Robust Facial Expression Recognition using Gabor and LDP Feature Fusion using CCA

MAHESH M. GOYANI
Assistant Professor, Department of Computer Engineering,
Government Engineering College, Modasa
Gujarat, India.

Facial Expression Recognition has become vital considering its numerous applications including Human Computer Interaction, security, gaming, animation, medical field etc. In order to effectively implement these applications, the proposed method aims to increase the overall accuracy and robustness of the recognition system. Fusion of two feature extraction methods, namely Gabor and Local Directional Pattern (LDP) that are complementary in nature is carried out. Gabor Features focuses on the structural details whereas LDP targets the textural and subtler details in a robust manner. Fusion is carried out using Canonical Correlation Analysis (CCA) based fusion, as it effectively utilizes the correlation between the two features for fusion. For the efficient training of the classifiers, optimal feature vector is obtained by projecting the original feature vector on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) subspace. To prove the robustness of the proposed method, it is tested on benchmark datasets like CK, JAFFE, TFEID and CASIA-VIZ under ideal as well as different conditions such as noisy environment, low resolution, small sample space, different facial components etc. The system is also tested on two spontaneous expression datasets called SFEW (standard) and WESFED (in-house). The proposed method has shown better performance than the state of the art methods.

Keywords: Facial Expressions, Gabor, Local Directional Pattern, Feature Fusion, Canonical Correlation Analysis.

1. INTRODUCTION

Facial Expressions are one of the most well-known types of non-verbal communication (Mehrabian [1968]) A single expression can perhaps convey more information than thousands of words combined as it reflects the emotional state of a person which a human brain can efficiently recognize and interpret. Automatic Facial Expression Recognition (AFER) intends to transfer a certain, if not full level of that ability to computers. With the advancements and increasing use of Human Computer Interaction (HCI) in many applications, the need for effective AFER is all the more necessary.

Apart from socially sensitive Human- Computer Interaction (Cid et al. [2013]), AFER can be utilized in many applications such as detection of mental disorders (Gaebel and Wlwer [1992]), safety against road rage (Nasoz et al. [2004]), security (Butalia et al. [2012]), animations and video games (Bartlett et al. [2003]), automation applications, etc. It can also be used to include emotion related information in automatic image captioning systems (Cowie et al. [2001]).

Irrespective of the perspective followed, a basic AFER system comprises of mainly the following stages as shown in Figure 1 (Sariyanidi et al. [2015]). *Pre-processing* includes enhancing the image in order to facilitate the further process and reduce any such possible conditions that might affect the recognition performance. It also includes *Face detection* and *face tracking* which involves detecting the face region from the frame and tracking it across subsequent frames. *Feature extraction* deals with extracting the expression related features from the face followed by reducing them in order to filter out redundant features and decrease the complexity. *Classification* step includes feeding the features to a classifier which then narrows down to one of the expression classes. The detail survey on facial expression recognition can be found in Pantic and Rothkrantz [2000], Fasel and Luettin [2003], Goyani and Patel [2017a], Sheth and Goyani [2018].

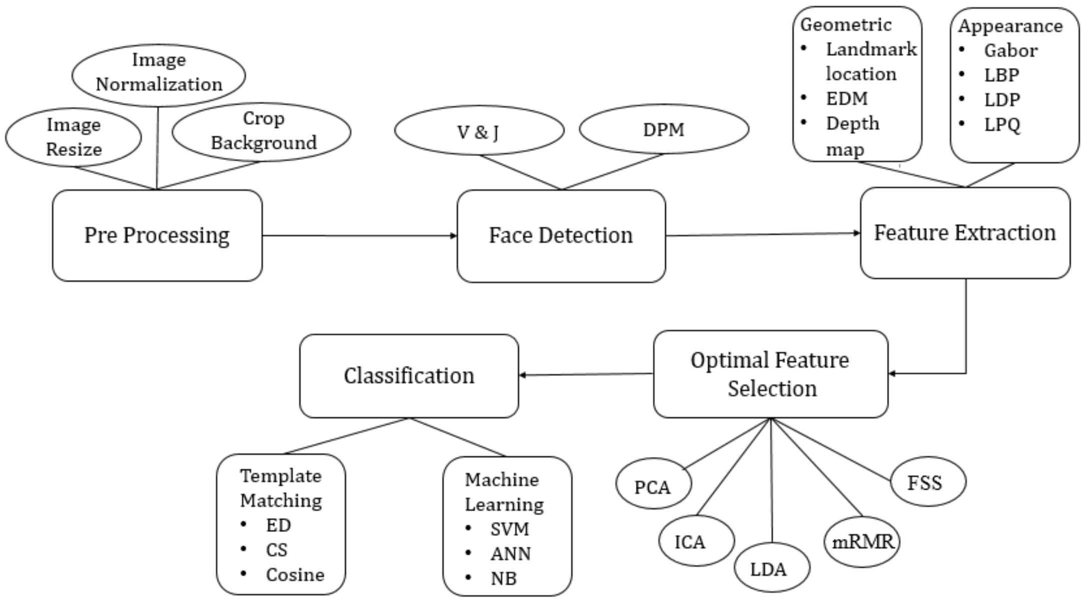

Figure 1. General Framework of Facial Expression Recognition System

The rest of the paper is structured as follows. Section 2 summarizes the prominent feature extraction techniques used by noteworthy approaches. Section 3 presents the proposed method and in depth discussion of the steps followed. Section 4 discusses the results and analysis of the experiments performed. Finally, section 5 concludes the overall performance of the system.

2. BACKGROUND AND RELATED WORK

There are two basic methodologies for feature extraction, namely geometric and appearance based methods. *Geometric methods* do not focus on face texture or intricate settings, rather they take feature indications from the geometry, deformation and tracking of fiducial points (Sariyanidi et al. [2015]). Geometric methods face many shortcomings such as dependency on face geometry and pose, failure to detect landmarks in case of occlusions, no tolerance against face representation errors, etc. Owing to the various pitfalls of geometric methods, appearance based feature extraction methods have gained much popularity in terms of higher accuracy and lower error conditions. *Appearance based* features take cue from the intensity levels of the face image to determine features contributing to facial expression. These features are extracted by generating a suitable filter which is the convolved around the face.

Ojala et al. [1996] formulated the Local Binary Patterns (LBP) which is an efficient and easy non-parametric method to describe mild and intricate texture information and summarize local structure of facial components. Among the pool of various appearance features, the ones with binarized local texture have shown promising results and effectiveness (Martinez et al. [2014]). Huang et al. [2011] provides a comprehensive study on various LBP techniques on facial image analysis along with their variations. LBP is computationally simple (Huang et al. [2011]) and is robust to monotonic illumination change and misalignment (Martinez et al. [2014]). Local Directional Pattern (LDP) is the technique proposed by Jabid et al. [2010], to address the nonrobustness problem of LBP towards random noise and non-monotonic illumination. It consists of directional information by comparing edge response values of each pixel in eight directions using Kirsch masks representing the impact of edge in each direction. The feature shows superior performance and can be represented in low-dimensional feature space with high accuracy for even low resolution images. However, the LDP codes can be problematic in smooth regions as they focus more on edge response values and produces inconsistent patterns in uniform regions (Ahmed and

Kabir [2010]). Goyani and Patel [2017c] proposed Local Mean Binary Pattern (LMBP) variant of LBP which thresholds the neighbor pixel with respect to mean of 3×3 patch. This approach is robust to noise and non monotonic gray level change compared to simple LBP.

Gabor filter is a linear parametric filter which analyses whether there are any specific frequency content in the image in specific directions in a localized region around the region of analysis (Donato et al. [1999]). Lyons and Akamatsu [1998] first used Gabor wavelets to generate facial expression features. It has proved to be an effective method and has since been used by many different approaches like in Lyons et al. [2000], Chen and Kotani [2007], Gu et al. [2012], Zhang et al. [2014], Xing and Luo [2016], for facial expression recognition.

