Evaluation of Various Classifiers for Expression Recognition using Multi Level Haar Features

MAHESH M. GOYANI
Assistant Professor, Department of Computer Engineering,
Government Engineering College, Modasa
Gujarat, India.
and
NARENDRA M. PATEL
Associate Professor, Department of Computer Engineering,
Birla Vishvakarma Mahavidyalaya, Vallabh Vidyanagar,
Gujarat, India.

Facial expressions play an equally important role as verbal communication and tonal expression. Recognition of facial expression is important in industrial automation, security, medical and many other fields. An image is a very rich and high dimensional data structure, which can result into a considerable computation when processed upon directly. Various feature extraction techniques have been proposed to represent the images efficiently in lower dimension which is understandable by the computer. Configuration and dynamics, both are crucial in the interpretation of facial expressions. This work is based on the configuration of facial texture, it does not account dynamics of muscle change. In this paper, we propose multilevel Haar wavelet-based approach, which extracts the features from prominent face regions at two different scales. The approach first segments most informative geometric components such as eye, mouth, eyebrows etc. using the AdaBoost cascade object detector. Haar features of segmented components are extracted. We have evaluated the performance of thirteen different classifiers (5 template matching and 8 machine learning classifiers). Machine learning based classifiers are more adaptive to features and hence outperforms template matching classifiers. Among all tested classifiers, LS-SVM and Discriminant Analysis Classifier provides the best results. Performance of classifiers is also evaluated in various scenarios like low resolution, noisy image, various mouth components etc.

Keywords: Affective computing, facial expression recognition, multi level haar, template matching,
machine learning.

1. INTRODUCTION

Face and Facial Expression Recognition (FER) have been an attractive field for the researchers since last decade. FER covers a wide range of applications including Human-Computer Interaction (HCI), pain detection for a patient, sentiment identification, surveillance systems, security monitoring, student engagement in teaching-learning process etc. Momentum of HCI is shifted from machine to human. Future of the computation is being more human-centric. However, HCI does not account the mental state of the person, which is projected on the face in the form of various expressions. Interpretation of facial expression through the machine can revolutionize user interfaces such as robotics, car driving etc.

Expressions play vital role in conveying emotions and the mood. Facial Expression provides an important clue for social communication. Mehrabian [1968] reported that facial expressions alone convey 55% of the clues, while vocal and verbal channel together carry only 45% of information. These clues and information lead towards a better understanding in a conversation. Facial expressions are the results of contractions of facial muscles, which in turn results in the deformation of various facial units such as eyes, nose, chin, lips, skin texture etc (Fasel and Luettin [2003]). Field of Facial Expression Recognition has been very old yet fascinating. Darvin [1872] reported the universality of facial expression between human and animals. Ekman and Friesen [1971] postulated judgment based six expressions which are universal across all cultures, races, age, gender and ethnicity. The basic six expressions are anger, disgust, fear, happy, sad and surprise. Later, Ekman and Friesen [1978] introduced descriptive coding, known as Facial Action Coding Scheme (FACS), which encodes the face using 46 facial action units. The first FER system proposed by Suwa et al. [1978] is considered to be a major turnaround in the field of FER. In this work, they tracked the motion of 20 facial points on image sequence to classify the expression. Expressions are not just a change in muscles position, rather it is a complex psycho-physiological process. At first, thoughts emerge in the human mind, which is a psychological process. The thoughts often end up as the rendering of expression on the face by means of muscle deformation, which is a physiological process. Changes of muscles typically last between 250 ms to 5 s, so spotting the exact class of expression from spontaneous expression is difficult than identifying it from posed expressions (Fasel and Luettin [2003]). The intensity of expression varies over time for an individual; similarly, it may vary for two different persons at a time. Subsequently, it is difficult to precisely determine the intensity of expression without referring to the neutral face of a given subject. Remarkable survey on facial expression recognition can be found in the literature of Fasel and Luettin [2003], Corneanu et al. [2016], Zeng et al. [2009].

Classification of FER approaches is manifold. FER system (FERS) operating on 2D image or 3D model is classified as image based or model based, respectively. Static FERS operates on the single static image. Features are extracted from the given input image without referring to any other image. While dynamic FERS takes an image sequence as an input and features are computed by observing the change in facial components over the time between successive frames. Feature extraction may be local or global. In contrast to local features, global features are computed holistically. The appearance-based approach typically uses filters to extract the subtle details like texture, wrinkles etc. from the face. While Geometry based approach classifies the expression based on the relationship between various fiducial points affected by expression change.

In this paper, we propose a novel approach using haar wavelets for facial expression classification. Haar wavelets are mathematically simple and quite fast. Many filter-based approaches operate at single scale, which may fail to capture salient yet useful features at different scales. Face and geometric facial components like an eye, mouth, eyebrow, etc. contain expressive features. Remaining facial parts contribute relatively less in recognition, while the background adds redundancy and computation. Proposed method first detects the geometric components and extracts local appearance features to effectively represent the face in a lower dimension. This representation becomes the input to the logistic regression classifier, this classifier outperforms the traditional classifiers like cross correlation, nearest neighbor and neural network.

Rest of the paper is organized as follows: Section II describes related work and literature survey. The proposed approach is discussed in section III, followed by discussion and comparison with different methods in section IV. Conclusions and future enhancements are highlighted in section V.

2. BACKGROUND AND RELATED WORK

The two important aspects of facial expression recognition include facial representation and design of the classifier. Facial representation refers to finding a set of features that can effectively represent the face image. The optimal feature should maximize the intra-class similarity and minimize the interclass similarity. The success of an appropriate classifier ultimately depends on an effective facial representation, which can lead to better recognition rate (Shan et al. [2009]).

2.1 Facial Representation

Many appearance and geometry based feature extraction methods have been investigated to date. Several attempts have been made to improve the robustness of such systems. Appearance-based methods try to express the face by utilizing the texture details of the face. Lyons and

Akamatsu [1998] and G. Wenfei and Lin [2012] Gabor-based approaches are used to model the facial expressions. Lyons and Akamatsu [1998] proposed a Gabor-based approach, in which Gabor response at manually selected 34 fiducial points was combined into a single vector. Principal Component Analysis PCA was applied to reduce the dimension of the feature vector of size 1020, and finally, LDA was used for classification. Liu and Wang [2006] suggested a similar approach, which collects Gabor response of 40 filters at pre-located 42 fiducial points. All Gabor filters are divided into 13 channels (corresponding to 5 scales and 18 orientations). Then PCA based classification method is adopted for classification of expression in each of 13 Gabor channel feature vector. T. Wu and Movellan [2010] proposed biologically inspired Gabor Motion Energy (GME) filter, which is implemented by adding 1D temporal filters on top of Gabor Energy (GE) filter. GME captures more information around eyebrow and upper eyelid, and it outperforms simple GE filters. Almaev and Valstar [2013] proposed LBP-TOP dynamic appearance descriptor-based approach which extends an idea of Local Gabor Binary Patterns (LGBP) of T. Senechal and Prevost [2012] to the temporal domain. LBP-TOP independently computes the GLBP features of all three orthogonal planes: the spatial x-y plane, and the temporal x-t and y-t plane. GLBP not only considers the dynamics, but also reduces the computation of Volume Local Binary Pattern (VLBP) by processing only the TOP orthogonal planes. Samad and Sawada [2011] presented edge-based feature extraction approach using Gabor filter and special convolution mask. Dimensions of extracted edge features are reduced using PCA and classification is done using SVM.

Local binary pattern is another extensively studied powerful texture analysis descriptor. Ojala et al. [1996] proposed LBP for analyzing macro and micro patterns of the image. Later on, a number of variations of LBP have been proposed. Due to its small kernel size, the applications of traditional LBP was limited. Ojala et al. [2002] proposed multi-scale, rotation invariant LBP operator, which works with any arbitrary size of the kernel by interpolating pixels along circle periphery. They also proposed uniform binary pattern to reduce the feature dimensions. LBP is attractive due to its simple computation and ability to detect the relationship between the current pixel and its local neighbors. Moore and Bowden [2011] used LBP for classification of multi-view facial expression. Shan et al. [2009] proposed boosted LBP-based approach to extract most discriminative LBP features. They also extend the experiment for low-resolution images. With linear programming and SVM classifiers, they achieved noticeable accuracy. To further reduce the computation, R. Hablani and Tanwani [2013] employed LBP on prominent regions of the face. LBP histograms of eyes, mouth, nose are concatenated to form the final feature vector. Expressions are classified using Chi-square test similarity. An in-depth survey of LBP and its application in facial analysis can be found in survey of D. Huang and Chen [2011].

