as far as possible from the grand mean) and minimize the within class scatter distance (signifying minimizing the compactness/spreadness of the data so that overlapping

with another class can be reduced) at the same time. As a result, the $SR_r$ elements are arranged in descending order.

Let $\mathbf{x_S}$ be the sorted 1D feature vector of a single expression of a particular class according to the descending order of $SR_r$, containing *L* number of moments. From the sorted $\mathbf{x}_S$, we select features upto a particular threshold after which taking any further features does not bear any more significance. We define the finally selected feature vector as $\tilde{\mathbf{x}}_s$ where

$$\tilde{\mathbf{x}}_s \leftarrow \mathbf{x}_s\{1,2,\cdots,\Gamma-1,\Gamma,\Gamma+1,\cdots,L\} \tag{3.9}$$

$\Gamma$ is the least of the selected feature numbers in $\tilde{\mathbf{x}}_s$ from $\mathbf{x}_S$ which can be used as selection threshold.

Such an arrangement of the moments in $\mathbf{x}_s$ ensures that only the first few elements of moments would be sufficient for obtaining appreciable accuracy in expression recognition.

#### 3.2.1.1 Forward Feature Selection

After the feature vector $\tilde{\mathbf{x}}_s$ is formed, we now apply SFFS algorithm to further reduce the dimensionality of the features. Proposed by Pudit et al. [80], SFFS algorithm is gaining widespread attention due to its contribution to substantial improvement in classification accuracy [81]. It must be noted that SFFS is an offline learning algorithm that is performed prior to feeding the features on the classifier. As a result, real time application is no way affected by this procedure. Instead of feeding the entire feature vector $\mathbf{X}_S$, we feed $\tilde{\mathbf{x}}_s$ since SFFS is a cumbersome process; even though it is an offline method, it can take significant amount of learning time if it has to select from huge number of features. So instead of $\mathbf{X}_S$, we give the truncated feature set $\tilde{\mathbf{x}}_s$ to SFFS for learning. Feature vector selected after SFFS is given by:

$$\mathbf{F}s = FSelect(\tilde{\mathbf{x}}_s) \tag{3.10}$$

SFFS is performed on the feature vector $\tilde{\mathbf{x}}_s$ based on a convenient fraction of the rest of the dataset (other than the training set used for forming the scattering ratio).The feature element with the maximum value of scattering ratio is first chosen. New

features from $\tilde{\mathbf{x}}_s$ are then added to a forward search procedure if they improve the classification rate. Feature that worsens the classification or does not change it, given the ones that are already selected, are not included. From the SFFS algorithm, only those elements from $\tilde{\mathbf{x}}_s$ are preserved that contribute to significant improvement in recognition accuracy and only those elements are used to form the feature space $\mathbf{F}_S$. These features are expected to boost recognition accuracy to a significant extent.

#### 3.2.1.2 Block Diagram of Feature Extraction from the Scattering Ratio

**Fig.3.2** gives an overall view of the steps used to extract features from the scattering ratio. It must be noted that the scatter ratio is calculated from training set. After calculating the scatter ratio, we scale the features according to the scatter ratio and then sort the features in descending order. Finally by forward feature selection, we select only the features that will be finally used for feeding the classifier.

### 3.2.2 Differential Expressive Moments from Neutral Subspace

In posed expressions as discussed in Chapter 1, the expressions from each class are highly exaggerated, and as a result, there are no significant correlations amongst the classes themselves. The moments of each class preserve unique information of each expression class by virtue of its orthogonality and independent properties, and so they can be easily selected by means of scattering ratio. But for spontaneous expressions, it is often found that differences between each expression class are very subtle and often hard to distinguish even by human visualization. There exist significant similarity between each class which is hard to separate by orthogonal projection of moments alone.

In such scenarios, we propose a new feature vector in addition to $F_S$ that can extract the difference information amongst the expression classes by decorrelating the features and utilizing the difference for improved recognition accuracy. In this thesis, the primary importance is given in decorrelating information between neutral and all other expressions since all expressions ensue from its corresponding neutral offset.

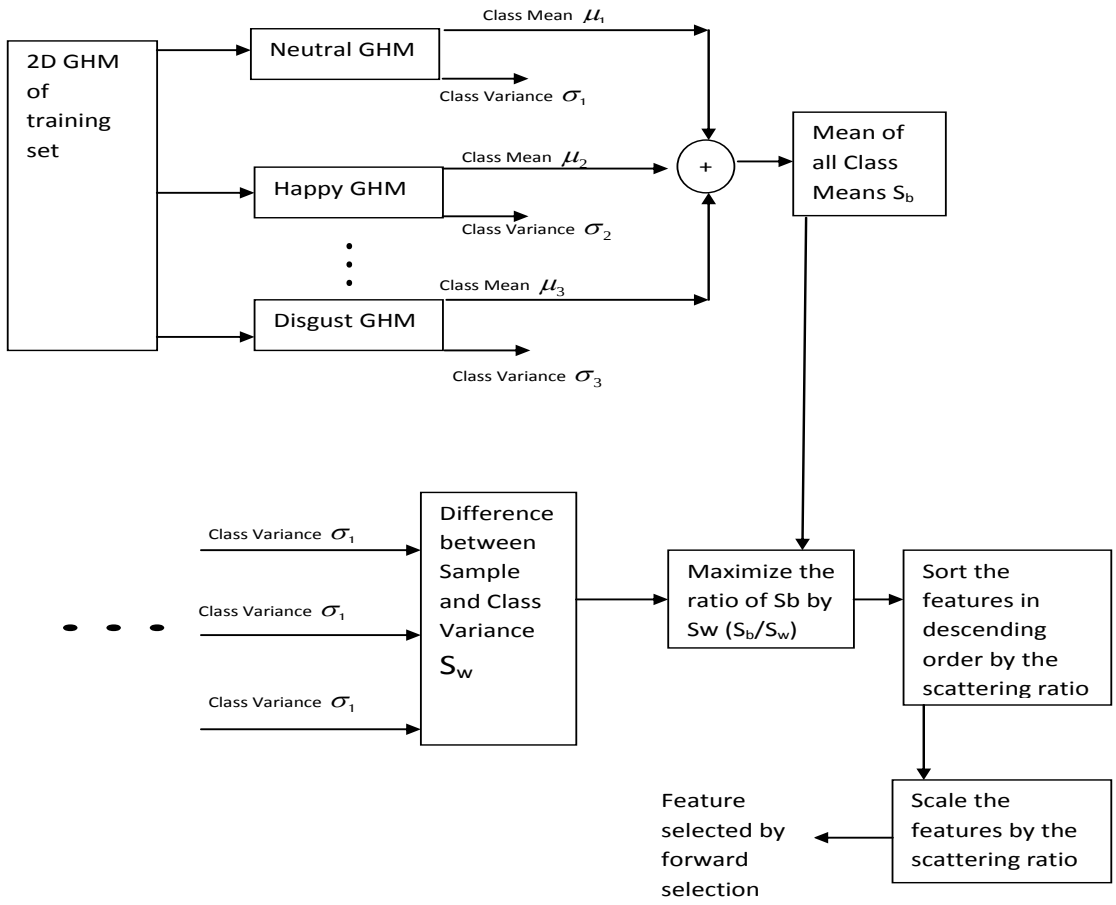



**Fig.3.2**: Block diagram of feature extraction method by novel scattering ratio.

We build a neutral reference space by performing principal component analysis on the neutral training set GHMs. By this process, a projection matrix is learnt that can take any expression into a neutral subspace with the mostly decorrelated information of each feature. Thus a neutral subspace is formed by eigenvector decomposition of neutral expression moments represented by $\mathbf{X}_{Nt}$ containing the variations present in the neutral moments. Whenever any moment is projected on this subspace, it is expressed as a linear combination of eigen-moments corresponding to neutral expression moments. Basically projection on neutral subspace splits the test image into two components; one representing its neutral component and the other representing the expression component.

3.2.2.1 Interpersonal Variability Reduction in Neutral Subspace

The goal of principal component analysis is to transform the GHM set of $L$ dimension of any expression to an alternative data set of dimension $l(l < L)$. In

order to obtain it, we have to obtain a transformation matrix from a neutral expression training set, and as a result the transformed space is also called the neutral space. In order to do this, we organize the GHM data set. The neutral training set are arranged a set of n GHM vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ with each $\mathbf{x_i}$ representing a single grouped observation of $L$ variables (the dimension of $x_i$ is $1 \times L$). So, the neutral training set is written as:

$$\mathbf{X}_{Nt} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \mathbf{x_3} \\ \vdots \\ \mathbf{x_n} \end{bmatrix} \tag{3.11}$$

where $\mathbf{X}_{Nt}$ is a $(n \times L)$ neutral expression GHM training matrix.

To find a principal component basis that minimizes the mean square error of approximating the data, we have to perform mean subtraction [82].Here we calculate the empirical mean along each dimension $j = 1, 2 \cdots, L$. The mean vector is then given as **u** of dimension $(1 \times L)$.

Subtracting the empirical mean vector **u** from each row of data matrix $\mathbf{X}_{Nt}$, we find the ( $L \times L$ ) empirical covariance matrix as follows:

$$\mathbf{C} = \frac{1}{n-1}[\mathbf{X}_{Nt} - \mathbf{u}]^T [\mathbf{X}_{Nt} - \mathbf{u}] \tag{3.12}$$

We now find the orthonormal matrix $\mathbf{W}_{Nt}$ of eigenvectors which diagonalizes the covariance matrix **C** as given by:

$$\mathbf{W}_{Nt}{}^T \mathbf{C} \mathbf{W}_{Nt} = \mathbf{\Lambda} \tag{3.13}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigen-values of **C.**

In other words, the empirical covariance matrix of the original GHM matrix can be also written as:

$$\mathbf{C} = \mathbf{W}_{Nt} \mathbf{\Lambda} \mathbf{W}_{Nt}{}^T \tag{3.14}$$

matrix $\mathbf{\Lambda}$ will take the form of( $L \times L$ )diagonal matrix containing the eigenvalues usually in sorted order. $\mathbf{W}_{Nt}$ can be also called the transformation matrix that switches from GHM space to the neutral space.

Since neutral expression must have the minimum amount of variation, we now seek to find the eigenvectors that represent the minimum variation in neutral space. For reducing variability of neutral features due to interpersonal difference (facial features like shape of eyes, nose etc may vary from one person to another in many regions of the face), we ignore the components with the highest energy of eigenvalues in the neutral space (expected to contain the maximum variations of interpersonal change in faces) and instead seek the eigen-vectors with lower variation. But we reject the lowest eigen-values that might represent noisy features. So we select band of eigenvectors corresponding to the eigenvalues where energy value is flat/variation is nil for a considerable number of eigen-values (found to occur in the mid-region of the eigen-value curve **Fig. 3.3**).

The new feature matrix of the neutral training set can be defined in the neutral space as:

$$\mathbf{F}_{Nt} = \mathbf{X}_{Nt} \widetilde{\mathbf{W}}_{Nt}^{T} \tag{3.15}$$

Where $\mathbf{F}_{Nt}$ has the dimension of *n x l*.

### 3.2.2.2 Differentially Expressive Moments

We define the generalized GHM matrix $\mathbf{X}_{Ex}$ as:

$$\mathbf{X}_{Ex} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} \tag{3.16}$$

Where $\mathbf{X}_{Ex}$ is a $(m \times L)$ all expression testing matrix, and

where $Ex \in (Sad, Happy, Angry, Fear, Surprise, Disgust, Neutral)$ and *m* is the number of samples in the testing set. We note that it also contains moment vector sets of all other expressions apart from the neutral training set used for

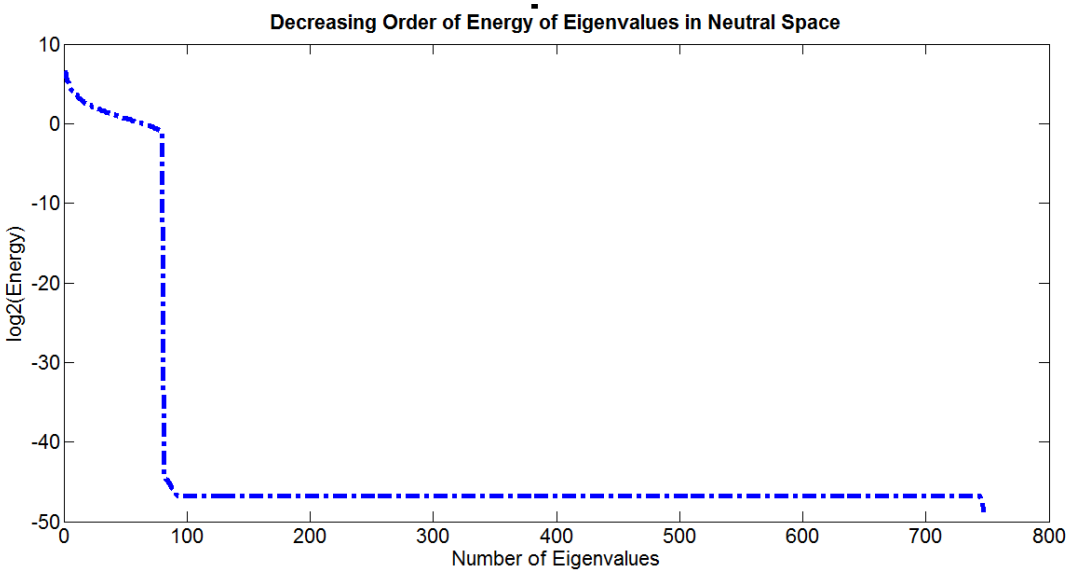



**Fig. 3.3**: The Energy-Eigenvalue curve in neutral/expression invariant subspace.

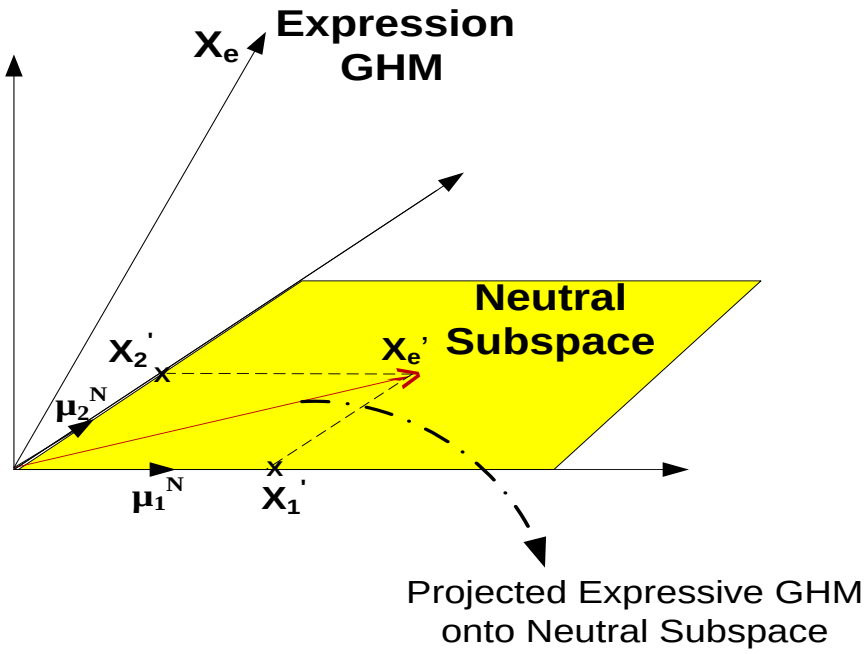



**Fig.3.4**: The concept diagram of projected expressive GHM into neutral subspace.

obtaining the transformation matrix, we can obtain the new differential expressive moment features given by

$$\mathbf{F}_{Exp}^{\Delta} = \mathbf{X}_{Ex} \widetilde{\mathbf{W}}_{Nt}^{T} - \boldsymbol{\mu}_{F_{Nt}} \tag{3.17}$$

where $\mu_{F_{Nt}}$ is the mean of the feature matrix $\mathbf{F}_{\mathrm{Nt}}$ in the neutral space and has dimension of $(1 \times l)$.

