

FACIAL EXPRESSION IDENTIFICATION BY USING FEATURES OF SALIENT FACIAL LANDMARKS

Abhilasha Shukla¹, D. D. Dighe²

¹Department of Electronics and Telecommunication, Savitribai Phule Pune University, India

²Department of Electronics and Telecommunication, Savitribai Phule Pune University, India

Abstract

Facial expression recognition/identification (FER) systems plays vital role in the field of biometrics. Localizing the facial components accurately is a challenging task in image analysis and computer vision. Accurate detection of face and facial components gives effective performance with classification of expressions. This paper proposes feature based facial recognition system using JAFFE and CK databases. 18 facial landmarks were located using Haar cascade classifier. The distances between 12 points were extracted as features. These features were classified using SVM and K-NN classifier and comparison based on accuracy and execution time is done. The proposed algorithm gives better performance.

Keywords: Facial expression recognition (FER) system, Face detection, Feature extraction, Expression classification

-----***-----

1. INTRODUCTION

Fundamental modes of communicating human emotions are facial expressions. Facial expression, body language, expressive features of speech, physiological signals (e.g. EMG, ECG, EOG, EEG, FMRI, etc.) are some rare examples of signals that are beneficial for identification purpose. The detection of six universal expressions like happiness, sadness, anger, fear, disgust and surprise is the base of the research on facial expression analysis. Since the last few decades, a number of facial expression identification techniques have been proposed. For effective expression analysis, dependencies are based upon the accurate representation of facial features. Using this feature we have several usages in the realm of human-computer interaction (HCI), like social signal processing, social robots, deceit detection, interactive video and behavior monitoring.

Automatic FER systems consist of three major steps: Face detection, Feature extraction and Facial expression classification. Firstly, in any FER system face detection is been done for detecting features like eyes, eyebrows, nose and mouth which is further followed by extraction of features. There are several methods used for feature extraction purpose; but most of the present algorithms are based on geometric and appearance based features. In geometric-based method the layout of the face and facial components is being tracked like eyes, eyebrows, nose, mouth (lip corners) etc., and classify the expressions based on the relative location of these facial features. Some researcher categories the facial expression using shape model. In many practical situations, it is difficult to achieve

tracking of facial landmarks, and for that these methods usually require exact and reliable detection. The Euclidean distance between facial features keeps varying with the people, thereby making the person independent expression identification less reliable. To overcome this appearance-based method can be applied which involves distinct filters such as Gabor, Wavelets, and LBP etc. These are implied on either entire face or at the précised part of it to encode the texture. PCA, ICA, LDA etc, are certain dimensionality reduction methods which are used in appearance based as in this high dimensional vectors are generated which further has to be delineated in lower dimensional subspace.

Lastly expression classification is performed in the learned subspace. Many researchers state that accurate extraction of features can be achieved by dividing the face into several components. But this approach fails with inappropriate face arrangement and occlusions. Based on training data, features from certain facial parts, mostly determined facial parts that contribute more towards expression discrimination. And the positions and sizes of the facial landmarks are differing in this kind of approaches, making it difficult to conceive a generic system.

2. RELATED WORKS

To increase the performance of facial expression identification it is very important to correctly detect the facial elements which are followed by feature extraction and classification of expressions. Pantic and Rothkrantz [1] used the geometric approach model based method in which 19 points from frontal view and 10 points from profile view were used to describe the face model. The frontal and profile combination has enhanced the face model quality. Hammal

[2] used the contour-based approach in which contours of facial features are been automatically extracted. Its major disadvantage is confusion errors arose between distinct facial expression classes. In [3] Black and Yacoob proposed a new approach using optical flow-based method in which facial expressions were detected from motion temporal video information.

These were the few geometric approach based methods [1], [2], [3], [4] where facial landmarks were detected by using the parametric geometric model. Some appearance based approach [5] Sirovich and Kirby used the Principal component analysis (PCA) also known as eigenface approach is widely used as reducing the dimensionality. Belhumeur [6] used FLD Fishers linear discriminant whose error rate is lower and recognition rate is more than PCA. Author [7] applied eigenface based algorithm to various images clicked under peculiar illuminations and backgrounds, where the size of an image is 180*200 and required durations is 4.5456 sec.