Optical flow is used to model the facial muscle changes in literature Yacoob and Davis [1996], Essa and Pentland [1997]. Optical flow estimates the change in facial feature points. However, optical flow can easily be disturbed by external parameters like intensity change, face registration and internal parameters like motion discontinuity and rigid non-rigid motion. Gao et al. [2003] present an approach which extracts line based caricature from the face image. Structural and geometric features are used to match the line edge map. In their work, they reported the average accuracy of 86.6% for neutral, smile and scream expressions. The authors noted interesting fact that recognition rate for female is 7.8% higher than that of male. Shih et al. [2008] analyzed the performance of holistic methods like 2D LDA representation of DWT, LDA, ICA, PCA and 2D PCA techniques on JAFFE dataset. They achieved the highest recognition rate of 95.71% by adopting the cross-validation strategy.

Another way of facial representation is to represent a face using the change in appearance. Holistic approaches including Principal component analysis (PCA) (Turk and Pentland [1991]), Linear Discriminant Analysis (LDA) (Belhumeur et al. [1997]), Independent Component Analysis (ICA) (M. S. Bartlett and Sejnowski [2002]), Gabor wavelet (Lyons and Akamatsu [1998]), 2D PCA (Yang et al. [2004]), $(2D)^2$ PCA (Oliveira et al. [2011]) are utilized with full face or the components of the face.

2.2 Facial Expression Classification

Classifying non-linearly separable classes is another major area of research. To classify the facial expression, various classification approaches have been proposed such as Nearest neighbor (Rahulamathavan et al. [2013], Zhi and Ruan [2008]), k-Nearest Neighbor (Oliveira et al. [2011], Wang et al. [2015]), Artificial Neural Network (ANN) (Owusu et al. [2014]), Support Vector Machine (SVM) (Shan et al. [2009], Shih et al. [2008]) Linear Programming (Shan et al. [2009]), Linear Regression (J. M. Guo and Wong [2016]). R. A. Khan and Bouakaz [2013] used Pyramid LBP features with classification techniques such as SVM, KNN, Random Forest, and Decision tree. Nearest neighbor is widely used, simple and effective classifier. ANN has emerged as a classical tool for pattern grouping. With the appropriate number of layers and number of hidden neurons, multi-layer neural network can classify almost any complex patterns. It forms hyperplane to separate different classes. In the recent year, SVM gained a lot of attention in pattern classification. Neural network classifies the patterns based on strict decision boundary, while SVM separates the classes using maximum margin leading to lesser classification error. SVM maps input to the high dimensional feature space and separates them by finding maximum margin hyperplane between two classes, using quadratic programming. Linear programming deals with the classification problem, where both the objective function to be optimized and all the constraints are linear in terms of decision variables. In the proposed approach, logistic regression classifier is used, which measures the relationship between the class label and the test sample by estimating the probability using logistic function.

3. PROPOSED SYSTEM

3.1 Preprocessing

Preprocessing of the face and locating Region of Interest (ROI) is a crucial step for robust feature extraction. Segmentation of local facial components can significantly reduce the computation cost of both the feature extraction and classification. Even images captured under controlled environment are not aligned, which introduces error in feature localization. Head pose variation, face size, illumination variations are the common causes of alignment error. We proposed eye-ball registration based face normalization technique for face normalization.

Following Tian [2004], Shan et al. [2009] and H. Boughrara and Chen [2016], we used fixed eye distance based approach to normalize the face. Shan et al. [2009] fix the distance between eyeballs to 52 pixels and face is cropped and normalized to 150 × 110 pixels using prior knowledge of facial geometry. Most of the literature have preferred manual or semi-automated approach for eye registration. However, our approach is completely automatic. In facial datasets, images are acquired under a controlled environment, so the global position of the face is always predictable. We used iterative approach for eye registration. At first, eye pair is detected using cascade object classifier proposed by Viola-Jones. Eye segment is thresholded and complemented using global threshold estimation.

The binary image contains some unwanted small regions which satisfy the global threshold. From the prior knowledge, areas with less than 65 pixels are removed, so that binary image contains only the eyeball region.

The thresholded eye region may not be connected due to the difference in skin tones of the subjects. Morphological erosion operation is applied to 3 × 3 structuring element with all 1s to connect areas around the eyeball. Let A represents the binary image of eye strip and B is the structuring element. In integer grid space E, erosion of the binary image A is defined as,

$$A \ominus B = \{z \in E \mid B_z \subseteq A\} \tag{1}$$

Where, Bz is the translation of B by the vector z, i.e.

$$B_z = \{b + z \mid b \in B\}, \forall z \in E \tag{2}$$

Centroid of both eyes is computed after applying erosion. Let (l_x, l_y) and (r_x, r_y) represents the spatial coordinates of the centroid of left and right eyes respectively. Even images are acquired in a controlled environment; head of certain subjects are not in exact upright frontal position. Such faces introduce alignment error, so we performed eyeball registration by measuring the angle of the line joining the eyeballs. If the face is perfectly vertically positioned, then the slope of the line joining eyeballs would be 0. Otherwise, it would be non-zero, and the face is aligned by performing negative rotation of the angle with the x-axis. Let x and y represent the difference of x and y coordinates of eyeballs. Thus, $\Delta x = r_x - l_x$ and $\Delta y = r_y - l_y$. Angle is estimated by taking tan-1 of the slope of the line,

$$\theta = tan^{-1}(\frac{\Delta y}{\Delta x}) \tag{3}$$

If the angle is greater than the prescribed threshold, then the image is rotated by negative rotation angle, and the process is reiterated from the eye pair detection phase. Figure 1 demonstrates the angle estimation for slant face.

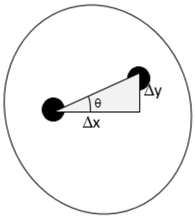

Figure 1. Angle estimation from eyeball for eye registration

Once the angle threshold is adjusted within the range, the image is rescaled such that distance between eyeballs maintained at 52 pixels. Scaling factor is computed by normalizing the required eye distance by actual eye distance.

$$ScalingFactor = \frac{52}{r_x - l_x} \tag{4}$$

Using advanced knowledge of facial geometry, we crop the facial components based on eyeball position and distance between them. Dimensions used for our experiments are portrayed in Figure 2.

This process will register the eye of all the images used in dataset. The registration process significantly improves the performance. The spatial features would be more correlated now. We evaluate the performance of upper and lower facial regions for expression recognition, and hence we also cropped top and bottom face regions. The entire process is explained in Figure 3.

3.2 Feature Extraction

Texture and geometry convey complementary yet valuable information for FER. Zeng et al. [2009] have shown that facial expression information is not only conveyed by geometric fiducial points but also through texture features. M. Song and Zhou [2010] also revealed that some expressions might have different geometric features although their texture features are similar, and vice versa. Pantic and Patras [2006] suggested that using both texture and geometric features might be the best choice for designing FER systems. In our work, we detected facial components like eye, mouth, eyebrows and haar features are derived from those regions.

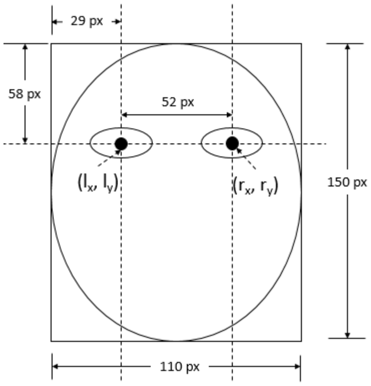

Figure 2. Facial dimensions estimated from eyeball distance.

