We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Nonverbal Communication through Facial Expression in Diverse Conditions

Mahesh Goyani

Abstract

In this chapter, we investigated computer vision technique for facial expression recognition, which increase both - the recognition rate and computational efficiency. Local and global appearance-based features are combined in order to incorporate precise local texture and global shapes. We proposed Multi-Level Haar (MLH) feature based system, which is simple and fast in computation. The driving factors behind using the Haar were its two interesting properties - signal compression and energy preservation. To depict the importance of facial geometry, we first segmented the facial components like eyebrows, eye, and mouth, and then applied feature extraction on these facial components only. Experiments are conducted on three well known publicly available expression datasets CK, JAFFE, TFEID and in-house WESFED dataset. The performance is measured against various template matching and machine learning classifiers. We achieved highest recognition rate for proposed operator with Discriminant Analysis Classifier. We studied the performance of proposed approach in several scenarios like expression recognition from low resolution, recognition from small training sample space, recognition in the presence of noise and so forth.

Keywords: facial expression recognition, multi-level Haar, local mean binary pattern, local Haar mean binary pattern, classification

1. Introduction

1.1 Background

Communication is not possible without some channel. Communication between human is modeled by two ways: multiplicity and modality. Multiplicity defines more than one way for the communication and modality defines the way human senses are used to perceive signals from outer world [1]. Speech and vocal information are communicated through the auditory channel, whereas facial expression is communicated via the visual channel. Organs such as nose, ear, skin provides the different modalities for the communication. Multi-channel communication is highly robust; failure of one channel can be compensated by another channel.

Facial expression provides an important behavioral measure for studies of emotion, cognitive processes, and social interaction [2]. For a human being, recognition of face and facial expression is a trivial task. We discriminate the faces with almost

no effort in a fraction of a second. But it is equally challenging to teach a machine to perform the same task.

Expressions are not mere changes in muscle position, rather a complex psycho-physiological process. The psychological process of thoughts emerging in mind is followed by a physiological process in which the thoughts render as expressions on the face by means of muscle deformation. The muscle movement lasts for a brief period of about 250 ms to 5 sec. Hence recognizing expressions from the spontaneous image is harder compared to posed still images [3].

Recognition of pure expression is difficult to wide range of expressions, as well as a same expression might have different intensities. Schmidt and Cohn [4] noted 18 unique classes of the expression smile. Intensity of expression can vary from gentle to peak.

1.2 Expression representation

According to the psychological and neurophysiological studies, there are six basic emotions. Each basic emotion is associated with one unique facial expression. Facial expressions can be represented using: discrete category model/Judgmental Coding Scheme (prototype expressions) or Facial Action Coding System (FACS) model.

1.2.1 Judgmental coding scheme

As the name suggests, the model classifies the expressions on subjective judgment. Prototypic expressions are a subjective measure of the texture like wrinkles, bulges, furs on the face, which is useful for judging the expression. Ekman and Friesen categorized expressions into six classes: Happiness (HA), Sadness (SA), Surprise (SU), Anger (AN), Fear (FE), and Disgust (DI) [5], which are portrayed in **Figure 1**. These six expressions serve as a ground truth label, and instead of distinguishing the comprehensive facial features, most FER systems attempt to recognize a small set of these prototypic expressions.

The other way of describing expression is using geometry of the face. Judgmental coding scheme based algorithms use appearance features for expression recognition. Descriptive coding scheme described in next section uses geometry of the face for expression recognition, which is more robust compared to judgmental coding scheme. However, extraction of exact Action Unit (AU) is challenging.

1.2.2 Descriptive coding scheme

Later, in 1978, Ekman and Friesen developed the Facial Action Coding System (FACS) [6], which describes and encodes the facial expressions based on the movements of the facial muscles. These codes are called action units (AUs). FACS
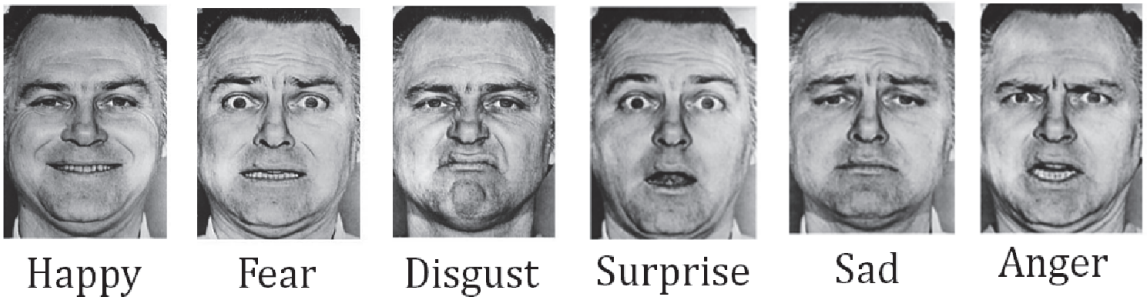

Happy   Fear   Disgust   Surprise   Sad   Anger

Figure 1.
Basic six expressions postulated by Ekman and Friesen.

Upper Face Action Units						Lower Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7	AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener	Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46	AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink	Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler

Figure 2.
Few of the upper and lower face action units.

identifies the facial muscles that cause changes in particular facial expression, thus enabling facial expression analysis. FACS consists of 44 action units describing the facial behaviors. FACS and the six prototypic expressions form the foundation for facial expression analysis and recognition research. **Figure 2** shows few of the upper and lower face action units.

Action Units can be additive or non-additive. If the appearance of AU is independent, then AU is said to be additive. When expression changes, different AU's get activated. In some expression, AU's are mixed and hence change the appearance of some AU during muscle deformation. AUs are said to be non-additive if they modify each other's appearance [7]. Each expression can be represented as a combination of one or more additive or non-additive AUs. For example, 'fear' can be represented as a combination of AUs 1, 2 4, 5, 7 and 26. Ekman and Friesen reported more than 7000 such combinations of the AUs [8]. In order to get the expression estimation, the FACS code needs to be converted into the Emotional Facial Action System. Even a good trained coder takes one to three hours of time to label one minute video on a frame by frame basis [9].

1.3 Applications

FER plays a vital role in many applications, such as human-computer interaction, indexing and retrieving images based on expressions, emotion analysis, image understanding, synthetic face animation, etc. A comprehensive study on recent advancements in affect recognition and its applications to HCI can be found in a survey by Zeng et al. [10].

Online Multiplayer Games (MOG) are increasingly becoming popular. Many FER based MOGs have been studied and proposed [11]. Applications of FER are not just limited to the physiological domain; rather it has touched many aspects of engineering, medical, social communication, entertainment, and automation. Application area of FER covers a broad spectrum, including grading of physical pain, smile detection [12–14], driver fatigue detection [15], patient pain assessment [16], video indexing, robotics and virtual reality [3], depression detection [10] etc.