In Zhang's work [8] the facial images are pre-processed and then the evaluation is done separately with different Gabor filters where these filters are functioned separately on different expression. The classifier assumes the discriminate function to be a linear function of the feature data. In this case the data is the feature vector obtained. Its advantage is reduction in dimension of feature space and computation complexity. In Shan's proposed work [9] analytic local features represent face, LBP is used for person sovereign expressions identification in which texture analysis is done using SVM.

In this paper we [1] proposed here a novel salient facial landmark detection technique based facial expression recognition framework, which gives significant performance at different image resolutions.

A multi-class classifier classifies the images into six basic expression classes. And it uses lower computational complexity to perform the state-of-art methods in near frontal images. To reduce the computation, the appearance features with lower number of histogram bins are used.

3. PROPOSED ALGORITHM

3.1 System Block Diagram

Facial expression involves the facial muscles contraction and expansion. The proposed facial expression recognition system is explained below-

3.2 Image Acquisition and Pre-processing

Facial expression images are gathered from the different databases such as JAFFE, Cohn-Kanade etc. JAFFE database consist of the image of different seven expressions of different Japanese female. The input image contains some distortion. To remove distortion and to get smoothen images; we applied the Gaussian filter of kernel size 5×5. In similar way CK database consist of images from 100 different universities from 18 to 30 years of age group. Image sequences from normal to target display were digitized into 640*490 pixels with 8bit precision for gray scale values.

3.3 Facial Landmark Detection

Facial part detection is important step to detect facial landmark and emotion recognition. To detect face and facial part Haar cascade classifier is used. The general object detector framework is proposed by Viola and Jones in 2004. It is mainly based on the basis of boosting cascade of weak classifiers. It trains the separate classifier for each target object and rejects non-object patterns. By using Haar cascade classifier, we detect left and right eyes, Nose, Mouth.

3.3.1 Eyes, Nose and Mouth Detection

The facial image is given to the Haar cascade classifier to detect both the eyes separately. Haar cascade classifier returns the vertices(x, y, w, h) of the eye position. By using Haar cascade classifier we can reduce the computational complexity of the algorithm. In similar manner, we can detect Nose as well as Mouth from the facial image.

Algorithm

1. Select the coarse ROI for eyes and nose
2. Detection of eyes using Haar Classifier which is trained separately for both the eyes
3. Center of eyes were computed using the coordinates of vertices which were provided by Haar classifier
4. Localization of nose and mouth in similar way that of eyes using Haar cascade
5. Localization of eyes was made using up-right face alignment as position of eyes do not change with expressions.

3.3.2 Lips and Eyebrows Corner Detection

Once the mouth is detected, we can extract the lip corners likewise with prior to eyebrow ROI, we can detect eyebrow corner. The upper lip and eyebrow produce the edge. It can be detected by the horizontal Sobel edge detector. The detailed process is shown in below flowchart Fig-2.