Another line of approach seen in literature is to cascade two or more effective methods one after the other. In this, one method is applied over the response of the other. One such pattern, LGBP (Local Gabor Binary Pattern) was proposed by Zhang et al. [2005] in which Gabor features capture orientations and scales while LBP focuses subtler texture details. LGBP is robust to variations in lighting and expressions. Among others, LGBP was also used by Loob et al. [2017] where expressions were classified into dominant and complementary emotions like happily surprised to cover a wider range of emotions instead of basic prototypes. In a recent approach, Sun and Yu [2017] proposed a novel method in which instead of applying LBP over the features described by Gabor, both Gabor and LBP features are extracted from the input image and then those features are fused together using feature fusion to form the final feature vector. The approach has yielded superior performance since features described using different perspectives are used together. Similar method was proposed in Luo et al. [2013] where Global features extracted by PCA are fused with local features extracted from the mouth area using LBP. PCA is an effective method as it reduces dimensionality while selecting distinct features, but the features extracted by it are subject to environment changes. This is combatted by fusing in LBP which focuses on texture details and hence the results show improved robustness. Since LBP is sensitive to random noise and non-monotonic illumination, this method was improved by Luo et al. [2016] where instead of LBP, LDP is used to extract local features from the eyes and mouth region. The method shows improved performance than PCA and LBP combined, since LDP has good stability against random noise. Rajesh et al. experimented various fusion of different methods in Kumari et al. [2016] such as fusion of LBP with LGC, HOG with LDP, and HOG with wavelets. The results showed that fusion of HOG with wavelets outperformed all others and fusion of LDP with HOG improved LDPs performance. Goyani and Patel [2018] have proposed local statistical approach which combines Haar and LBP features. The proposed method has proven robust against illumination, noise and low resolution.

3. PROPOSED SYSTEM

We propose to increase the overall accuracy of the facial expression recognition system by combining Gabor and LDP appearance features using Canonical Correlation based feature level fusion. The combined feature vector proposes to provide a more robust and stable feature vector with more discriminatory information than previous state of the art methods.

3.1 Gabor Features

Local Gabor features are extracted from the original image by applying Gabor filters of varying scale and orientations. Gabor features are selected as they represent facial shape and appearance over multiple scales, locations and orientations. Since Gabor is a parametric filter, its parameters such as size, angles, wavelength etc. can be tuned to suit the corresponding scenarios. Usually 8 orientations and 5 scales are used to generate 40 filters. Another advantage of using Gabor filters is that it can simultaneously encode localized spatial and frequency information.

However, Gabor Features are not able to detect subtle and mild texture changes that occur during facial expressions, thus missing out vital discriminatory information. Gabor filters suffer from identity bias and are not able to efficiently discriminate between expressions related and face related characteristics. Stability of performance in low resolution images also remains an issue. Also, Gabor filters are computationally complex as 40 filters have to be applied at each neighborhood, which increases both dimensions and processing complexity. Since our method fuses two features, the dimensions as well as complexity is expected to increase, which is not a favorable condition. Therefore in our method, instead of filtering each neighborhood with all the filters, we take the average of the 40 filters and use the average filter obtained to generate feature vector as shown in Figure 2 below.

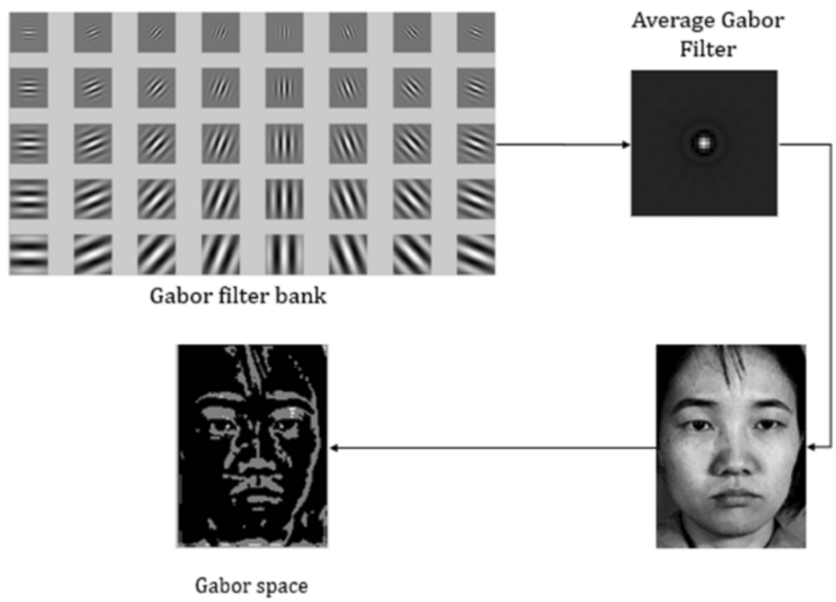

Figure 2. Averaged Gabor Filter Derivation

Hence, for one image only one corresponding image is generated as opposed to 40 images, thus largely reducing the image size and subsequent computation time. The obtained Gabor filtered image is down sampled by 4 times both length and breadth wise before generating the feature vector to further reduce the vector size. Therefore for a 150 110 image instead of a feature vector of size 6, 66, 000 we generate a feature vector of size 1064 thereby decreasing the size to a great extent. If only Gabor features are used, this practice might reduce the discriminatory information but since in the proposed approach another feature vector is fused, it not only compensates for the loss but generates an even more discriminatory feature vector.

3.2 LDP Features

LDP features are extracted by encoding 3 highest edge response values using Kirsch edge masks in 8 directions. Since they are based on gradient values instead of just gray scale pixel information, the resulting feature vector is more stable and robust than LBP features. Computationally, the process is more expensive than LBP features generation, but significantly lower than Gabor feature generation. The strength of LDP lies in the ability to detect subtle texture changes and local primitives such as curves, corners and junctions. They are efficiently robust to monotone/ non-monotone illumination changes and random noise. Also stability in low resolution and person independent conditions is higher as compared to Gabor features. However, LDP features are not able to encode bigger structural information as a result of operating over small neighborhoods. It also lacks contrast information that can be useful in providing more discriminative information. Another issue with LDP is that it tends to produce unstable patterns when extracting information from smooth regions as they focus on edge related information.

By fusing Gabor and LDP features, we intend to combine features that would address different challenges by combining both their advantages to form a more stable and discriminating feature vector. Many requirements of one method are fulfilled by the other, thus suggesting better recognition rates

3.3 CCA Fusion

We intend to use feature level fusion as its advantage lies in the fact that it can derive the most discriminatory information from multiple feature sets involved in fusion. It is able to gain the most effective and least dimensional feature vector sets. This practice makes the subsequent classification easier and improves the final recognition accuracy.

Canonical Correlation Analysis (CCA) fusion is based on finding the correlation between two sets of features. Correlation is a measure that indicates the degree to which two or more variables change together. CCA is a way of determining the linear relationship between two multidimensional variables corresponding to the same observation. It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations. In other words, CCA transforms the multidimensional variables into a subspace that are correlated to each other. The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables.

Consider two vectors $X = [x_1, x_2, ..., x_m]$ and $Y = [y_1, y_2, ..., y_n]$.Canonical correlation analysis can be defined as the problem of finding two sets of basis vectors, one for X and the other for Y, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Let us look at the case where only one pair of basis vectors are sought, namely the ones corresponding to the largest canonical correlation: Consider the linear combinations $x = X^T \hat{W}_x$ and $y = Y^T \hat{W}_y$. The function to be maximized is

$$\rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \tag{1}$$

where,

$$C_{xx} = \frac{\sum (X_i - \mu_x)(X_j - \mu_x)}{min(n-1)} \tag{2}$$

$$C_{yy} = \frac{\sum (Y_i - \mu_y)(Y_j - \mu_y)}{min(n-1)} \tag{3}$$

$$C_{xy} = \frac{\sum (X_i - \mu_x)(Y_i - \mu_y)}{min(n-1)} \tag{4}$$

The maximum of ρ with respect to w_x and w_y is the maximum canonical correlation. The projections onto w_x and w_y, i.e. x and y, are called canonical variants. In our methodology, we first transform the two feature vectors into the correlated subspace and then serially join the two vectors to form the CCA based fused feature vectors.