Bartlett et al. [17] have successfully used their face expression recognition system to develop an animated character that mirrors the manifestations of the user. They have also managed to deploy the recognition system on Sony's Aibo Robot and ATR's RoboVie [17]. Anderson and McOwen developed an interesting application called the 'EmotiChat' [18]. It provides set of emojis for the easier and quick communication. The FERS is connected to this chat application, and it automatically inserts emoticons based on the user's facial expressions. Recently, Microsoft developed a fascinating Emotion API [19], which detects the face from an image and finds the weight of each expression.

1.4 Scopes and challenges

The main issue in the design of ideal automated expression analyzer is the degree of automation. All the stages - face detection, facial representation, and expression classification – should be fully automated. However, incorporation of these operations in the system depends on an application where the analyzer is to be used. Real-time performance is not expected if the analyzer is to be utilized for the study of behavioral science. Whereas, running time of the systems is an important issue for advanced user interfaces, in which delay of few seconds makes the system non-effective or non-usable [20].

Expression recognition in the low-resolution environment is almost unaddressed. Real-time videos like conference recordings, surveillance videos etc. are normally available in low resolution. Precise recognition of expression in such environment is challenging task. In 2004, Tian [21] used geometric and appearance based features to perform expression recognition in low-resolution images. Bartlett et al. [8] evaluated the performance of Gabor features and achieved noticeable accuracy. Later, in 2009 Shan et al. [22] investigated Gabor and LBP features for FER in a similar environment. Jabid et al. [23] evaluated the performance of Local Directional Pattern (LDP) features for low-resolution images.

To provide the standardized platform, Facial Expression Recognition Analysis (FERA) challenge events are being held by Social Signal Processing Network (SSPNET) in conjunction with Face and Gesture Recognition Group. Two such editions of FERA were held in 2011 at Santa Barbara [24], California and in 2015 at Ljubljana, Slovenia [25]. FERA 2017 is to be held in Washington, USA in March 2017. FERA brings the researchers across the globe under a common roof to understand and solve the issues of FER.

Based on the study of previous work, we list out following challenges in the field of facial expression recognition:

- Approaches are evaluated for person dependent databases only.

- Generalizing approach for spontaneous expressions is still an open area.

- The Very little contribution is made for the occluded facial expression recognition.

- Work cited in the different literature is addressing only one or two databases.

- Facial expression recognition under noisy environment is rarely addressed.

- The system is expected to run effectively on profile views.

- Expression recognition under low-resolution environment is still an almost unaddressed issue.

2. Multi level Haar wavelet based system

Texture and geometry convey complementary yet important information for FER. Studies [10] have shown that facial expression information is equally conveyed by geometric fiducial points and texture features. It has been observed that expression might have similar texture features bud different geometric features and

vice versa [26]. Experiments have shown that combination of both the type of features could prove better for implementation of FERS [27].

The proposed method detects the facial components, which in turn, effectively reduces the computation and improves the accuracy. A prototype human face is shown in **Figure 3**. Rectangle with dotted border indicates the region of interest which can further be used for feature extraction. The Larger region itself may contain smaller regions of interest within it.

2.1 Pre-processing

Preprocessing of the face and locating Region of Interest (ROI) is a crucial step for robust feature extraction. Segmentation of local facial components leads to significant reduction in computation cost of both - the feature extraction and classification. Following Tian [21], Shan et al. [22] and Baughrara et al. [28], we used fixed eye distance-based approach to normalize the face. Shan et al. [22] 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. We used iterative approach for eye registration. At first, eye pair is detected using cascade object classifier proposed by Viola-Jones (Refer **Figure 4**). Eye segment is thresholded and complemented using global threshold estimation.
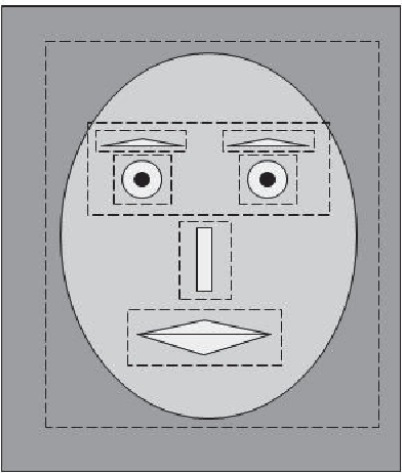

Figure 3.
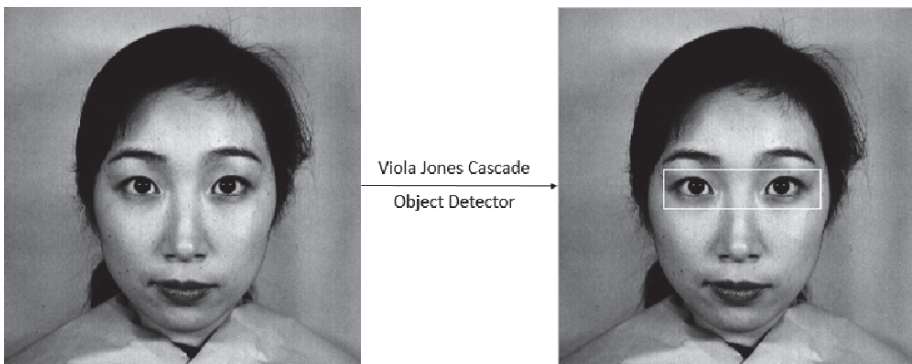*Face prototype and few regions of interest.*

Figure 4.
Eye detection using Viola-Jones cascade object detector.

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 with 3×3 structuring element having all 1's 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,

$$A \oplus B = \{z \in E \mid B_z \subseteq A\} \qquad (1)$$

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

$$B_z = \{b + z \mid b \in B\}, \forall x \in E \qquad (2)$$

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 θ. Otherwise, it would be non-zero, and the face is aligned by performing negative rotation of the angle around the z-axis. Let Δx and Δy 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,

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

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 5** demonstrates the angle estimation for slant face.

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.
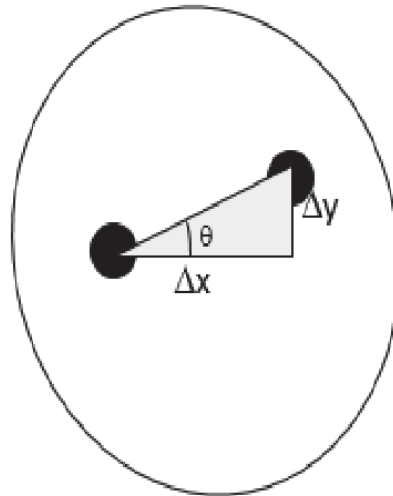
Figure 5.
Angle estimation from eyeball for eye registration.

