Rajiv Kapoor, Pallavi Mathur / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 4, Jul-Aug 2013, pp. 82-95

# Face Recognition Using Moments and Wavelets

**Rajiv Kapoor, Pallavi Mathur**
(HOD, Department of Electronics and Communication, Delhi Technological University, Delhi-42
(Department of Electronics and Communication, University of Pune, Pune

## ABSTRACT

**The field of face recognition has been explored a lot and the work is still going on. In this paper we have proposed a novel approach for face recognition using moments. Four feature extraction methods have been used: Hu moments, Zernike moments, Legendre moments and Cumulants. Hu moments include a set of seven moments which are derived from the conventional geometric moments. These moments are invariant against rotation, scaling and translation. Legendre and Zernike moments have an orthogonal basis set and can be used to represent an image with minimum amount of information redundancy. These are based on the theory of orthogonal polynomials and can be used to recover an image from moment invariants. Cumulants are sensitive to image details and therefore are suitable for representing the image features. For feature extraction, moments of different orders are calculated which form the feature vectors. The feature vectors obtained are stored in the database and are compared using three different classifiers. In case of cumulants, we have calculated the bispectrum of images and compressed it using wavelets.**

***Keywords*** **–** Bispectrum, Biwavelant, Face Recognition, Moments, Wavelets,

## I. INTRODUCTION

A face recognition system can be defined as a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. The most convenient way to do this is by comparing selected facial features from the image and a facial database. The face recognition approaches can be classified as: Holistic approach, Region based approach and Hybrid and multimodal approach. Holistic approach involves matching of faces on the basis of entire face. This makes it sensitive to face alignment. Region based approaches work on the principle of applying different processing methods to distinct face regions. They filter out those regions that are mostly affected by expression changes or spurious elements.

Hybrid and multimodal approaches involve greater architectural complexity and hence provide highest accuracy. In hybrid approaches, different approaches like holistic or region based are combined or are used to perform matching separately and then their results are fused.

There are many variations in the facial images that need to be considered for face recognition. These are due to changes in illumination, direction of viewing, facial expression and aging etc. Also, the face images have similar geometrical features which make discriminating one face from the other in the database a challenging task. All these factors make it difficult to represent face images with distinct feature vectors that are invariant to transformation.

In this paper, feature extraction techniques using moments are considered. In image processing, computer vision and related fields, an image moment is a particular weighted average (moment) of the image pixels' intensities, or a function of such moments, generally chosen to have some attractive property or interpretation. These are useful to describe objects after segmentation. Simple properties of the image which can be found via image moments include area (or total intensity), its centroid and information about its orientation. Moments and their functions have been utilized as pattern features in a number of applications. These features can provide global information about the image. Cumulants and wavelets (wavelants) could result in easy distinction among the various faces.

Section 2 mentions the literature review, Section 3 discusses about the databases and the normalization of images, section 4 discusses about the features extraction and section 5 gives the classification. Section 6 is for results and section 7 is for conclusion and future scope.

## II. LITERATURE SURVEY

A lot of work has been done in the field of face recognition using Linear Discriminant Analysis. In [1], the author has proposed a novel Bayesian logistic discriminant (BLD) model which addresses normality and heteroscedasticity (problem in which the LDA algorithm assumes the sample vectors of each class which are generated from underlying multivariate normal distributions of common covariance matrix with different means). Chao-Kuei Hsieh, Shang-Hong Lai and Yung-Chang Chen [2], proposed face recognition using an optical flow based approach. A single 2-D face image with facial expression is used. Information from the computed intrapersonal optical flow and

the synthesized face image are combined in a probabilistic framework for face recognition. However, the proposed integrated system is more computationally costly. Color Space Normalization to enhance the Discriminating Power of Color Spaces for Face Recognition was used in [3]. Here the authors explain the concept of color space normalization (CSN) and two CSN techniques for enhancing the discriminating power of color spaces for face recognition.

In [4], the author has combined Gabor features within the scope of diffusion-distance calculation. This strategy starts from the Gabor filtering that consists of three scales and six orientations. It is followed by the calculation of diffusion distance based on a Bayesian model. The recognition rate of the proposed algorithm reduces while handling the occlusions due to dramatical pose changes. Zhen Lei, Shengcai Liao, Matti Pietikäinen and Stan Z. Li [5], proposed a face representation and recognition approach by exploring information jointly in image space, scale and orientation domains.

In [6], Zhiming Liu and Chengjun Liu present a novel face recognition method by means of fusing color, local spatial and global frequency information. Specifically, the proposed method fuses the multiple features derived from a hybrid color space, the Gabor image representation, the local binary patterns (LBP) and the discrete cosine transform (DCT) of the input image.

Weiwen Zou and Pong C. Yuen [7], proposed multi-image face recognition, instead of using a single still-image-based approach in order to handle complex face image variations in face recognition.

Xiaoyang Tan and Bill Triggs [8] focused mainly on robustness to lighting variations. For this purpose, the author has combined robust illumination normalization, local texture-based face representations, distance transform based matching, kernel based feature extraction and multiple feature fusion. Richard M. Jiang, Danny Crookes and Nie Luo [9], concentrated on feature selection process. In the proposed scheme, global harmonic features instead of disconnected pixels are used, which represent information about 2-D spatial structures. Baochang Zhang, Yongsheng Gao, Sanqiang Zhao and Jianzhuang Liu in [10] worked on high-order local pattern descriptor, local derivative pattern (LDP), for face recognition.

Gui-Fu Lu, Zhong Lin and Zhong Jin [11], proposed a discriminant locality preserving projections based on maximum margin criterion (DLPP/MMC). DLPP/MMC seeks to maximize the difference, rather than the ratio, between the locality preserving between class scatter and locality preserving within class scatter. This method is a little rough and can be improved. In [12], Saeed Dabbaghchian, Masoumeh P. Ghaemmaghami and Aliaghagolzadeh proposed a new category of coefficient selection approach in the Discrete Cosine Transform domain for face recognition. The approach called Discrimination power analysis (DPA) is a statistical analysis based on the DCT coefficients properties and discrimination concept. In [13], Miao Cheng, Bin Fang, Yuan Yan Tang, Taiping Zhang and Jing Wen devised a supervised learning method, called local discriminant subspace embedding (LDSE), to extract discriminative features for face recognition. Imran Naseem, Roberto Togneri and Mohammed Bennamoun [14], in their research, proposed a linear Regression based classification (LRC) for the problem of face identification. Wen-Chung Kao, Ming-Chaihsu and Yueh-Yiing Yang [15] concentrated on recognizing human faces in various lighting conditions.

Shuicheng Yan, Huan Wang, Jianzhuang Liu, Xiaoou Tang and Thomas S. Huang [16], in their paper worked on providing solution to the face recognition problem under the scenarios with spatial misalignments and/or image occlusions. A comprehensive system was presented by Ritwik Kumar, Angelos Barmpoutis, Arunava Banerjee and Baba C. Vemuri [17], for capturing the reflectance properties and shape of the human faces using Tensor Splines. But the method requires at least nine input images with known illumination directions. Also, accurate recovery of Apparent Bidirectional Reflectance Distribution functions ABRDF field from a single image with cast shadows and specularities with no lighting information remains a challenge. In [18], Tae-Kyun Kim, Josef Kittler and Roberto Cipolla have addressed the problem of face recognition by matching image sets. Each set of face images is represented by a subspace (or linear manifold) and recognition is carried out by subspace to subspace matching.