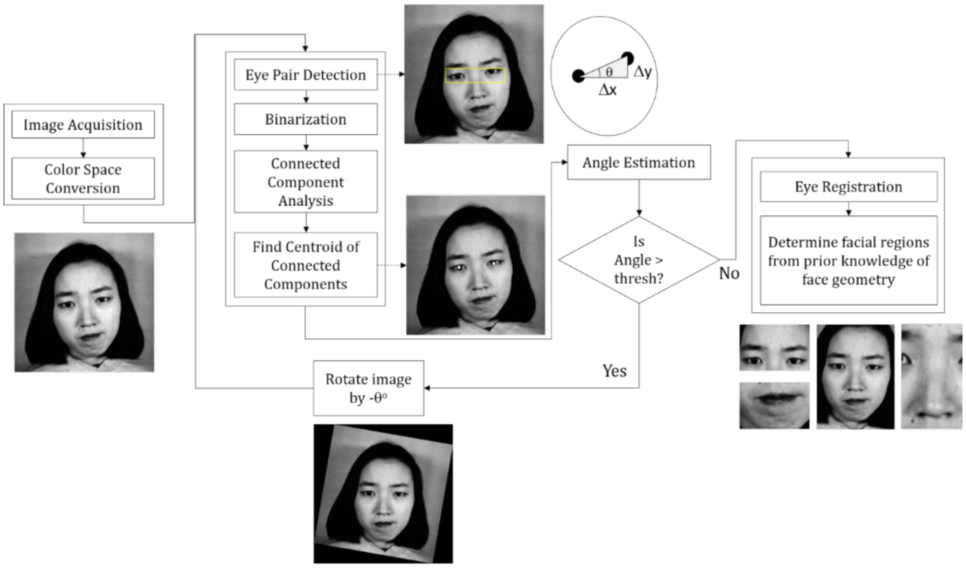

Figure 3. Preprocessing flow of proposed system

Haar functions were introduced by mathematician Alfred Haar. A Haar wavelet is the simplest type of wavelet, and it serves as a prototype for all other wavelet transforms. Haar transform provides a natural mathematical structure for describing the patterns (Papageorgiou and Poggio [2000]). Let us represent the discrete signal of length N as $f = (f_1, f_2, ..., f_N)$. Like all other wavelet transforms, haar also decomposes a signal into two sub-signals of half its length: *running average* or *trend* (a^1) and *running difference* or *fluctuation* (d^1). Components of the first trend signal $a^1 = (a_1, a_2, ..., a_{N/2})$ are computed as,

$$a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}} \tag{5}$$

and values of first fluctuation signal $d^1 = (d_1, d_2, ..., d_{N/2})$ are produced according to the following formula:

$$d_m = \frac{f_{2m-1} - f_{2m}}{\sqrt{2}} \tag{6}$$

Reconstruction of original signal f from a^1 and d^1 is done by the following formula:

$$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, ..., \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}} \right) \tag{7}$$

For Multi-Level Haar feature extraction, the same decomposition is repeatedly applied to latest trend signal in each iteration. The process of multi-level haar analysis is described in Figure 4.

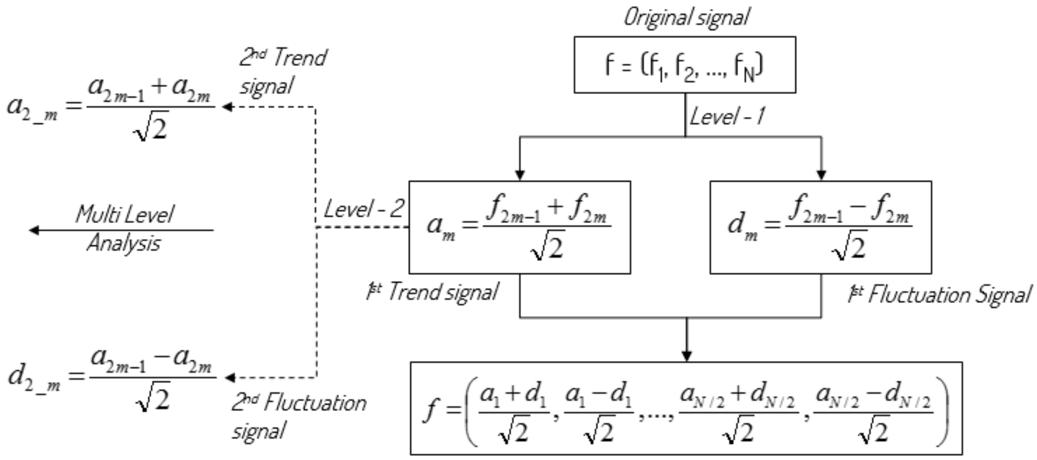
Original signal
f = (f1, f2, ..., fN)
Level - 1
$a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}}$
$d_m = \frac{f_{2m-1} - f_{2m}}{\sqrt{2}}$
1st Trend signal
1st Fluctuation Signal
$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, ..., \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}} \right)$
Level - 2
2nd Trend signal
$a_{2\_m} = \frac{a_{2m-1} + a_{2m}}{\sqrt{2}}$
Multi Level Analysis
$d_{2\_m} = \frac{a_{2m-1} - a_{2m}}{\sqrt{2}}$
2nd Fluctuation signal

Figure 4. Multi Level Haar feature extraction

Original signal and its level -1 haar transform are presented in Figure 5 (left) and Figure 5 (right), respectively. Magnitude of most of the fluctuation component is close to 0. Also notice that the trend sub-signal and original signal are alike. Trend signal is shrunk by half in length and expanded by a factor of $\sqrt{2}$ vertically.

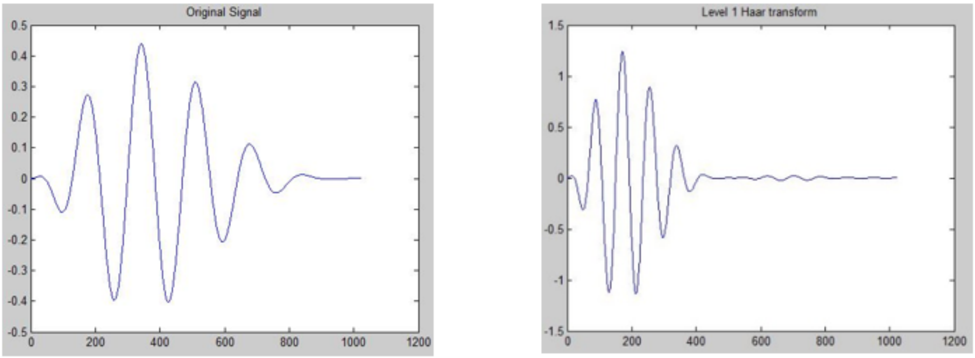

Figure 5. Original Signal (left), and Level-1 haar transform (right)

Haar wavelets are a sequence of rescaled square-shaped signals which form a wavelet basis (Stankovir and Falkowski [2003]). Haar transform is similar to Fourier transform but they are more intuitive and useful. Computation of haar is very simple. The mother wavelet function $\psi(t)$ is defined as,

$$\psi(t) = \begin{cases} 1 & 0 \le t < 0.5, \\ -1 & 0.5 \le t < 1, \\ 0 & otherwise \end{cases} \tag{8}$$

The scaling function (father wavelet) is given by,

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1, \\ 0 & otherwise \end{cases} \tag{9}$$

Typically, the structures that we want to recognize may have very different size. And hence it is not possible to define optimal resolution for analyzing images (Mallat [1989]). In his studies, Mallat has shown the effectiveness of multiresolution analysis to extract the information from the image. The face has an uneven size of features. If we use fixed size kernel, it may miss the features which are smaller or larger than the kernel. And hence we applied two level decompositions to the facial component. Block diagram of proposed architecture is shown in Figure 6. Preprocessing step extracts the region of interest from input image. In order to align the features, eye-ball registration and face alignment is done. Multi-level approximation haar coefficients are extracted from the various facial components. Concatenated coefficients form the feature vector. Dimensions of the generated feature vector is reduced by projecting it in PCA subspace followed by LDA subspace. Various template matching and machine learning classifiers are used to evaluate the system performance.