**Fig.3.4** illustrates the phenomena when any expression GHM is projected into the neutral subspace for *l=2*. Let us consider an expression feature vector $\mathbf{X}_e = [X_1, X_2]$ to be projected on a neutral subspace. After projection, the transformed feature vector becomes $\mathbf{X}_e^{'} = [X_1^{'}, X_2^{'}]$. If we consider the mean neutral vector in the neutral subspace as $\boldsymbol{\mu}^N = [\mu_1^N, \mu_2^N]$ , then the differential expressive component of the expression in the neutral subspace is given by:

$$\mathbf{F}_{Exp}^{\Delta} = \begin{bmatrix} X_1^{'} - \mu_1^N & X_2^{'} - \mu_2^N \end{bmatrix} \tag{3.18}$$

Eq. (3.18) is expected to extract useful differential information for each expression for improving the recognition accuracy. The concept can be extended to $l$ dimensions where $l$ is the maximum number of dimensions selected in the neutral subspace.

**Fig.3.5** gives the block diagram of the process of feature extraction from evaluation of expressive differential moments. The neutral expression from the 2D GHM training set is selected for forming a feature transformation matrix by PCA that would take GHM moment vectors of any expression (both training and testing set) to the neutral space. The difference component $\mathbf{F}_{Exp}^{\Delta}$ is then obtained by subtracting the mean of the components of the neutral expression of training set in neutral space from the components of expression in the neutral space. $\mathbf{F}_{Exp}^{\Delta}$ is then preserved for later forming the joint feature vector with the features found from the scattering ratio, and the combination is explained in the later section.

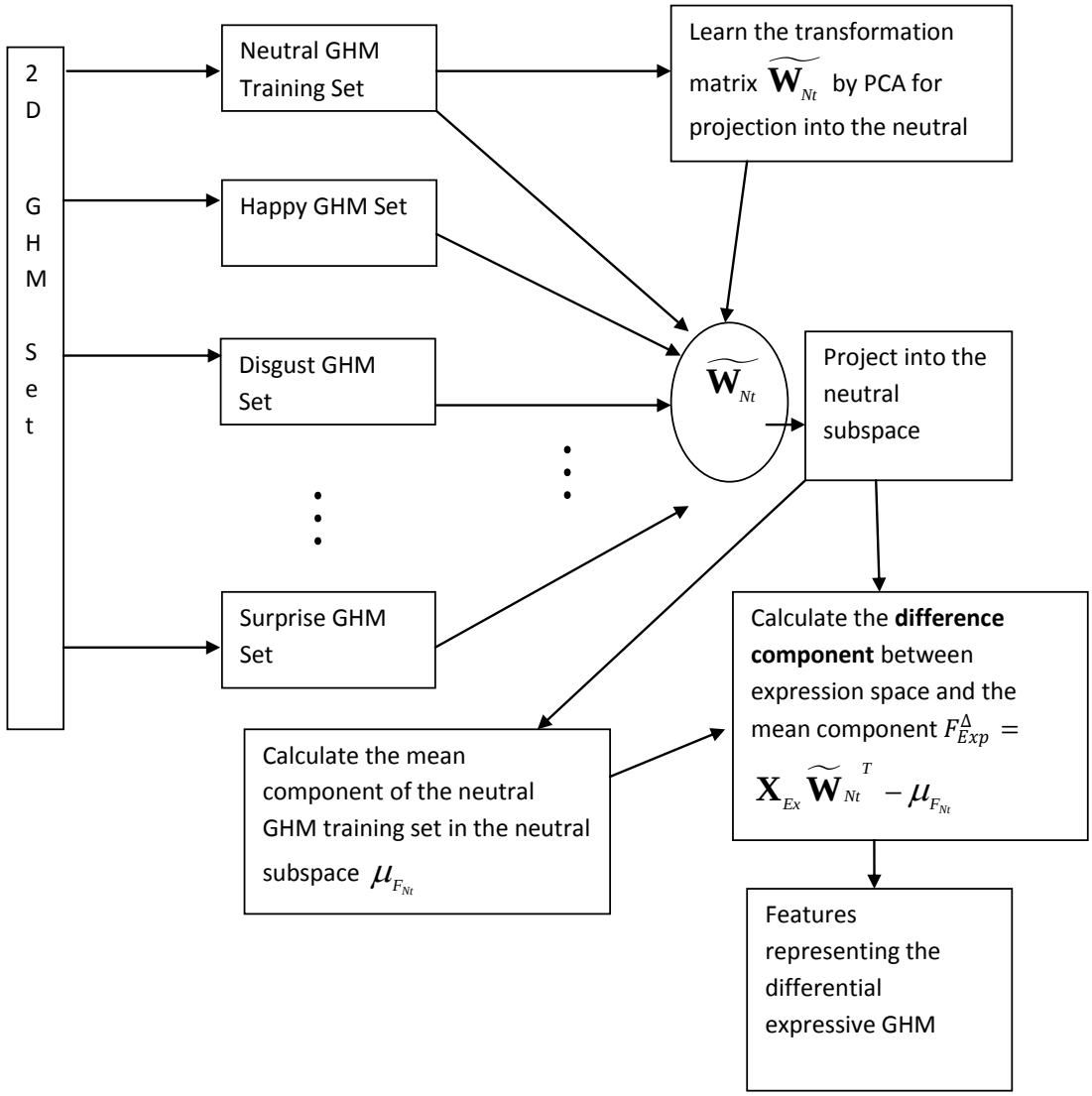



**Fig.3.5**: Block diagram of feature extraction by differential expressive moments

### 3.2.3 Feature Fusion

After we obtain the two feature vectors $\mathbf{F}_S$ and $\mathbf{F}_{Exp}^{\Delta}$, one from the sorted scattering ratio and the other from the difference component between neutral GHM and expression GHM in the neutral subspace respectively, we form the resultant feature vector $\mathbf{F}_S^{\Delta}$ for all expressions by combining the two feature vectors $\mathbf{F}_S$ and $\mathbf{F}_{Exp}^{\Delta}$, as defined by

$$\mathbf{F}_s^{\Delta} = [\mathbf{F}_S \quad \mathbf{F}_{Exp}^{\Delta}];$$

The resultant feature vector $\mathbf{F}_s^{\Delta}$ is named as the expressive GHM (EGHM). For all the expressions, its feature vector $\mathbf{F}_S^{\Delta}$ is then used to feed the classifier to detect the expression classes. In order to evaluate the discriminative capacity of the proposed expression oriented moment-based feature, we present the scatter plot of typical samples of expression in well known Cohn-Kanade [83] and FRGC database [84] respectively. The scatter plot is obtained in reduced two dimensions by means of multi-dimensional scaling [85], a very common and effective data visualization technique that can produce representation of data in reduced dimensions. **Fig.3.6** and **Fig.3.7** show how the distribution of features for each class formed from the feature vector $\mathbf{F}_S^{\Delta}$ looks like in compressed two dimensions. As can be seen, the expression class 'happy' and 'sad' are in the opposite side of the neutral expression, which is considered to be the offset for every expression. Since 'happy' and 'sad' are two opposite expressions (happiness is mouth bent on the upper side while sad means mouth bent on the down side), it is very natural to fall on opposite side of each other. It must be noted that the scatter plot is built taking features of some samples from the feature space. Of course classes have overlaps, and as a result, the classification accuracy for both the database are found to be around 80%. The details of the results are discussed in the next chapter.

## 3.3 Feature Classification

There are two types of classification; supervised and unsupervised. Supervised classification involves using the samples of known identity to identify the samples of unknown identity. We use supervised classifier for classification purpose since we have all the emotion labels at our disposal and we have extensive prior knowledge of expression class for each static image. Amongst the various classification schemes that have been employed to recognize facial expression the famous ones include the minimum distance classifier [14,43], the support vector machine (SVM) [35,44], the naïve-Bayes classifier [45-46], and the artificial neural network [47-48]. We choose the SVM classifier for its simplicity, faster classification capacity, lesser time required for training the dataset, and ability to handle nonlinear data at the same time [35,44,86-87].

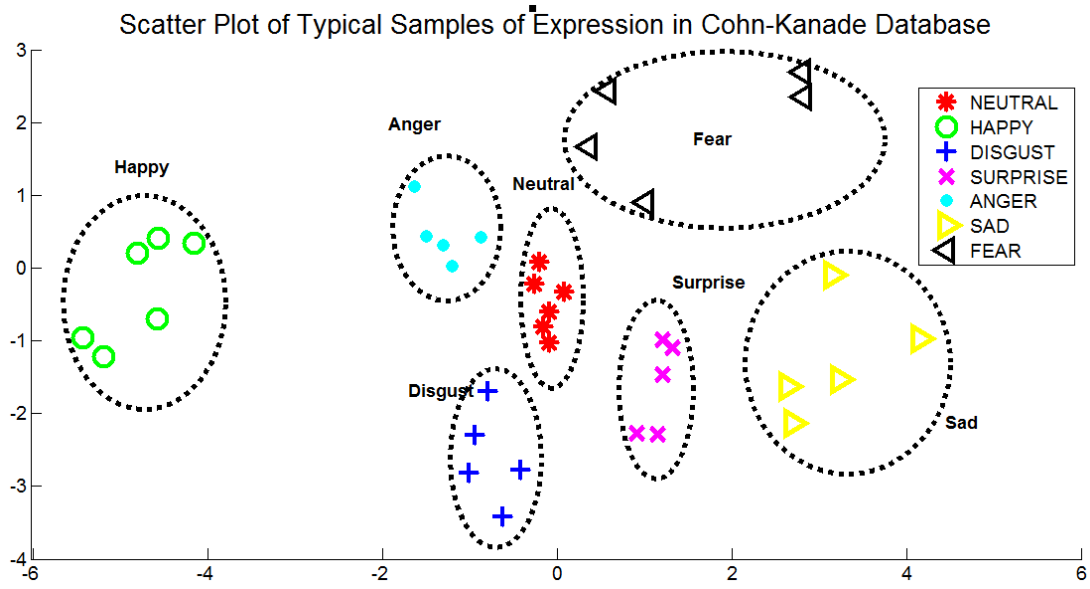



**Fig.3.6**: Scatter plot of typical samples of expression in Cohn-Kanade database.

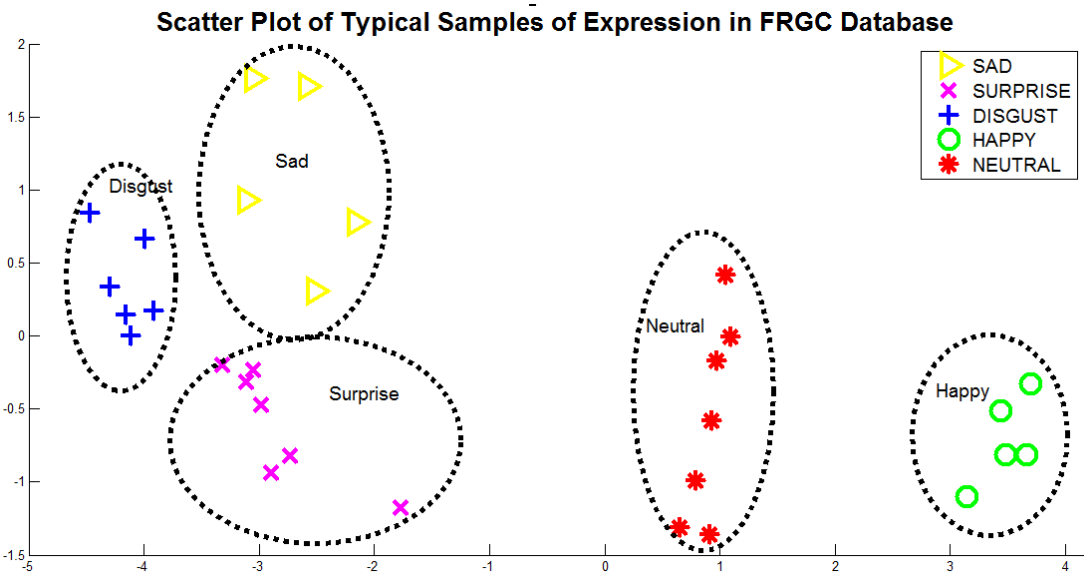



**Fig.3.7**: Scatter plot of typical samples of expression in FRGC database

SVM is a well founded statistical learning theory that has been successfully applied in various classification tasks in computer vision. SVM performs an implicit mapping of data into a higher dimensional feature space and finds a separating hyper-plane with maximal margin to separate the data. Recognized as a binary classifier, multi-class classification can be achieved by adopting the one-against-rest or one-against-one approach. One against one approach is found to be more accurate and competitive approach for multiclass classification [88]. We selected a suitable

kernel and carried out grid-search on the hyper-parameters in a cross-validation approach for selecting the parameters, as suggested in [88].

## 3.4 Conclusion

This chapter explains in depth the concept and use of the proposed methodology for feature extraction of facial expression of still images. It describes how two different feature space $F_S$ and $F_{Exp}^{\Delta}$, the former generated from the concept of scattering ratio and the later formed by the differential information between meaningful expression and neutral expression in the neutral subspace, are formed and how the two could be combined to boost recognition accuracy of the expression classification. Mathematical expressions are used with the help of several illustrations and block diagrams to make the concept meaningful to the reader.

# 4 : EXPERIMENTAL SETUP AND RESULTS

## 4.1 Introduction

In Chapter 3, the proposed methodology to be used for expression recognition is being described in detail with illustrations and intermediate results. This chapter highlights how the datasets are organized keeping the real-life and online implementation in perspective. The chapter also divulges the achieved results and go on to elaborate further the effectiveness of the achieved results with adequate figures, plots and illustrations.

## 4.2 Database Description

We selected the two renowned and standard datasets for our experiments; namely the Face Recognition Grand Challenge (FRGC) dataset (ver. 2), and the Cohn-Kanade (CK) dataset (ver. 1) respectively. By using these two datasets, we ensure that the subjects to be used in our experiments have blends of different ethnicity, age and sex in our setup which ensure that our methods have the capacity to tackle these variations. Considering these variations by itself is a very challenging task in expression recognition [22].

### 4.2.1 Face Recognition Grand Challenge Dataset

Mostly used as a defacto standard for face recognition, Face Recognition Grand Challenge (FRGC v.2) is a huge collection of colored still images consisting of 50,000 recordings. Assembled at the University of Notre Dame, the data corpus contains high resolution still images taken under controlled lighting conditions and with unstructured illuminations, 3D scans, and contemporaneously collected still images. The subject's session consists of four controlled still images, two uncontrolled still images and one 3D image. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions (two or three studio lights) and with several facial expressions. The uncontrolled images were taken in various illumination conditions; hallways, atria or even outdoors. Still images were taken by a 4 Mega-pixel Canon Power Shot G2. Images are either $1704 \times 2272$ pixels or $1200 \times 1600$ pixels. Images are in JPEG format and storage size ranges from 1.2 Mbytes to 3.1 Mbytes. The dataset contains blend of subjects of

different ethnicity, age groups and sex groups (Male and Female); Asian 22%, White 68%, Other 10%, of different ages; 65% subjects within 18-22 years old, 18% subjects within 23-27 years old and 17% above 28 years old, and finally male constituting 57% and female 43% of the subjects respectively [84].