Jiwen Lu and Yap-Peng Tan propose in their paper [19] a parametric regularized locality preserving projections (LPP) method for face recognition. The proposed method is designed to regulate LPP features corresponding to small and zero eigen values, i.e., the noise space and null space, and exploit the discriminant information from the whole space. In [20], a semi-supervised dimensionality reduction method called sparsity preserving discriminant analysis (SPDA) was developed by Lishan Qiao, Songcan Chen and Xiaoyang Tan. This algorithm models the ‘‘locality” and improves the performance of typical LDA. Sang-Ki Kim, Youn Jung Park, Kar-Ann Toh and Sangyoun Lee [21], proposed a feature extraction algorithm called SVM-based discriminant analysis (SVM-DA). Through his approach, the author aims at overcoming the limitations of LDA. Bidirectional principal component analysis (BDPCA) has been used for face recognition in [22] by Chuan-Xian

Ren and Dao-Qing Dai attempt to solve the small sample size problem of PCA.

In [23], Yong Wang and Yi Wu used Complete Neighborhood preserving embedding (CNPE) for the purpose of face recognition. CNPE aims at removing the singularity problem of eigen matrix suffered by Neighborhood preserving embedding (NPE).

## III. DATABASE AND NORMALIZATION

In the presented work, standard ORL database is used. It consists of images of 40 subjects. There are 10 images of each subject in different orientations. This makes a total of 400 images. These images are in gray scale and consist of pixel values ranging from 0 to 255. The size of each image is 92 X 112 pixels. The subjects in the database are shown in Fig 1.

Figure 1: Subjects of the ORL database

The orientations of each subject are like:

Figure 2: Different orientations of a particular subject in ORL database

Before proceeding with the method, image normalization is done. For normalizing an image, its pixel values are divided by the maximum pixel value contained in that image.

## IV. FEATURE EXTRACTION

In image processing, feature extraction is a special form of dimensionality reduction. It involves simplifying the amount of resources required to describe a large set of data accurately. In the presented work, features are extracted from images using four methods: Hu moments, Zernike moments, Legendre moments and cumulants.

### 4.1 Hu moments

Hu moments have been derived from the geometric moments. Geometric moments are also known as regular or raw moments. These are nonnegative integers, which can be computed by equation (1):

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)dxdy \quad (1)$$

where p,q = 0,1,2,3……

In equation (1), $m_{pq}$ is the moment of order $(p+q)$ and $f(x,y)$ is a two dimensional, real and continuous density function of which the moment has to be calculated. For digital image we need a discrete version of the above equation represented as in equation (2):

$$m_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} x^p y^q f(x,y) \quad (2)$$

Here, M×N is the size of the image. For the database used in the presented work, the values of M and N will be 92 and 112 respectively.

The central moment of order *(p+q)* of the image is defined by equation (3):

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy \quad (3)$$

Here, $\bar{x}$ and $\bar{y}$ are the centroids of the image and are defined by:

$$\bar{x} = \frac{m_{1,0}}{m_{0,0}} \text{ and } \bar{y} = \frac{m_{0,1}}{m_{0,0}} \quad (4)$$

Now, if f(x, y) is the density function for a digital image, equation (3) in discrete format becomes:

$$\mu_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \quad (5)$$

The above defined central moments are origin independent and therefore they are translation invariant. But these moments are not invariant to scale or rotation in their original form.

Moments of order two or more can be constructed to be invariant to both translation and scale. Scale invariance can be achieved by dividing the corresponding central moment by the scaled energy of the original i.e. the $00^{th}$ moment as shown in equation (6):

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{oo}^{\left(1+\frac{i+j}{2}\right)}} \quad (6)$$

Hu, in 1962 introduced 'Hu moments' which are invariant to translation, scale as well as rotation. Hu moments are a set of seven moments which are nonlinear combinations of normalized central moments up to order three. The seven Hu moments are shown by equation (7):

$$\begin{aligned}
M_1 &= \eta_{20} + \eta_{02} \\
M_2 &= (\eta_{20} - \eta_{20})^2 + 4\eta_{11}^2 \\
M_3 &= (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \\
M_4 &= (\eta_{30} - \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
M_5 &= (3\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\
M_6 &= (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
M_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + \\
&\quad (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{30})^2\right]
\end{aligned} \quad (7)$$

The first moment, $M_1$, is analogous to the moment of inertia around the image's centroid, where the pixels' intensities are equivalent to physical density. The last one, $M_7$, is skew invariance. It enables it to distinguish mirror images of otherwise identical images. In our experiment conducted using Hu moments, all the seven Hu moments are calculated for each facial image.

Table 1 consists of 10 feature vectors for 10 orientations of subject 2 in the database.

Table 1: Hu moments of subject 2

| Image no | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_5$ | $M_6$ | $M_7$ |
|---|---|---|---|---|---|---|---|
| 2-1 | 0.0012 | 1.1145 | 1.7522 | 2.7004 | 1.0397 | 2.354 | -2.9259 |
| 2-2 | 0.0012 | 0.9822 | 1.6067 | 1.7262 | -0.6716 | 1.909 | 0.1064 |
| 2-3 | 0.0012 | 0.9967 | 2.451 | 3.2528 | -2.0731 | 2.449 | -0.4848 |
| 2-4 | 0.0012 | 1.0766 | 1.7206 | 1.4823 | -1.1005 | 0.287 | 0.4849 |
| 2-5 | 0.0012 | 0.8636 | 0.2311 | 1.3562 | -0.2185 | -4.34 | -0.0452 |
| 2-6 | 0.0012 | 0.9797 | 2.0262 | 1.8414 | -2.2178 | -1.646 | 1.6047 |
| 2-7 | 0.0012 | 1.1682 | 3.7579 | 4.4563 | 5.4096 | 13.215 | 1.0396 |
| 2-8 | 0.0012 | 1.0005 | 2.2362 | 2.5456 | 0.842 | 5.723 | 0.0691 |
| 2-9 | 0.0012 | 1.5421 | 2.3615 | 2.0594 | -3.0265 | -3.042 | 2.1438 |
| 2-10 | 0.0012 | 0.9115 | 2.5089 | 2.6352 | -2.7682 | 1.922 | 0.2446 |

In the above table, we see that the first moment i.e. $M_1$ is same for all the orientations of the subject. Also, the variations in the lower order moments are less but in the higher order moments the variations are high. In $M_5$, $M_6$ and $M_7$, there are certain values which are negative also and hence the variation is too large. This affects the performance of the classifier and results in poor classification. The feature vector is formed using the calculated moment values which makes the length of the feature vector seven.

**4.2 Zernike moments**

Zernike in 1934 introduced a set of complex polynomials called the Zernike polynomials. These polynomials form a complete orthogonal basis set defined on the interior of the unit disc, i.e., $x^2 + y^2 = 1$. Let us denote the set of these polynomials by $V_{nm}(x,y)$. In polar coordinates, these are then represented as in equation (8):

$$V_{nm}(x,y) = V_{nm}(r,\theta) = R_{nm}(r)\exp(jm\theta) \quad (8)$$

Where:
'n' is positive integer or zero
'm' is positive and negative integer subject to constraint ( n-|m|) is even and $|m| \le n$
'r' is length of vector from origin to ( x , y) pixel
'θ' is angle between vector ρ and x axis in counter clockwise direction
In the above equation, $R_{mn}$ is the Radial polynomial. It is orthogonal and is defined as in equation (9):

$$R_{nm}(r) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)! r^{n-2s}}{s!\left[\frac{n+|m|}{2}-s\right]!\left[\frac{n-|m|}{2}-s\right]!} \quad (9)$$

It is to be noted that $R_{n-m}(r) = R_{nm}(r)$.