3.4 Proposed System

Block diagram of proposed method is depicted in Figure 3. The input image is first normalized both in terms of size and intensity. All the images are resized to 150 × 110 pixel resolution and histogram equalization is performed. This step is taken so that all images are uniform in size and covers the entire range of intensity levels. Since facial expressions occur only on the face region, the next logical step is to detect the face from the whole image frame. Haar Cascade face detection methodology is used for this purpose. In this method, rectangular Haar features are extracted that are then fed to a cascade of AdaBoost learning algorithm, which is then used to determine the face region. An extra portion is included from the top and bottom of face so that the entire forehead and chin areas are included and no vital facial expression related structure is missed out.

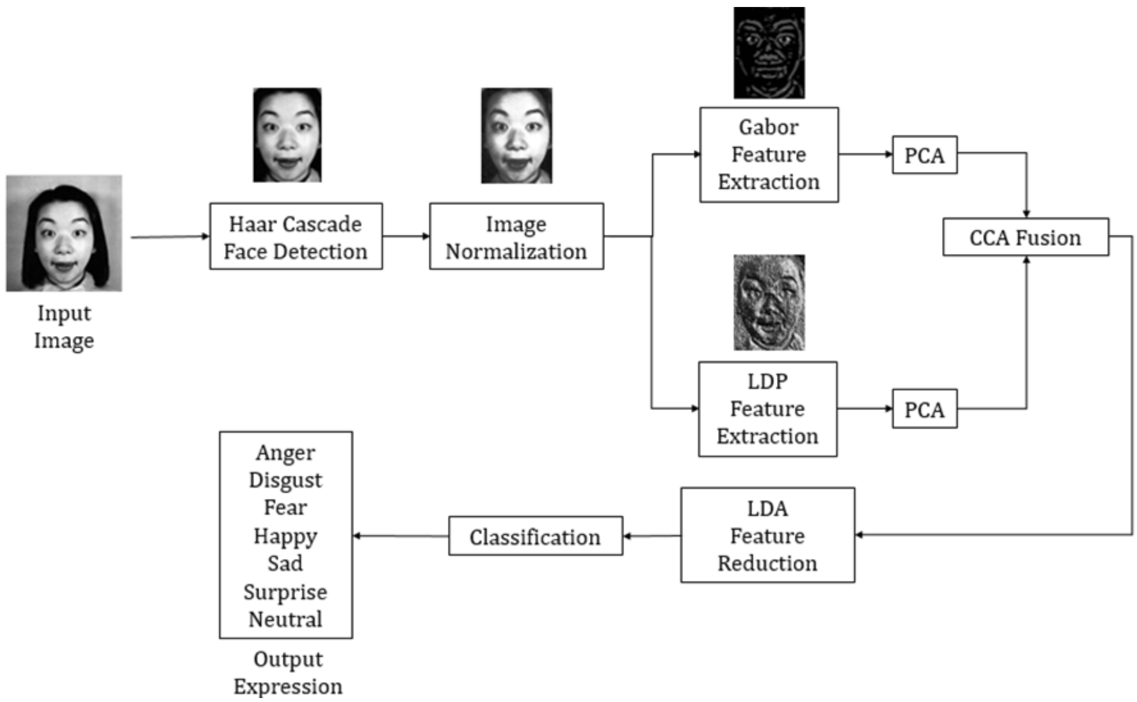

Figure 3. Block Diagram of Proposed System

In the next step, Gabor and LDP features are extracted individually from the original image. Gabor features are extracted by convolving the average Gabor filter across the image, and the results are stored in a feature vector of dimensionality 1064. LDP features are computed by operating on a local neighborhood and encoding edge response value with the help of Kirsch masks. To capture local spatial information, the generated LDP map is divided into 8 × 7 regions and the feature vector is generated by concatenating the histogram of each region. The region size is selected through experiments ranging from whole face to 9 × 9 regions, taking both accuracy and feature ex-traction time in consideration. The generated feature vector is of size 14336.

Both the generated feature vectors dimensions are reduced using PCA technique that captures the variance in the data and reduces the feature size. The number of Eigenvectors selected are 95 for Gabor features and 95 for LDP features which is determined through experiments. Hence the new dimension for both Gabor and LDP features is 95.

The reduced Gabor and LDP features are then fused together using CCA based feature fusion. Both the Gabor and LDP features are transformed to a correlated sub-space and then concatenated together. The dimension of each transformed feature vector is the minimum of the two feature size, i.e. in this case 95.

For further feature selection, LDA is applied for increasing the between class variance and final dimensionality reduction. The final feature vector of 6 dimensionality is then fed to the classifier. In the proposed methodology both Template Matching and Machine Learning classifiers are used and their performances are compared. In template matching, Correlation, Cosine, Chi-Square, Euclidean distance and k-Nearest Neighbors measures are used, of which Euclidean distance and Chi-Square are observed to give the overall best performance. In Machine Learning, Support Vector Machine (SVM), Least Square SVM (LS-SVM), Discriminant Analysis (DA), Random

Forest Classifier (RFC) and Logistic Regression classifiers are used for classification of vectors into one of the seven expression classes. LS-SVM and LR are selected as they display high performance.

4. RESULTS AND DISCUSSION

This section shows the experimental results and analysis on benchmark datasets, namely CK (Kanade et al. [2000]), JAFFE (Lyons [1999]), TFEID (Chen and Kotani [2007]), CASIA-VIS (Zhao et al. [2011]), SFEW (Dhall et al. [2011]) and in-house dataset called WESFED. We first select optimal parameters for the model by conducting experiments on the CK dataset. After the selection we display the results of tests on all the datasets under normal conditions. Thereafter the performance of the model is evaluated under varying conditions such as noisy environment, low resolutions, low sample space etc. by simulating the conditions on CK, JAFFE and TFEID datasets. Finally the section concludes with comparison of the model with the state of the art methods

4.1 Experimental Data

CK dataset is one of the standard datasets used by many researches to evaluate their performance. It consists of image sequences of six basic expressions from 97 subjects under various illumination conditions, age groups, head positions and ethnic diversities. In our experiments we have used 100 images from each expression class and 100 neutral images. We use equal number of images from each class so that the dataset is not biased towards a particular expression. JAFFE dataset is also one of the benchmark dataset used for testing facial expressions. It consists of 213 images from 10 Japanese fe-males of similar age group with frontal head poses and single illumination conditions. We have used all the 213 images for 7-class expression recognition. The TFEID dataset consists of a total of 267 images of 40 Taiwanese male and female subjects performing 7 expressions. The images are direct gaze, front view and of high intensity. We have used all 267 images for our experiments. Sample expressions of the datasets used in experiments are shown in Figure 4.

CASIA-VIS is a dataset consisting 480 sequences of facial expressions performed by 80 university students in classroom settings through camera lens of laptops in visible light. This dataset is useful for evaluating facial expression based feedback based smart tutoring systems. In our experiments, we use 3 peak frames from each sequence of expression, which leads to 240 images from each expression and 240 neutral face images.

Spontaneous Facial Expressions in the Wild (SFEW) is a comprehensive benchmark dataset consisting of real time expressions with all sorts of challenges such as occlusions, head pose variations, ethnic diversity, different illuminations etc. We use 729 training images and 366 images for validation. Web Enabled Spontaneous Facial Expression Dataset (WESFED) is spontaneous dataset consisting of real time expressions taken from the web, consisting of images from varying backgrounds, ethnicity, age, gender and face angles. It contains a total of 920 images from six basic expressions along with neutral images. We use all the images for our experiments.