Ranges of expression category in the dataset include Sad, Disgust, Surprise, Happy/Smile, Blank-Stare and Neutral. Based on prototypic expressions set by Ekman et. al [7] which are defined to be constant among all cultures and races, we selected the available four expressions (Sad, Disgust, Surprise and Happy) besides Neutral subset from the dataset to incorporate in our experimental setup and the results are established based on these five classes (including Neutral). **Fig.4.1**displays the five expression images for five typical subjects used in our experiments. It can be clearly seen the expressions vary from one person to another, some subjects exhibit happy faces almost explicitly (opening of mouth), while others show subdued form of this expression, bending up but closing their mouth at the same time. Differences are also seen for other expressions, for example in Sad expression, some show a frozen dejected look while others bend down their mouth to express this emotion. Though it is a huge dataset consisting of large number of individuals, we consider a generic FRGC dataset ensuring that the mutually exclusive subjects found in the expressions ***Sad***, ***Disgust*** and ***Surprise*** combined are selected from the **Neutral** and **Happy** classes of the entire dataset to form our own classes ***Neutral*** and ***Happy*** to be used later in our experiment. The generic dataset contains 161 image samples in **Sad**, 241 samples in **Surprise**, 179 samples in **Disgust**, 272 samples in **Happy** and **Neutral** Expression sets respectively.

### 4.2.2 Cohn-Kanade Dataset

The best known database for facial expression analysis has been developed at Carnegie Mellon University, and is known as CMU-Pittsburgh Action Unit Coded (AUC) Facial Expression Database, or Cohn-Kanade (CK v.1) Database [86,83]. It provides a large, representative test-bed for comparative studies of different approaches to facial expression analysis. Version 1.0, the initial release of this database, which has been used for our research, includes approximately 486 image sequences from 97 subjects. Each sequence begins with a neutral expression and proceeds to a peak expression. The subjects are university students enrolled in

introductory psychology classes. They ranged in age from 18 to 30 years. All the subjects are trained to display specific particular expression. The subject distribution across genders is 69% female and 31% male. In terms of racial distribution, 81% of the subjects are Euro-American, 13% are Afro-American, and 6% belong to other races. The Cohn-Kanade database was created in an observation room equipped with a chair for the subject and two Panasonic WV3230 cameras, each connected to a Panasonic S-VHS AG-7500 video-recorder with a Horita synchronized time-code generator. The cameras were located directly in front of the subject.

We obtain six expressions in this version of dataset besides Neutral, viz. Happy, Surprise, Disgust, Sad, Anger and Fear, all classes falling in the Ekman et al.'s [7] universal category of facial expressions for exhibiting emotions. Instead of the class labels tagged for every expression in the FRGC dataset, this version of dataset supplies Facial Action Codes (FACs) for Action Units (AUs) of facial muscles of peak expression for each subject from which the expressions have to be interpreted and classified. The dataset contains the variants of FACs for every expression available in its corpus. By translating the AUs for every expression by means of FACs, the expressions are classed into seven categories (including Neutral). **Fig.4.2** shows the seven expressions of five typical subjects found in the dataset. Since the subjects are trained prior to each session, it can be found all the subjects exhibit nearly similar styles of expression, i.e. for Happy class, all the subjects open their mouth and smile, for Sad class, all the subjects bend down their mouth and give a dejected look. In all the classes, it is clearly seen that the expressions are clearly exaggerated by the subjects. Although each class of expression have different number of members in each set (Fear class having only 35 members while Neutral class having 97 members), we ensure that set of members in each class have mutually exclusive subjects. We form a generic dataset containing 97 image samples in **Neutral**, 89 samples in **Happy**, 51 samples in **Surprise**, 73 samples in **Disgust**, 42 samples in **Anger**, 69 samples in **Sad** and 32 samples in **Fear** expression set respectively.

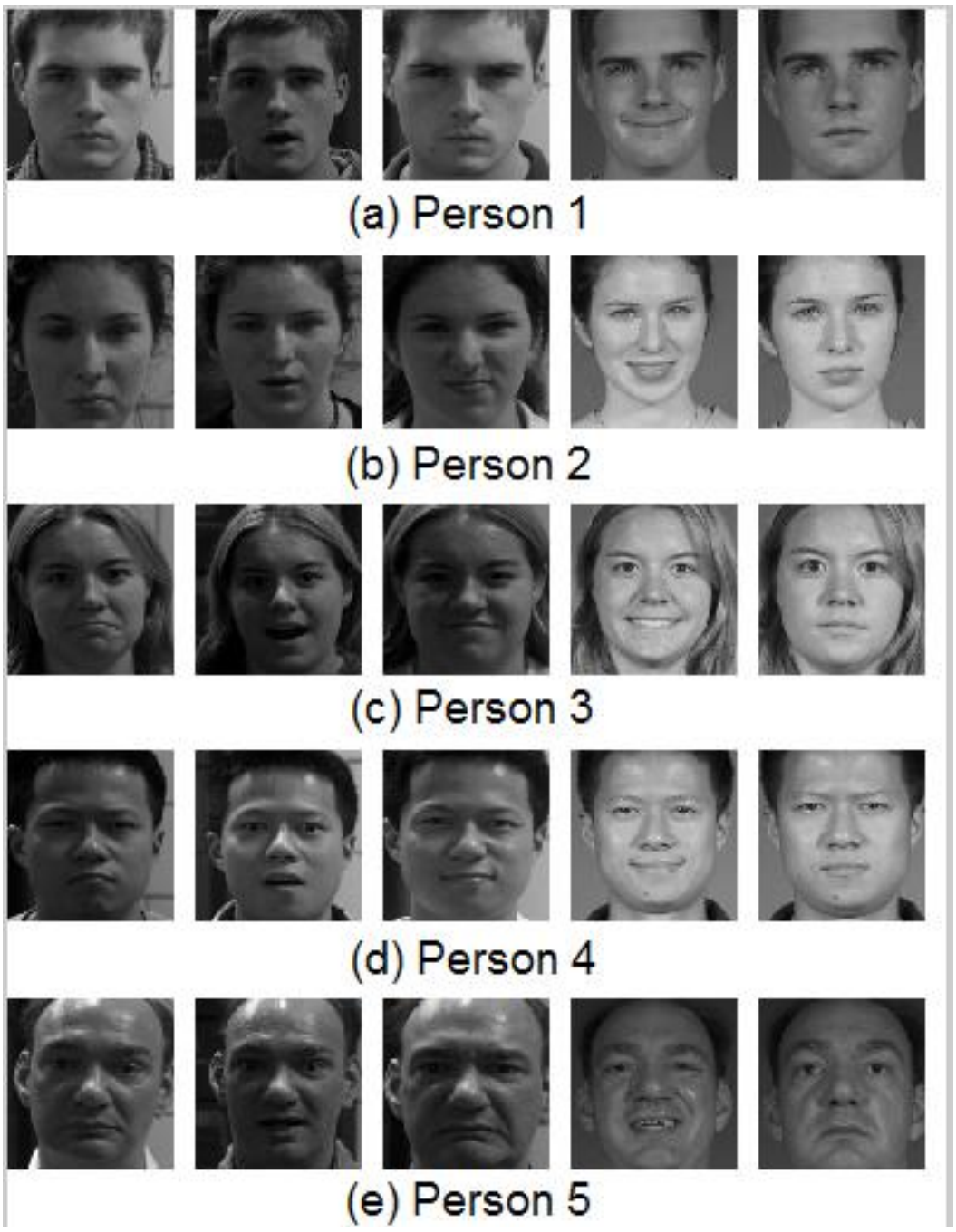



**Fig.4.1**: Typical expressions of FRGC dataset for five subjects. Each of (a),(b),(c),(d) and (e) represent specific persons with different expressions. Left to right columns in the figure represent Sad, Surprise, Disgust, Happy, and Neutral expressions, respectively.

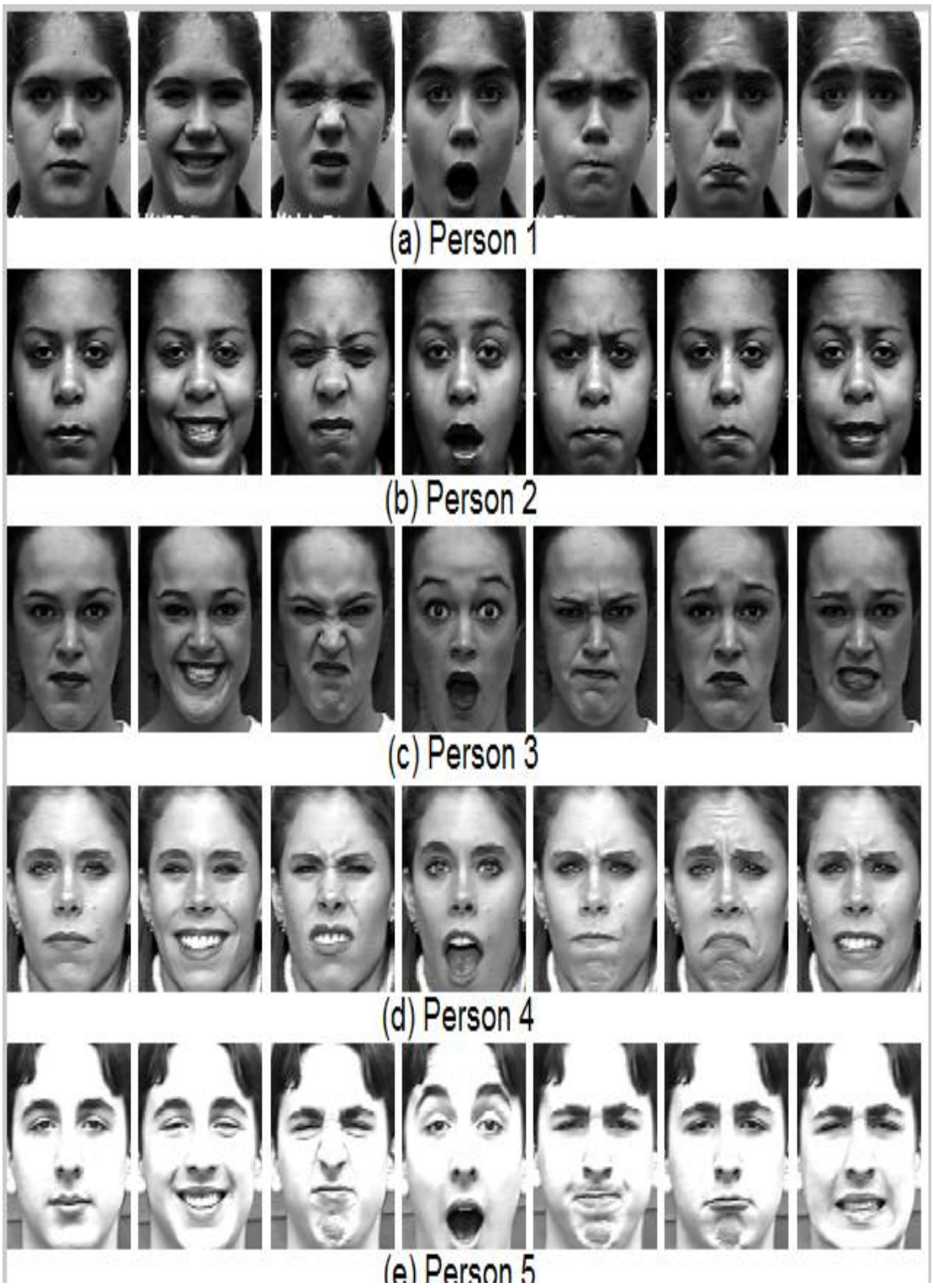



**Fig.4.2**: Typical expressions of Cohn-Kanade dataset for five subjects. Each of (a),(b),(c),(d) and (e) represent specific persons with different expressions. Left to right columns in the figure represent Neutral, Happy, Disgust, Surprise, Anger, Sad, and Neutral expressions, respectively.

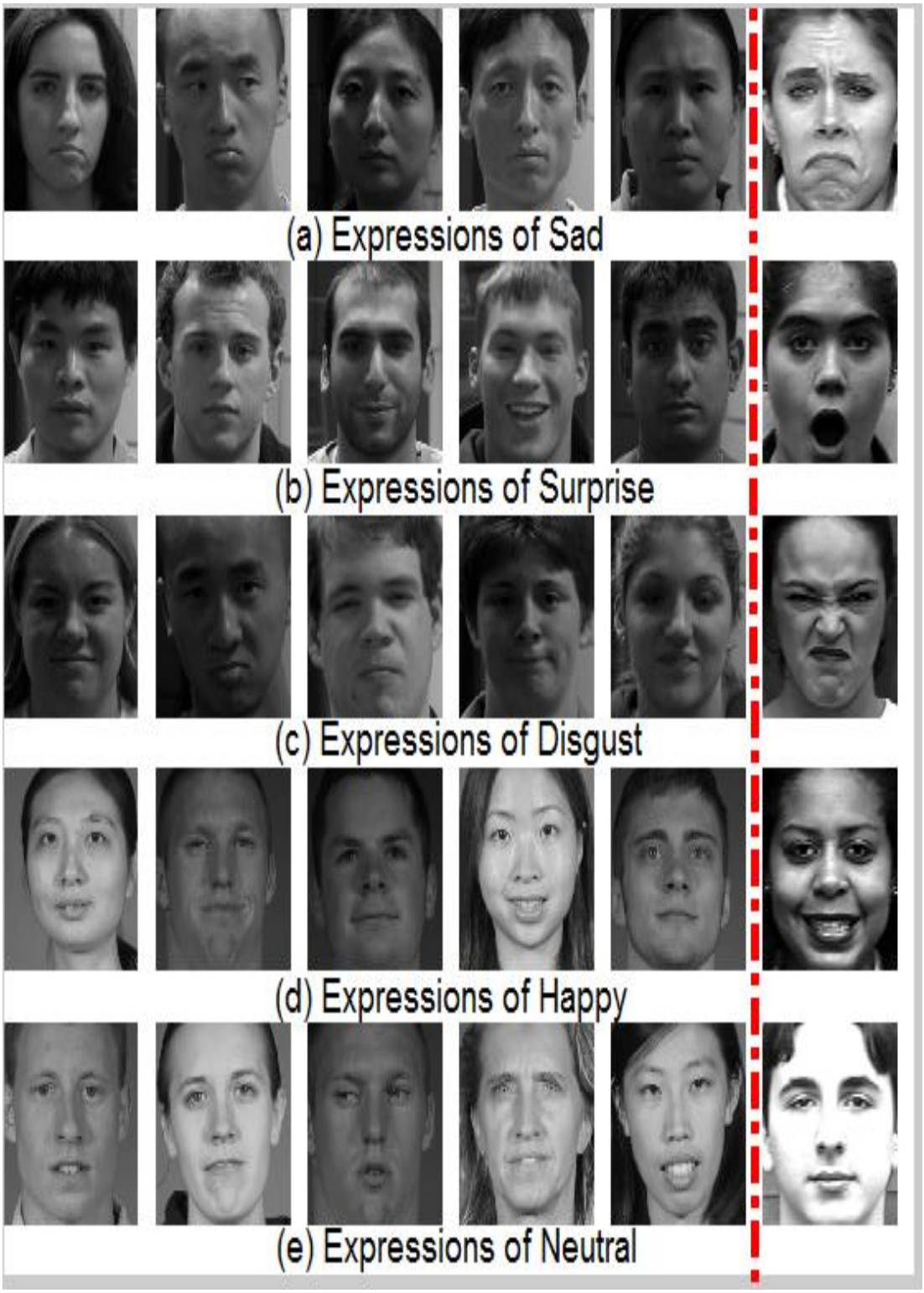



**Fig.4.3**: Comparison of typical variants of each expression in FRGC dataset with Cohn-Kanade dataset. (a), (b), (c), (d) and (e) represent expressions of Sad, Surprise, Disgust, Happy and Neutral respectively. Col. 1-5 represents images from FRGC Dataset while Col. 6 represents the concerned expression image from Cohn-Kanade Dataset.