Complex Zernike moments are constructed using the set of Zernike polynomials. Zernike moments are the projection of the image density function on these orthogonal basis functions. The two dimensional Zernike moment for a continuous image function f (x, y) that disappears outside the unit circle is defined as in equation (10):

$$Z_{nm} = \frac{n+1}{\pi} \iint_{\text{unit disk}} f(x,y) V_{nm}^*(x,y)\, dxdy \quad (10)$$

Since in our case, the image is digital, so the integrals are replaced by summations as shown in equation (11):

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(x,y) \quad (11)$$

Here, $x^2 + y^2 \le 1$

To calculate the Zernike moments of a particular image, we first take the centre of image as the origin and map the region of interest to the range of unit disc. Those image pixels which do not fall inside the unit disc are not used for computation. Thereafter, the image coordinates are described in terms of the length of the vector from the origin, the angle from the axis to the vector etc.

In the experiment conducted by us, Zernike moments of different orders (2 to 8) are calculated.
Table 2 consists of 10 feature vectors for 10 orientations of subject 2 in the database.

Table 2: Zernike moments of subject 2

| Image no | $Z_{20}$ | $Z_{31}$ | $Z_{42}$ | $Z_{40}$ | $Z_{51}$ | $Z_{53}$ | $Z_{62}$ |
|---|---|---|---|---|---|---|---|
| 2-1 | -1.102 | -6.4727 | -1.3869 | -7.4029 | -3.2025 | 3.3126 | -0.4755 |
| 2-2 | -1.099 | -6.6799 | -1.1644 | -7.2521 | -3.5483 | 2.2785 | 0.078 |
| 2-3 | -1.0975 | -6.825 | -1.3694 | -7.1614 | -3.6634 | 2.586 | -0.0683 |
| 2-4 | -1.0854 | -6.7228 | -1.2138 | -7.1546 | -3.5915 | 2.2853 | 0.046 |
| 2-5 | -1.1187 | -6.7846 | -0.9088 | -7.494 | -3.6191 | 2.7186 | 0.4117 |
| 2-6 | -1.1037 | -6.9002 | -1.3914 | -7.3231 | -3.8162 | 2.0686 | -0.2327 |
| 2-7 | -1.1259 | -7.0017 | -2.4854 | -7.3039 | -3.4261 | 2.067 | -1.2901 |
| 2-8 | -1.1072 | -7.0425 | -1.7104 | -7.2503 | -3.7064 | 2.4051 | -0.3873 |
| 2-9 | -1.1033 | -7.0592 | -1.9559 | -7.0575 | -3.772 | 1.2888 | -0.687 |
| 2-10 | -1.142 | -7.1451 | -1.7801 | -7.4172 | -3.7097 | 2.1011 | -0.3757 |

The results using Zernike moments are better than those using Hu moments. This is because we can see in the table above that the variations in the Zernike moments are less as compared to those in the Hu moments. Here again, the variations in lower order moments are less but those in higher order moments are large. We also calculated moments of order higher than 8 but they only degraded the performance of the classifier. The feature vector is formed using these calculated moments. The size of the feature vector is 7.

### 4.3 Legendre Moments

Legendre moments were derived from Legendre polynomials as kernel function. Legendre polynomials were first proposed by Teague. These are orthogonal moments which can represent an image with minimum information redundancy. Thus the moments represent the independent characteristics of an image.

The two-dimensional Legendre moments of order (p+q), are defined as in equation (12):

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{1}\int_{-1}^{1} P_p(x) P_q(y) f(x,y)\, dxdy \quad (12)$$

where p,q = 0,1,2,3…….∞ and x,y ∈ [-1,1].
$P_p$ and $P_q$ are Legendre polynomials and f(x, y) is the continuous image function. These Legendre polynomials define a complete orthogonal basis set over the interval [-1, 1]. In order to maintain orthogonality in the moments as well, the image

function is also defined over the same interval i.e. [-1, 1]. The Legendre polynomial of order 'n' is defined as in equation (13):

$$P_n(x)=\sum_{k=0}^{n}\left\{(-1)^{\frac{n-k}{2}}\frac{1}{2^n}\frac{(n+k)!x^k}{\left(\frac{n-k}{2}\right)!\left(\frac{n+k}{2}\right)!k!}\right\} \quad (13)$$

In the above definition, (n-k) should be even. Now, the Legendre moments for a discrete image consisting of pixels $M\times N$ with intensity function $f(i, j)$ are defined as in equation (14):

$$L_{pq}=\lambda_{pq}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}P_p(x_i)P_q(y_j)f(i,j) \quad (14)$$

where $\lambda_{pq}$ is called the normalizing constant and is defined as in equation (15):

$$\lambda_{pq}=\frac{(2p+1)(2q+1)}{MN} \quad (15)$$

Where, $P_0(x)=1$ and $P_1(x)=x$.

The image is scaled in the region [-1, 1] since the definition of Legendre polynomials exists in this region only. So, $x_i$ and $y_j$ can be normalized as in equation (16):

$$x_i=\frac{2i}{(M-1)}-1 \quad \text{and} \quad y_j=\frac{2j}{(N-1)}-1 \quad (16)$$

Legendre moments are orthogonal moments and can be used to represent an image with minimum redundancy.

In the experiment conducted using Legendre moments, moments of different orders (2 to 4) are calculated. Table 3 shows the values of Legendre moments calculated for subject 2 in the database.

Table 3: Legendre moments subject 2

| Image no | $L_{02}$ | $L_{20}$ | $L_{11}$ | $L_{21}$ | $L_{12}$ | $L_{30}$ | $L_{03}$ | $L_{22}$ | $L_{40}$ | $L_{04}$ | $L_{31}$ | $L_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2-1 | -233.06 | -234.63 | 3.79 | -0.42 | 5.59 | 12.5533 | 19.38 | 368.88 | 273.13 | 279.06 | -4.08 | 2.32 |
| 2-2 | -230.32 | -239.42 | -2.10 | -13.65 | 4.43 | 10.4496 | 14.48 | 364.07 | 277.42 | 277.39 | 0.632 | 0.50 |
| 2-3 | -236.67 | -242.65 | 0.74 | -1.77 | 5.21 | 12.2524 | 18.89 | 376.32 | 281.02 | 284.91 | -2.59 | 0.96 |
| 2-4 | -226.34 | -236.17 | -2.81 | -13.22 | 1.64 | 11.1439 | 15.09 | 355.67 | 273.73 | 277.14 | 1.585 | 0.18 |
| 2-5 | -234.10 | -240.16 | 0.74 | -9.22 | 5.77 | 12.5979 | 9.798 | 365.07 | 281.92 | 278.91 | -2.87 | 1.59 |
| 2-6 | -228.53 | -235.15 | -5.08 | -23.25 | 3.69 | 9.8463 | 17.82 | 353.84 | 281.71 | 276.22 | 3.230 | -0.11 |
| 2-7 | -246.27 | -250.07 | 1.91 | 8.76 | 5.21 | 14.7457 | 27.45 | 387.81 | 296.92 | 307.27 | -1.36 | 3.082 |
| 2-8 | -237.68 | -242.04 | 0.33 | 0.23 | 3.06 | 13.5121 | 22.09 | 375.32 | 283.58 | 290.96 | -0.79 | 2.924 |
| 2-9 | -235.63 | -251.38 | -5.90 | -13.74 | 9.56 | 8.7386 | 24.65 | 377.40 | 295.51 | 299.23 | 4.193 | -0.67 |
| 2-10 | -250.66 | -252.53 | 1.08 | 2.03 | 8.07 | 14.0975 | 23.26 | 389.54 | 304.00 | 305.08 | -1.71 | 2.035 |

The results using Legendre moments are better than those using Hu moments and comparable to those of Zernike moments. We can see in the table above that the variations are less. Here again, the variations in lower order moments are less but those in higher order moments are large. The variations in higher order moments were large which degraded the performance of the classifier. Hence, for the construction of feature vectors we have calculated only lower order moments (up to four). The feature vector is formed using these calculated moments. The size of the feature vector is 12.