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

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 6**.

This process registers the eye of all the images used in the 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
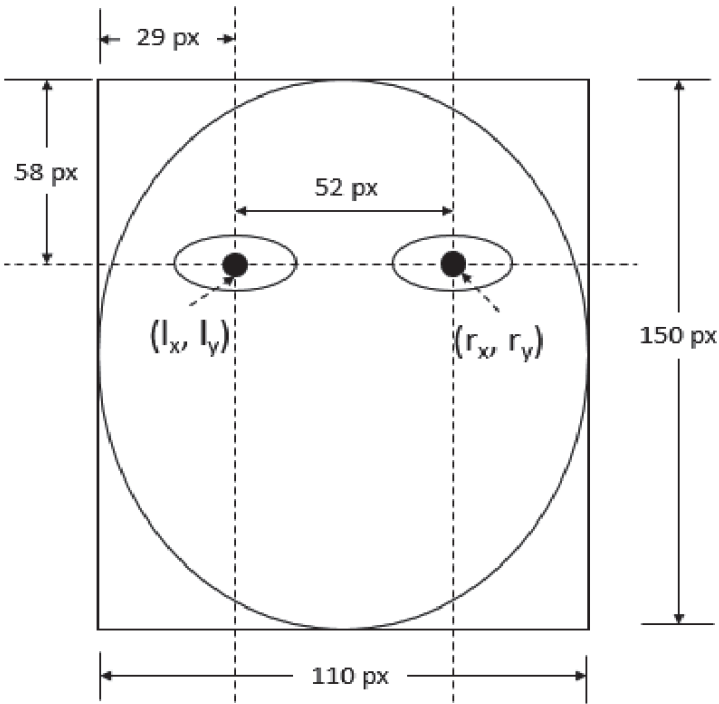
Figure 6.
Facial dimensions estimated from eyeball distance.
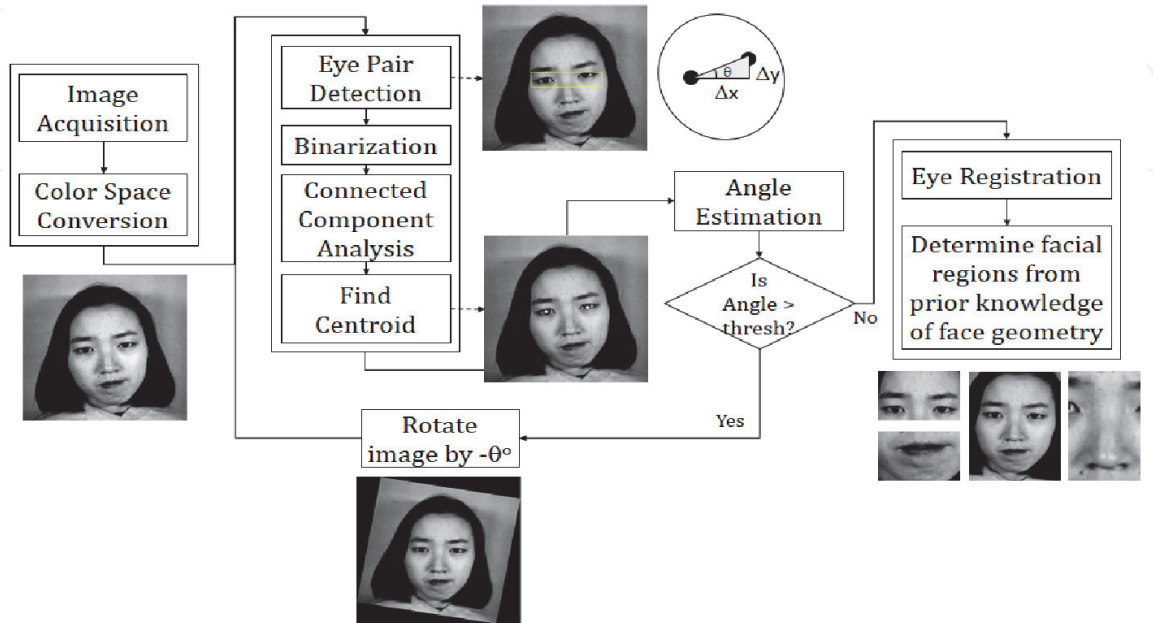
Figure 7.
Preprocessing flow for face and facial component extraction.

Component	Face	Eye	Mouth	Nose
Resolution	150×110	60×90	40×60	70×40

Table 1.
Dimensions of cropped facial components.

regions for expression recognition, and hence we also cropped top and bottom face regions. The entire process is explained in **Figure 7**.

Extracted geometric components are normalized and send to feature extraction module. The normalized size of the individual component is listed in **Table 1**.

2.2 Feature extraction

Haar functions were introduced by mathematician Alfred Haar [29]. A Haar wavelet is the simplest type of wavelet. It decomposes the image into one low-frequency band and number of high-frequency bands, known as coarse signal and detail signals respectively. Results are analogs to the output of low pass and high pass filters. Corse signal is an approximation of luminance and chrominance distribution of the original signal. In discrete form, Haar wavelets are related to a mathematical operation called the Haar transform. The Haar transform serves as a prototype for rest of all wavelet transforms. It provides a natural mathematical structure for describing the patterns [30].

The digital image is a discrete signal, which is a function of time with values occurring at discrete positions or time intervals. A discrete signal of length N is represented as $f = (f_1, f_2, \dots, f_N)$. The values $f_1, f_2, \dots, f_N$ are the approximation of analog signal g, measured at the time intervals $t = t_1, t_2, \dots, t_N$. Components of signal f are obtained as,

$$f_1 = g(t_1), \quad f_2 = g(t_2), \quad \dots, \quad f_N = g(t_N) \tag{5}$$

Haar wavelet decomposes the signal into two sub signals called *running average* or *trend* and *running difference* or *fluctuation*. The first trend sub signal, $a1 = (a_1, a_2, \dots, a_{N/2})$, for the signal f, is computed by taking a running average of a pair of components of f. Mathematically, we can compute a_i as,

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

for m = 1, 2, …, N/2. Multiplication of average by $\sqrt{2}$ is needed in order to ensure that the Haar transform preserves the energy of a signal.

The other sub signal is called the first fluctuation, which is denoted by $d1 = (d_1, d_2, \dots, d_{N/2})$, and it is computed by taking a running difference of a pair of values of f. In general,

$$d_m = \frac{f_{2m-1} - f_{2m}}{\sqrt{2}} \tag{7}$$

for m = 1, 2, …, N/2. The Haar transform is performed in several stages, or levels. The first level is the mapping H_1 defined by,

$$f \xrightarrow{H_1} (a^1 \mid d^1) \tag{8}$$

The mapping H_1 in Eq. (8) has an inverse. Its inverse maps the transformed signal $(a^1|d^1)$ back to the signal f, via the following formula:

$$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, \dots, \frac{a_{N/2} + d_{N/2}}{\sqrt{2}} \right) \tag{9}$$

For Multi-Level Haar feature extraction, the same decomposition is repeatedly applied to latest trend signal in each iteration.