### 4.2.3 Posed versus Spontaneous Expression

In chapter 1 and 3, the concept of posed and spontaneous form of expression is already explained. This section now illustrates these two types of expression with the help of two datasets used in our experiments.

**Fig.4.3** shows the comparison of typical variants of spontaneous expression in the FRGC dataset with a typical expression image found in the Cohn-Kanade dataset. There are many other variants of spontaneous expression, but only typical ones of significance have been chosen to represent FRGC dataset. Subjects from both the datasets are made aware that they are being monitored. But unlike a designated style of expression as often found in the Cohn-Kanade dataset, FRGC dataset contains many variants of a particular expression since no trainer is been employed to specify what look an expression is bound to show. As a result, the subject has the independence to choose their style of expression, which brings spontaneity to these expression images. On the other hand, in Cohn-Kanade dataset, each subject has been trained to undergo a particular expression, as a result, instead of the spontaneous look, a touch of exaggeration exist for all the expressions and for all subjects. Difference between spontaneous and posed expression can be clearly seen in this figure. In (d) of **Fig.4.3** for example, it is seen that the subjects show happy faces (as labelled by the dataset), but their degree of expressing the happy faces vary, some opening their mouth and exhibiting their teeth to express happiness, others are more reserved and show a subtle change of facial muscles around their mouth to express happiness. The degrees of correlation amongst several classes are more enhanced in spontaneous expression than in posed expression. For example, col.1 image of *IV* (Happy Expression) and col.2 image of *V* (Neutral Expression) have high degree of correlation amongst themselves in the lip region of the face.

**Fig.4.4** and **Fig.4.5** show the scattering ratio of moments calculated from the training set for FRGC and CK dataset respectively. While the ratio of maximum discriminant moment is around 1, it is seen that for CK the ratio of maximum discriminant moment goes as high as 80. It clearly shows the degree/extent of discrimination among expression classes in the CK dataset. To counter this low discrimination power in FRGC dataset, we hence propose the concept of differential expressive GHM as explained in Chapter 3. We argue that the differential expressive GHM does not only provide a de-correlated neutral feature space, it captures the subtle

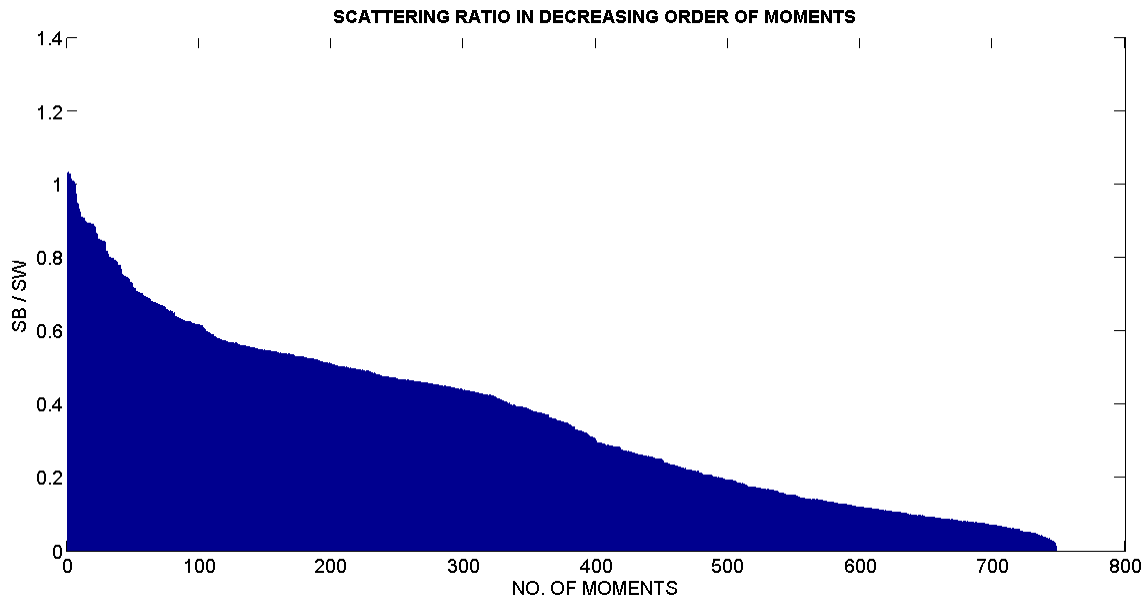



**Fig.4.4**: Calculated scattering ratio of moments (In decreasing order) for FRGC dataset

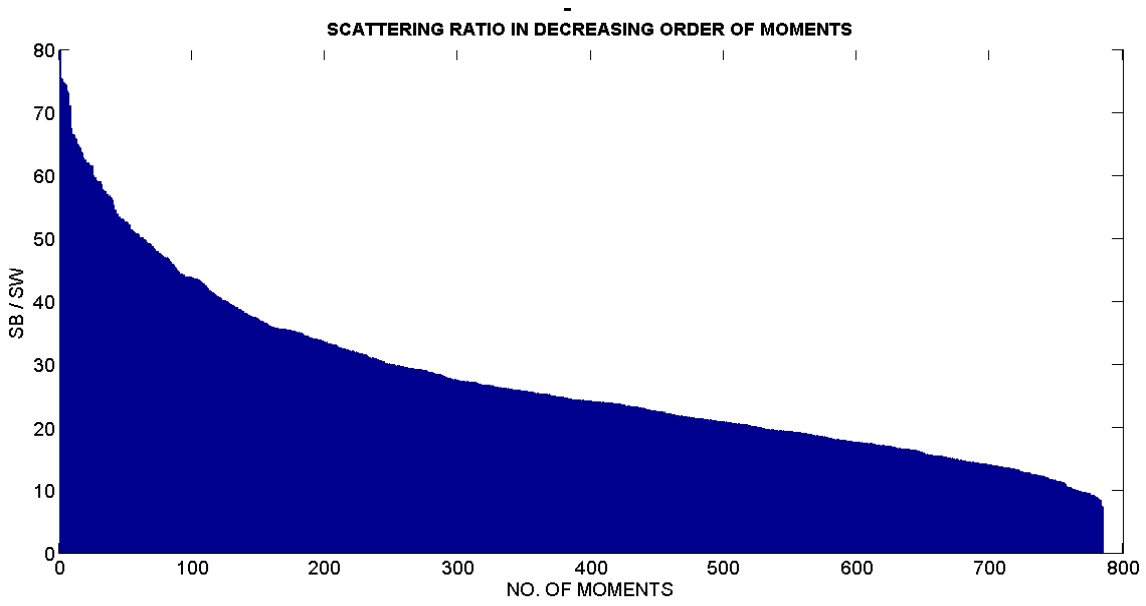



**Fig.4.5**: Calculated scattering ratio of moments (in decreasing order) for Cohn-Kanade dataset

difference of expression that may be objectively present and prominent amongst the several expression classes; and as a result it increases the robustness of expression classification.

## 4.3 Experimental Setup

### 4.3.1 Dataset Setup

Having set the methodology, images from the dataset are assembled for experimental setup. Each expression for both the datasets contains mutually exclusive subjects, indicating each expression set contains no duplicate subject but common subjects may be found across different expression sets. By this setup, the influence of same person in a particular expression set has been eliminated by excluding multiple number of samples of the same subject; since it is possible that it might enhance the unique characteristics/features of the person rather than the expression characteristics itelf; resulting in a convoluted result. The concept of person independency and person dependency is also introduced in our setup. We consider the neutral mug-shot images of all the persons to be available beforehand for person dependency setup. This means that the neutral subspace is created from the neutral GHM set of all the available persons in the dataset. For person independency setup, we consider that only one third (neutral expression GHM of all the subjects from the training set to be discussed later) of the subjects are known beforehand to the system.

Having generated the expression sets for both the datasets, the pre-processing is then carried out so that the region of interest (face) is now extracted. In both the datasets, the background of the image is found to be almost uniform throughout all the images. Also as a result, no background subtraction is considered necessary. Based on the nose coordinates as supplied by the datasets, the images are cropped so that the approximate face regions (from forehead to chin vertically and from left ear to right ear horizontally) are highlighted. Since geometric normalization is avoided, we hope to counter the misalignment of faces (eye and nose at fixed coordinates) by introducing rotation and translation invariant RTIGHM in the feature space. Histogram equalization is then carried out to combat any illumination variation. Finally the image is scaled and converted to a $160 \times 160$ 2-D image block by bicubic interpolation.

The pre-processed images are divided into two major sets; one for training and the other for testing. One third of the images are used for training and the rest used for testing. The testing set comprises again of two sub-sets; gallery and the probe images. The training set is basically acquired to learn the discriminating power and order of scattering ratio indices. From the training set, the neutral expression GHMs are selected for learning the transformation matrix that would take any expression GHM features into the neutral subspace. On the other hand, the testing set is basically acquired for classification purpose; the gallery set used to train the classifier and the probe set used to test the classifier performance and thereby the overall performance of the methodology. Gallery set contains 80% of the testing set while probe set contains 20% of the testing set respectively. Forward feature selection algorithm is applied on a random 30% of the testing set to provide a threshold of feature selection from the GHMs. These features sets are finally used in evaluating the classifier performance.

### 4.3.2 Classifier Setup

As explained in Chapter 3, we use Support Vector Machine (SVM) as a classifier for evaluating the performance of our feature extraction methodology. The testing set (two-thirds of the data samples divided into gallery set and the probe set) is used exclusively for training and testing the classifier. We use LibSVM toolbox [89], a freely available toolbox to implement SVM. Out of the several kernels (linear, polynomial, Radial Basis Function (RBF), sigmoid etc) used in SVM, we use the RBF kernels since it has the capacity to handle nonlinear data by mapping nonlinearly samples into higher dimensional space, and also because the feature dimension in our case is moderate compared to number of instances/observations [90]. Values of hyper-parameters such as the regularization parameters and width of radial basis functions are selected from cross-validation performed on30% of the testing set as suggested by [90], and samples in this set are selected randomly. Complexity that may arise from unbalanced number of members in each class is prudently handled by introducing a weighting factor for each class to compensate the biasness of classifier towards classes having greater members in their sets [89]. The best parameters selected from tuning the cross-validation set are then used for training the whole gallery set in the classifier.

## 4.4 Performance Evaluation

The performance evaluation of the proposed methodology is now explained in this section. The section is divided into two sub-topics; one containing short description of the existing methods with which the performance is evaluated and the other section contains the tools used to assess its merit.

### 4.4.1 Comparison with existing methods

We have compared our method with three other existing methods used for holistic facial expression recognition.

Method 1 [91]: MCM of training set are used for expression recognition. The MCM are formed by linear combination of sample group and pooled covariance matrixes. The classifier used in this method is Gaussian Maximum Probability classifier.

Method 2 [92]: LDA is applied to the PCA components of face images. LDA creates a linear combination of given feature dimensions that yields large mean differences between the desired classes. Finally, SVM is used as the classifier for expression classification.

Method 3 [93]: Zernike Moments (ZM), one kind of orthogonal invariant moment, are used as feature extractor. Moments are selected heuristically. Classification works optimum for moments selected for 52 order (number of dimensions 756 for each instance of expression). Instead of Naive Bayes Classifier, we use SVM for expression recognition.

Method 4: 2-D Gauss-Hermite Moments, another kind of orthogonal moment, are used as feature extractor. Moments are selected heuristically. Classification works best for order N around 25~30. SVM is used as classifier for expression recognition.

### 4.4.2 Evaluation Criteria and Results

We used three evaluation criteria to compare the classification performance of the methods considered in the experiments; Classification Accuracy amongst all the methods. Since accuracy is not a reliable metric for the real performance of a

classifier (it might yield misleading results if the data set is unbalanced (that is, when the number of samples in different classes vary greatly), we incorporate Receiver Operating Characteristics (ROC) curve to evaluate the performance of each expression, and finally Confusion Table to indicate which expression are confused for other expressions in this setup.

#### 4.4.2.1 Classification Accuracy

Based on the experimental setup ( one random testing set containing gallery set of 80% of test images and probe set comprising the rest 20% of test images), the classification accuracy is evaluated for all the available expressions in each dataset and for all the methods. The two tables (**Table 4.1** for FRGC dataset and **Table 4.2** for CK dataset) clearly show the proposed method outperforms the existing methods used for comparison. Accuracy for each expression is calculated by noting how many correct members are classified for each class (Number of Correct Members in each Expression Class/Total number of members in each Expression Class). The overall accuracy is calculated by considering the ratio of overall correct members for all the classes by total number of members in the probe set used for testing the classifier. **Table 4.1**for FRGC dataset indicates that for each expression, except for Neutral Expression Training Set, all other methods outperform the existing methods. This can be explained by the fact due to large number of variation in the neutral expression training set (as explained in Sec 4.2.3) our method gets more biased towards classifying meaningful expression images rather than neutral images. **Table 4.2** for CK dataset confirms our observation. Since in CK the expression set in each class are more or less clearly defined and each class have far lesser variation within themselves, the classification accuracy for each expression (including neutral expression) beats all other methods.

**TABLE 4.1**: CLASSIFICATION ACCURACY SHOWING COMPARISON OF SEVERAL METHODS WITH OUR PROPOSED METHOD (FRGC DATASET):

| Expression | MCM | PCA+LDA+SVM | ZM+SVM | GHM+SVM | DGHM+SVM |
|---|---|---|---|---|---|
| Sad | 18.75 | 31.25 | 25 | 50 | **62.5** |
| Surprise | 43.75 | 81.25 | 75 | 81.25 | **93.75** |
| Disgust | 58.82 | 70.59 | 58.82 | 64.71 | **81.25** |
| Happy | 73.53 | 79.41 | 58.82 | 82.35 | **85.29** |
| Neutral | **86.67** | **86.66** | 83.33 | 83.33 | 83.33 |
| Overall Accuracy | 62.83 | 73.45 | 62.83 | 75.22 | 81.41 |

**TABLE 4.2**:CLASSIFICATION ACCURACY SHOWING COMPARISON OF SEVERAL METHOD WITH OUR PROPOSED METHOD (CK DATASET)

| Expression | MCM | PCA+LDA+SVM | ZM+SVM | GHM+SVM | DGHM+SVM |
|---|---|---|---|---|---|
| Neutral | 21.05 | 73.68 | 68.42 | 68.42 | **78.95** |
| Happy | 58.33 | **92.30** | 83.33 | **92.30** | **92.30** |
| Disgust | 33 | 66.67 | 66.67 | 66.67 | **100** |
| Surprise | 58.33 | 75 | 91.67 | 83.33 | **100** |
| Sad | 33.33 | 41.66 | **58.33** | **58.33** | **58.33** |
| Fear | 14.28 | 50 | 50 | 50 | **66.67** |
| Anger | 14.28 | 50 | 33.33 | 58.33 | **66.66** |
| Overall Accuracy: | 32.86 | 65.71 | 68.57 | 70.42 | **78.87** |

#### 4.4.2.2 Confusion Table

A confusion matrix, also known as a contingency table or an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two or multiple classes (i.e. commonly mislabelling one as other).