Sample images from each dataset along with exact number of images used for each expression are shown in Figure 4.

4.2 Selection of Optimal Parameters

Selection of number of eigenvectors (NOEV) while applying PCA for feature reduction plays an important role as it determines the amount of information that would be encoded for further processing. We choose same number of eigenvectors for both Gabor and LDP features as CCA fusion strategy considers only minimum of the two feature dimensions. We start experimenting from the value 55 as below this the results are non-significant, and the value ranges up to 105, as the results start declining thereafter.

Table I and Table II shows the results of our experiments on CK dataset for five Template Matching and five Machine Learning Classifiers, respectively. The results are obtained by 10 Fold

Figure 4. Sample Dataset Images JAFFE, CK, TFEID, CASIA-VIS, SFEW, WESFED (Top to Bottom)

Table I: Effect of number of eigenvectors on CK dataset for template matching classifiers

#Eigenvector	L2-Norm	Correlation	Cosine	Chi-Square	K-NN	Average
55	**95.43**	87.00	95.71	**95.43**	95.43	93.80
60	**95.14**	95.57	96.14	**95.14**	95.57	96.51
65	**96.43**	92.29	96.57	**96.43**	95.86	95.51
70	**97.71**	92.43	97.71	**97.71**	98.57	96.83
75	**97.00**	95.43	97.71	**97.00**	96.86	96.80
80	**95.86**	97.14	96.29	**95.86**	96.57	96.34
85	**98.14**	96.71	97.86	**98.14**	98.14	98.49
90	**98.14**	96.43	98.00	**98.14**	97.71	97.69
95	**98.57**	**97.29**	**98.29**	**98.57**	**98.00**	**98.14**
100	**97.57**	95.86	96.43	**97.57**	98.14	97.11
105	**99.43**	91.71	99.14	**99.43**	99.43	97.83

cross validation and 90-10 validation, which means that 90 percent of the images in the dataset are used for training and the rest 10 percent is used for testing, and this process is repeated 10 times to achieve final average accuracy. Along with NOEV, we also observe the overall performance of various classifiers.

As it is observed from the Table I and Table II, NOEV 95 provides the best results for all clas-

Table II: Effect of number of eigenvectors on CK dataset for Machine Learning based classifiers

#Eigenvector	L2-Norm	Correlation	Cosine	Chi-Square	K-NN	Average
55	89.86	**96.57**	96.00	95.00	**96.43**	94.77
60	88.57	**95.71**	95.71	94.29	**94.86**	93.83
65	94.00	**98.00**	97.00	96.43	**97.14**	96.51
70	95.29	**98.29**	98.29	97.71	**98.57**	97.63
75	95.29	**97.14**	97.71	96.14	**97.86**	96.83
80	95.29	**97.29**	98.00	97.71	**97.86**	97.23
85	96.43	**99.14**	99.29	98.29	**99.29**	98.49
90	97.29	**98.71**	98.14	98.14	**98.43**	98.14
95	**97.29**	**99.14**	**98.71**	**98.14**	**99.29**	**98.51**
100	97.86	**98.29**	98.57	98.14	**98.57**	98.29
105	96.43	**99.14**	99.29	98.29	**99.29**	98.49

sifiers and hence we have selected NOEV as 95 for both Features. Among the Template Matching classifiers, L2-norm and Chi-Square gives the best results, while among Machine Learning classifiers, LS-SVM and Logistic Regression gives the best overall performance. Also, the observation of Template Matching and Machine Learning classifiers clearly suggest that Machine Learning Classifiers gives better results.

Histogram of LDP map is treated as LDP feature vector. Thus it does not include any spatial information. Therefore, we divide the LDP map into more than one regions and generate the histogram for each region to encode spatial information from the LDP map generated. This practice ensures that local structural information such as minor deformations, wrinkles, furs, etc. are preserved spatially. We have conducted experiments for regions ranging from 1 × 1 (whole face) to 9 × 9 parts as shown in Figure 5.

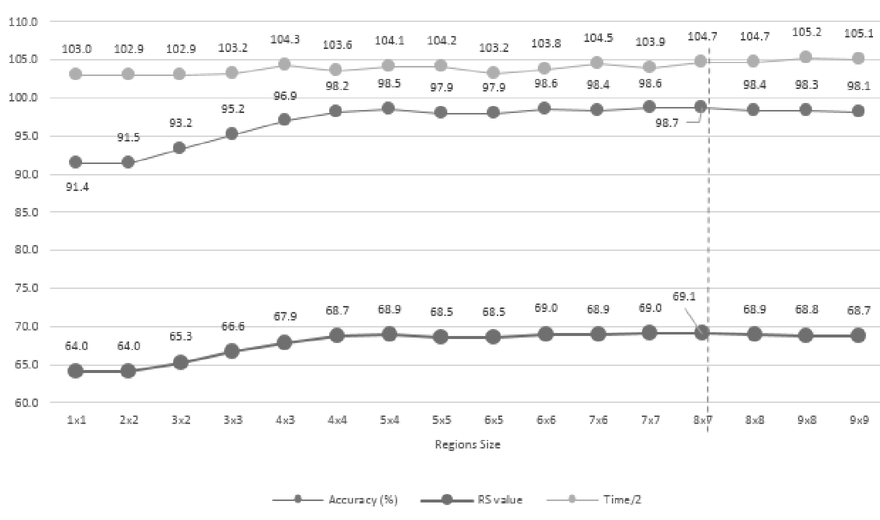

Figure 5. Selection of optimal number of regions

Region sizes beyond 9 × 9 are not used as each region would then encode lesser amount of information and increase the feature size unnecessarily. The results obtained are from the average of all the 10 classifiers. Along with the average accuracy, the feature extraction time is also taken into consideration for selection of optimal number of regions. The selection of final region size is performed by obtaining the RS values corresponding to the following equation, and selecting the region size with the maximum value.

$$RSValue_i = (\alpha * Acc_i) + \frac{1-\alpha}{T_i} \tag{5}$$

where, $\alpha = 0.7$, Acc is average accuracy and T is feature extraction time. The parameter α controls the trade-off between feature extraction time and accuracy.

Using this methodology, we obtain the best Accuracy/Time trade-off by giving 70 per cent weightage to accuracy and 30 per cent to feature extraction time. Is clearly observed from the Figure 5 that region size of 8 × 7 gives the best RS value among all the other region sizes, and hence 8 × 7 is selected as the optimal number of regions.

4.3 Performance under Ideal Conditions

In this section we evaluate the performance of our method on six datasets namely CK, JAFFE, TFEID, CASIA-VIS, SFEW and WESFED under normal conditions. The accuracy of the system is evaluated by 10 Fold Cross Validation Scheme. Table III shows the accuracy obtained by the four selected classifiers.

Table III: Accuracy (%) under Ideal Conditions for selected classifiers

	Datasets					
Classifier	CK	JAFFE	TFEID	CASIA-VIS	SFEW	WESFED
L2-norm	98.43	100.0	100.0	91.19	54.37	92.07
Chi-Square	98.57	100.0	100.0	91.73	56.83	92.28
LS-SVM	99.14	100.0	100.0	91.96	72.13	93.70
LR	98.86	100.0	100.0	87.44	74.32	93.26

It is observed that LS-SVM classifier gives the highest accuracy compared to other classifiers. The confusion matrix for all the datasets for LS-SVM classifier is shown in Table IV.