2.3 Experimental setup

We conduct the experiments on three widely used comprehensive datasets, Cohn-Kanade (CK) [31], Japanese Female Facial Expression (JAFFE) [32] and Taiwanese Facial Expression Image Database (TFEID) [33]. Existing datasets rarely address the issues of spontaneous expressions. Most of the time, images are acquired under a static environment with a 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. WESFED dataset was created by collecting the images from google. Random images have been processed, faces were detected using Viola Jones Cascade face detector. Each face was manually labeled by 10 persons. SVM based model was also created to classify the expression of cropped face. Majority voting based scheme was employed to label the face. WESFED dataset contains subjects from various country, different age groups, different head positions, varying illumination condition and so on. Details of a number of images used for the experiment from all datasets are listed in **Table 2**.

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 8**.

2.4 Result analysis

2.4.1 Optimal parameter selection

The performance of the algorithm is bound to many parameters like a number of features, the number of images used to train the model, regions size used to compute the features, etc. Derivation of optimal combination of parameters follows here:

A. *Estimation of Number of Eigenvectors*

More often than not, kernel methods generate a large feature vector. It has few obvious disadvantages. Training the classifier with such a large input vector is time-

	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 2.
Number of images used for experiment from datasets.

consuming and often leads to poor generalization. Dimensions of the feature vector has been reduced by Linear Discriminant Analysis (LDA). To handle the issue of singular matrix problem, we have employed Principal Component Analysis (PCA). We applied PCA on original feature vector, and LDA is applied on PCA subspace. For C class problem, LDA produces (C – 1) features.

To find the optimal number of features, we varied the number of eigenvectors from 20 to 200 in step of 20. **Table 3** 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 (CS) and Cosine distance (CO), and two machine learning classifiers –Least Squares Support
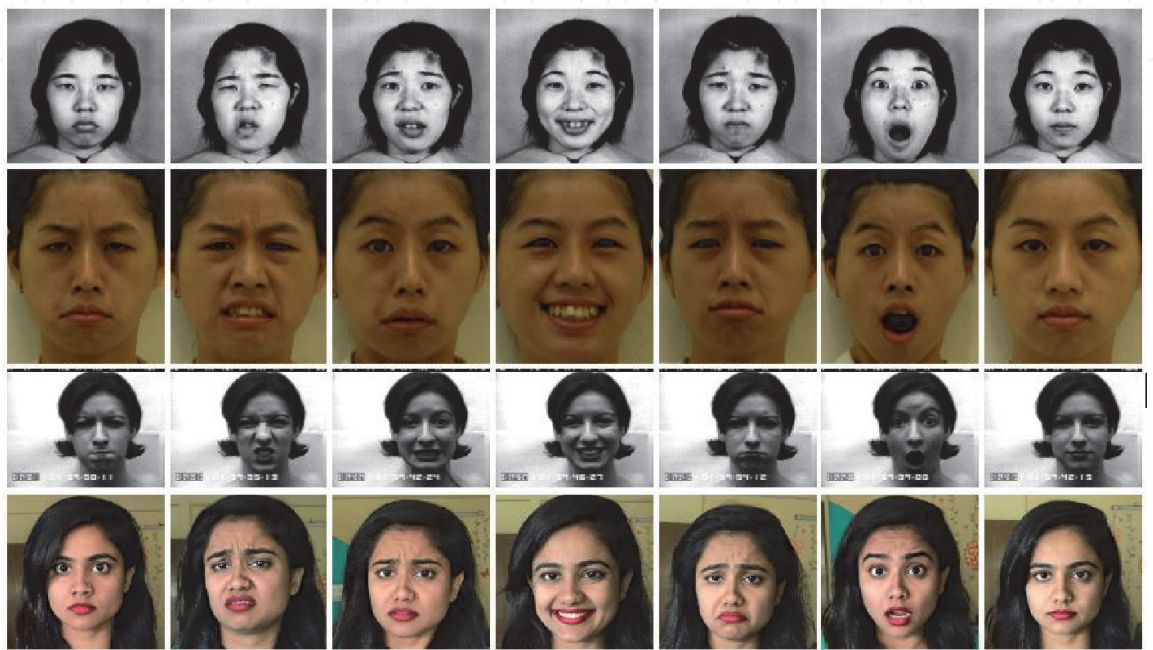
Figure 8.
Snapshots of various expressions from JAFFE (first row), TFEID (second row), CK (third row) and WESFED (fourth row) datasets.

	Template Matching		Machine Learning		AP[a]	API[b]
#	CS	CO	LS-SVM	DA		
20	76.43	76.24	71.57	73.48	74.43	0.00
40	91.10	90.33	90.90	91.86	91.05	16.62
60	94.62	93.95	93.19	94.90	94.17	3.12
80	96.05	95.86	95.38	95.76	95.76	1.60
100	96.62	96.52	96.52	96.62	96.57	0.81
120	96.62	96.62	96.43	96.62	96.57	0.00
140	97.00	97.00	97.00	97.00	97.00	0.43
160	97.00	97.00	97.00	97.00	97.00	0.00
180	97.00	97.00	97.00	97.00	97.00	0.00
200	97.00	97.00	97.50	98.00	97.38	0.38

[a]*Average Performance.*
[b]*Average Performance Improvement.*

Table 3.
Effect of eigenvectors on performance (%), dataset: JAFFE.

Vector Machine (LS-SVM) with RBF kernel and Discriminant Analysis (DA) classifier. Performance of all four classifiers is averaged to find Average Performance (AP) and Average Performance Improvement (API) is analyzed to select the optimal number of directions for PCA projection.

To choose the optimal number of eigenvectors, we averaged the performance of all four classifiers. To balance the accuracy-computation trade-off, we choose 140 eigenvectors for the further analysis.

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. 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 three datasets with common parameters and results are shown in **Figure 9**.