Typically, it is a table with two rows and two columns that reports the number of *false positives (type I error)*, *false negatives (type II error)*, *true positives*, and *true negatives*. This allows more detailed analysis than mere proportion of correct guesses (accuracy).

This section shows the confusion matrix of our proposed method for both the datasets used in the experimental setup. The confusion matrix are generated selecting one random testing set containing gallery set of 80% of test images and probe set comprising the rest 20% of test images. Data in the confusion matrix show the normalized correct and misclassification rates for each expression respectively.

As can be seen in **Table 4.3**, the Sad Expression is confused often with surprise and disgust. Since all these expressions show negative expressions, they are very close to each other in the feature space and so their jumbling with each other is often justified. Happy and Neutral Expression are confused amongst themselves, this can be explained by the fact that in some of the neutral mug-shot face images, some of the subject's mouth are found to be open which is often confused with Happy expression class and vice versa.

**Table 4.4** for CK dataset shows a slightly different picture. As can be seen from the table, Surprise and Disgust expressions can be recognized with 100% accuracy. Unlike the FRGC dataset, the neutral expression are being confused more often with the Sad faces instead of the Happy faces; The expression suffering the most confusion is Sad faces and it is confusing themselves with Surprise, Disgust and Neutral; it shows that the method fails to recognize the bent down mouth of the subjects in Sad expression (the chief property of the Sad expression). As a result, so many misclassifications occur. Similar things also occur for Anger and Fear. We presume that the misclassification may occur due to lower number of samples compared to other expressions which have ample members in the training set (see Sec 4.3 for number of samples in each class).

**TABLE 4.3**:CONFUSION MATRIX FRGC:

| | Sad | Surprise | Disgust | Happy | Neutral |
|---|---|---|---|---|---|
| Sad | **0.625** | 0.250 | 0.125 | 0 | 0 |
| Surprise | 0.062 | **0.938** | 0 | 0 | 0 |
| Disgust | 0.059 | 0.176 | **0.765** | 0 | 0 |
| Happy | 0 | 0 | 0 | **0.853** | 0.147 |
| Neutral | 0 | 0 | 0 | 0.167 | **0.833** |

**TABLE 4.4**:CONFUSION MATRIX CK:

| | Neutral | Happy | Disgust | Surprise | Anger | Sad | Fear |
|---|---|---|---|---|---|---|---|
| Neutral | **0.789** | 0 | 0 | 0 | 0 | 0.211 | 0 |
| Happy | 0.077 | **0.923** | 0 | 0 | 0 | 0 | 0 |
| Disgust | 0 | 0 | **1.000** | 0 | 0 | 0 | 0 |
| Surprise | 0 | 0 | 0 | **1.000** | 0 | 0 | 0 |
| Anger | 0 | 0.167 | 0.167 | 0 | **0.667** | 0 | 0 |
| Sad | 0.167 | 0 | 0.167 | 0.083 | 0 | **0.583** | 0 |
| Fear | 0.167 | 0.167 | 0 | 0 | 0 | 0 | **0.667** |

#### 4.4.2.3 Receiver Operating Characteristics (ROC) Curve

Receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the total actual positives (TPR = true positive rate) vs. the fraction of false positives out of the total actual negatives (FPR = false positive rate), at various threshold settings.

An ROC curve demonstrates several things [94]:

- It shows the trade-off between sensitivity (true positive rate) and specificity (1-false positive rate) (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.

- The closer the curve comes to the 45-degree diagonal of the linear ROC space, the less accurate the test.
- The area under the curve is a measure of classification accuracy; i.e. it evaluates the performance of the classifier to correctly classify a single expression from several expression classes.

Since our method employs a multi-class classifier (several expression classes needs to be classified), we build a customized binary class classifier for each expression to generate ROC space. Such cases are handled by considering all other expressions as the negative class, while the expression of interest is considered to be positive class.ROC curve is obtained by using the probability estimates given by the classifier on each test instance of the dataset during classification. Instead of choosing a random testing set (gallery set comprising 80% of test set and probe set comprising 20% of test set)for generating the ROC curve, we jumble samples in the testing set several times (around 14-20 times) so that different samples have the chance to form the gallery and the probe sets while evaluating the classifier. This ensures that more or less all the samples are used for training and testing the classifier and thus give a better picture of the classifier performance (in case we have numerous outliers/bad samples in the gallery or probe sets at a time that can seriously belie the performance of all the methods used alongside our proposed method). The following figures (**Fig.4.6**-**Fig.4.19**) show the ROC curve for each expression. To evaluate ROC curve objectively for all the methods in each of the dataset, the area under the curves (**Table 4.5**-**Table 4.6**) are also calculated to determine the margins of elevation of the ROC curve from the horizontal line. Finally the overall ROC curve which indicates the average classification performance of all the binary classifiers is obtained to determine the overall performance of the proposed methodology with the other methods.

The following illustration shows except for the neutral and disgust expression (**Fig.4.6** and **Fig.4.10**) in the FRGC dataset, ROC curves for all other expressions (for both the datasets) show satisfactory performance of our proposed methodology. While we reason our method's biasness towards classifying expressive images than neutral mug shot images (see Sec 4.4.2.1) for the down-performance in ROC curve of Neutral Expression, the case for Disgust expression is hard to explain. We assume due to the large variation in Disgust expression, it is more often confused with Sad

and Surprise expression (as evident in the confusion matrix, see **Table 4.1**). As a result, the ROC curve slightly lowers with respect to the method 2 [92], the most competitive method after our proposed method.

The ROC curves for other expressions (both for FRGC and CK) reveal the domination of our method. While **Fig.4.13** shows that method 3 [93] is slightly ahead of our method upto a certain range, for the rest of the range it is then outplayed by the proposed method in significant margin. It is also evident from **Table 4.6**, which show that curve of our method bounds more area under it compared to method 3, which is a clear indication of superiority of our method.