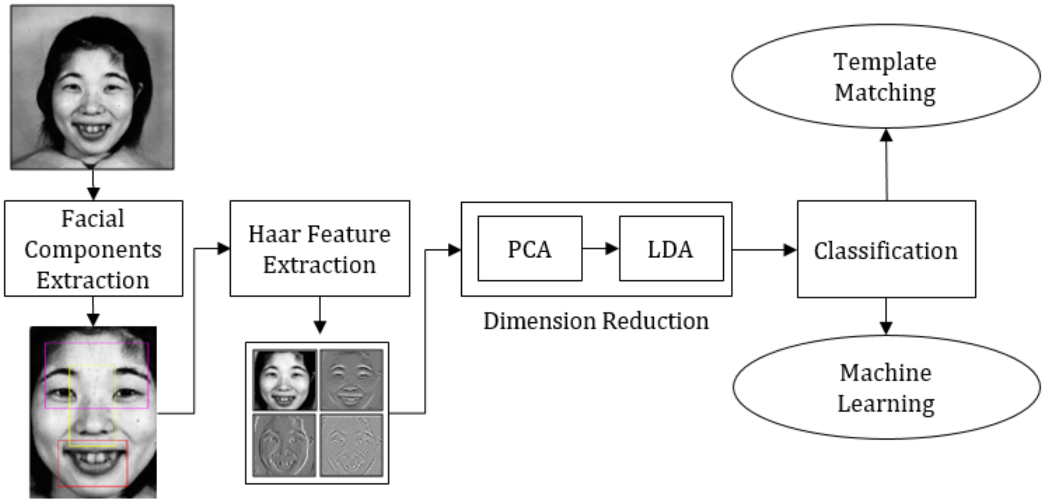

Figure 6. Architecture of Multilevel haar wavelet based FER system

4. RESULTS AND DISCUSSIONS

4.1 Experimental Data

We conduct the experiments on three widely used comprehensive datasets, Cohn-Kanade (Kanade et al. [2000]), Japanese Female Facial Expression (Lyons [1999]) and Taiwanese Facial Expression Image Database (Chen and Yen [2007]). We also conduct the test on our inhouse dataset called WESFED (Web Enabled Spontaneous Facial Expression Dataset). CK dataset contains image sequence of 97 subjects having a 7:13 Male: Female ratio. The dataset contains people of different ethnicity, of age group 18-30 years. Subject were trained to perform series of 23 facial displays,

	AN	DI	FE	HA	SU	SA	NE	Total
CK	110	120	100	280	130	220	320	1280
JAFFE	30	29	32	31	31	30	30	213
TFEID	34	40	40	40	39	36	39	268
WESFED	130	60	66	204	133	145	182	920

Table I: Number of images used for experiment from datasets

six of them were based on prototypical expression description. Each sequence contained 12-16 frames. Each image sequence starts with a neutral expression and ends at the apex level.

JAFFE database was planned and assembled by Lyons [1999] in 1998. JAFFE contains 213 images of Ten Japanese female, with 3 or 4 samples of each of the seven basic expressions. Numbers of images corresponding to each of the seven expressions are roughly identical. Images in JAFFE are recorded under uniform illumination.

The Taiwanese Facial Expression Image Database (TFEID) was designed by Chen and Yen [2007] at Brain Mapping Laboratory, Taiwan in 2007. It consists of 40 subjects from the same ethnicity with an equal proportion of male and female. Subjects in TFEID are instructed to perform eight facial expressions: neutral, anger, contempt, disgust, fear, happiness, sadness and surprise. Models were asked to gaze at two different angles ($0^o \ and \ 45^o$). Each expression included two kinds of intensities (high and slight) and was captured by two CCD-cameras simultaneously with different viewing angles.

We also designed our own Web Enabled Spontaneous Facial Expression Dataset (WESFED), in which Images are acquired from the internet sources, preprocessed and normalized to 150 × 110. Region of Interest is first detected using Viola Jones face detector. For the robust class label, user feedback based scheme was utilized. Each cropped face is presented to 10 different persons and they are asked to vote them from one of the seven predefined classes. Class with maximum votes is assigned to the face.

Existing datasets rarely address the issues of spontaneous expressions. Most of the time, images are acquired under static environment with fixed illumination source. On the other hand, real life scenarios are very different. We addressed all possible issues by considering images of different ethnicity, age, pose, illumination, and occlusion in WESFED. Details of a number of images used for the experiment from all datasets are listed in Table I.

We considered basic seven expressions anger (AN), disgust (DI), fear (FE), happy (HA), sad (SA), surprise (SU) and neutral (NE), for our experiment. Subjects from all four datasets with all seven expressions are depicted in Figure 7.

4.2 Optimal Parameter Selection

The performance of the algorithm is bound to many parameters like number of features, the number of images used to train the model, regions size used to compute the features, etc. In this section, we will discuss the selection of optimal parameters which affects both performance and computation.

4.2.1 *Estimation of Number of Eigenvectors.* Many times, kernel methods generate a large feature vector. It has few obvious disadvantages. Training the classifier with such a large input vector is time-consuming and often leads to poor generalization. In the case of template matching approaches, the comparison of test feature vector with stored template vectors is computationally intensive. Various feature selection and dimension reduction methods have been studied over a period of time to handle the curse of dimensionality. In our experiment, we used Linear Discriminant Analysis (LDA) subspace to reduce the dimensions of the feature vector. Usually, a number of training images are quite less than the number of features, which leads to singular matrix problem while computing generalized eigenvector problem in LDA. To tackle this issue, as a pre-processing step we applied Principal Component Analysis (PCA) on original feature vector, and LDA is applied on PCA subspace. For the C-class problem, LDA produces C - 1 features.

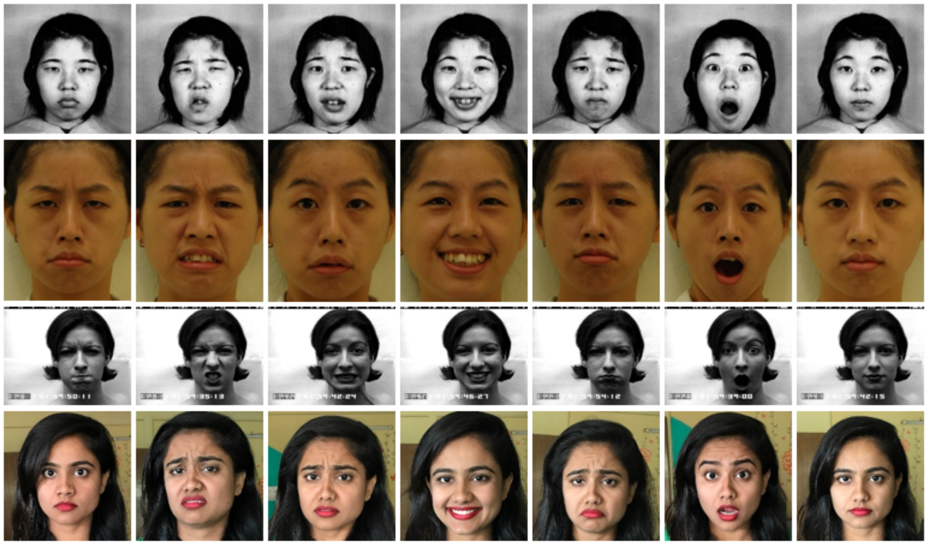

Figure 7. Snapshots of various expressions from JAFFE (first row), TFEID (second row), CK (third row) and WESFED (fourth row) datasets.

Reconstruction of face and recognition accuracy depends on direction and number of projection axis. The optimal direction of projection axis in PCA is derived by finding eigenvectors of the covariance matrix of feature vector. Dimensions of the projected features are directly proportional to a number of eigenvectors used for projection. Eigenvectors with very less eigenvalue really do not contribute to discrimination; rather they increase the computation. In order to determine the optimal feature dimension, we conduct the experiment by varying the number of eigenvectors from 20 to 200 in step of 20. Table II shows the performance of discussed approach on JAFFE against various classifiers with 2-fold cross-validation strategy. Performance is reported for two template matching strategy Chi Square and Cosine distance, and two machine learning classifiers LSSVM with RBF kernel and Discriminant Analysis classifier.