B. Estimation of Region Size.

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,
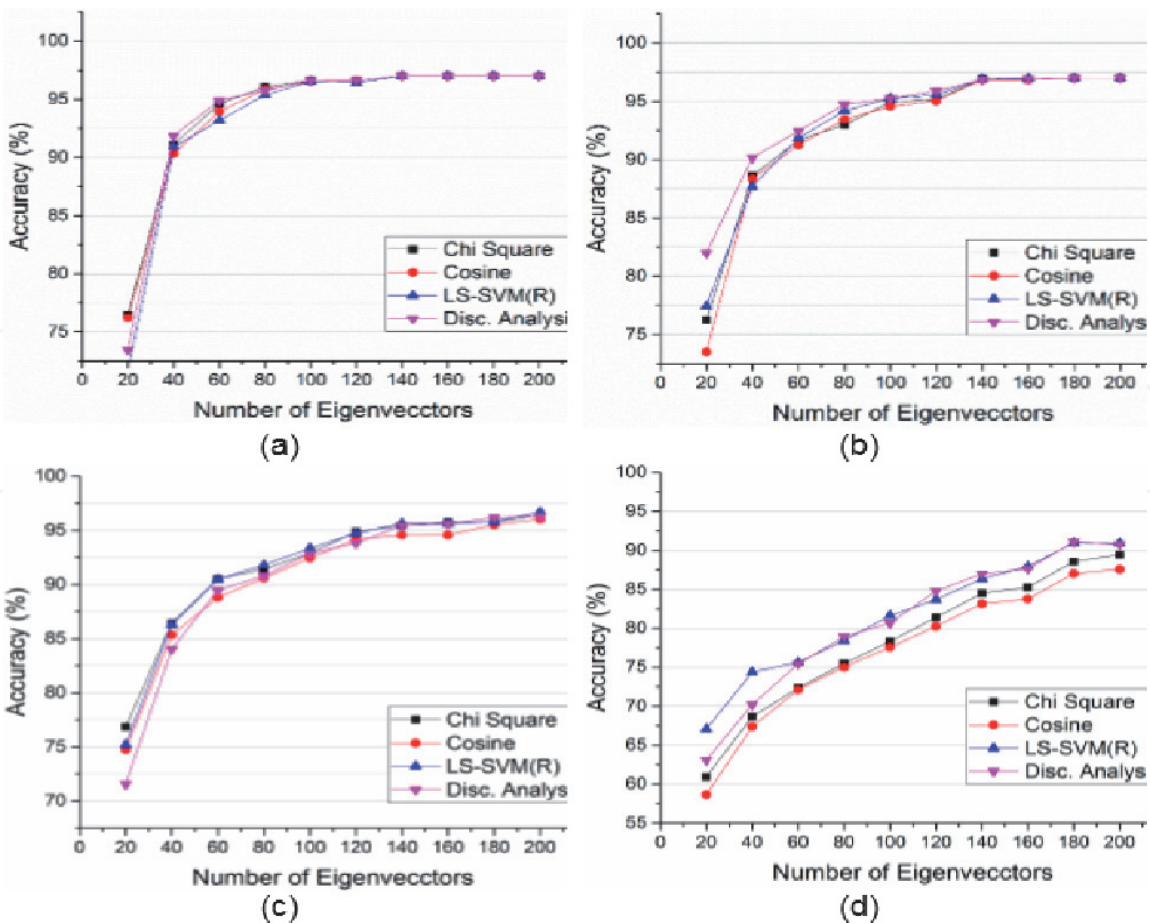

Figure 9.
Plot of a number of eigenvectors vs. accuracy (%) for (a). JAFFE, (b). TFEID, (c). CK and (d). WESFED datasets.

$5 \times 5, 7 \times 6$ and 9×8 blocks. Bartlett et al. [8], Tian [21], Shan et al. [22], Jabid et al. [23] have also conducted experiments in these neighborhoods. Larger regions size fails to capture the texture of small size. Small regions effectively capture local and spatial relationship. However, after a 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 10**.

With 1×1 region, we can only derive the holistic features, which also suffers from the feature localization. From the **Figure 10**, we observed that algorithm performs well for 7×6 regions on both datasets. From above results, we chose a number of regions to be 7×6, as it gives a proper balance between accuracy and computational time. With more regions, dimensions of feature vector grow tremendously and PCA also takes more time to compute covariance matrix, eigenvector, and eigenvalues of huge feature matrix.

C. 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 a 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

Figure 10.
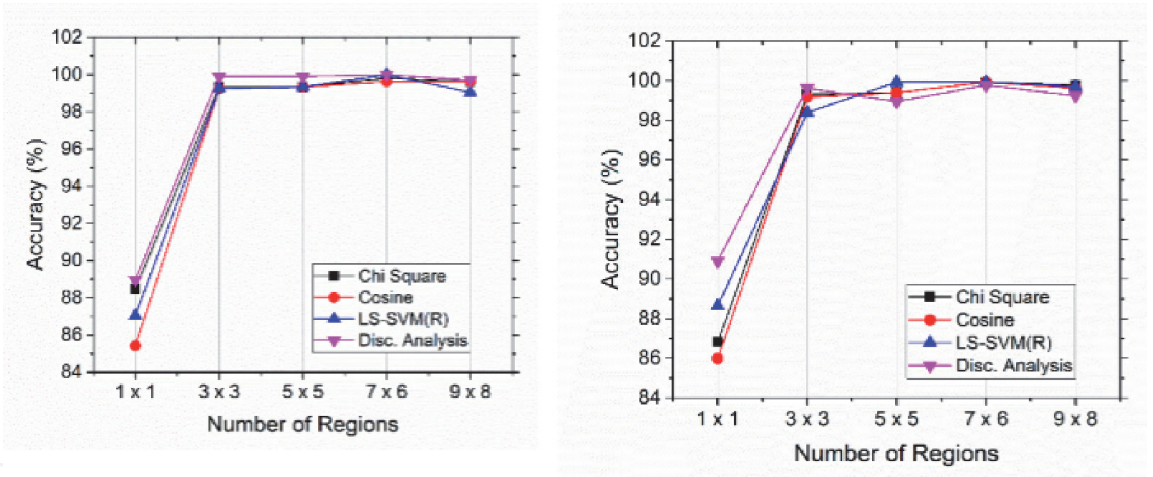*Effect of a number of regions on performance on JAFFE (left) and TFEID (right) dataset. Response for 7×6 regions is better compared other tested regions.*
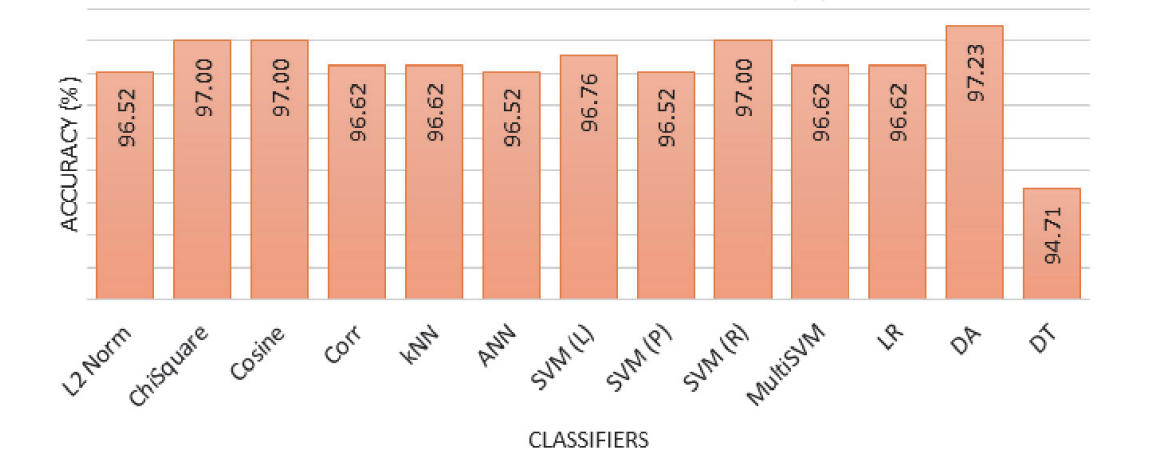

Figure 11.
Performance comparison of different classifiers, dataset: JAFFE.