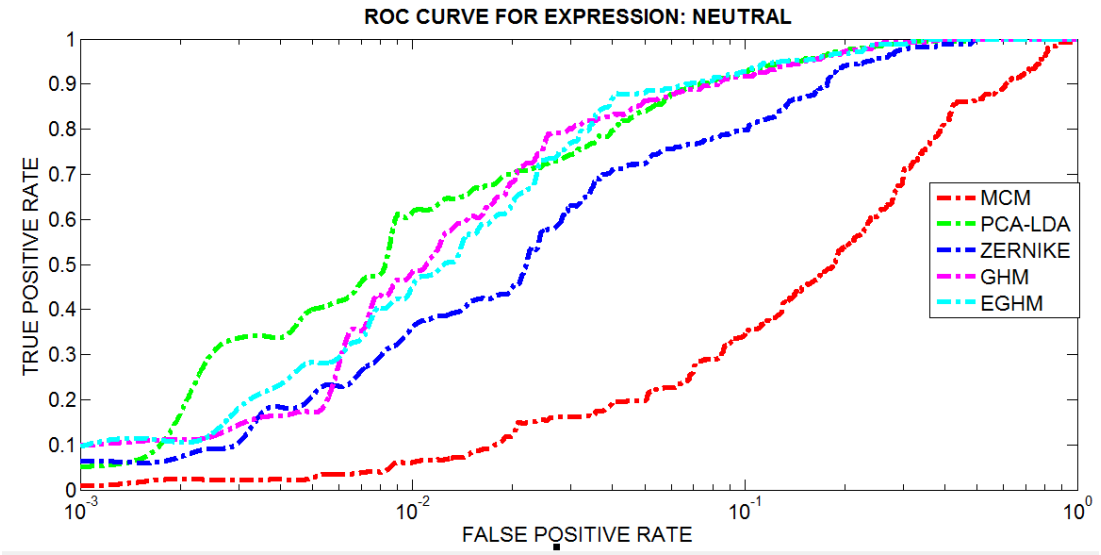



**Fig.4.6**: ROC curve for neutral expression in FRGC dataset

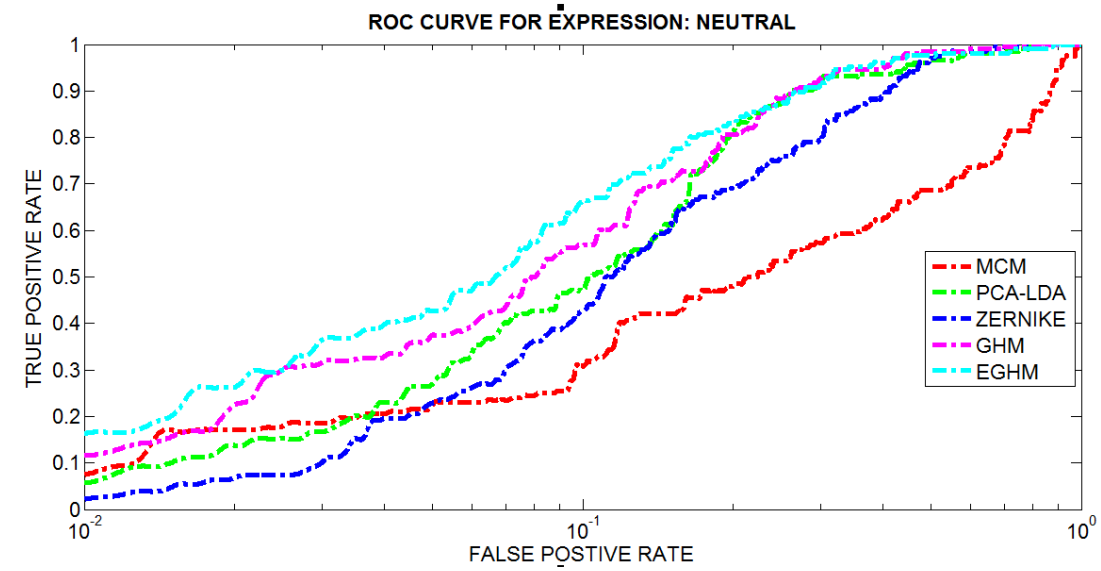



**Fig.4.7**: ROC curve for neutral expression in CK dataset

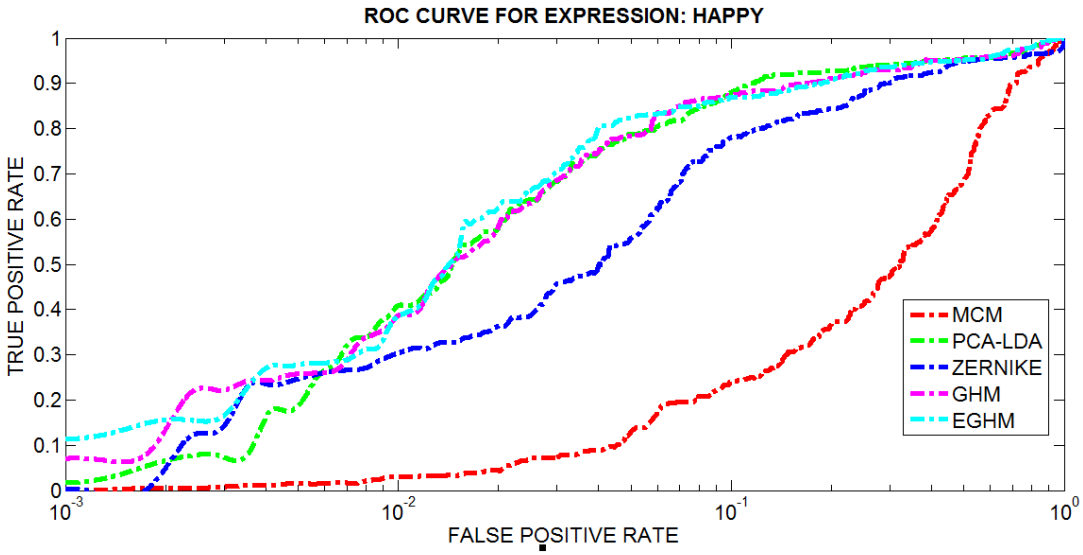



**Fig.4.8**: ROC curve for happy expression in FRGC dataset

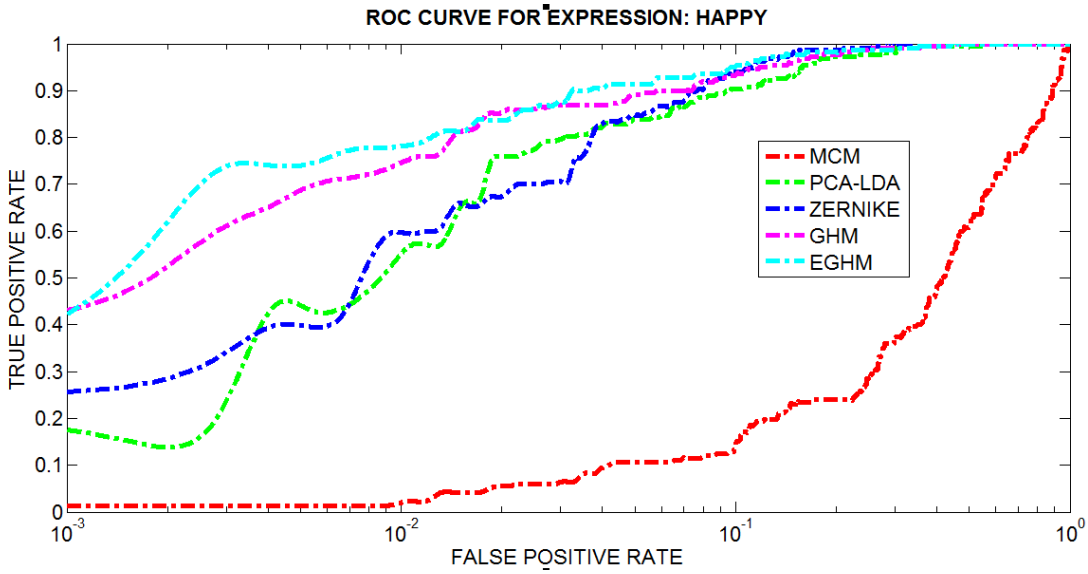



**Fig.4.9**: ROC curve for happy expression in CK dataset

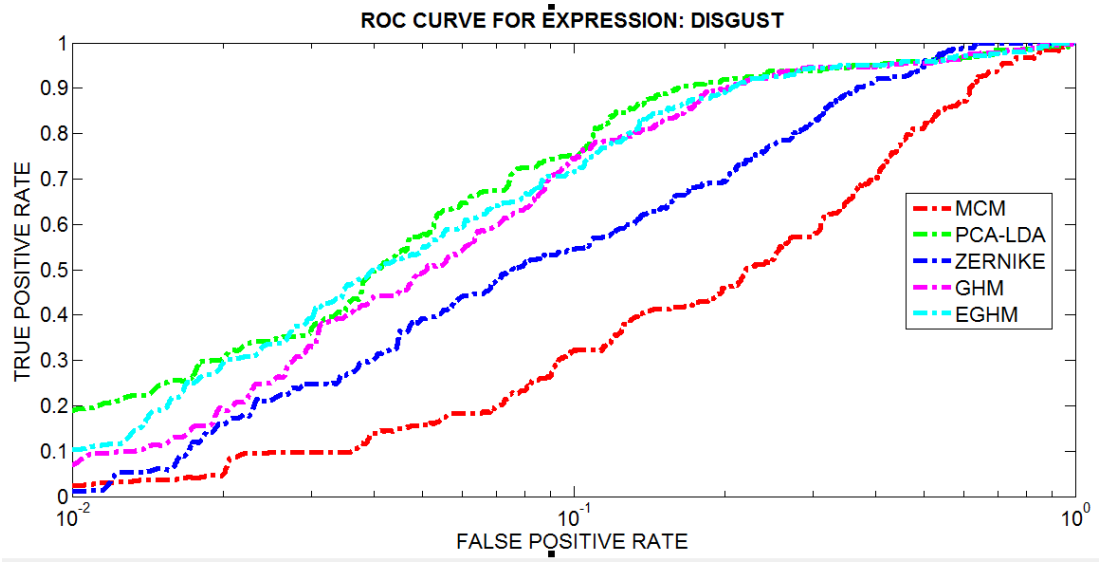



**Fig.4.10**: ROC curve for disgust expression in FRGC dataset.

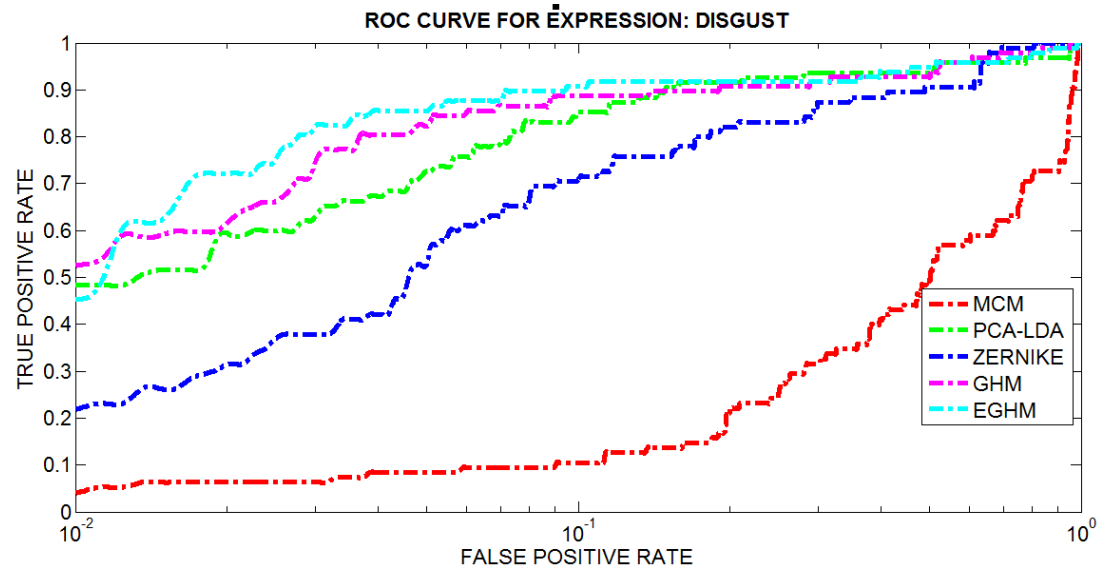



**Fig.4.11**: ROC curve for disgust expression in CK dataset.

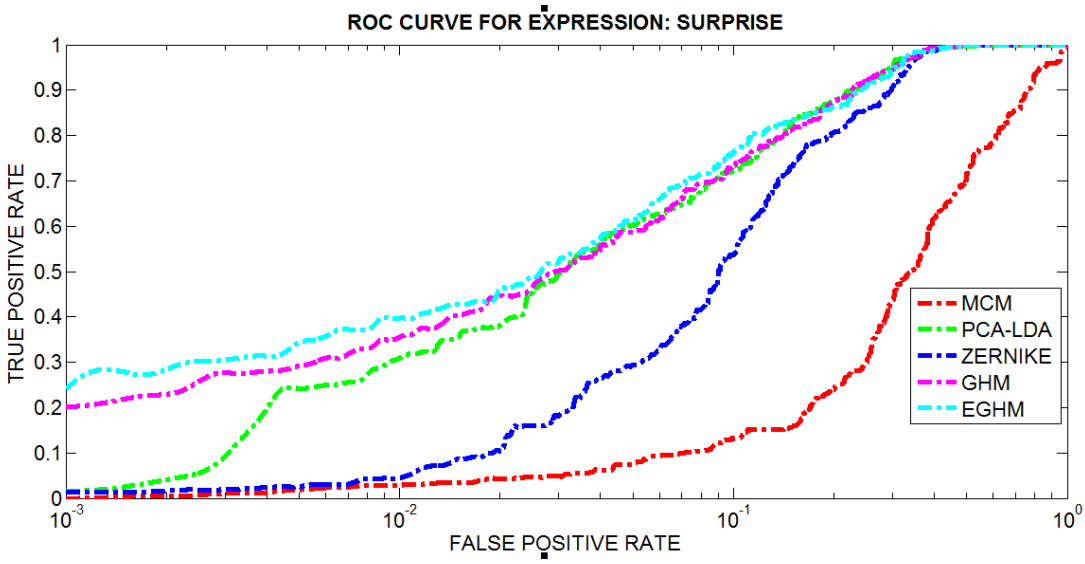



**Fig.4.12**: ROC curve for surprise expression in FRGC dataset.

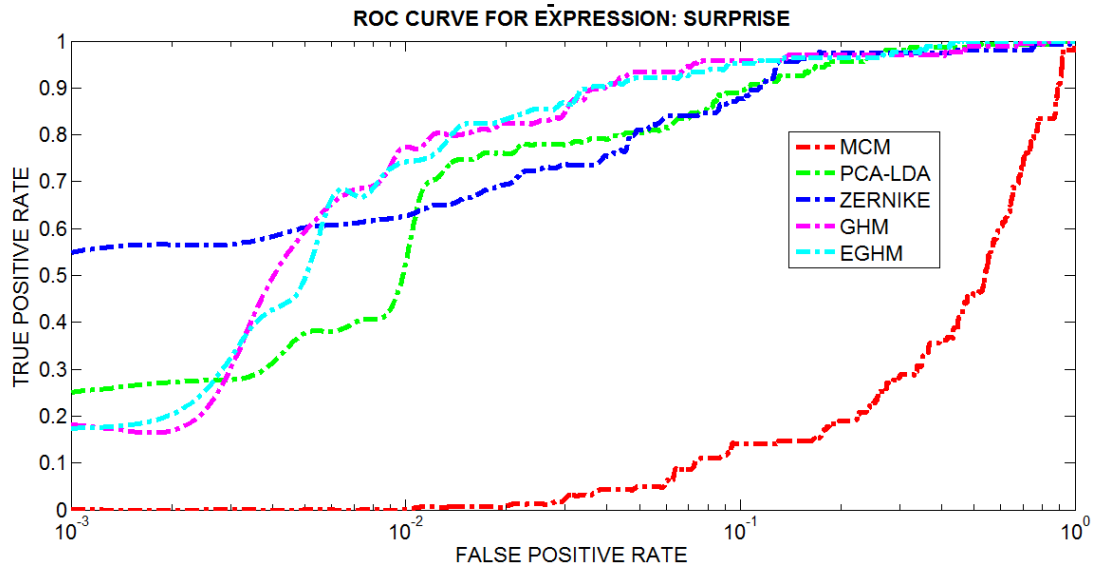



**Fig.4.13**: ROC curve for surprise expression in CK dataset.

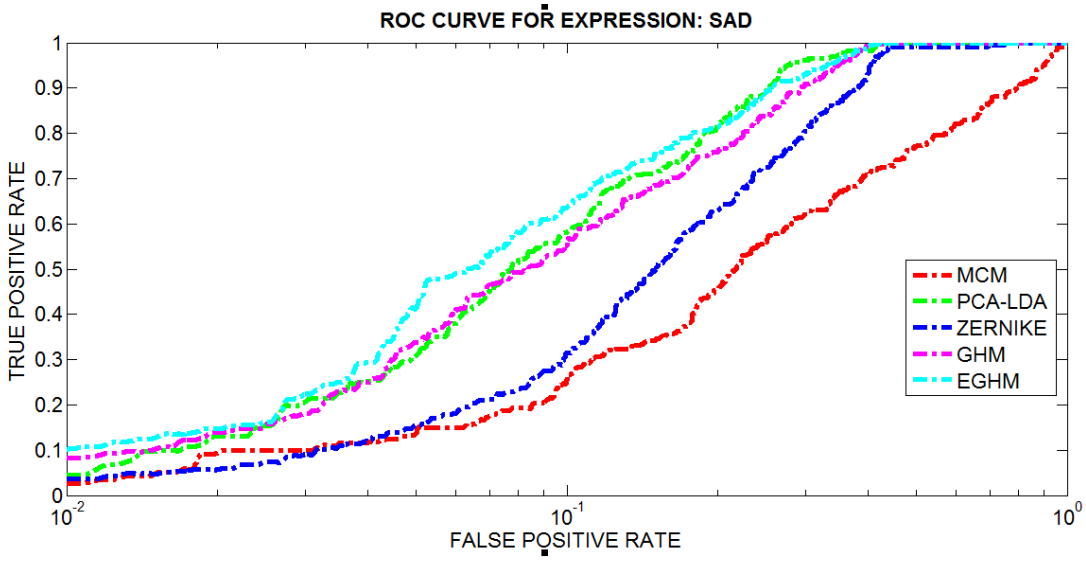



**Fig.4.14**: ROC curve for sad Expression in FRGC dataset.

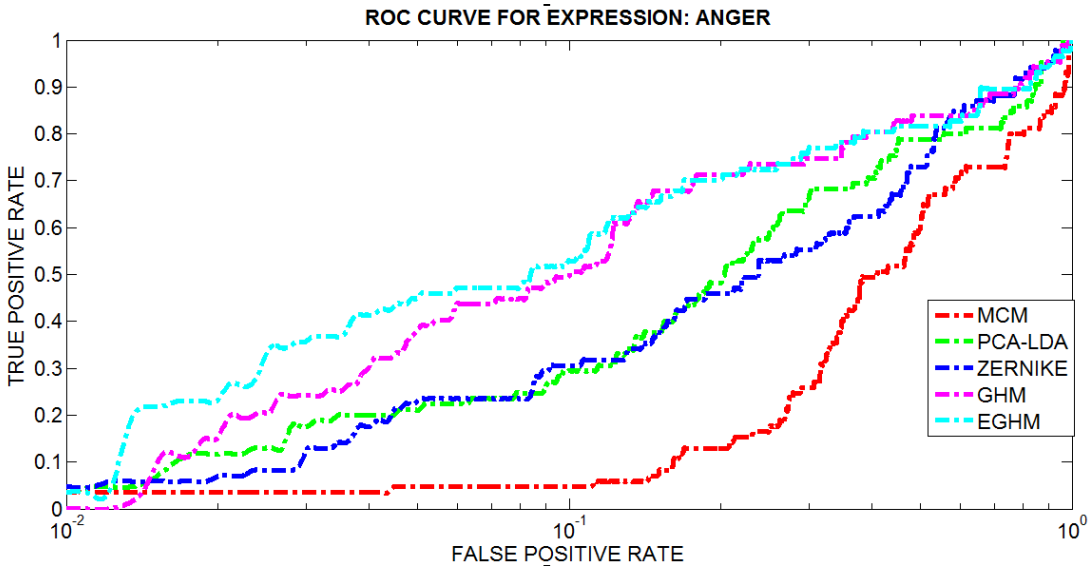



**Fig.4.15**: ROC curve for sad Expression in CK dataset.

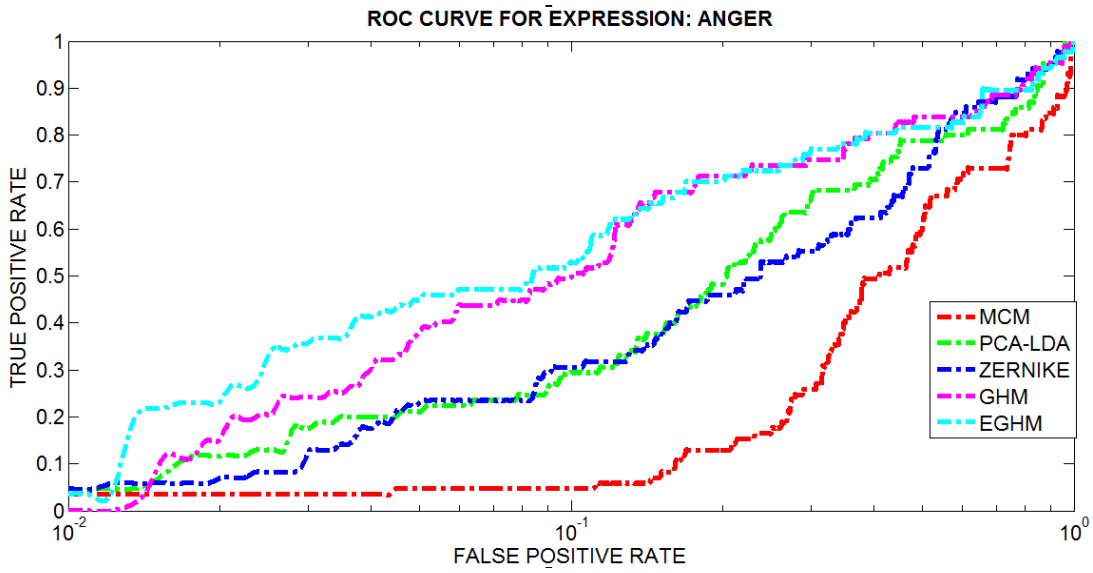



**Fig.4.16**: ROC curve for anger expression in CK dataset.

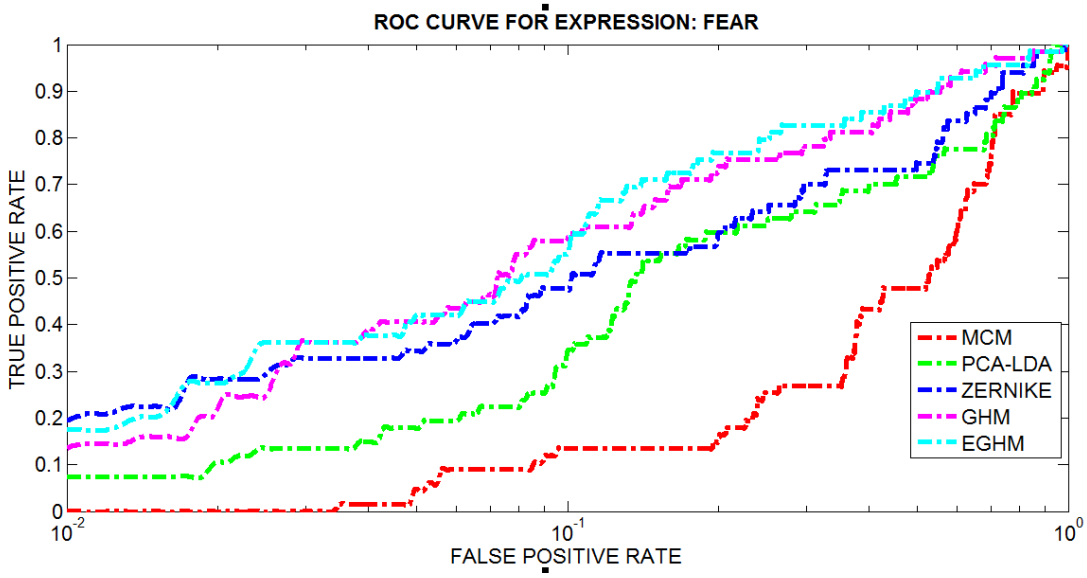



**Fig.4.17**: ROC curve for fear expression in CK dataset.

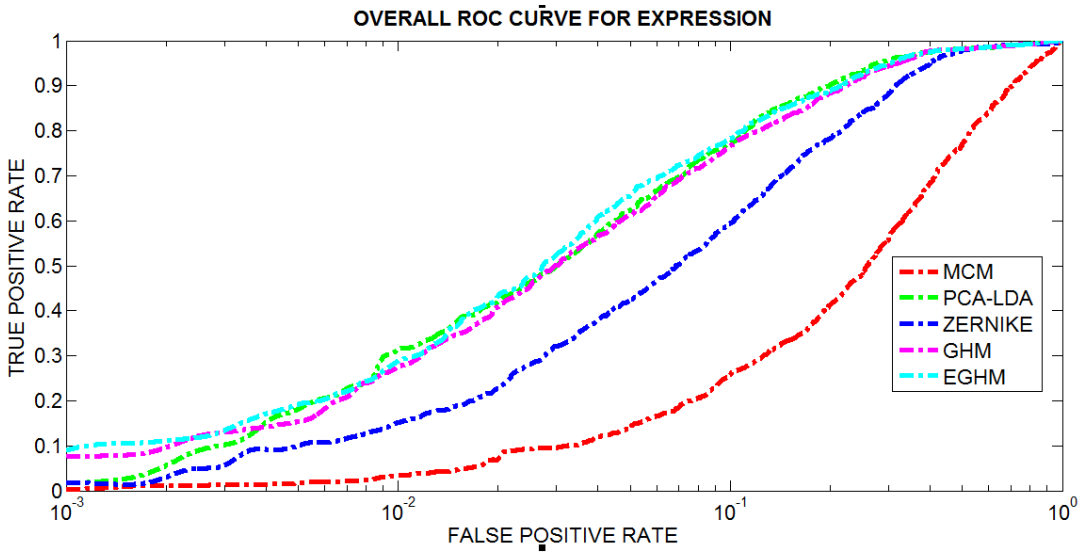



**Fig.4.18**: ROC curve for overall expression in FRGC dataset.

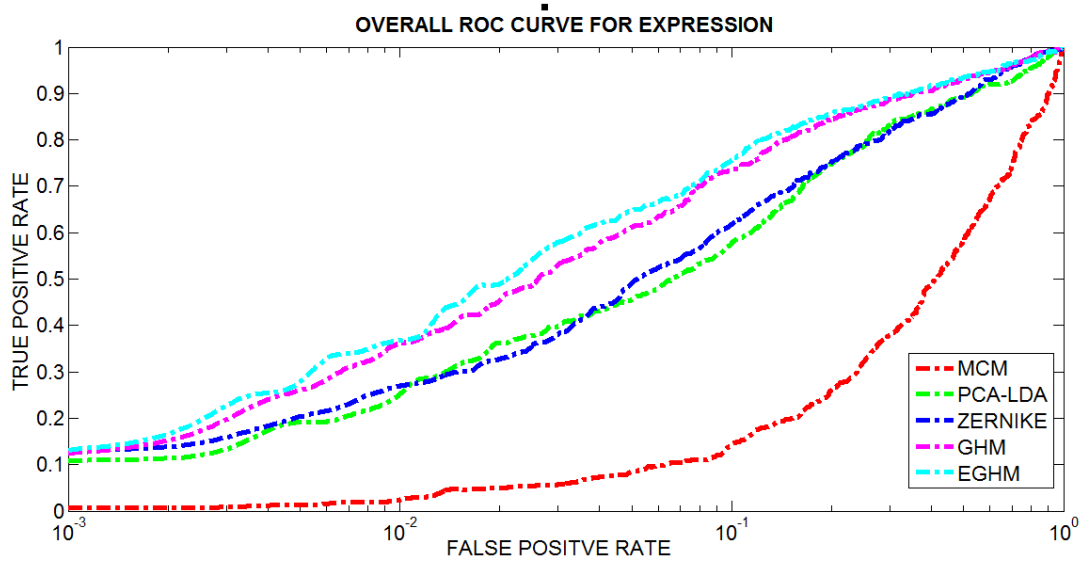



**Fig.4.19**: ROC curve for overall expression in CK dataset.

**TABLE 4.5**:AREA UNDER ROC CURVE (FRGC):

| Expression | MCM | PCA+LDA+SVM | ZM+SVM | GHM+SVM | DGHM+SVM |
|---|---|---|---|---|---|
| Sad | 0.6859 | 0.8893 | 0.8191 | 0.8772 | **0.8956** |
| Surprise | 0.6152 | 0.9238 | 0.8803 | 0.9250 | **0.9286** |
| Disgust | 0.7193 | **0.9102** | 0.8514 | 0.8995 | 0.9038 |
| Happy | 0.6499 | 0.9219 | 0.8890 | 0.9232 | **0.9303** |
| Neutral | 0.7570 | **0.9713** | 0.9438 | 0.9696 | 0.9706 |
| Overall | 0.6855 | 0.9250 | 0.8767 | 0.9189 | **0.9258** |

**TABLE 4.6:**AREA UNDER ROC CURVE (CK):

| Expression | MCM | PCA+LDA+SVM | ZM+SVM | GHM+SVM | DGHM+SVM |
|---|---|---|---|---|---|
| Neutral | 0.6541 | 0.8601 | 0.8313 | 0.8809 | **0.8901** |
| Happy | 0.5589 | 0.9673 | 0.9745 | 0.9796 | **0.9831** |
| Surprise | 0.4677 | 0.92 | 0.8709 | 0.9276 | **0.9331** |
| Disgust | 0.4970 | 0.9611 | 0.9589 | 0.9716 | **0.9761** |
| Anger | 0.5120 | 0.6927 | 0.6815 | 0.7763 | **0.7810** |
| Sad | 0.6509 | 0.7471 | 0.8493 | 0.8597 | **0.8694** |
| Fear | 0.5109 | 0.6948 | 0.7506 | 0.8231 | **0.8349** |
| Overall | 0.5502 | 0.8347 | 0.8453 | 0.8884 | **0.8954** |

4.4.2.4 Significance of using Combined Feature Space

The following table (**Table 4.7**) shows the average classification accuracy using the features from each feature space ($\mathbf{F}_s$ *and* $\mathbf{F}_{Exp}^{\Delta}$ ) individually and in combined state. It can be clearly seen that though $\mathbf{F}_S$ outperforms $F_{Exp}^{\Delta}$ individually by significant

**TABLE 4.7**: AVERAGE ACCURACY USING THE PROPOSED FEATURE SPACE:

| Data-Set | Features from Scattering Ratio $\mathbf{F}_S$ | Features from transformation in neutral space $\mathbf{F}_{Exp}^{\Delta}$ | Combined Feature Space $\mathbf{F}_s^{\Delta} = \left[ \mathbf{F}_s \quad \mathbf{F}_{Exp}^{\Delta} \right]$ |
|---|---|---|---|
| FRGC | 76.16 | 75.35 | 78.37 |
| CK | 77.53 | 46.16 | 77.37 |

margin, in combined space it provides significant improvement for FRGC dataset, but reduces for CK dataset. It is assumed that since there is significant variation amongst within class members, the de-correlated differential expressive information provides significant information to the classifier and as a result, the accuracy increases. However, for CK dataset, since the variations between within class members are negligible (apart from subjects' unique appearance, no significant variation is seen in expression), the differential expressive features provide redundant information and as a result, instead of boosting the classification accuracy, the accuracy slightly falls. It must be noted that the average classification accuracy is attained by considering different (around 14-20 sets) gallery and probe sets for the classifier, where instances/observations in the gallery and probe sets are selected randomly for each case from the testing set.

#### 4.4.2.5 Performance comparison between Person Independent and Person Dependent Setup

In this setup, experiments are carried out to determine what effect the proposed method would have had it known all the subjects' neutral mug-shot images prior to recognizing expression. This is set by taking into consideration the neutral images of all the subjects to build the neutral space. Since the expressions of the neutral subjects are considered, it is expected that the neutral space contains all the information that results in minimum variation across several regions of face images. As a result, the differential expressive features provide significant boost to classification accuracy. As shown in **Table 4.8**, in both the datasets, it is seen that for person dependent setup, the accuracy figures rise up around 5-10%, compared to the classification rate

**TABLE 4.8**: AVERAGE ACCURACY USING PERSON INDEPENDENCY AND PERSON DEPENDENCY:

| Database | Person Independent Setup | Person Dependent Setup |
|---|---|---|
| FRGC | 78.37 | 83.35 |
| CK | 77.43 | 87.61 |

during person independent setup. In this case also, the average classification accuracy is attained by considering different gallery/probe sets for the classifier.

## 4.5 Conclusion

The chapter touched several achievements of the proposed methodology by highlighting the results from different aspects. Different state-of-the-art performance metrics are used for evaluating its effectiveness. From all the results and illustrations, it is clearly seen that the proposed methodology has the capacity to outperform many of the existing holistic facial expression recognition approaches.

# 5 : CONCLUSIONS

## 5.1 Conclusions and Discussions

Automatic facial expression recognition has turned into an emerging and challenging field of research for scientists and engineers. An effective expression recognition strategy can go a long way in making successful automatic mood-meters and lay foundation of successful human-machine interactive systems or even introducing artificial intelligence in the social media. In this thesis, considerable attention has been given to making a sound, robust and low complexity based expression recognition. Out of the two most recognized methods used for recognizing static facial expression images; viz. holistic and landmark based approach, we selected holistic based approach due to its low computational burden in locating and selecting features, less demanding face acquisition tools and low pre-processing requirement of face images. The limitations found from existing holistic expression recognition methodologies have been tried to overcome in this proposition. 2D-GHM moment coefficients, which have ability to capture local texture of an image have been calculated and selected based on the higher discrimination capacity amongst the expression classes by means of scattering ratio. For spontaneous expressions, a neutral de-correlated subspace has been created and features have been selected from that subspace based on the expectation to minimizing the variation caused due to interpersonal change of facial images and highlighting the differential content amongst each expression and the neutral expression.