EV	CS	Cosine	LSSVM	DA	AP	API	CAPI
20	76.43	76.24	71.57	73.48	74.43	00	0
40	91.10	90.33	90.90	91.86	91.05	16.62	16.62
60	94.62	93.95	93.19	94.90	94.17	3.12	19.74
80	96.05	95.86	95.38	95.76	95.76	1.60	21.33
100	96.62	96.52	96.52	96.62	96.57	0.81	22.14
120	96.62	96.62	96.43	96.62	96.57	0.00	22.14
140	97.00	97.00	97.00	97.00	97.00	0.43	22.57
160	97.00	97.00	97.00	97.00	97.00	0.00	22.57
180	97.00	97.00	97.00	97.00	97.00	0.00	22.57
200	97.00	97.00	97.50	98.00	97.38	0.38	22.95

Table II: Effect of eigenvectors on performance (%), Dataset: JAFFE

To choose the optimal number of eigenvectors, we averaged the performance of all four classifiers. The plot of cumulatively averaged performance improvement is shown in Figure 8. The result shows that average performance is increasing up to 140 eigenvectors, and after that, there is a slight effective performance gain. With 140 eigenvectors, an algorithm achieves 97.0% average accuracy, and after that, there is no significant improvement. As we increase the number of eigenvectors, recognition rate improves. But after a certain level, the improvement in recognition rate would be very less with additional computation cost. Thus there is a tradeoff between a number of eigenvectors and accuracy. To balance the accuracy-computation trade-off, we choose 140 eigenvectors for the further analysis.

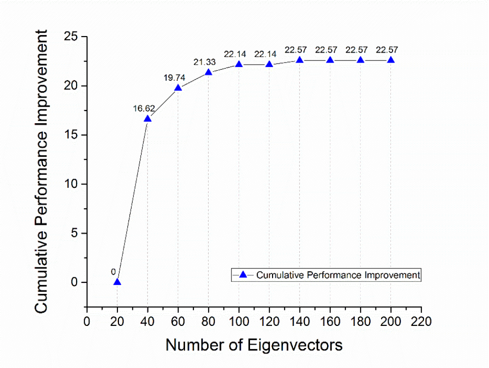

Figure 8. Number of eigenvectors vs. cumulative performance improvement, Dataset: JAFFE.

JAFFE dataset contains only female subjects. To add the gender specific variation, we also performed the same experiment on TFEID dataset, which includes 50% male and an equal number of female. Although TFEID contains male and female both gender, JAFFE and TFEID do not have ethnicity diversion. All subjects in both datasets belong to the same ethnicity. JAFFE contains 10 Japanese female models whereas TFEID contains 20 male and 20 female Taiwanese subjects. To test the robustness of algorithm against various diversities, we also conduct the experiment on comprehensive CK dataset. In CK, 65% of the subjects are female, and 35% are male. 15% of subjects belong to African-American background and 3% subjects belong to Asian or the Latino-American background. Images in CK contains large variations in illumination. We also test the accuracy of the system for our in-house dataset WESFED. We conducted all experiments on all four datasets with common parameters and results are shown in Figure 9 and Figure 10.

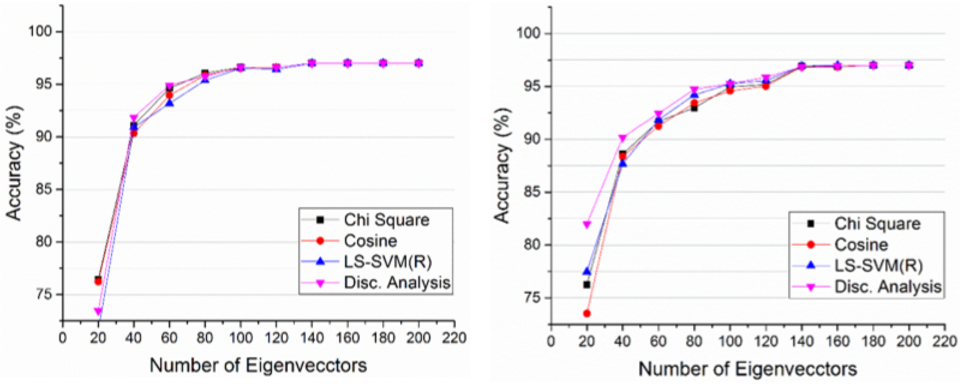

Figure 9. Plot of a number of eigenvectors vs. Accuracy (%) for JAFFE (left) and TFEID (right)

4.2.2 *Estimation of Region Size.* Though global features have proven good for the image analysis, local features provide a good estimation of locality of the features, which can be considered the building blocks of many pattern recognition algorithms. Compared to their counterparts; local features are robust to scale changes, rotation, and occlusion better. Local feature efficiently encodes the local structures such as a point, edge, or small image patch.

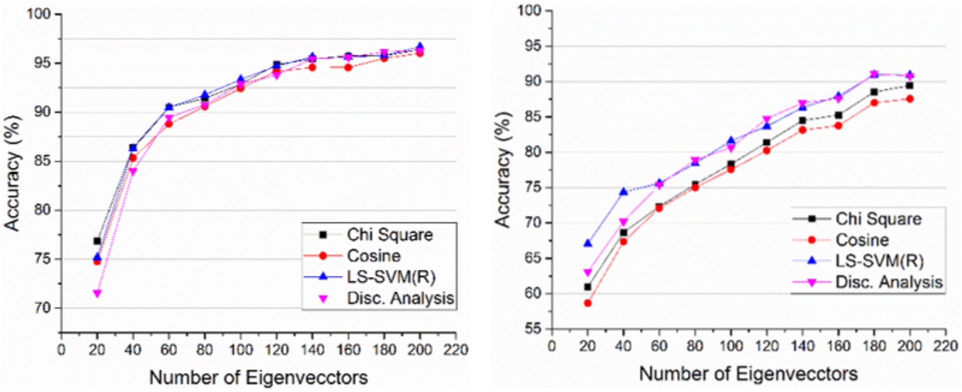

Figure 10. Plot of a number of eigenvectors vs. Accuracy (%) for CK (left) and WESFED (right)

In the prototypic facial expression, textures such as wrinkles, bulges, furs play a crucial role. To extract the local texture features, we divided face image into M × N regions. To find the optimal number of regions, we divide images into 1 × 1, 3 × 3, 5 × 5, 7 × 6 and 9 × 8 blocks. Bartlett et al. [2005], Tian [2004], Shan et al. [2009], Jabid et al. [2010] have conducted experiments in these neighborhoods. To compare the results of proposed approaches with state of the art methods, we also employed the same dimensions in our work. Larger regions size fails to capture the texture of small size. Small regions effectively capture local and spatial relationship. However, after certain point, the smaller regions introduce unnecessary computation and feature vector becomes too large to train the classifier efficiently. We used 100 eigenvectors in the experiment. Performance behavior on JAFFE and TFEID dataset for a different number of blocks is stated in Figure 11. Response for 7 × 6 region is better compared to other tested regions.

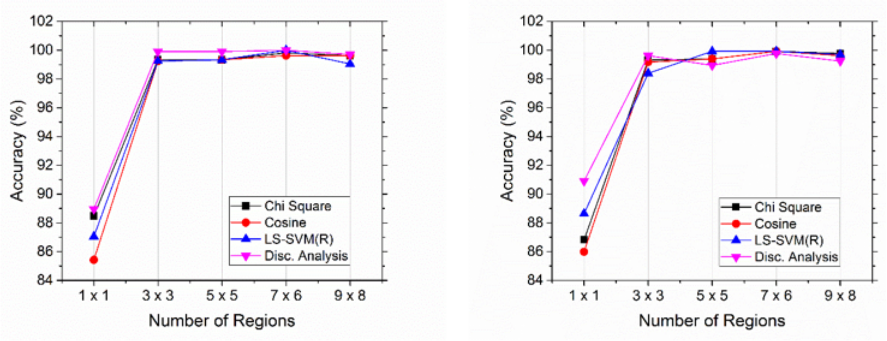

Figure 11. Effect of a number of regions on performance on JAFFE (left) and TFEID (right) dataset. Response for 7 x 6 regions is better compared other tested regions.