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 11**.

Chi-square and Cosine measure gives the best classification results among all template matchers. 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.

D. 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. K-fold cross-validation is the most commonly used validation technique. 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 a 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. **Figure 12** exhibits the behavior of the system for various validation strategies. Varying number of sample
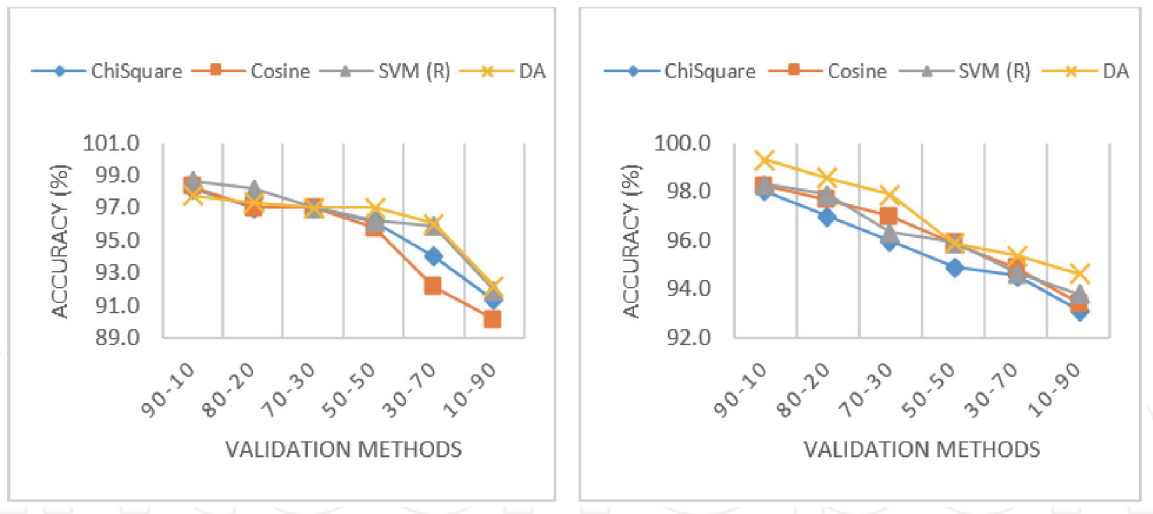
Figure 12.
Performance of various classifiers against different validation methods for JAFFE (left) and TFEID (right) dataset.

Parameter	Chosen value
Number of eigenvectors	140
Number of regions	7×6
Validation method	2-Fold (50% training – 50% testing)
Template matching classifier	Chi-square, cosine measure
Machine learning classifier	LS-SVM, discriminant analysis

Table 4.
Optimal parameters chosen for result analysis.

size is used to see the generalization of algorithm. If algorithm can give better results even with small number of training samples implies proposed algorithm is able to effectively capture the discriminating features of the image.

Based on the experiments, we choose parameters for the further analysis as shown in **Table 4**.

2.4.2 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. We performed expression recognition using MLH features extracted from the only eye, only mouth, eye + mouth, and face. Results are stated in **Figure 13** for JAFFE and TFEID datasets.

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 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, these changes are also incorporated and highest recognition rate is achieved.

2.4.3 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
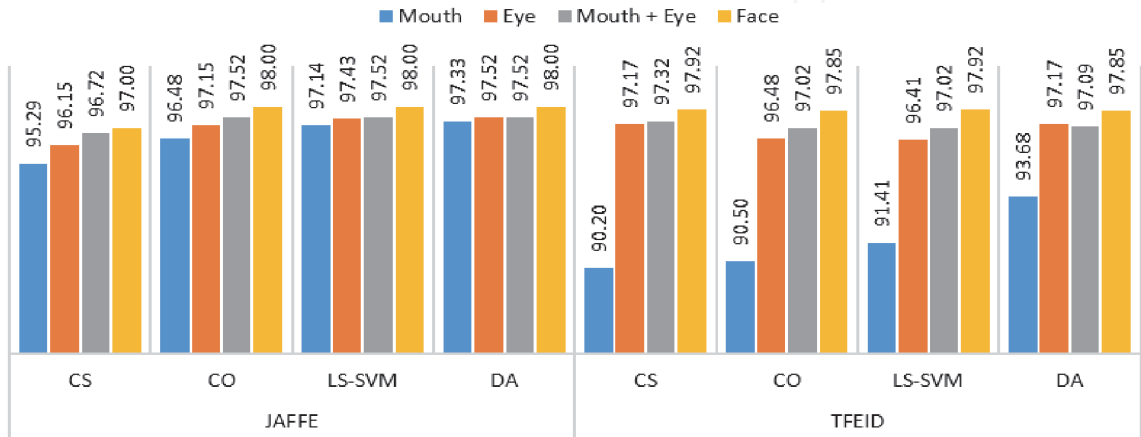