### 4.4 Cumulants

Cumulants are quantities that are similar to moments and can be used to extract the inherent features of images. Cumulants can extract features which are otherwise very difficult to extract. Simple moments are used to derive cumulants. The $r^{th}$ moment of a real valued continuous random variable $X$ with probability density function $f(x)$ can be defined as shown in equation (17):

$$\mu_r=E(X^r)=\int_{-\infty}^{\infty}x^r f(x)dx \quad (17)$$

where $r$ is a finite integer i.e. $r = 0,1, \ldots .$

All the moments can be represented with a single expression if the moment generating function has a Taylor expansion about the origin which is defined as in equation (18):

$$M(\xi)=E(e^{\xi X})=E(1+\xi X+.....+\xi^r X^r/r!+.....)=\sum_{r=0}^{\infty}\mu_r\xi^r/r! \quad (18)$$

The $r^{th}$ derivative is the $r^{th}$ moment of $M$ at the origin.

Similarly, the cumulants represented as $\kappa_r$ are the coefficients of the Taylor expansion of the

cumulants generating function about the origin defined as in equation (19):

$$K(\xi) = \log M(\xi) = \sum_r \kappa_r \xi^r / r! \quad (19)$$

The second and third order cumulants of a zero-mean stationary process *u* (n) are defined as in equation (20):

$$\kappa_2(\tau) = E\{u(n)u(n+\tau)\}$$

$$\kappa_3(\tau_1, \tau_2) = E\{u(n)u(n+\tau_1)u(n+\tau_2)\} \quad (20)$$

The first-order cumulant of a stationary process is the mean of the process and the second order cumulant is its variance. The higher-order cumulants are equivalent to the central moments; they are invariant to the mean shift and hence defined under the assumption of zero mean.

4.4.1 Polyspectra

The k[th] order polyspectrum or the k[th] order cumulants spectrum is the Fourier transform of the corresponding cumulants.

A polyspectrum exists when the corresponding k[th] order cumulants are absolutely summable. Special cases of the k[th] order polyspectrum are the power spectrum, bispectrum and the trispectrum [24].

For k = 2, we have the ordinary power spectrum defined by:

$$C_2(\omega_1) = \sum_{\tau_1=-\infty}^{\infty} \kappa_2(\tau_1) \exp(-j\omega_1\tau_1) \quad (21)$$

For k = 3, we have the bispectrum, defined by:

$$C_3(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \kappa_3(\tau_1, \tau_2) \exp[-j(\omega_1\tau_1 + \omega_2\tau_2)] \quad (22)$$

While the power spectrum is real-valued, bispectrum, and other higher order polyspectra are complex in nature.

For this experiment using cumulants, we first find the third order cumulant of the image and then take its Fourier transform to get the bispectrum. On calculating the bispectrum of the different facial images, we find that the patterns obtained are similar for the images of the same person (in different orientations) and different for the images of different persons. Few orientations of subject 1 are shown in Fig 3:

Figure 3: Four orientations of subject 1

Fig 4 shows the bispectrum patterns of the subject in Fig 3.

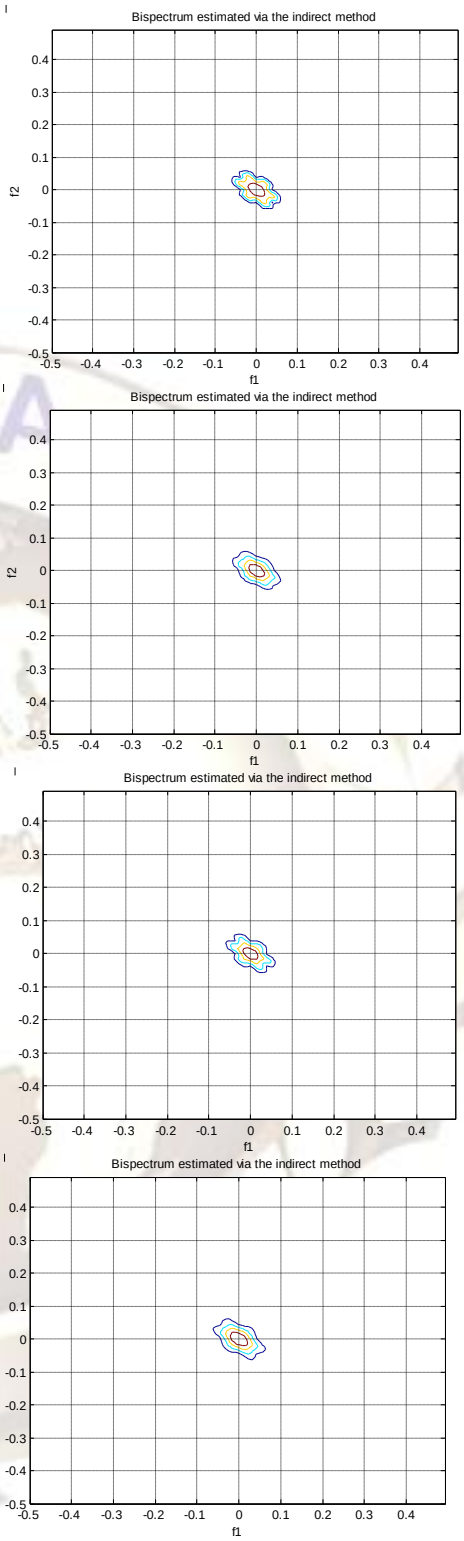



Figure 4: Bispectrum patterns of some orientations of subject 1

All the above patterns are similar to each other since they belong to the images of same person with different orientations. Fig 6 shows the bispectrum patterns of subject 4 in Fig 5.