Table IV: Confusion Matrices for various dataset using LS-SVM Classifier

	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**
AN	**100**	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0	0.0	0.0
DI	1.96	**98.04**	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0	0.0
FE	0.0	0.0	**98.90**	0.0	0.0	0.0	1.10	0.0	0.0	**100**	0.0	0.0	0.0	0.0
HA	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0
SA	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0
SU	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0
NE	0.0	0.0	0.0	0.0	2.86	0.0	**97.14**	0.0	0.0	0.0	0.0	0.0	0.0	**100**
	Average Accuracy (CK) - 99.14 %							**Average Accuracy (JAFFE) 100.0 %**						
	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**
AN	**100**	0.0	0.0	0.0	0.0	0.0	0.0	**87.34**	6.33	0.0	0.0	5.91	0.0	0.42
DI	0.0	**100**	0.0	0.0	0.0	0.0	0.0	7.86	**91.70**	0.0	0.0	0.44	0.0	0.0
FE	0.0	0.0	**100**	0.0	0.0	0.0	0.0	1.41	0.0	**92.02**	0.94	4.69	0.94	0.0
HA	0.0	0.0	0.0	**100**	0.0	0.0	0.0	1.65	0.0	1.24	**96.28**	0.83	0.0	0.0
SA	0.0	0.0	0.0	0.0	**100**	0.0	0.0	11.92	0.77	2.69	0.38	**83.85**	0.0	0.38
SU	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.42	0.0	2.92	0.42	1.67	**94.58**	0.0
NE	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.77	0.0	0.0	0.0	1.16	0.0	**98.07**
	Average Accuracy (TFEID) 100.0 %							**Average Accuracy (CASIA-VIS) 91.96 %**						
	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**	**NE**
AN	**77.61**	1.49	1.49	2.99	4.48	1.49	10.45	**97.90**	1.40	0.0	0.0	0.70	0.0	0.0
DI	15.0	**35.00**	5.0	5.0	15.0	5.0	20.0	1.75	**85.96**	3.51	3.51	5.26	0.0	0.0
FE	12.82	0.0	**53.85**	2.56	7.69	15.38	7.69	1.64	0.0	**88.52**	0.0	4.92	4.92	0.0
HA	4.84	0.0	1.61	**87.10**	4.84	0.0	1.61	0.50	1.99	0.0	**95.02**	1.99	0.50	0.0
SA	7.14	7.14	0.0	3.57	**66.07**	0.0	16.07	0.82	0.82	0.0	0.82	**88.52**	0.0	9.02
SU	15.0	2.50	2.50	7.50	7.50	**47.50**	17.50	0.0	0.0	2.60	0.65	0.0	**96.75**	0.0
NE	1.22	1.22	0.0	1.22	6.10	0.0	**90.24**	0.0	0.55	0.0	0.55	4.95	0.0	**93.96**
	Average Accuracy (SFEW) 72.13 %							**Average Accuracy (WESFED) 93.70 %**						

4.4 Performance under Low Resolution Conditions

The previous section evaluated the performance of the system on six datasets under ideal and unchanged conditions. Datasets such as SFEW and WESFED contain real time images with various experimental conditions embedded in them. On the contrary CK, JAFFE, TFEID and CASIA datasets do not contain much variations. Hence to test the model under various conditions, we simulate such conditions on the three datasets and evaluate them. Validation method of 10-90 is used for our experiments to analyze and study maximum variations under lowest sample space.

It is not always possible to get high resolution images for recognition. In many cases such as surveillance, smart tutoring systems etc. we have to perform recognition under low resolution conditions. Therefore we evaluate the performance of our method under low resolution conditions by simulating different resolutions on CK, JAFFE and TFEID datasets and study the stability of our model. We choose three different resolutions: 150 × 110, 75 × 55, 48 × 36 and 37 × 27 pixels as shown in Figure 6.

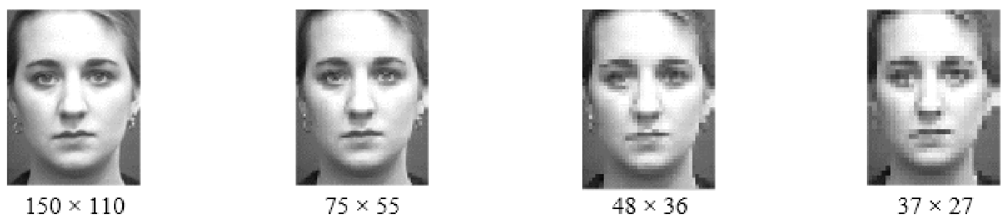

Figure 6. Test image samples for different resolution

Performance of proposed method under low resolution on all three datasets is depicted in Figure 7.

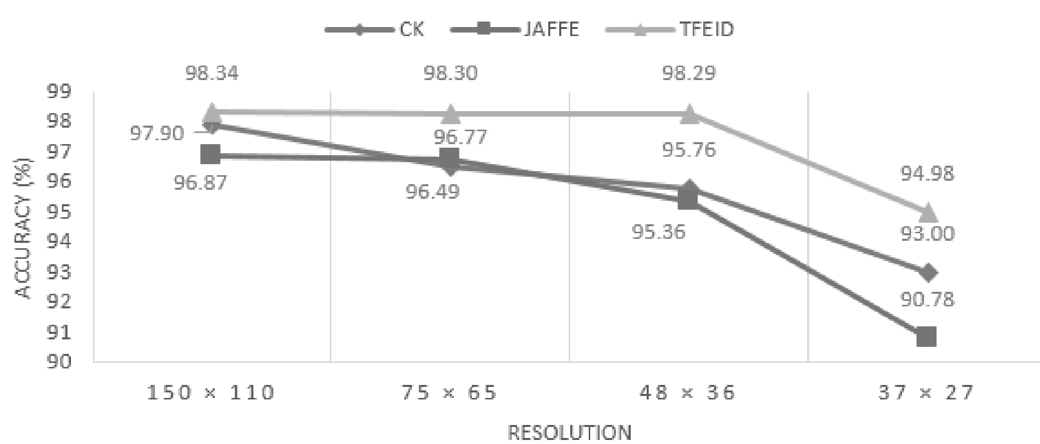

Figure 7. Performance on different dataset in low resolution

It can be seen from the above plot that performance until 48 × 36 is somewhat similar and stable but takes a dip for 37 × 27. This is because as the resolution decreases, the amount of information encoded also decreases and discriminatory information is lost. However, the results are still acceptable as the outcomes do not fall below 90 per cent for any dataset.

Table V shows a performance comparison of our model with other state of the art methods for low resolution conditions on CK dataset, using LS-SVM classifier.

Table V: Comparison with State of the Art methods in Low Resolution

		Resolution			
Feature	VM	150 × 110	75 × 55	48 × 36	37 × 27
Gabor (Shan and Gong [2009])	90-10	89.1 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
Gabor (Xing and Luo [2016])	70-30	89.8 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
LBP (Zhang et al. [2014])	90-10	92.6 ± 2.9	89.9 ± 3.1	87.3 ± 3.4	84.3 ± 4.1
LDP (Jabid et al. [2010])	90-10	96.4 ± 0.9	95.5 ± 1.6	93.1 ± 2.2	90.6 ± 2.7
Proposed	90-10	99.1 ± 1.4	98.9 ± 1.1	98.6 ± 1.7	96.7 ± 2.2
Proposed	10-90	97.9 ± 1.9	96.5 ± 1.5	95.8 ± 2.0	93.0 ± 1.9

It is seen that our method outperforms the highest recorded performance with a standard deviation as low as 2.2. Even in 10-90 validation method the proposed model performs better than the state of the art methods.

4.5 Performance under Noisy Environment

Images are often corrupted by noise due to numerous conditions such as weather conditions, temperature changes, and noise during transmissions or corrupt sensor etc. Therefore it is important to test the effect of noise on the proposed method. We test our method on three different types of noises, namely Gaussian noise, Salt and Pepper and Speckle noise, and for different parametric values (μ - Mean, v - Variance, d - Density). The corresponding noisy images are as shown in Figure 8.