Figure 13.
Performance evaluation of various facial components on JAFFE and TFEID dataset.
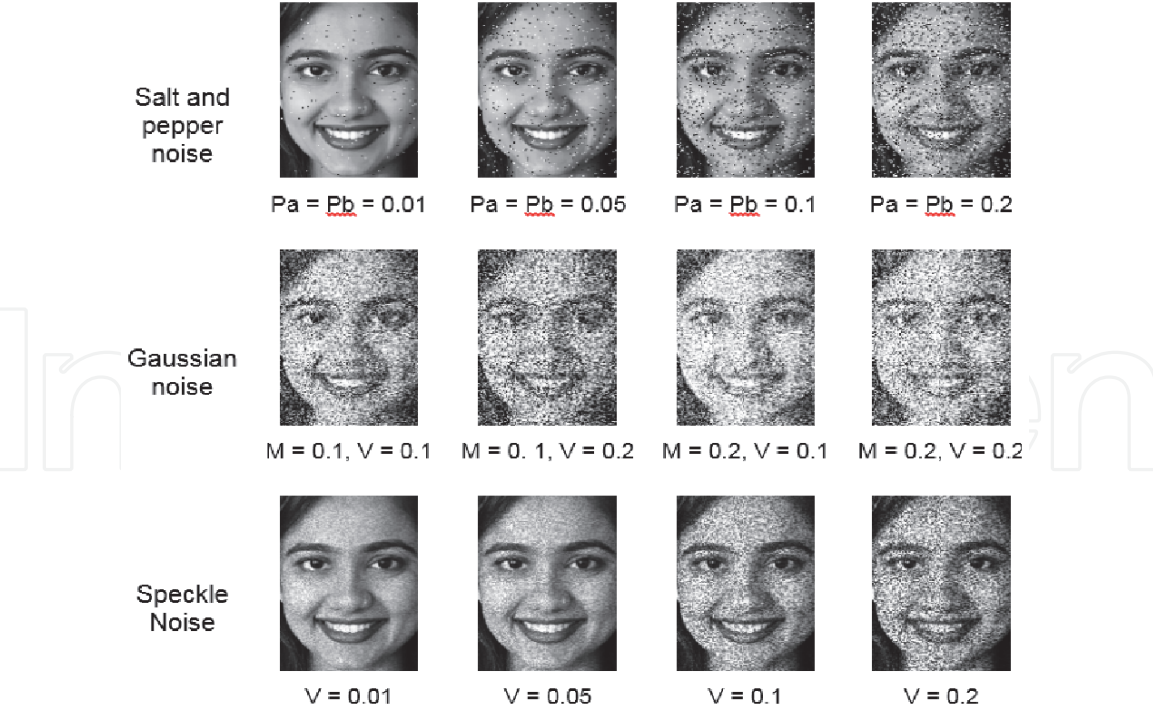
Figure 14.
Performance evaluation of various facial components on JAFFE and TFEID dataset.

	Template matching		Machine learning	
Variance	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 5.
Results on JAFFE in noisy environment, noise: Salt & Pepper.

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 14**.

Wavelets have shown good applications to noise removal. The selection of wavelet depends on the energy conservation in approximation subband. Haar possesses the nice property of signal compaction and energy preservation and hence they can prove an ideal choice for noise reduction. Salt and pepper noise has very high impact on the illumination of affected pixels. Robustness to noise is inherent in Haar. Performance of proposed method in presence of salt and pepper noise is shown in **Table 5**.

Gaussian noise is controlled by two parameters, mean and variance. As can be seen from the **Figure 14**, Gaussian noise corrupts images visually higher than other two noises. And hence it has a more diverse effect on accuracy and performance degrades compared to the presence of salt and pepper noise. Intensity disturbance created by speckle noise is less compared to Gaussian and hence the effect on performance is also less compared to Gaussian.

Due to wrinkles and aging, skin of aged people have more texture than younger ones. Suck skin texture may introduce noise effect in feature vector due to high variability in skin texture. Perhaps, the noise reduction works better with younger people.
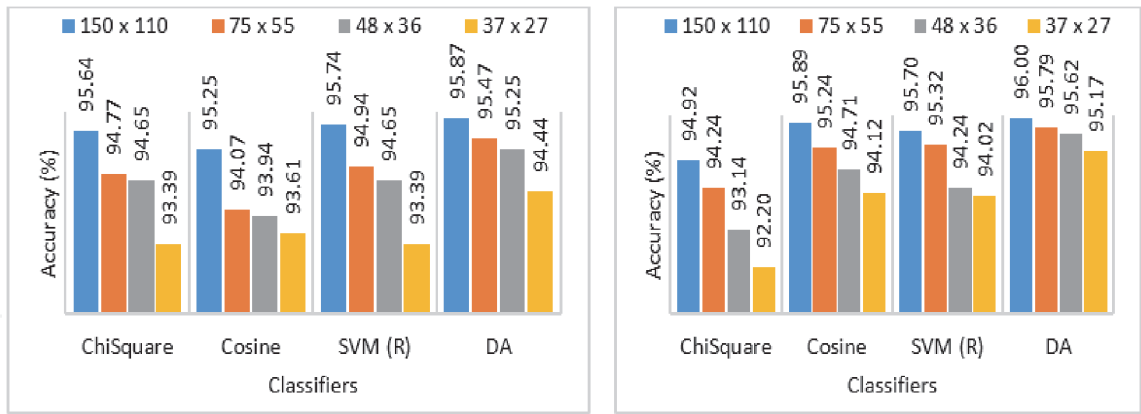
Figure 15.
Recognition rate of MLH in low resolution on JAFFE (left) and TFEID (right) datasets.

Resolution	150 × 110	75 × 55	48 × 36	37 × 27
Chi-square	95.6	94.8	94.7	93.4
Cosine	95.3	94.1	93.9	93.6
LSSVM (R)	95.7	94.9	94.7	93.4
DA	95.9	95.5	95.3	94.4
Average	95.6	94.8	94.6	93.7

Table 6.
Average recognition rate over all four classifiers.

2.4.4 Expression recognition in low resolution

Resolution can have significant effect on quality of the image. High resolution images may not be available always. Applications such as surveillance applications, home monitoring, smart meeting produces low resolution videos, which makes facial expression recognition difficult [21]. Very little work has been done on low resolution images. 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 15**.

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 6**. Results confirm that the performance decreases with lower resolution.

It is apparent that recognition of expression becomes difficult from low-resolution images. Even for a human it gets difficult. **Table 6** shows that the performance degradation for 75 × 55 is 0.8% but it is as high as 1.9% for 75 × 55 resolution.

3. Conclusions

This chapter presents preprocessing technique for face registration. Head pose angle is estimated and the head is rotated if needed to make it up-right frontal pose. Eyeballs are aligned in order to register the face. In this chapter, we have also

proposed Multi-Level Haar (MLH) based facial expression recognition system. The proposed method extracts level-1 and level-2 approximation coefficients of various facial components and feature vector are derived by concatenating these coefficients. Dimensions of the obtained feature vector are reduced by projecting it in PCA subspace followed by LDA subspace. Performance of the algorithm is evaluated in different scenarios like low resolution, noisy environment, low training sample space etc. Due to nice properties of Haar, and proper alignment of features due to preprocessing method, proposed method is able to achieve high recognition rate in diverse scenarios.

The work could be extended to various real world applications. For example, expression based song selection application can help users to create playlist of songs based on their mood and expressions. Face recognition based bio-metrics can be made more robust and secure by incorporating expression along with face. In class room, engagement level of the students can also be analyzed based on their facial expression, which could help teachers to understand the mood of students and change the teaching style. Facial expression based surveillance systems in shopping mall can be helpful to understand the customer's feedback from their expressions. Recommendation system based on facial expression can auto suggest the products to the customers. Facial expression plays very crucial role in nonverbal communication in society at different levels.