Figure 5: Four orientations of subject 4

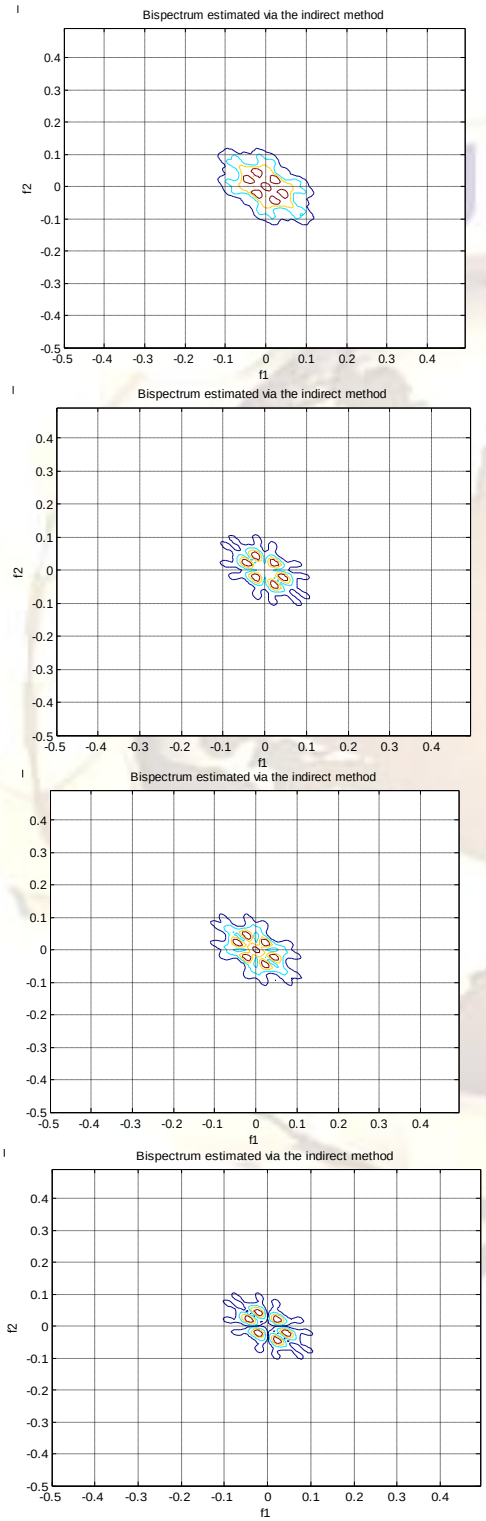


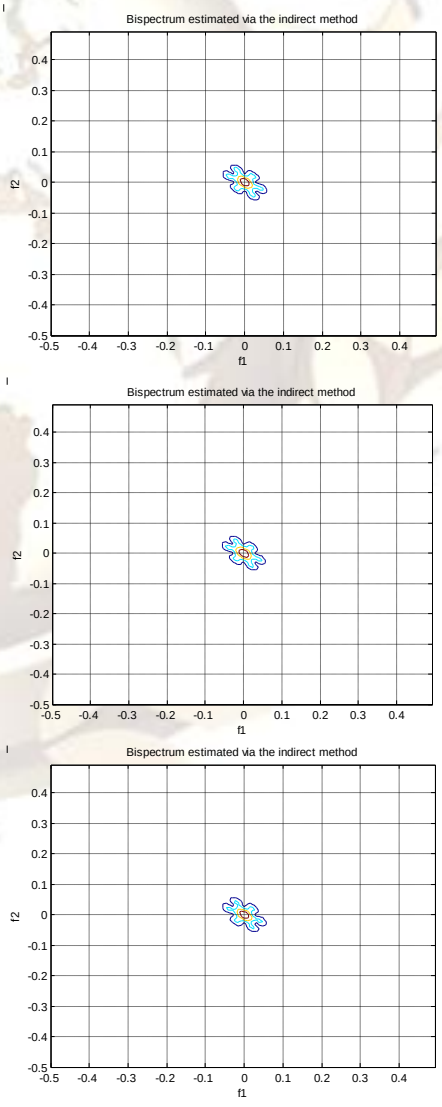

Figure 6: Bispectrum patterns of some orientations of subject 4

Again we notice that all the above patterns are similar to each other since they belong to the images of same person with different orientations. Also, the patterns of subject 1 are different from those of subject 4, which shows that they can be classified easily.

To ensure that the bispectrum patterns for all the subjects are different, we present below one more example.

Figure 7: Four orientations of subject 10

Fig 8 shows the bispectrum patterns of the orientations of subject shown in Fig 7.



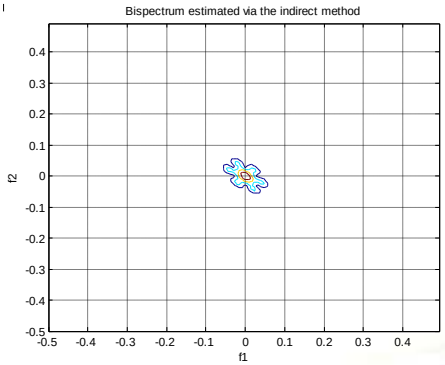



Figure 8: Bispectrum patterns of some orientations of subject 10

Again we see that the orientation patterns of subject 10 are different both from subject 1 and subject 4. Hence, we can conclude that these patterns of different orientations of same subjects are similar while those of different subjects are different. This fact can be used to classify the different subjects. For classification, we will form the biwavelant of the above patterns.

4.4.2 Biwavelant

For using bispectra for face recognition, the redundant information i.e. the high frequency components have to be removed and only the low pass information which corresponds to the significant features are retained. The two dimensional bispectrum matrix is converted into one dimensional matrix by concatenating the rows horizontally. The frequency response is given as:

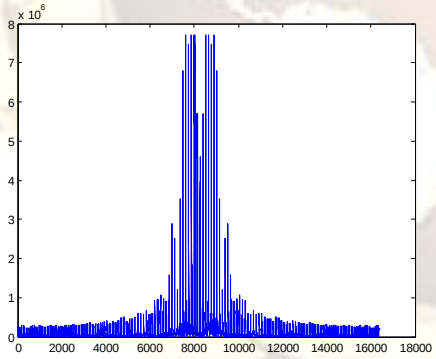



Figure 9: One dimensional frequency response of bispectrum

The frequency response above consists of very large number of values. To reduce the amount of computation, we try to find out its envelope (by removing the high frequency components). For this, we need to carry out low pass filtering which can be done using wavelets. Smoothing filters can also be used but they are not able to protect the precious details while removing the high pass information.

Wavelet transform was explained in [25] and [26]. It is the integral transform defined as in equation (23).

$$W_{\psi f}(a,b)=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty}\overline{\psi\left(\frac{x-b}{a}\right)}f(x)dx \tag{23}$$

where *f(x)* is the signal being transformed and $\psi(x)$ is the 'analyzing wavelet' $\psi(x)$ satisfies the admissibility condition $\int|\psi(x)|^2\frac{d\omega}{\omega}<\infty$ which is equivalent to $\int\psi(t)dt=0$ i.e. a wavelet has zero mean.

The wavelet coefficients are given by equation (24):

$$c_{jk}=W_{\psi f}\left(2^{-j},k2^{-j}\right) \tag{24}$$

Here, $a=2^{-j}$ is called the binary dilation or dyadic dilation, and $b=k2^{-j}$ is the binary or dyadic position.

When we use the continuous wavelet transform, we can detect a signal buried in Gaussian noise. This fact has been used in the concept of wavelants. Before understanding wavelants, we need to know two properties of cumulants [27]:

1) The third order cumulant of a Gaussian (or any symmetrically distributed) random process is zero.
2) If a subset of random variables $\{x_i\}$ is independent of the rest, then the third-order cumulants is zero.

The above formulation exhibits properties closely related to those of cumulants [28]. The ideas from the two preceding topics have led to the motivation for the development of wavelants which is a combination of wavelet and cumulant theory. In the following section we shall consider only the third order wavelant which is defined as in equation (25):

$$W^3_{xxx}(b_1,a_1;b_2,a_2)=\frac{1}{\sqrt{a_1a_2}}\int\int x(t)x\left(\frac{t-b_1}{a_1}\right)x\left(\frac{t-b_2}{a_2}\right)dt \tag{25}$$

$$W^3_{xyz}(b_1,a_1;b_2,a_2)=\frac{1}{\sqrt{a_1a_2}}\int\int x(t)y\left(\frac{t-b_1}{a_1}\right)z\left(\frac{t-b_2}{a_2}\right)dt \tag{26}$$

Equation (25) represents the third order auto wavelant while equation (26) represents the third order cross wavelant. Now we apply the wavelet transform (db4) at different levels on the frequency response obtained earlier (example of which is shown in Fig 9). The result of applying the transform on an image of subject 1 is shown in Figure 10.