Although the feature extraction methodology is the major topic of research, several effective and less demanding automatic face acquisition techniques (e.g. Viola Jones) can be easily incorporated with the method. We ensured the person independency of the methodology (method does not depend on prior knowledge of person/subjects) by including mutually exclusive subjects in each expression class set of images. The challenging concept of recognizing posed and spontaneous expressions are also being addressed by using two exhaustive datasets; CK dataset for posed expressions and FRGC dataset for spontaneous expressions respectively. We have also selected the datasets since they represent subjects of diverse ethnicity, age, sex and colors; another challenging aspect of expression recognition problems. And it is found that

the method works equally effective for both posed and spontaneous facial expression images.

For performance evaluation, we have used three existing and popular methods used in facial holistic expression recognition systems. State of the art performance metrics such as the likes of ROC curves and type I and type II errors are used alongside classification accuracy tables to illustrate the possible impact and acceptance of the proposed methodology. Results are also shared what impact the individual feature space and the combined feature space would have on the posed and spontaneous expression data corpus respectively. Based on the results, it is found that the proposed methodology outperforms many of the existing facial expression recognition methods.

## 5.2 Future Works

There are tremendous scopes of future works by analyzing facial expression in Gaussian Hermite Moment space. Some of the scopes are listed out:

- We also would like to investigate feature extraction using GHMs locally on eye, nose and mouth regions to see whether localizing these areas increase chances of expression recognition. For example, Pentland et. al. have used modular eigen-spaces for face recognition and have found some very promising results using this approach [95]. It is yet to be seen whether assigning a higher "priority" to the upper face features than to the lower face features have any significant effect on our method since it is found that upper face features play a more important role in facial expression interpretation as opposed to lower face features [96].
- Although humans can recognize facial expressions quite well from static graylevel images, expression recognition improves with motion information [97]. Static images in many cases, do not clearly reveal subtle changes in faces and it is therefore essential to measure also the dynamics of facial expressions. In our future endeavour, we plan to explore the motion sequences of facial images and like to see if the differences in projected moments of video sequences of particular subjects have any impact on boosting recognition capacity.
- Our future endeavour will also include recognizing expression of faces at different poses and orientation of the faces. It is yet to be seen completely or

to which extent utilizing the rotation and scale invariant properties of GHM have any effect on changes in pose or orientation of the faces since the datasets used in the thesis are all frontal face images.

- We are yet to see how the Gauss-Hermite polynomials would behave if instead of finding the moments, we treat the basis functions as moment filters and convolute the image with some of the orders of these polynomials. In addition to rotation, scale invariance and orthogonal properties, these polynomials could be a very competitive basis functions compared to gabor functions/filters.
- Since many expressions often come in several blended emotions (surprise and happiness, sad and angry, etc), an emotion index is likely to be created that would objectively measure the level of a particular emotion in an expression by using the probability estimates of the classifier about each class.

## REFERENCES


[1] Mehrabian A., "Communication without words", *Psychology Today*, vol. 2, no. 4, pp. 53-56, April 1968.

[2] Ekman P., Sullivan M.O, Friesen W.V., Scherer K.R.,"Face, voice, and body in detecting deceit", *Journal of Nonverbal Behavior*, vol. 15, no. 2, pp. 125-135, Summer 1991.

[3] Derks D., Bos A.E.R., Grumbkow J.V.," Emoticons and social interaction on the Internet: the importance of social context", *Computers in Human Behavior, Elsevier*, vol. 23, no. 1, pp. 842-849, Jan. 2007.

[4] Turk M, "Multimodal human computer interaction", *Real-Time Vision for Human-Computer Interaction*, Springer, pp.269-283, 2005.

[5]Sharma R., Pavlovic V.I., Huang T.S., "Towards multimodal human-computer interface", *Proc. of IEEE*, vol. 86, no. 5, pp. 853 – 869, May 1998.

[6] Bettadapura V.,' Facial expression recognition and analysis: The state of the art', *ArXiv e-prints*, Mar 2012.

[7] Ekman P., Friesen W. V., 'Constants across cultures in the face and emotion', *Journal Personality Social Psychol.*, vol. 17, no.2, pp. 124-129, 1971.

[8] Kaur M., Vashisht R., "Comparative study of facial expression recognition techniques", *Int'l Journal of Computer Applications*, vol. 13, no.1, Jan 2011.

[9] Bassili J.N., 'Facial motion in the perception of faces and of emotional expression', *Journal Experimental Psychology 4*, pp. 373-379, 1978.

[10] Kobayashi H. and Hara F., "Recognition of six basic facial expressions and their strength by neural network", *Proc. Int'l Workshop Robot and Human Comm.*, pp. 381-386, 1992.

[11]Pantic M. and Rothkrantz L.J.M., “Expert system for automatic analysis of facial expression”, *Image and Vision Computing Journal*, vol. 18, no. 11, pp. 881-905, 2000.

[12]Edwards G.J., Cootes T.F., and Taylor C.J., “Face recognition using active appearance models”, *Proc. European Conf. Computer Vision*, vol. 2, pp. 581-595, 1998.

[13] Padgett C., Cottrell G., ‘Representing face images for emotion classification’, *Advances in Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA, pp. 894-900.

[14] Huang C.L. and Huang Y.M., “Facial expression recognition using model-based feature extraction and action parameters classification”, *Journal Visual Comm. and Image Representation*, vol. 8,no. 3, pp. 278-290, 1997.

[15] Alugupally N., Samal A., Marx D., Bhatia S., ‘Analysis of landmarks in recognition of face expressions’, *Pattern Recognition and Image Analysis, Pleiadas Publishing Ltd*, vol. *21,no.4, 2011*.