From above results, we chose a number of regions to be 7 × 6, as it gives a proper balance between accuracy and standard deviation of results from its mean.

4.2.3 *Selecting Classifier* . Certain classifiers are good at classifying specific features only. We evaluated the performance of MLH feature descriptor against various template matching and machine learning based classifiers. We tested out system for L2 norm, Chi-Square, Cosine, Correlation and k-NN based template matching classifiers. We also measured the performance using various machine learning classifiers like Artificial Neural Network, Least Square Support Vector Machine (with linear, polynomial and RBF kernel), Multi-SVM (extension of binary SVM to multi-class SVM), Logistic Regression, Discriminant Analysis and Decision Tree. Results of all classifiers are compared in Figure 12.

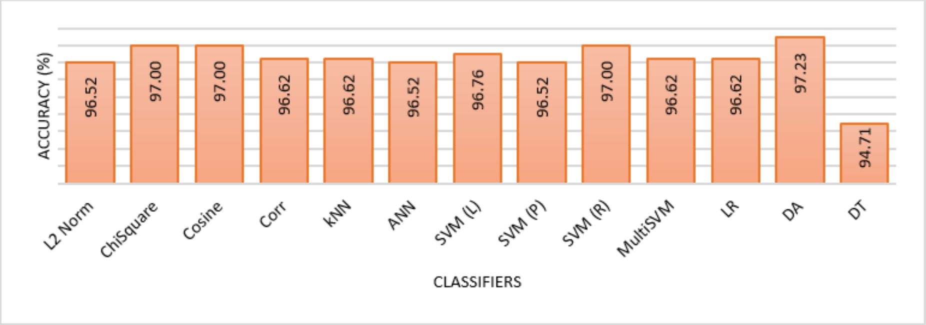

Figure 12. Performance comparison of different classifiers, Dataset: JAFFE

Chi-square and Cosine measure gives the best classification results among all template matching techniques. Discriminant Analysis classifier achieves the highest accuracy among used machine learning classifiers for chosen parameters. A particular instance of execution is shown here; in general, the performance of LS-SVM is very close to that of Discriminant Analysis. For further analysis, we used two template matching (Chi-Square and Cosine) and two machine learning (Discriminant Analysis and LS-SVM with RBF kernel) classifiers.

4.2.4 *Performance Analysis in Small Training Sample Space.* An important aspect of learning methods is that they should generalize well on unknown data. The success of any classifier depends on how quickly adapts to new and unseen patterns. It guides the choice of learning method and gives a measure of the quality of the method. A lot of training samples leads to better recognition rate. K-fold cross validation is the most commonly used validation technique for predicting accuracy and to measure the generalization performance of the algorithm. A lot of work done in the past reports the use of 10-fold validation, wherein 90% samples are used for training and the rest are used for testing. Reduction in number of training samples has shown to negatively impact the performance. Discrimination capability of the proposed methods has been evaluated with six different cross-validation methods, varying the training samples from 90%, 80%, 70%, 50%, 30% and 10%. Even with 10% training samples, it exhibits far better accuracy compared to many state of the art methods. Fig. 3.16 exhibits the behavior of the system for various validation strategies.

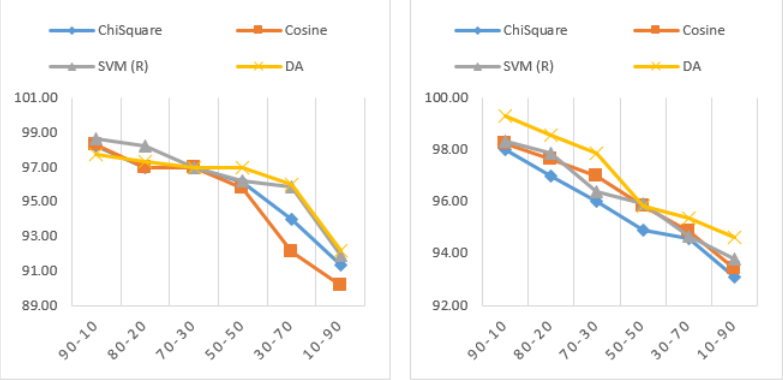

Figure 13. Performance of various classifiers against different validation methods for JAFFE (left) and TFEID (right) dataset

Parameter	Chosen value
Number of eigenvectors	140
Number of regions	7 x 6
Validation method	2-Fold (50% training 50% testing)
Template matching classifier	Chi-square, Cosine measure
Machine learning classifier	LS-SVM, Discriminant Analysis

Table III: Optimal parameters chosen for result analysis

Generalization of the model with very few training samples is indeed a challenging task. It largely depends on the discriminative capability of features and the training set. Larger training set helps the model to learn the underlying pattern represented by a feature vector quickly. The goodness of any model can be defined based on how well it classifies unseen samples. As shown in Figure 13, with 30% training samples, discriminant analysis reports recognition rate as high as 95%. Proposed feature descriptor is so robust that even with just 10% training samples, it achieves more than 90% accuracy.

Based on the experiments, we choose parameters for the further analysis as shown in Table III.

4.3 Expression Recognition using Facial Components

It is observed that beauty is a factor which affects the reminiscence of the face. Faces with higher beauty factor are remembered for a long time. Similarly, certain facial regions have more influence on recognition rate. We evaluated the importance of upper and lower facial regions in expression recognition. Eye, eyebrow and forehead lines show different geometrical movement during certain expressions. The texture on facial component surface carries essential discrimination information. In anger state, eyebrows pulled down, upper and lower lids pulled up, and lips may be tightened. In the fear state, eyebrows and upper eyelids are pulled up, and mouth is stretched. During disgust state, eyebrows are pulled down; nose gets wrinkled and upper lip is pulled up. Similar changes can be observed in other expressions too. Thus the effective representation of skin texture is crucial. We performed expression recognition using MLH features extracted from the only eye, only mouth, eye + mouth, and face. Results are stated in Figure 14 for JAFFE and TFEID datasets.

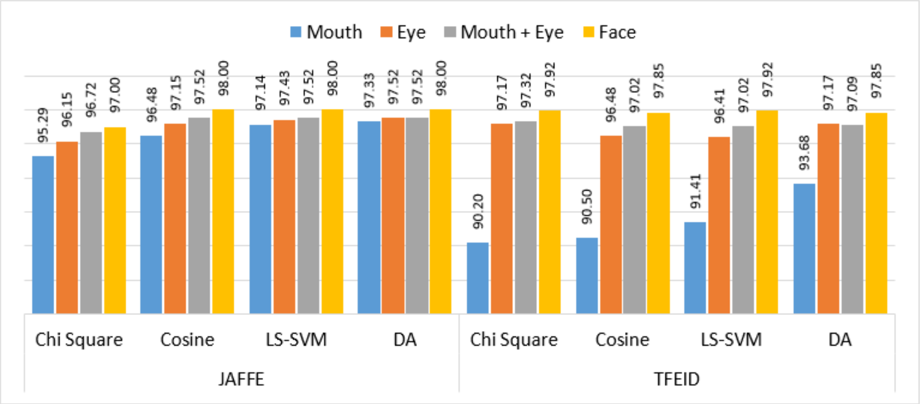

Figure 14. Performance of various classifiers against different validation methods for JAFFE (left) and TFEID (right) dataset

Results show that performance of FER system with features extracted from upper face regions is slightly better than features extracted from mouth region. However, a fusion of both the

	Template Matching		Machine Learning	
Noise Probability	Chi-Square	Cosine	LS-SVM (RBF)	Discriminant Analysis
Pa = Pb = 0.01	93.43	93.43	93.52	93.81
Pa = Pb = 0.05	93.52	93.43	93.24	93.52
Pa = Pb = 0.1	94.00	94.00	93.81	94.00
Pa = Pb = 0.2	93.43	93.14	91.05	92.38

Table IV: Results on JAFFE in noisy environment, Noise: Salt & Pepper

features outperforms results of individual components. Although nose remains in almost same shape and position, for few expressions like disgust and anger, its appearance changes. While the full face is used for feature extraction, when these changes are also incorporated we achieve highest recognition rate.