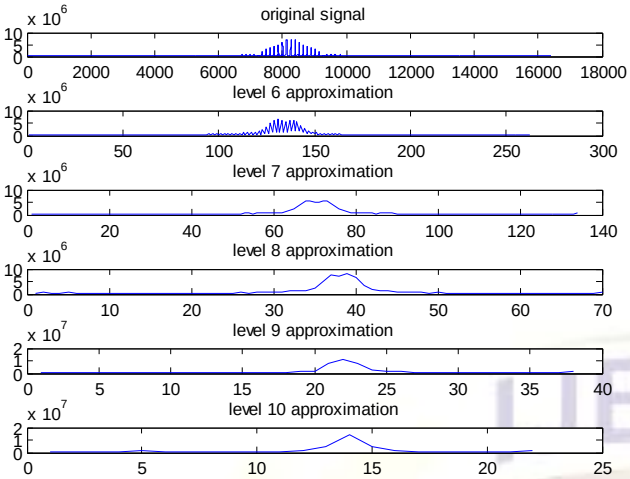



Figure 10: waveforms for subject 1 at various approximations]

We notice from the above figure, that we obtain a nice envelope at Level 7 approximation. This seems to be the most appropriate approximation level for classification because in further approximations, details are being lost. The similar waveforms for subjects 4 and 10 are shown below.

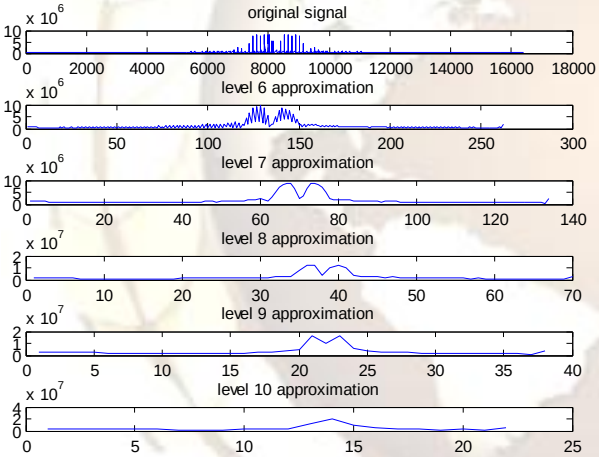



Figure 11: waveforms for subject 4 at various approximations

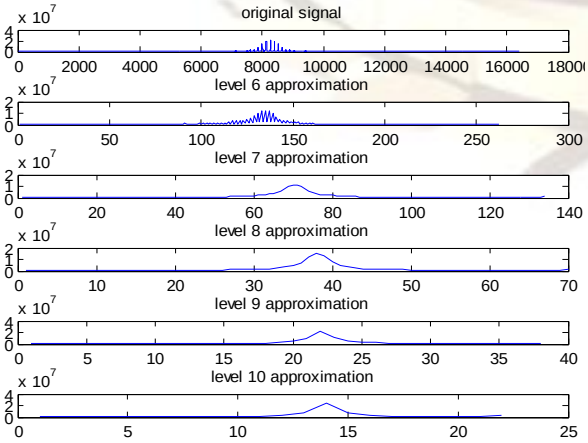



Figure 12: waveforms for subject 10 at various approximations

In all the above waveforms, level 7 approximation is most appropriate and also at this level all the three waveforms can easily be distinguished from one another. Let us now have a closer look at the level seven approximation of the frequency response for few orientations of all the three subjects.

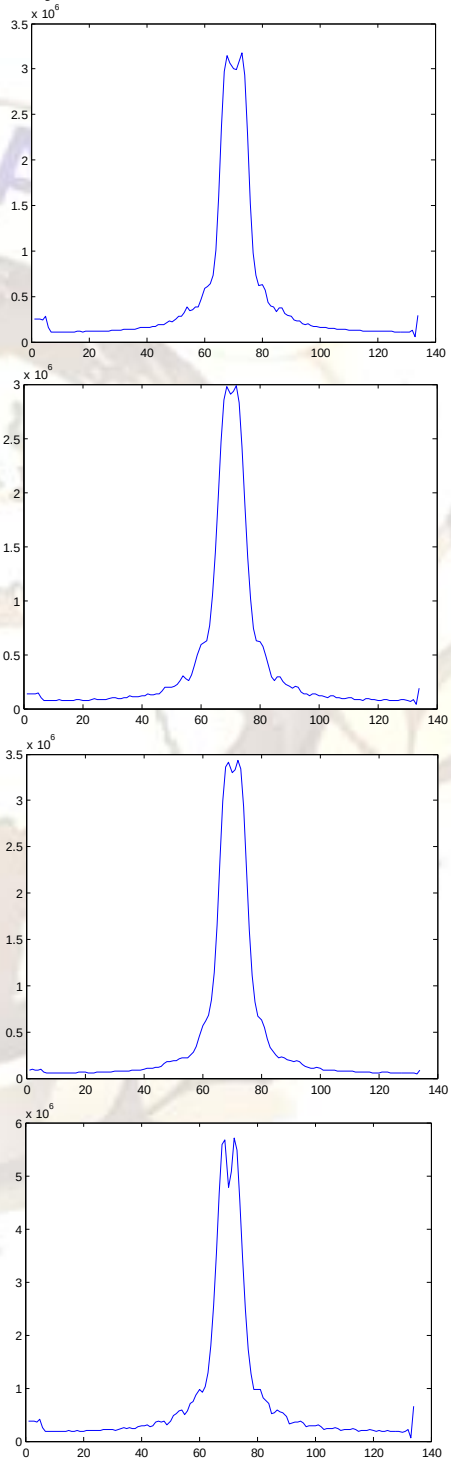


Figure 13: Level 7 approximations of subject 1

All the above waveforms belong to subject 1 and are similar to each other.

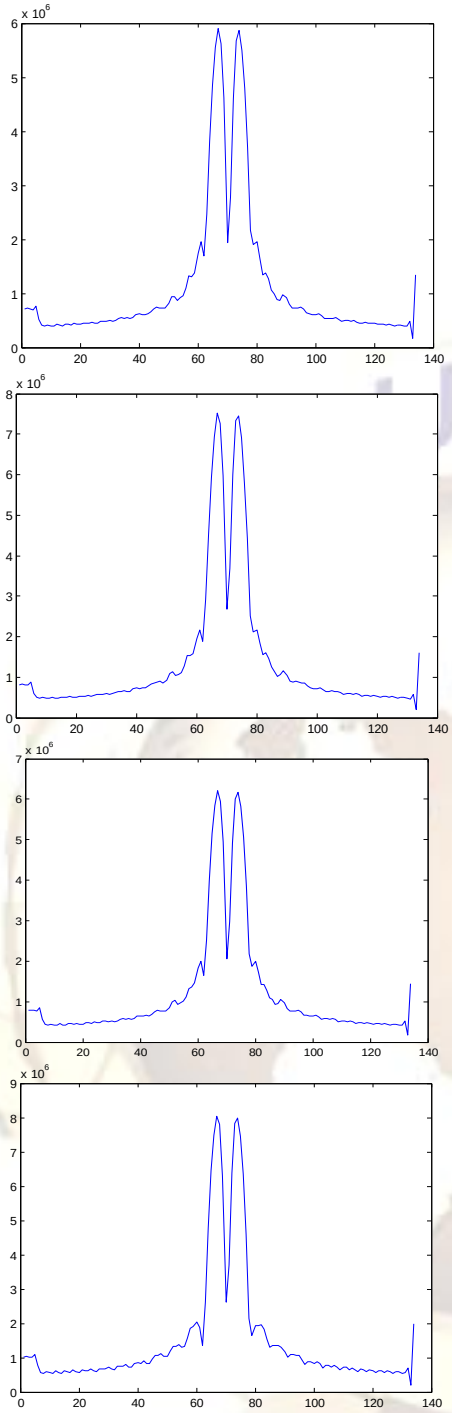
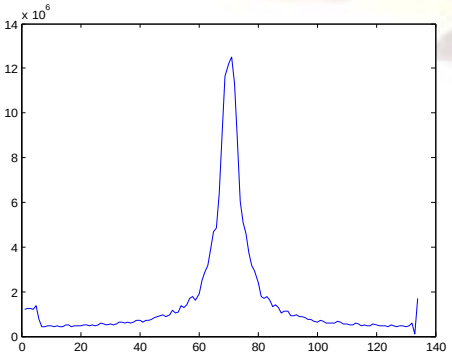