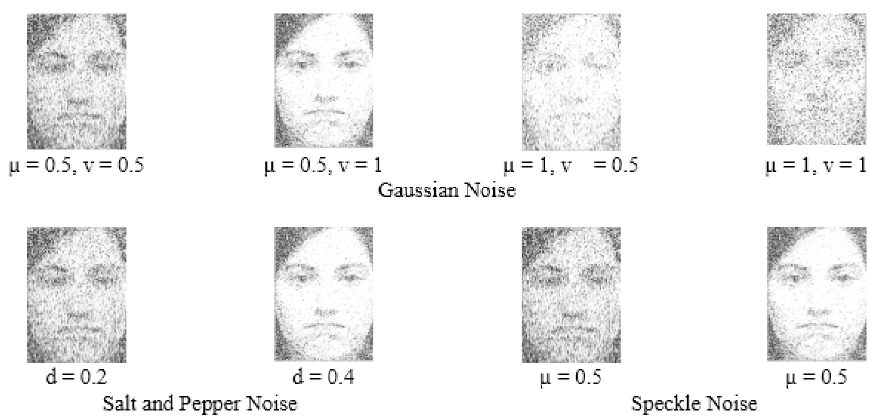

Figure 8. Images corrupted with various noise

We have used 10-90 validation method and the 4 selected classifiers to evaluate the performance on CK, JAFFE and TFEID datasets, the results of which are shown in Tables VI, Table VII and Table VIII respectively.

It is obvious that performance under noisy conditions will decrease since the pixel values are corrupted and expression related information might get lost or get confused with noisy pixels. It is observed that under noisy conditions Machine Leaning classifiers give much better performance.

Table VI: Accuracy (%) Under Noisy Conditions (CK dataset)

Noise	Parameters	ED	Chi-Square	LS-SVM	LR
Gaussian	$\mu = 0.5$ v = 0.5	79.14	80.89	86.17	87.37
	$\mu = 1.0$ v = 1.0	76.80	77.23	82.69	84.37
Salt & Pepper	d = 0.2	94.43	94.00	95.91	95.89
	d = 0.5	80.97	80.77	87.14	86.49
Speckle	v = 0.5	85.26	85.63	89.97	89.97
	v = 1.0	84.74	84.74	89.00	88.69

Table VII: Accuracy (%) Under Noisy Conditions (JAFFE dataset)

Noise	Parameters	ED	Chi-Square	LS-SVM	LR
Gaussian	$\mu = 0.5$ v = 0.5	93.65	93.65	93.48	93.65
	$\mu = 1.0$ v = 1.0	96.72	96.72	96.72	96.72
Salt & Pepper	d = 0.2	97.96	97.96	98.65	98.65
	d = 0.5	93.64	93.64	93.64	93.64
Speckle	v = 0.5	98.43	98.43	98.38	98.43
	v = 1.0	96.88	96.88	96.88	96.88

Table VIII: Accuracy (%) Under Noisy Conditions (TFEID dataset)

Noise	Parameters	ED	Chi-Square	LS-SVM	LR
Gaussian	$\mu = 0.5$ v = 0.5	97.72	96.64	95.15	97.80
	$\mu = 1.0$ v = 1.0	97.97	96.39	95.77	97.84
Salt & Pepper	d = 0.2	98.34	98.34	98.33	98.34
	d = 0.5	98.26	97.50	96.76	98.23
Speckle	v = 0.5	98.34	98.34	97.68	98.34
	v = 1.0	98.34	98.29	97.30	98.29

4.6 Performance under Small Training Sample Space

It is not always possible that we get enough training samples and the model might have to predict for more number of testing data. Hence we have used various validation methods to test the performance for different ratios of training and testing data.

Figure 9 shows the results on all the datasets for various validation methods, using LS-SVM classifier. It is apparent that accuracy decreases as the number of training data decreases. But the performance decrease is very low and the model gives higher accuracy for even low number of training samples.

Performance decrease of CK is greater than JAFFE and TFEID, since the latter two datasets have very low number of images relative to the former. Also, CK dataset is more complex in nature, owing to its diverse illumination, gender, age and ethnicity variations, making it difficult to achieve higher accuracy for small sample space, when compared to other datasets

4.7 Performance on Different Facial Components

We analyze the performance of our model by testing on different facial components. The experiment is performed on CK, JAFFE and TFEID datasets by 10-90 validation method and using LS-SVM classifier. Upper face includes eyes and lower face consists of mouth region. Figure 10 shows the results of the experiments performed.

The upper face gives the lowest accuracy, since the expression is conveyed only by the eyes and forehead area. Expressions such as happy, sad and surprise fail to get recognized as vital

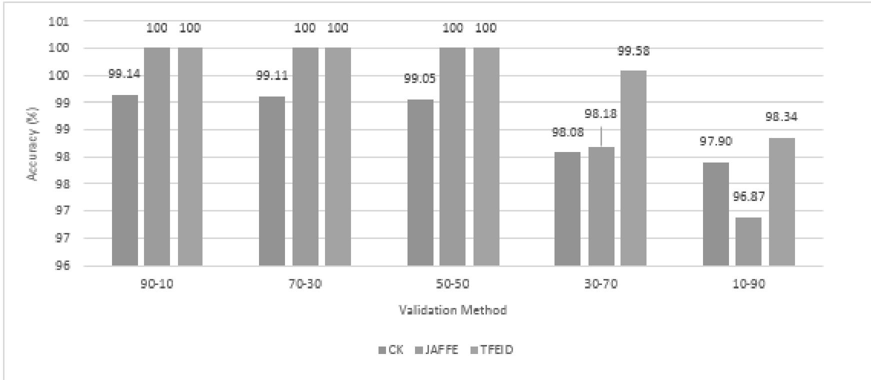

FIgure 9. Accuracy (%) of different validation methods

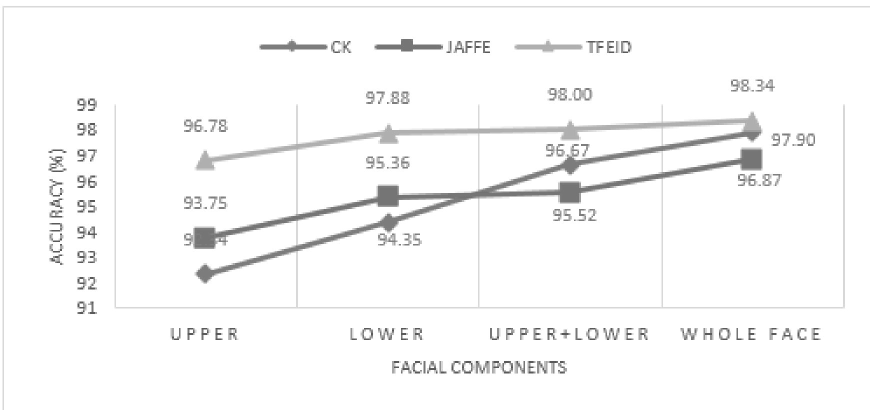

Figure 10. Accuracy (%) on different facial components

mouth region for information encoding is not present. The lower face gives better results, but misses out on expressions like anger or fear since eye brow in-formation is lost. The combination of upper and lower surpasses the previous two, as both eyes and lips are preserved. However, the results of combined are lower than that of whole face, indicating that the structure of whole face is necessary for expression recognition, rather than just eyes and lips, especially in case of expressions such as disgust, where nose region conveys vital information.

4.8 Performance under the Effect of Occlusions

Real time applications might contain images with occluded regions such as eye patches, sun glasses, scarfs, masks, or any such props that block a part of the face that restricts expression related information. There is no benchmark dataset available that explicitly exhibits different types of occlusions. Hence to test the robustness of the pro-posed model against occlusions, we simulate such conditions manually by adding white masks of different sizes at different positions. We simulate five different types of occlusions, (i) Eye Patches, (ii) Mouth Patch, (iii) R_8 Patch, (iv) R_{16} Patch, and (v) R_{24} Patch, where Rs patch means randomly placed $S \times S$ patch on the face region. A sample set of such occlusions is shown in Figure 11.