4.4 Expression Recognition from Noisy Images

Images acquired in real-time are often noisy. A robust system should be able to handle the noise. Salt and pepper, gaussian and speckle noise are the common noise introduced in the image. We conducted the experiment by manually adding noise in the images. Noise is added in half of the randomly selected images. The performance of the system in a noisy environment is evaluated with various noise parameters like mean and variance. The amount of various noise is controlled by the probability of salt (Pa), the likelihood of pepper (Pb), variance (V) and mean (m). Effect of different types of noises with varying probability is shown in Figure 15, and the corresponding recognition rate is reported in Table IV, Table V and Table VI.

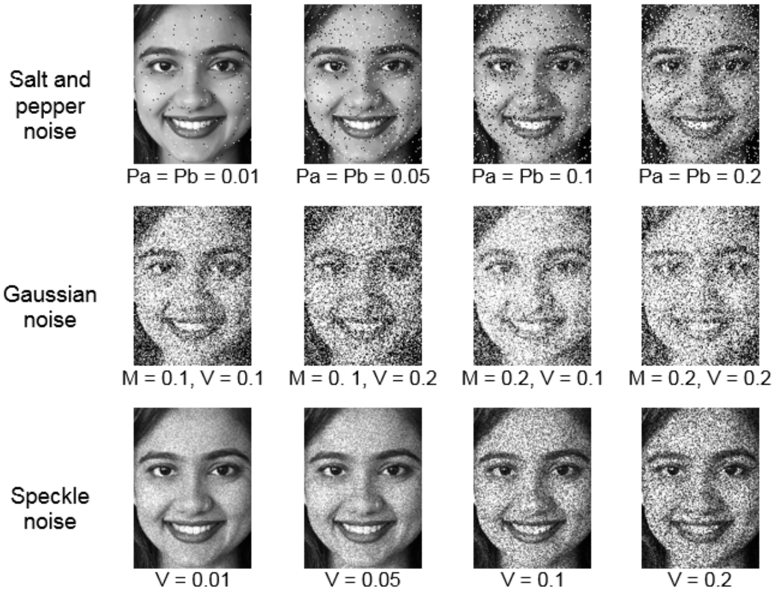

Figure 15. Images corrupted with different noises

Gaussian noise is controlled by two parameters, mean and variance. As can be seen from the Fig. 3.18, gaussian noise corrupts images visually higher than other two noises. And hence it has more diverse effect on accuracy.

		Template Matching		Machine Learning	
Mean	Variance	Chi-Square	Cosine	LS-SVM (RBF)	Discriminant Analysis
0.1	0.1	90.76	90.29	90.48	91.05
0.1	0.2	92.19	91.81	90.76	92.19
0.2	0.1	92.76	92.95	92.10	92.57
0.2	0.2	92.95	93.24	92.38	92.48

Table V: Results on JAFFE in noisy environment, Noise: Gaussian

	Template Matching		Machine Learning	
Variance	Chi-Square	Cosine	LS-SVM (RBF)	Discriminant Analysis
0.01	93.33	93.33	93.33	93.33
0.05	94.00	94.00	94.00	94.00
0.1	93.71	93.52	93.24	93.43
0.2	92.67	92.67	92.00	93.05

Table VI: Results on JAFFE in noisy environment, Noise: Speckle

4.5 Expression Recognition in Low Resolution

It is not always possible to acquire high-quality images. In certain applications such as home monitoring, surveillance applications, smart meeting, only low-resolution videos are available (Tian [2004]). Expression recognition in low resolution is an almost unaddressed area. area. Even for a human being, recognizing facial expressions in low resolution objects is a challenging task. In our experiment, we studied the performance of MLH operator in four different resolutions: 150 × 110, 75 × 55, 48 × 36 and 37 × 27. Low-resolution images are derived by down-sampling the original images. Results on JAFFE and TFEID are portrayed in Figure 16.

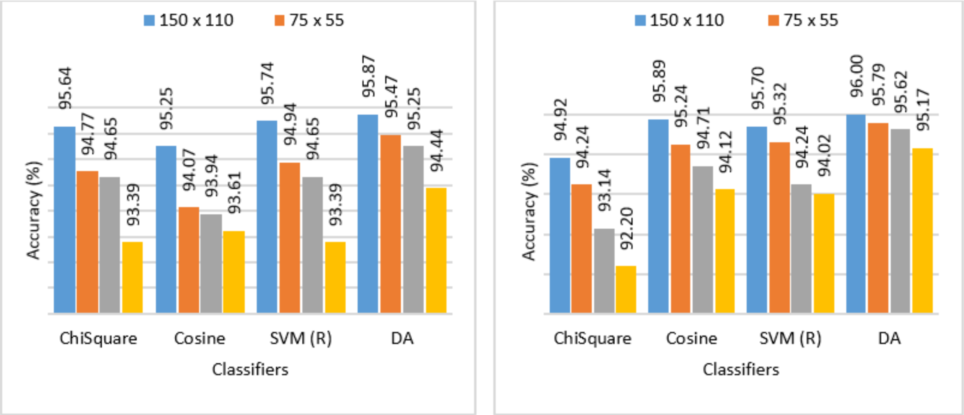

Figure 16. Recognition rate of MLH in low resolution on JAFFE (left) and TFEID (right) datasets.

For JAFFE dataset, the average recognition rate of all four classifiers for 150 × 110 resolution is 95.6%, which is 1.9% higher than the recognition rate in case of 37 × 27 resolution, which has an average recognition rate of 93.7%. Performance degradation with lower resolution is stated in Table VII. Results confirm that the performance decreases with lower resolution.

Resolution	150 × 110	75 × 55		48 × 36		37 × 27	
Classifiers	Acc. (%)	Acc. (%)	Degradation	Acc. (%)	Degradation	Acc. (%)	Degradation
Chi-Square	95.6	94.8	0.8	94.7	0.9	93.4	2.2
Cosine	95.3	94.1	1.2	93.9	1.3	93.6	1.6
LS-SVM(R)	95.7	94.9	0.8	94.7	1.1	93.4	2.3
DA	95.9	95.5	0.4	95.3	0.6	94.4	1.4
Average	95.6	94.8	0.8	94.6	1.0	93.7	1.9

Table VII: Average recognition rate over all four classifiers

5. CONCLUSION AND FUTURE WORK

Till date, Gabor has been considered to be one of the prominent texture analysis tools. Many researchers have exploited the ability of Gabor to extract the features of different scale and orientation. But Gabor is computationally intensive and time-consuming. In this work, we present a simple yet effective, haar wavelet based facial expression recognition technique. Haar can effectively express the signal with comparatively less features. Ability of haar to preserve the energy of the signal even with very few coefficient makes it a choice for pattern recognition. Since full facial image provides redundant and less valuable information, the method first detects facial components dominating facial expressions. Haar features are computed only for these components rather than the entire image. This reduces feature dimension without losing much information. The performance of the method is tested on three different datasets. We prove the effectiveness of coding of facial expression through haar wavelets. Many times, kernel-based techniques may fail to capture features of different size. The proposed method extract features at various scales via multi-level haar decomposition.

Due to its non linear classification property and kernel trick, SVM maps the non separable samples into separable classes by mapping it into high dimensional space. Discriminant analysis classifier separates the classes by projecting the class samples on the axises which gives the optimal value for the between class scatter to within class scatter. And the theoretical foundation of the classifiers is also supported by the practical proofs shown in Figure 12, where LS -SVM and DA achieves recognition rate of 97.00% and 97.23% for the 2 fold validation on JAFFE dataset.

References

Almaev, T. R. and Valstar, M. F. 2013. Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. In *Humaine Association Conference on Affective Computing and Intelligent Interaction.* pp.356–361.