Figure 14: Level 7 approximation of subject 4



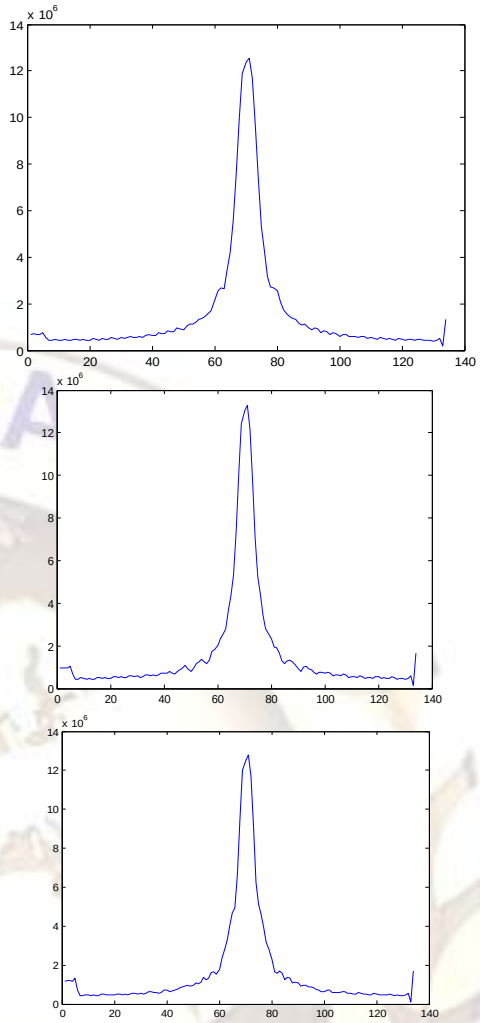


Figure 15: Level 7 approximations of subject 10

We notice from the above figures, that the waveforms for different orientations of one subject are similar while for different subjects they are different. Hence it is possible to classify them.

## V. CLASSIFICATION

Image classification involves analyzing the numerical properties of various image features and organizing the data into categories. In this paper, we have used three classifiers for classifying the data: minimum distance classifier, support vector machine and k nearest neighbor. In all the three methods, out of the ten orientations of each subject, eight were used for training the classifier and two were used as test cases.

### 5.1 Minimum distance classifier

The classification was done using the minimum distance classifier. Euclidean distance has been used for calculating the distance. Since the total number of subjects is 40, there were 40 classes. After classification, following results were obtained:

Table 4: Classification accuracy using minimum distance classifier

| Sr. No | Feature extraction method | Classification accuracy |
|---|---|---|
| 1. | Hu moments | 60% |
| 2. | Zernike moments | 90% |
| 3. | Legendre moments | 90% |

The results of Hu moments are not as good as those of Zernike and Legendre moments as can be seen from the table.

### 5.2 Support Vector Machine

Classification using non linear SVM was done using the following parameters:

Table 5: Various parameters used to design the support vector machine

| Sr. No. | Parameters | Parameter value |
|---|---|---|
| 1. | Classifier Type | Nonlinear multi class SVM |
| 2. | Kernel | Poly |
| 3. | C | 1000 |

The results of various feature extraction methods using SVM are shown below:

Table 6: Classification accuracy SVM

| Sr. No | Feature extraction method | Classification accuracy |
|---|---|---|
| 1. | Hu moments | 60% |
| 2. | Zernike moments | 85% |
| 3. | Legendre moments | 95% |

Here again, the results of Hu moments are the poorest. Also, the results of Legendre moments improved as compared to the ones from minimum distance classifier, while those of Zernike moments have degraded.

### 5.3 K Nearest Neighbor

When classified using the k nearest neighbor algorithm, most accurate results were obtained when the value of k = 3. The following figure shows the classification results obtained for face recognition using the nearest neighbor classifier for Hu moments:

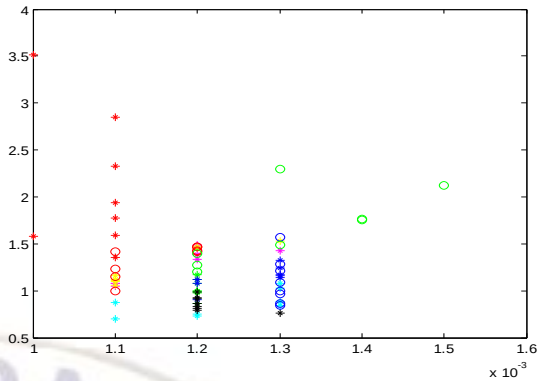



Figure 16: Classification results obtained using Hu moments

Each point in Fig 16 represents the feature vector corresponding to an image. Each sample is classified and associated to one of the ten clusters (subjects) i.e. to a particular subject. The clusters are shown with different colors and markers for easy understanding. In the above figure we see that the classes are not very distinct and kind of overlap. Hence the performance of Hu moments is not very good. Figure 17 shows the classification results obtained for face recognition using the nearest neighbor classifier for Zernike moments:

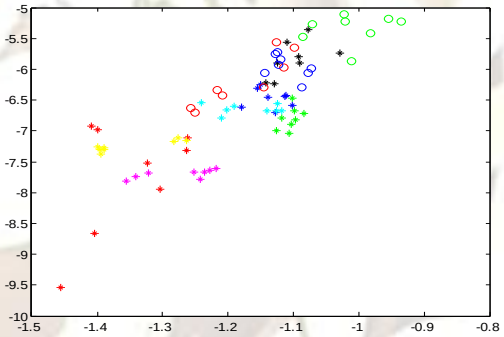



Figure 17: Classification results obtained using Zernike moments

In above figure, the cluster formation is better than that obtained from Hu moments. Hence the classification is better. Fig 18 shows the results obtained for face recognition using nearest neighbor classifier for Legendre moments:

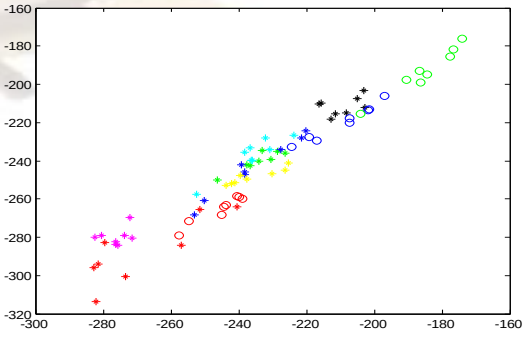



Figure 18: Classification results obtained using Legendre moments

Here also the cluster formation is good. The results of using k nearest neighbor classifier on the three feature extraction methods are shown in Table 7:

Table 7: Classification accuracy of the feature extraction methods using KNN

| Sr. No | Feature extraction method | Classification accuracy |
|---|---|---|
| 1. | Hu moments | 44% |
| 2. | Zernike moments | 85% |
| 3. | Legendre moments | 95% |

Once again the performance of Hu moments is the poorest. The results of Zernike and Legendre are same as those obtained from SVM.