We distribute the five types of occlusions equally among each dataset and take the results on 10-90 validation method using LS-SVM classifier. The results obtained for each dataset are tabulated in Table IX.

From the Table IX, it can be deduced that the model is sufficiently robust against occlusions,

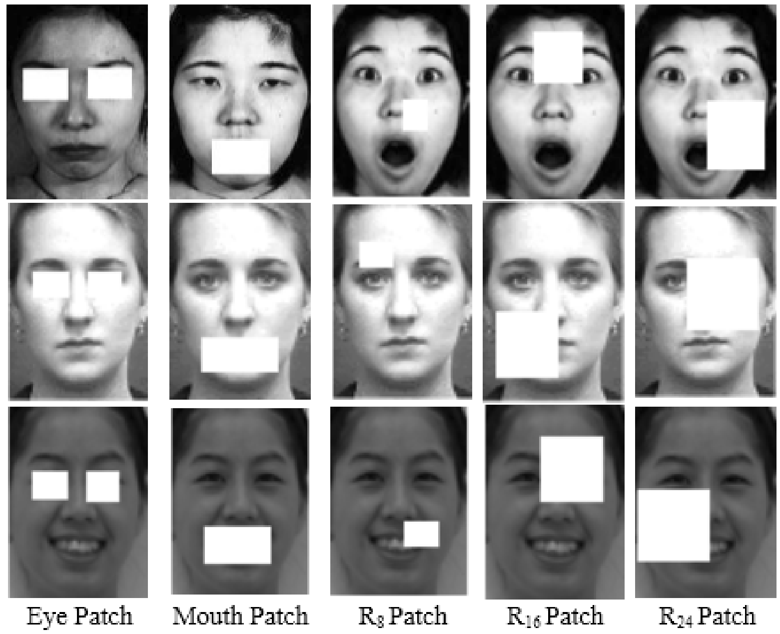

Figure 11. Sample Occluded Images

Table IX: Accuracy (%) under Occlusion Conditions

Accuracy (%)	CK	JAFFE	TFEID
Ideal Conditions	97.90	96.87	98.34
Faces with occlusion	95.37	95.05	93.53

since the performance drop is not more than 5 per cent for any dataset.

4.9 Feature Extraction Time

The feature extraction time of the proposed method is calculated for each sub step of the extraction process, using an image size 150 × 110 and is tabulated in Table 10 below.

Table X: Comparative analysis of proposed model with state-of-the-art methods

Feature	Gabor	LDP	PCA	CCA Fusion	LDA	**Total**
Time (Sec)	0.0067	0.2859	0.0037	1.44×10^{-4}	1.8×10^{-4}	**0.2967**

Since we use average Gabor filter instead of 40 Gabor filters, the Gabor feature ex-traction time is substantially reduced. LDP feature extraction takes maximum time as it involves generating local directional pattern of 8 neighbors for each neighborhood. The time taken for Fusion and Dimensionality reduction is relatively much lower than the generation process.

4.10 Comparison with State of the Art Methods

We now compare our results with the state of the art methods present in literature. Comparisons on CK, JAFFE, CASIA and SFEW datasets are shown in Table XI, Table XII, Table XIII and Table XIV, respectively. Results for TFEID and WESFED dataset for any state of the art methods are not available for comparison.

Table XI: Performance Comparison with State of the Art Methods (CK database)

Literature	Features	VM	Accuracy (%)
Shan et al. (Shan and Gong [2009])	LBP	90-10	88.9
Zhang et al. (Zhang et al. [2014])	Gabor	90-10	90.8
Jabid et al. (Jabid et al. [2010])	LDP	90-10	93.4
Bashar et al. (Bashar et al. [2013])	MTP	90-10	94.2
Xing et al. (Xing and Luo [2016])	Gabor	70-30	95.1
Proposed	Gabor + LDP + CCA	90-10	99.1
Proposed	Gabor + LDP +CCA	10-90	97.9

Table XII: Performance Comparison with State of the Art Methods (JAFFE database)

Literature	Features	VM	Accuracy (%)
Zhang et al. (Zhang et al. [2014])	Gabor	90-10	78.4
Shan et al. (Shan and Gong [2009])	LBP	90-10	80.7
Jabid et al (Jabid et al. [2010])	LDP	90-10	84.9
Ryu et al. (Sariyanidi et al. [2017])	LDTP	90-10	94.3
Holder et al. (Holder and Tapamo [2017])	GLTP	90-10	95.2
Siddiki et al (Siddiqi et al. [2015])	SWLDA	90-10	96.4
Yu et al. (Yu and Gu [2017])	ASM	70-30	98.3
Proposed	Gabor + LDP + CCA	90-10	100.0
Proposed	Gabor + LDP + CCA	10-90	96.9

Table XIII: Performance Comparison with State of the Art Methods (CASIA-VIZ database)

Literature	Features	VM	Accuracy (%)
Klaser et al. (Klaser and Marszalek [2008])	HOG 3D	90-10	70.6
Zhao et al. (Zhao et al. [2011])	AdaLBP	90-10	73.5
Guo et al. (Guo et al. [2012])	Atlases	70-30	75.5
Liu et al. (Liu et al. [2014])	STM-ExpLet	90-10	74.6
Proposed	Gabor + LDP + CCA	90-10	92.0
Proposed	Gabor + LDP + CCA	10-90	85.8

Table XIV: Performance Comparison with State of the Art Methods (SFEW database)

Literature	Features	Accuracy (%)
Eleftheriadis et al. (Eleftheriadis et al. [2015])	DSGP	24.7
Goyani et al. (Goyani and Patel [2017b])	LHMBP	52.0
Meng et al. (Meng et al. [2017])	CNN	54.3
Mollahosseini et al. (Mollahosseini et al. [2016])	Deeply Learnt Features	62.1
Proposed	Gabor + LDP + CCA	72.1

From the above comparisons, it is clear that the proposed method gives us better performance than state of the art methods on all datasets in both large and small sample space conditions, and is robust in nature.

5. CONCLUSION AND FUTURE WORK

Various methods for facial expression recognition are available that address different issues. The study of these techniques reveal that in order to overcome various is-sues that hinder the performance of recognition system, expression related features that focus on different aspects of representation need to be fused together to fulfill the short-comings of one method by the other. We proposed the fusion of Gabor and LDP features since Gabor focuses on bigger structural details while LDP focuses on subtle edge based texture information and together they provide a robust, stable and discriminative set of expression features. The features are fused based on canonical correlation analysis, which has shown better performance as compared to other fusion techniques.

The proposed model yields superior performance on benchmark datasets when com-pared to state of the art methods. A detailed study of various confusion matrices reveals that the expressions happy and surprise give highest recognition rate, whereas expressions anger and disgust give the lowest recognition rate. Anger is confused mostly with neutral and disgust, whereas disgust gets confused with fear or sadness. The model shows reasonable robustness in conditions such as noise and occlusions, and also outperforms the state of the art methods in low resolution conditions. Superior results are obtained even in small sample space conditions. Experiments conducted on different facial components revealed that the structure of the whole face is important for effective recognition, rather than just eyes or lips. Also, it is observed that Machine Learning Classifiers give overall better results than Template Matching Classifiers.

In future, we plan to extend the feature extraction phase to GPU, so that the system can extract the features concurrently from multiple images. This would definitely boosts the execution time of the system. Other opportunity is to project the features using locality preserving projection, which might outperform the PCA based projection, which simply tries to maximize the variance in the projection space.

References

Ahmed, F. and Kabir, M. 2010. Directional ternary pattern (dtp) for facial expression recognition. In *IEEE International Conference on Consumer Electronics*. IEEE, 265–266.

Bartlett, M., Littlewort, G., Fasel, I., and Movellan, J. 2003. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*. Vol. 5. 53–53.

Bashar, F., Khan, A., Ahmed, F., and Kabir, M. 2013. Robust facial expression recognition based on median ternary pattern (mtp). In *International Conference on Electrical Information and Communication Technology*. 1–5.