Bartlett, M. S., Littleworт, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J. 2005. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* pp.568–573.

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence Vol.19,* No.7.

Chen, L.-F. and Yen, Y.-S. 2007. Taiwanese facial expression image database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.

Corneanu, C. A., Marc, O., Cohn, J. F., and Sergio, E. 2016. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.38,* No.8.

D. Huang, C. Shan, M. A. Y. W. and Chen, L. 2011. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews Vol.41,* No.6.

Darvin, C. 1872. The expression of the emotions in man and animals. J. Murray, London.

Ekman, P. and Friesen, W. V. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology Vol.17,* No.2.

Ekman, P. and Friesen, W. V. 1978. Facial action coding system: A technique for measurement of facial movement. Consulting Psychologists Press.

Essa, I. A. and Pentland, A. P. 1997. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.19,* No.7, pp.757–763.

Fasel, B. and Luettin, J. 2003. Automatic facial expression analysis: A survey. *Pattern Recognition Vol.36,* No.1.

G. Wenfei, C. Xiang, Y. V. V. D. H. and Lin, H. 2012. Facial expression recognition using radial encoding of local gabor features and classifier synthesis. *Pattern Recognition Vol.45,* No.1.

Gao, Y., Leung, M., Hui, S. C., and Tananda, M. 2003. Facial expression recognition from line-based caricatures. *IEEE Transactions on Systems, Man, and Cybernetics - Part A:Systems and Humans Vol.33,* No.3.

H. Boughrara, M. Chtourou, C. B. A. and Chen, L. 2016. Facial expression recognition based on a mlp neural network using constructive training algorithm. *Multimedia Tools and Applications Vol.75,* No.2, pp.709–731.

J. M. Guo, S. H. T. and Wong, K. 2016. Accurate facial landmark extraction. *IEEE Signal Processing Letters Vol.23,* No.5, pp.605–609.

Jabid, T., Kabir, M. H., and Chae, O. 2010. Robust facial expression recognition based on local directional pattern. *ETRI Journal Vol.32,* No.5.

Kanade, T., Cohn, J. F., and Tian, Y. 2000. Comprehensive database for facial expression analysis. In *4th IEEE International Conference on Automatic Face and Gesture Recognition.* pp.46–53.

Liu, W. and Wang, Z. 2006. Facial expression recognition based on fusion of multiple gabor features. In *18th International Conference on Pattern Recognition.* pp.536–539.

Lyons, M. and Akamatsu, S. 1998. Coding facial expressions with gabor wavelets. In *3rd IEEE Conference on Automatic Face and Gesture Recognition.* pp.200–205.

Lyons, M. J. 1999. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.21,* No.12.

M. S. Bartlett, J. R. M. and Sejnowski, T. J. 2002. Face recognition by independent component analysis. *IEEE Transactions on Neural Network Vol.13,* No.6.

M. Song, D. Tao, Z. L. X. L. and Zhou, M. 2010. Image ratio features for facial expression recognition application. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics Vol.40,* No.3, pp.779–788.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.11,* No.7, pp.674–693.

Mehrabian, A. 1968. Communication without words. *Psychology Today Vol.2*, pp.53–55.

Moore, S. and Bowden, R. 2011. Local binary patterns for multi-view facial expression recognition. *Computer Vision and Image Understanding Vol.115,* No.4.

Ojala, T., Pietikainen, M., and Harwood, D. 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition Vol.29,* No.1.

Ojala, T., Pietikainen, M., and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.24,* No.7.

Oliveira, L., Koerich, A., Mansano, M., and Britto, A. 2011. 2d principal component analysis for face and facial-expression recognition. *Computing in Science and Engineering Vol.13*, pp.9–13.

Owusu, E., Zhan, Y., and Mao, Q. R. 2014. A neural-adaboost based facial expression recognition system. *Expert Systems With Applications Vol.41,* No.7, pp.3383–3390.

Pantic, M. and Patras, I. 2006. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics Vol.36,* No.2, pp.433–449.

Papageorgiou, C. and Poggio, T. 2000. A trainable system for object detection. *International Journal of Computer Vision Vol.38,* No.1, pp.15–33.

R. A. Khan, A. Meyer, H. K. and Bouakaz, S. 2013. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters Vol.34,* No.10, pp.1159–1168.

R. Hablani, N. C. and Tanwani, S. 2013. Recognition of facial expressions using local binary patterns of important facial parts. *International Journal of Image Processing Vol.7,* No.2.

Rahulamathavan, Y., Phan, R. C.-W., Chambers, J. A., and Parish, D. J. 2013. Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Transactions on Affective Computing Vol.4,* No.1, pp.83–92.

Samad, R. and Sawada, H. 2011. Edge-based facial feature extraction using gabor wavelet and convolution filters. In *IAPR Conference on Machine Vision Applications.* pp.430–433.

Shan, C., Gong, S., and Mcowan, P. W. 2009. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing Vol.27,* No.6.

Shih, F. Y., Chuang, C.-F., and Wang, P. 2008. Performance comparisons of facial expression recognition in jaffe database. *International Journal of Pattern Recognition and Artificial Intelligence Vol.22,* No.3.

Stankovir, R. S. and Falkowski, B. J. 2003. The haar wavelet transform: Its status and achievements. *Computers and Electrical Engineering Vol.29,* No.1, pp.15–33.

Suwa, M., Sugie, N., and Fujimora, K. 1978. A preliminary note on pattern recognition of human emotional expression. In *International Joint Conference on Pattern Recognition.* pp.408–410.

T. Senechal, V. Rapp, H. S. R. S. K. B. and Prevost, L. 2012. Facial action recognition combining heterogeneous features via multikernel learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics Vol.42,* No.4.

T. Wu, M. S. B. and Movellan, J. R. 2010. Facial expression recognition using gabor motion energy filters. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.*

Tian, Y. 2004. Evaluation of face resolution for expression analysis. In *CVPR Workshop on Face Processing in Video.*

Turk, M. A. and Pentland, A. P. 1991. Face recognition using eigenfaces. *Journal of Cognitive Neuroscience Vol.3,* No.1.

Wang, X., Liu, A., and Zhang, S. 2015. New facial expression recognition based on fsvm and knn. *Optik - International Journal for Light and Electron Optics Vol.126,* No.21, pp.3132–3134.

Yacoob, Y. and Davis, L. S. 1996. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.18,* No.6.

Yang, J., Zhang, D., Frangi, A. F., and Yang, J. 2004. Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.26,* No.1.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.31,* No.1.

Zhi, R. and Ruan, Q. 2008. Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing Vol.71,* No.7, pp.1730–1734.

Mahesh Goyani has obtained Bachelors Degree in field of Computer Engineering from Veer Narmad South Gujarat University, Surat, India in 2005, and did post-graduation from Sardar Patel University, V.V.Nagar, India in 2007. At present, he is serving as an Assistant Professor at Department of Computer Engineering, Government Engineering College, Modasa, India. He has published number of books in area of Computer Graphics and Analysis of Algorithm. He also has published many research paper in reputed journals and conferences. His area of interest includes Pattern Recognition, Machine Learning and Image Processing. He is also a Research Scholar at Charotar University of Science and Technology, Changa, India. Prof. Goyani has also served as a reviewer in many international journals. He has been a committee member in many national and international conferences. He is a life time member of Indian Society of Technical Education. He was also invited as a session chair in International Conference on Computer Science, Engineering and Information Technology, Tirunveli, Tamilnadu, India in 2011.

Prof. Narendra Patel received his B.E. and M.E. degrees in computer engineering from the M.S. University, India, in 1995 and 1998, respectively, and the Ph.D. degree in facial animation from SVNIT, India in 2011. Currently, he is an associate professor in the Department of Computer Engineering at BVM Engineering College, India. He has published about 60 refereed journal and conference papers. His research interest covers Image Processing, Cloud Computing and Distributed Operating Systems. Prof. Author received research award from Science Foundation, and the Best Paper Award of the IS International Conference in 2000 and 2006, respectively. He is a member of SICE, IEE and IEEE.