## VI. RESULTS

Four experiments were performed for extracting the features using four approaches: Hu moments, Legendre moments, Zernike moments and cumulants. The feature vectors obtained using these moments were classified using three methods: minimum distance classifier, SVM and k nearest neighbor. The average performance of the feature extraction methods can be tabulated as below:

Table 8: Average performance of the feature extraction methods

| Sr. No | Feature extraction method | Classification accuracy |
|---|---|---|
| 1. | Hu moments | 53.33% |
| 2. | Zernike moments | 86.66% |
| 3. | Legendre moments | 93.33% |

From the above table, we can infer that the performance of Hu moments was not satisfactory and hence they cannot be used for the purpose of face recognition using complex facial images. On the other hand, the results obtained using Zernike and Legendre moments are good.

For the experiment performed using cumulants, we have obtained waveforms which are similar for different orientations of same subject and different for different subjects. These waveforms can be classified using any suitable method.

Many techniques have been used for face recognition like LDA (Linear Discriminant Analysis), BLD (Bayesian logistic discriminant) etc. While the results of these techniques on ORL database are 87% and 91.9% respectively. The results of orthogonal moments like Zernike and Legendre are better than these (95%).

## VII. CONCLUSION

In this paper a novel method for face recognition using moments and cumulants has been presented. Four experiments were conducted using Hu moments, Zernike moments, Legendre moments and cumulants. Features were extracted using the above mentioned moments and feature vectors were formed. The size of the feature vector for Hu moment was 7, for Zernike 7 and for Legendre moments it was 12. In all the experiments, three different classifiers were used for classification: minimum distance classifier, SVM and k nearest neighbor. From the experimental results it was observed that the feature representation using Hu moments provides a maximum recognition rate of 60%. Orthogonal moments like Zernike and Legendre which are also invariant in nature are capable of representing the image features with minimum number of coefficients and the percentage of accuracy is also superior. Hence these moments have been found very much suitable for feature representation of complex face images which are almost similar with respect to variations in size, pose, illumination and orientation within a smaller area. In this paper, we also worked with cumulants and used them for face recognition. The method can be used to successfully identify different faces. The proposed technique using cumulants is sensitive to the structural changes in the images and is able to distinguish them successfully. But, the same sensitivity makes the method vulnerable to noise in the samples, so the images have to be noise free for expected results.

## REFERENCES

[1] R. Ksantini, B. Boufama, Djemelziou, Bernardcolin, A Novel Bayesian Logistic Discriminant Model: An Application to Face Recognition, *Pattern Recognition*, 43(4), 2010, 1421–1430.

[2] C.K. Hsieh, S.H. Lai, and Y.C. Chen, An Optical Flow-Based Approach to Robust Face Recognition Under Expression Variations, *IEEE Transactions on Image Processing*, 19(1), 2010, 233- 240.

[3] J. Yang, C. Liu, L. Zhang, Color Space Normalization, Enhancing the Discriminating Power of Color Spaces for Face Recognition, *Pattern Recognition*, 43(4), 2010, 1454–1466.

[4] H. Zhou, and A.H. Sadka, Combining Perceptual Features with Diffusion Distance for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications And Reviews*, 41(5), 2011, 577- 588.

[5] Z. Lei, M. Pietikäinen, S.Z. Li, S. Liao, Face Recognition by Exploring Information Jointly in Space, Scale and Orientation, *IEEE Transactions on Image Processing*, 20(1), 2011, 247- 256.

[6] Z. Liu, J. Yang, C. Liu, Extracting Multiple Features in the Cid Color Space for Face Recognition, *IEEE Transactions on Image Processing*, 19(9), 2010, 2502-2509.

[7] W. Zou, P.C. Yuen, Discriminability and Reliability Indexes: Two New Measures to Enhance Multi-Image Face Recognition, *Pattern Recognition*, 43(10), 2010, 3483–3493.

[8] X. Tan, B. Triggs, Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions, *IEEE Transactions On Image Processing*, 19(6), 2010, 168-182.

[9] R.M. Jiang, D. Crookes, N. Luo, Face Recognition in Global Harmonic Subspace, *IEEE Transactions on Information Forensics and Security*, 5(3), 2010, 416- 424.

[10] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor, *IEEE Transactions on Image Processing*, 19(2), 2010, 533-44.

[11] G.F. Lu, Z. Lin, Z. Jin, Face Recognition Using Discriminant Locality Preserving Projections Based on Maximum Margin Criterion, *Pattern Recognition*, 43(10), 2010, 3572–3579.

[12] S. Dabbaghchian, M. P. Ghaemmaghami, A. Aghagolzadeh, Feature Extraction Using Discrete Cosine Transform and Discrimination Power Analysis with a Face Recognition Technology, *Pattern Recognition* 43(4), 2010, 1431–1440.

[13] M. Cheng, B. Fang, Y.Y. Tang, T. Zhang, J. Wen, Incremental Embedding and Learning in the Local Discriminant Subspace with Application to Face Recognition, *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications And Reviews*, 40(5), 2010, 580- 591.

[14] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11), 2010, 2106- 2112.

[15] W.C. Kao, M.C. Hsu, Y.Y. Yang, Local Contrast Enhancement and Adaptive Feature Extraction for Illumination-Invariant Face Recognition, *Pattern Recognition*, 43(5), 2010, 1736–1747.

[16] S. Yan, H. Wang, J. Liu, X. Tang, T. Huang, Misalignment-Robust Face Recognition, *IEEE Transactions on Image Processing*, 19(4), 2010,1- 6.

[17] R. Kumar, A. Barmpoutis, A. Banerjee, B.C. Vemuri, Non-Lambertian Reflectance Modeling and Shape Recovery of Faces Using Tensor Splines, *IEEE Transactions on Pattern Analysis and Machine Intelligence*,33(3), 2011, 553-567.

[18] T.K. Kim, J. Kittler, R. Cipolla, On-Line Learning of Mutually Orthogonal Subspaces for Face Recognition by Image Sets, *IEEE Transactions on Image Processing*, 19(4), 2010, 1067-74.

[19] J. Lu, Y.P. Tan, Regularized Locality Preserving Projections and its Extensions for Face Recognition. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics*, 40(3), 2010, 958- 963.

[20] L. Qiao, S. Chen, X. Tan, Sparsity Preserving Discriminant Analysis for Single Training Image Face Recognition, *Pattern Recognition Letters*, 31(5), 2010, 422–429.

[21] S.K. Kim, Y.J. Park, K.A. Toh, S. Lee, SVM Based Feature Extraction for Face Recognition, *Pattern Recognition*, 43(8), 2010, 2871–2881.

[22] C.X. Ren, D.Q. Dai, Incremental learning of bidirectional principal components for face recognition, *Pattern Recognition*, 43(1), 2010, 318 – 330.

[23] Y. Wang, Y. Wu, Complete neighborhood preserving embedding for face recognition, *Pattern Recognition*, 43(3), 2010, 1008 -1015.

[24] S. Haykin, *Adaptive Filter Theory*, (Prentice Hall, 2001)

[25] A. Grossman, R.K. Martinet, Time and scale representation obtained through continuous wavelet transform, *Signal Processing IV, Theories and Applications*, 1988, 475-482.

[26] S.G. Mallat, A theory for multiresolution signal decomposition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 1989, 674- 693.

[27] R. Kapoor, and A. Dhamija, A New Method for Identification of Partially Similar Indian Scripts, *International Journal of Image Processing, 6(2)*, 2012.

[28] M, Durrani TS. Wavelants - An Introduction to Higher Order Wavelets