Butalia, A., Ingle, M., and Kulkarni, P. 2012. Facial expression recognition for security. *International Journal of Modern Engineering Research 2,* 4, 1449–1453.

Chen, F. and Kotani, K. 2007. Facial expression recognition by svm-based two-stage classifier on gabor features. In *IAPR Conference on Machine Vision Applications*. 453–456.

Cid, F., Prado, J., Bustos, P., and Nunez, P. 2013. A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering. In *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2188–2193.

Cowie, R., Douglas-Cowie, E., and Tsapatsoulis, N. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine 18,* 1, 32–80.

Dhall, A., Goecke, R., Lucey, S., and Gedeon, T. 2011. Static facial expressions in tough conditions: Data, evaluation protocol and benchmark. In *IEEE International Conference on Computer Vision, Barcelona, Spain.*

Donato, G., Bartlett, M., Hager, J., Ekman, P., and Sejnowski, T. 1999. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21,* 10, 974–989.

Eleftheriadis, S., Rudovic, O., and Pantic, M. 2015. Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition. *IEEE Transactions on Image Processing 24,* 1, 189–204.

Fasel, B. and Luettin, J. 2003. Automatic facial expression analysis: A survey. *Pattern Recognition 36,* 1, 259–275.

Gaebel, W. and Wlwer, W. 1992. Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience 242,* 1, 46–52.

Goyani, M. and Patel, N. 2017a. Judgmental feature based facial expression recognition systems and fer datasets-a comprehensive study. *International Journal of Next-Generation Computing 8,* 1, 62–81.

Goyani, M. and Patel, N. 2017b. Robust facial expression recognition using local mean binary pattern. *Electronic Letters on Computer Vision and Image Analysis 16,* 1, 54–67.

Goyani, M. and Patel, N. 2017c. Template matching and machine learning-based robust facial expression recognition system using multi-level haar wavelet. *International Journal of Computers and Applications*, 1–12.

Goyani, M. and Patel, N. 2018. Robust facial expression recognition using local haar mean binary pattern. *Journal of Information Science and Engineering 34,* 5, 1175–1186.

Gu, W., Xiang, C., Venkatesh, Y., Huang, D., and Lin, H. 2012. Facial expression recognition using radial encoding of local gabor features and classifier synthesis. *Pattern Recognition 45,* 1, 80–91.

Guo, Y., Zhao, G., and Pietikainen, M. 2012. Dynamic facial expression recognition using longitudinal facial expression atlases. In *European Conference on Computer Vision*. Vol. 7573.

Holder, R. and Tapamo, J. 2017. Improved gradient local ternary patterns for facial expression recognition. *EURASIP Journal on Image and Video Processing 2017,* 42, 1965–1978.

Huang, D., Shan, C., Ardabilian, M., Wang, Y., 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 41,* 6, 765–781.

Huang, D., Shan, C., Ardebilian, M., and Chen, L. 2011. Facial image analysis based on local binary patterns: A survey. *IEEE Transactions on Image Processing 41,* 6, 765–781.

Jabid, T., Kabir, M., and Chae, O. 2010. Facial expression recognition using local directional pattern (ldp). In *17th IEEE International Conference on Image Processing*. IEEE, 1605–1608.

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

Klaser, A. and Marszalek, M. 2008. A spatio-temporal descriptor based on 3d-gradients. In *19th British Machine Vision Conference*. 275–285.

Kumari, J., Rajesh, R., and Kumar, A. 2016. Fusion of features for the effective facial expression recognition. In *International Conference on Communication and Signal Processing*. 457–461.

Liu, M., Shan, S., Wang, R., and Chen, X. 2014. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Loob, C., Rasti, P., Lsi, I., Julio, C., Junior, J., Bar, X., Escalera, S., Sapinski, T., Kaminska, D., and Anbarjafari, G. 2017. Dominant and complementary multi-emotional facial expression recognition using c-support vector classification. In *IEEE 12th International Conference on Automatic Face and Gesture Recogition*. IEEE, 833–838.

Luo, Y., Wu, C., and Zhang, Y. 2013. Facial expression recognition based on fusion feature of pca and lbp with svm. *International Journal for Light and Electron Optics 124,* 17, 2767–2770.

Luo, Y., Zhang, T., and Zhang, Y. 2016. A novel fusion method of pca and ldp for facial expression feature extraction. *International Journal for Light and Electron Optics 127,* 2, 718–721.

Lyons, M. 1999. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21,* 12, 1357–1362.

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

Lyons, M., Budynek, J., Plante, A., and Akamatsu, S. 2000. Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis. In *4th International Conference on Automatic Face and Gesture Recognition*. 202–207.

Martinez, B., Valstar, M., Jiang, B., and Pantic, M. 2014. Automatic analysis of facial actions: A survey. *IEEE Transactions on Affective Computing 13,* 9, 1–22.

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

Meng, Z., Liu, P., Cai, J., Han, S., and Tong, Y. 2017. Identity-aware convolutional neural network for facial expression recognition. In *12th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 558–565.

Mollahosseini, A., Chan, D., and Mahoor, M. H. 2016. Going deeper in facial expression recognition using deep neural networks. In *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 1–10.

Nasoz, F., Alvarez, K., Lisetti, C., and Finkelstein, N. 2004. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology and Work 6,* 1, 4–14.

Ojala, T., Pietikainen, M., and Harwood, D. 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition 29,* 1, 51–59.

Pantic, M. and Rothkrantz, L. 2000. Automatic analysis of facial expressions: the state of the art. *IEEE Transaction on Pattern Analysis and Machine Intelligence 22,* 12, 1424–1445.

Sariyanidi, E., Gunes, H., and Cavallaro, A. 2015. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37,* 6, 1–22.

Sariyanidi, E., Gunes, H., and Cavallaro, A. 2017. Learning bases of activity for facial expression recognition. *IEEE Transactions on Image Processing 26,* 4, 1965–1978.

Shan, C. and Gong, S. 2009. Facial expression recognition based on local binary patterns: a comprehensive study. *Image and Vision Computing 27,* 6, 803–816.

Sheth, N. and Goyani, M. 2018. A comprehensive study of geometric and appearance based facial expression recognition methods. *International Journal of Scientific Research in Science, Engineering and Technology 4,* 2, 163–175.

Siddiqi, M., Ali, R., Khan, A., Park, Y., and Lee, S. 2015. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Transactions on Image Processing 24,* 4, 1386–1398.

Sun, Y. and Yu, J. 2017. Facial expression recognition by fusing gabor and local binary pattern features. In *International Conference on Multimedia modeling*. Springer, 209–220.

Xing, Y. and Luo, W. 2016. Facial expression recognition using local gabor features and adaboost classifiers. In *IEEE Internation al Conference on Progress in Informatics and Computing*. IEEE, 1–5.

Yu, T. and Gu, X. 2017. Facial expression recognition using double-stage sample-selected svm. In *International Conference on Intelligent Computing*. Springer.

Zhang, L., Tjondronegoro, D., and Chandran, V. 2014. Random gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing 145*, 451–464.

Zhang, W., Shan, S., Gao, W., and Chen, X. 2005. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *IEEE 10th International Conference on Computer Vision*. Vol. 1. IEEE, 786–791.

Zhao, G., Huang, X., Taini, M., Li, S., and Pietikainen, M. 2011. Facial expression recognition from near-infrared videos. *Image and Vision Computing 29,* 9, 607–619.

Dr. 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. He earned the Ph.D. Degree from at Charotar University of Science and Technology, Changa, India. 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, Data Mining, Data Analytics and Image Processing. Prof. Goyani has also served as a reviewer in many international journals. He was also invited as a session chair in International Conference on Computer Science, Engineering and Information Technology, Tirunveli, Tamilnadu, India in 2011. He has been a committee member in many national and international conferences.