Author details

Mahesh Goyani
Gujarat Technological University, Gujarat, India

*Address all correspondence to: mgoyani@gmail.com

IntechOpen

References

[1] Pantic M, Rothkrantz LJM. Automatic analysis of facial expressions-the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000:1424-1445. DOI: 10.1109/34.895976

[2] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski: Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999;974–989. DOI: http://dx.doi.org/10.1109/34.799905

[3] B. Fasel and J. Luettin: Automatic facial expression analysis: A survey. Pattern Recognition. 2003;259–275. DOI: http://dx.doi.org/10.1016/S0031-3203(02)00052-3

[4] Schmidt KL, Cohn JF. Dynamics of facial expression: Normative characteristics and individual differences. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '01); 22–25 August 2001; Tokyo. Japan: IEEE; 2001. pp. 547-550

[5] P. Ekman and W. V Friesen: Constants across cultures in the face and emotion. Journal of Personality and Social Psychology. 1971;124–129. DOI: http://dx.doi.org/10.1037/h0030377

[6] Ekman P, Friesen WV. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press; 1978

[7] Ekman P, Friesen WV. Facial Action Coding System. A Human Face: Salt Lake City; 1977

[8] Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J. Recognizing facial expression: Machine learning and application to spontaneous behavior. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition (CVPR '05); 20-25 July 2005; San Diego, CA, USA. IEEE. 2005:568-573

[9] Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J. Fully automatic facial action recognition in spontaneous behavior. In: Proceedings of IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR '06); 10–12 April 2006; Southampton. UK: IEEE; 2006. pp. 223-230

[10] Zeng Z, Pantic M, Roisman GI, Huang TS. A survey of affect recognition methods: Audio, visual. and spontaneous expressions: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009:39-58

[11] Zhan C, Li W, Ogunbona P, Safaei F. Facial expression recognition for multiplayer online games. In: Proceedings of ACM International Conference Proceeding Series (Australasian Conference on Interactive Entertainment); 04–06 December 2006; Perth. Australia: ACM; 2006. pp. 52-58

[12] Freire D, Santana MC, Déniz-Suárez O. Smile detection using local binary pattern and support vector machine. In: Proceedings of 4th International Conference on Computer Vision Theory and Applications; 5–8 February. Vol. 2009. Portugal: Lisboa; 2009. pp. 398-401

[13] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan: Toward practical smile detection: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009;2106–2111. DOI: http://dx.doi.org/10.1109/TPAMI.2009.42

[14] Shan C. Smile detection by boosting pixel differences. IEEE transactions on image processing. 2012:431-436

[15] Q. Ji, Z. Zhu, and P. Lan: Real-time nonintrusive monitoring and prediction of driver fatigue: IEEE Transaction on Vehicular Technology. 2004;1052–1068. DOI: http://dx.doi.org/10.1109/TVT.2004.830974

[16] B. Gholami, W. M. Haddad, and A. R. Tannenbaum: Relevance vector machine learning for neonate pain intensity assessment using digital imaging: IEEE Transactions on Biomedical Engineering. 2010;1457–1466. DOI: http://dx.doi.org/10.1109/TBME.2009.2039214

[17] Bartlett MS, Littlewort G, Fasel I, Movellan J. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In: Proceedings of Computer Vision and Pattern Recognition Workshop; 16–22 June. Vol. 2003. USA: Madison, Wisconsin; 2003. pp. 53-53

[18] K. Anderson and P. W. McOwan: A real-time automated system for the recognition of human facial expressions: IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics. 2006; 96–105. DOI: http://dx.doi.org/10.1109/TSMCB.2005.854502

[19] Azure Cognitive Services [Internet]. 2016. Available from: https://azure.microsoft.com/en-in/services/cognitive-services/face/ [Accessed: 2020-06-05]

[20] M. Pantic and L. J. M. Rothkrantz: Automatic analysis of facial expressions: The state of the art: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;1424–1445. DOI: http://dx.doi.org/10.1109/TSMCB.2005.854502

[21] Tian YTY. Evaluation of face resolution for expression analysis. In: Proceedings of Computer Vision and Pattern Recognition Workshop; 27 June-02 July. Vol. 2004. Washington, D.C.: USA; 2004. pp. 1-7

[22] C. Shan, S. Gong, and P. W. Mcowan: Facial expression recognition based on local binary patterns - a comprehensive study: Image and Vision Computing. 2009;803–816. DOI: http://dx.doi.org/10.1016/j.imavis.2008.08.005

[23] T. Jabid, M. H. Kabir, and O. Chae: Robust facial expression recognition based on local directional pattern: ETRI Journal. 2010;784–794. DOI: http://dx.doi.org/10.4218/etrij.10.1510.0132

[24] Valstar M, Jiang B, Mehu M, Pantic M, Scherer K. The first facial expression recognition and analysis challenge. In: Proceedings of IEEE International Conference on Face & Gesture Recognition; 27 June-02 July. Vol. 2004. Washington, D.C.: USA; 2004. pp. 921-926

[25] Valstar MF, Almaev T, Girard JM, McKeown G, Mehu M, Yin L, et al. FERA 2015 - second facial expression recognition and analysis challenge. In: Proceedings of 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; 04–08 May 2015; Ljubljana, Slovenia. 2015. pp. 1-8

[26] M. Song, D. Tao, Z. Liu, X. Li, and M. Zhou: Image ratio features for facial expression recognition application: IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 2010; 779–788. DOI: http://dx.doi.org/10.1109/TSMCB.2009.2029076

[27] Pantic M, Patras I. 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. 2016: 433-449

[28] H. Boughrara, M. Chtourou, C. Ben Amar, and L. Chen: Facial expression recognition based on a mlp neural network using constructive training

algorithm: Multimedia Tools and Applications. 2016;709–731. DOI: http://dx.doi.org/10.1007/s11042-014-2322-6

[29] Stankovir RS, Falkowski BJ. The Haar wavelet transform: Its status and achievements. Computers and Electrical Engineering. 2003:25-44

[30] Papageorgiou C, Poggio T. A trainable system for object detection. International Journal of Computer Vision. 2000:15-33

[31] T. Kanade, J. F. Cohn, and Y. Tian: Comprehensive database for facial expression analysis: Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition; 28–30 March 2000; Grenoble, France; 2000. p. 46–53. DOI: http://dx.doi.org/10.1109/AFGR.2000.840611

[32] M. J. Lyons: Automatic classification of single facial images: IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999;1357–1362. DOI: http://dx.doi.org/10.1109/34.817413

[33] Chen L-F, Yen Y-S. Taiwanese Facial Expression Image Database: Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University. Taiwan: Taipei; 2007