[16] Zhang Z., Lyons M., Schuster M. and Akamatsu S., ‘Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron’, *Proc. 3rd IEEE Int’l Conf. on Automatic Face and Gesture Recognition*, IEEE Computer Society, pp. 454–459, 1998.

[17] Beretti S., Amor B.B., Daoudi M., Bimbo A.D.,‘3D facial expression recognition using SIFT descriptors of automatically detected keypoints’, *Vision Computer*, vol. 27, no. 11, pp. 1021-1036, 2011.

[18] Li H., Morvan J.-M, Chen L., ‘3D Facial Expression Recognition based on Histograms of Surface Differential Quantities’, *Advanced Concepts for*

*Intelligent Vision Systems*, Lecture Notes in Computer Science, vol. 6915, pp. 483-494, 2011.

[19]Essa I. A. and Pentland A. P., 'Coding, analysis, interpretation and recognition of facial expressions', *IEEE Trans. Pattern Anal. And Mach. Intell.*, vol. 19, pp. 757-763, 1997.

[20] Essa I.A., Pentland A.P., 'Facial expression recognition using dynamic model and motion energy', *Proc. IEEE Computer Vision*, Cambridge, MA, pp. 360-367, 1995.

[21] Ekman P. and Friesen W. V., 'Facial Action Coding System: A Technique for the Measurement of Facial Movement', *Consulting Psychologists Press,* Palo Alto, 1978.

[22] Lanitis A., Taylor C.J. and Cootes T.F., 'Automatic interpretation and coding of face images using flexible models', *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 19, no. 7, July 1997.

[23]Tao H. and Huan T. S., "Connected vibrations: A modal analysis approach to non–rigid motion tracking," *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 753–750, Santa Barbara, 1998.

[24]Bartlett M.S., 'Face Image Analysis by Unsupervised Learning', *Kluwer Academic Publisher*,1998.

[25]Movellan J., 'Visual speech recognition with stochastic networks', In G. Tesauro, D. Touretzky, and T. Leen, editors, 'Advances in Neural Information Processing Systems', *MIT Press,* Cambridge, MA, vol. 7, pp. 851–858, 1995.

[26] Cottrell G. And Fleming M., 'Face recognition using unsupervised feature extraction', *Proc. Int'l Neural Network Conf.*, Dordrecht, Kluwer, pp. 322–325, 1990.

[27] Turk M. and Pentland A., 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience,* vol.3 no. 1,pp.71–86, 1991.

[28] Padgett C., Cottrell G, Adolphs R., 'Categorical perception in facial emotion classification', *Proc. 18th Annual Cognitive Science Conf.*, San Diego, CA, 1996.

[29] Cottrell C.W., Metcalfe J., 'EMPATH: face, gender and emotion recognition using holons', *Advances in Neural Information Processing Systems*, CA, vol. 3, pp. 564-571, 1991.

[30]Belhumeur P., Hespanha J., andKriegman D., 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. and Machine Intell.*, vol.19 no.7, 1997, pp. 711–720.

[31] Etemad K., Chellappa R., 'Discriminant analysis for recognition of human face images', *J. Opt. Soc. America*, vol. 14, no. 8, pp. 1724-1733, 1997.

[32] Pollen D., Ronner S., 'Phase relationship between adjacent simple cells in the visual cortex', *Science 212,* pp. 1409-1411, 1981.

[33] Daugman J., 'Complete discrete 2D Gabor transform by neural networks for image analysis and compression', *IEEE Trans. Acoustics, Speech Signal Process.*, vol. 36, pp. 1169-1179, 1988.

[34] Shan C., Gong S. and McOwan P.W., 'Facial expression recognition based on local binary patterns', *Image and Vision Computing*, vol. 27, no. 9, pp.803-816, 2009.

[35] Rivera A.R., Castillo J.R. and Chae O., 'Local directional number pattern for face analysis: face and expression recognition', *IEEE Trans. Image Processing*, vol. 22, no., 5, May 2013.

[36] Lajevardi S.M., Hussain Z.M., 'Higher order orthogonal moments for invariant facial expression recognition', *Digital Signal Processing Elsevier*, vol. 20, pp. 1771-1779, 2010.

[37] Sridhar D., Krishna I.V.M., 'Face recognition using tchebichef moments', Int'l *Journal of Information and Network Security (IJINS)*, vol.1, no. 4, pp. 243-254, Oct 2012.

[38]Londhe R., Pawar V., 'Facial expression recognition based on affine moment invariants', *Int'l Journal of Computer Science Issues*, vol. 9, no. 2, Nov 2012.

[39] Rani J.S, Devaraj D., 'Face recognition using krawtchouk moments', vol. 37, no. 4, pp. 441-460, *Sadhana, Indian Academy of Science,* Aug 2012.

[40] Pantic M., Rothkrantz L.J.M., 'Automatic analysis of facial expressions: the state of the art', *IEEE Tran. On Pattern Analysis and Machine Intell.*, vol.22, no.12, Dec. 2000.

[41] Bruce.V, 'Recognizing Faces', Hove, East Sussex: Lawrence Erlbaum Assoc., 1986.

[42]V. Bruce, V. 'Human face perception and identification', In H. Wechsler, P. Phillips, V. Bruce, F. Fogelman-Soulie, and T. Huang, editors, 'Face Recognition: From Theory to Applications', *NATO ASI Series F.*, Springer Verlag, 1998.

[43] Yang J., Yang J.Y., 'Why can LDA be performed in PCA transformed space?', *Pattern Recognition*, Elsevier, vol. 36, no. 2, pp. 563–566, Feb. 2003.

[44] Kazmi S.B., Ain Q., Jafar M.A., 'Wavelets-based facial expression recognition using a bank of support vector machines', *Soft Computing, Springer-Verilag*, vol. 16, pp. 369-379, May 2011.

[45] Lajevardi S.M., "Facial expression recognition from image sequences using optimized feature selection," *23rd Int'l Conf. on Image and Vision Computing New Zealand (IVCNZ),*Nov. 2008.

[46] Lajevardi S.M., Hussain Z.M., 'Automatic facial expression recognition: feature extraction and selection', *Signal, Image and Video Processing, Springer*, no. 6, pp. 159-169, Aug 2010.

[47]Zhang Z., Lyons M., Schuster M. and Akamatsu S., 'Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron', *Proc. 3rd IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, IEEE Computer Society, pp. 454–459, 1998.

[48] Ma L., "Facial expression recognition using constructive feed-forward neural networks" , *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, no. 3, pp. 1588 – 1595, June 2004.

[49] Belkasim S.O., Shridhar, Ahmadi M., 'Pattern recognition with moment invariants: A comparative study and new results', *Pattern Recognition*, vol. 24, no. 12, pp. 1117-1138, 1991.

[50] Flusser J., Suk T., 'Pattern recognition by affine moment invariants', *Pattern Recognition*, vol. 26 no. 1, pp. 167-174, 1993.

[51] Luo L.M., Lamitouche C., Dillenseger J.L, Coatrieus J.L, 'A moment based three-dimensional edge operator', *IEEE Trans. Biomedical Engineering*, vol. 40, no. 7, pp. 693-703, 1993.

[52] Liu S.T., Tsai W.H., 'Moment-preserving corner detection', *Pattern Recognition*, vol. 23 no. 5, pp. 441-446, 1990.

[53] Yokoya N., Levine M.D., 'Range image segmentation based on differential geometry: a hybrid approach', *IEEE Trans. Pattern Anal. And Machine Intell.*, vol. 11 no. 6,pp. 643-649, 1989.

[54] Tuceryan M., 'Moment-based texture segmentation', *Pattern Recognition Letters*, vol. 15 no. 7, pp. 659-668, 1994.

[55] Teh C.H., Chin R.T., 'On image analysis by the methods of moments', *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 10, no. 4, pp. 496-513, 1988.

[56] Teage M.R., 'Image analysis via general theory of moments', *Journal Opt. Soc. America*, vol.7, pp. 920-930, 1980.

[57] Lo C.H., Don H.S., '3-D moment forms: their construction and application to object identification and positioning', *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 11 no. 10, pp. 1053-1064, 1989.

[58] Abu-Mostafa Y.S., Psaltis D., 'Recognitive aspects of moment invariants', IEEE Trans. *Pattern Anal. and Machine Intell.*, vol. 6, no. 6, pp. 698-706, 1984.

[59] Hu M. K., 'Pattern recognition by moment invariants', *Proc. IRE 49*, pp. 1428, 1961.

[60] Mukundan R., Ong S.H., and Lee P.A., 'Discrete vs. continous orthogonal moments for image analysis', *Proc. IEEE Int'l Conf. on Imaging Science, Systems and Technology (CISST)*, pp. 23-19, Las Nevada (USA), June 2011.

[61] Kotoulas L., Andreadis I., 'Image analysis using moments', *Proc. IEEE Intl. Conf. on Technology and Automation (ICTA 05)*, pp. 360-364, 2005.

[62] Padilla-Vivanco A., Martiber-Ramirer A., and Granados-Agustin F., 'Digital image reconstruction by using Zernike moments', *Proc. SPIE (5237)*, pp. 281-289, 2004.

[63] Mukundan R., Ongt S.H. and Lee P.A., 'Image analysis by tchebichef moments', *IEEE Trans. Image Processing,* vol. 10, no. 9, pp. 1357-1364, 2001.

[64] Shen J., 'Orthogonal Gaussian-Hermite moments for image characterization', *Proc. of SPIE Intell. Robots Computer Vision XVI*, Pittsburg, 1997.

[65] Shen J., Shen W. and Wu Y.F., 'Orthogonal moments and their application to motion detection in image sequences', *Int'l Journal of Information Acquis.,* vol. 1, no.1, pp. 77-87, 2004.

[66] Wu Y.F., Shen J., 'Properties of orthogonal Gaussian-Hermite moments and their application', EURASIP Journal of Applied Signal Proc., vol. 4, pp. 588-599, 2005.

[67] Wu Y.F., Shen J. and Dai M., 'Traffic object detections and its action analysis', *Pattern Recognition Letters,* vol. 26 no. 13, pp. 1963-1984, 2005.

[68] Wang L., Dai M., 'Application of a new type of singular points in fingerprint classification', Pattern *Recognition Letters*, vol. 28 no. 13, pp. 1640-1650, 2007.

[69] Wang L., Wu Y.F., Dai M., 'Some aspects of Gaussian-Hermite moments in image analysis', *Proc. Int'l Conf. Natural Computation*, Haikou, 2007.

[70] Ma L., Tan T.N., Wang Y. H., Zhang D.X., 'Local intensity variation analysis for iris recognition', *Pattern Recognition,* vol. 37, no. 6, pp. 1287-1298, 2004.

[71] Sun L., Yang J.L, 'Gaussian Hermite moments segmentation of SAR images using the Pearson system distributions', *Proc. European Conf. on Synthetic Aperture Radar*, Dresden, 2006.

[72] Shen W., Xiao Y., 'Stereo matching based on orthogonal Gaussian-Hermite moments', *Proc. SPIE Int'l Symp. Multispectral Image Processing and Pattern Recognition*, Yichang, 2009.

[73] Wu Y., Zhou G. and Wu J., "Recognizing characters based on Gaussian-Hermitemoments and BP Neural Networks," , *Proc. Int'l Conf. Intell. Comput. Tech. Autom.*, pp.992-995, Guiyang 2010.

[74] Wang L., Dai M., "An effective method for extracting singular points in fingerprint images," *Int'l Journal Electron. Commun.*, vol. 60, no. 9, pp. 671–676, Oct. 2006.

[75] Ma L., Tan T., Wang Y. and Zhang D., "Local intensity variation analysis for iris recognition," *Pattern Recognition*, vol. 37, no. 6, pp. 1287–1298, June 2004.

[76] Wang L., Pan X., Niu Z. and Ma X., 'Application of Gaussian-Hermite moments in license', Recent Advances in Document Recognition and Understanding', Dr. Minoru Mori (Ed.), *InTech,* 2011.

[77] Yang B., Li G., Zhang H., Dai M., 'Rotation and translation invariants of Gaussian-Hermite moments', *Pattern Recognition Letters,* vol. 32, pp. 1283-1298, 2011.

[78] Fernandez-Garcia N.L. and Medina-Carnicer R., 'Characterization of empirical discrepancy evaluation measures', *Pattern Recognition Letters,* vol. 25, no. 1, pp. 35-47, 2004.

[79] Flusser J., 2000. 'On the independence of rotation moment invariants', *Pattern Recognition*, vol. 33, no. 9, pp. 1405-1410, 2000.

[80] Ferri F., Pudil P., Hatef M., and Kittler J., 'Comparative study of techniques for large scale feature selection', *Pattern Recognition in Practice IV,* E. Gelsema and L. Kanal, eds., pp. 403-413. Elsevier Science B.V., 1994.

[81] Jain A. andZongker D., 'Feature selection: evaluation, application and small sample performance', *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 19, no. 2, pp. 153-158, Feb 1997.

[82] Miranda A.A., Le Borgne Y.-A. and Bontempi G., 'New routes from minimal approximation error to principal components', *Neural Processing Letters, Springer*, vol. 27, no.3, June 2008.

[83] http://www.pitt.edu/~emotion/ck-spread.htm.

[84] Philips P.J., Flynn P..J., Scruggs T., Bowyer K.W., 'Overview of Face Recognition Grand Challenge', *IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[85] Borg I. and Groenen P.J.F.,'Modern Multidimensional Scaling', *2nd Edition, New York: Springer-Verilag*, 2005.

[86] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Proc. 4th IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 46–53, Grenoble, 2000.

[87] Chuang C-Fa. and Shih F.Y., 'Recognizing Facial Action Units using independent component analysis and support vector machine', *Pattern Recognition*, vol. 39, no. 9, pp. 1795-1798, 2006.

[88] Hsu C-W. and Lin C-J., 'A comparison of methods for multi-class Support Vector Machines', IEEE Trans. Neural Networks, vol. 13, no. 2, pp. 415-425, 2002.

[89] Chang C-C., Lin C-Jen, 'LIBSVM- A library for Support Vector Machines', *ACM Trans. on Intelligent Systems and Technology*, vol. 2, no. 3, April 2011. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[90] Hsu C-Wei, Chang C-C, Lin C-Jen, 'Practical guide to Support Vector classification', Dept. of Computer Science, National Taiwan University, April 2010.

[91] Thomas C.E., Gillies D.F., Feitosa R.Q., 'Using mixture covariance matrixes to improve face and facial expression recognitions', *Pattern Recognition Letters, Elsevier*, vol. 24, pp. 2159-2165, 2003.

[92] Mitra S., Lazar N.A. and Liu Y., 'Understanding the role of facial asymmetry in human face identification', *Stat. Computing*, vol. 17, pp. 57-70, Jan 2007.

[93] Lajevardi S.M., Hussain Z.M., 'Higher order orthogonal moments for invariant facial expression recognition', *Digital Signal Processing, Elsevier*, vol. 20, pp. 1771-1779, 2010.

[94] Fawcett T.,' An Introduction to ROC analysis', *Pattern Recognition Letters, Elsevier*, vol. 27, pp. 861-874, Dec 2005.

[95] Pentland A., Moghaddam B. and Starner T., 'View based and modular eigenspaces for facial recognition', *IEEE Conf. Computer Vision and Pattern Recognition*, 1994.

[96] P. Ekman, 'Emotion in the Human Face'. *Cambridge Univ. Press*,1982.

[97]Wallbott H., 'Effects of distortion of spatial and temporal resolution of video stimulation emotion attributions', *Journal of Nonverbal Behavior*, vol. 15 no. 6, pp. 5–20, 1992.