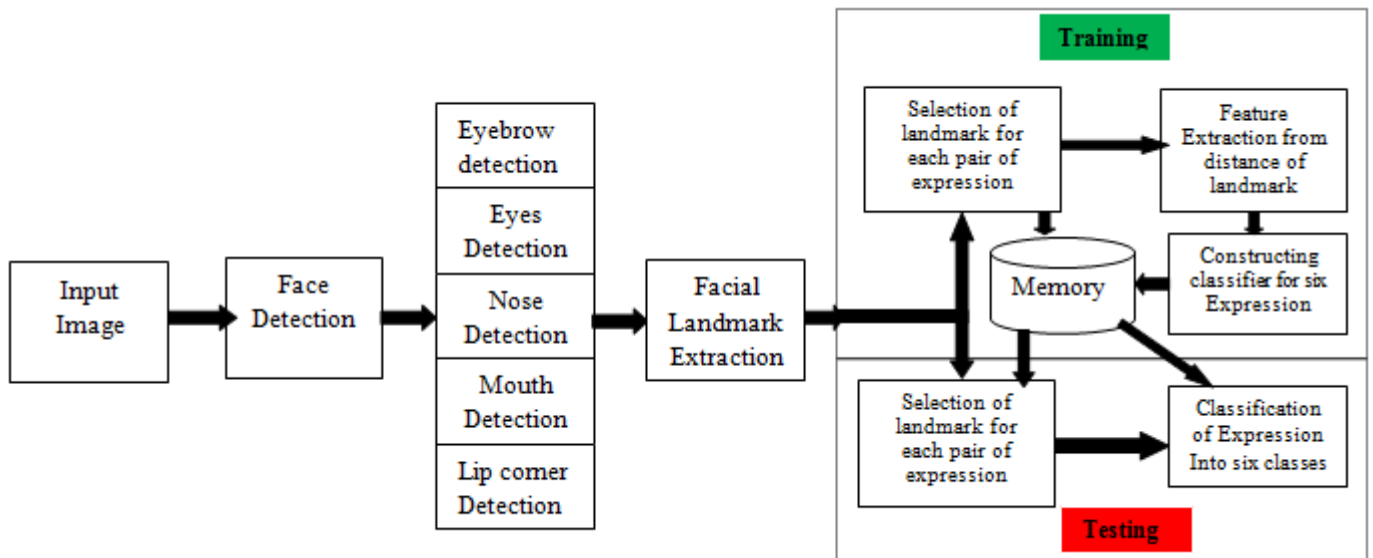


Fig-1 System Block Diagram

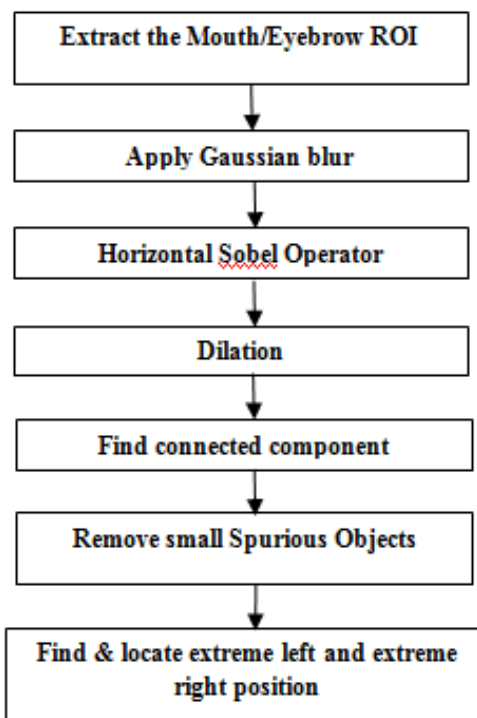


Fig-2: Flowchart for Lip corner and Eyebrow corner detection Algorithm

3.4 Facial Landmark Extraction

During different facial expression different facial points are active. So we need to detect such a facial points from the facial image and corresponding position of different facial part. The detailed active facial points are shown in below fig.3. As shown below in Fig-3, the 18 landmark points (L1-L18) were localized using which 12 features were extracted whose selection is explained later in this paper.



Fig-3: Position of facial landmarks.

3.5 Selection of Features

The different facial expression involves a different pair of facial landmarks, e.g. for Happy, the lip corner distance(L5-L6 & L17-L18) getting increased as well as distance between cheek point(L12-L13) and eye center(L8-L9) get decreased. Below Table-1 described the facial expressions and involve facial points.

Table 1: Facial Expressions and Involved Facial points

Expressions	Involved facial Landmark Points
ANGER	L5,L6,L17,L18
DISGUST	L5, L6, L8, L9, L2, L13, L14, L15, L18
FEAR	L1, L4, L5, L6, L8, L9, L14, L15, L17, L18
HAPPY	L5, L6, L8, L9, L12, L13, L16, L17, L18
NORMAL	None
SAD	L16, L17, L18
SURPRISE	L5, L6, L7, L8, L9, L11, L12, L13, L15, L16, L17, L18

Different facial expressions involve some common landmarks. So we need to optimize those landmarks. After

optimizing the landmarks 6 horizontal distance and 6 vertical distance feature were obtained. Final feature selection is tabulated below in Table-2.

Table 2: Final Feature Selection

Horizontal Distances	Vertical distances
1.A1(q)=L16(1)-L5(1),	7.A7(q)=L12(2)-L8(2),
2.A2(q)= L16(1)-L17(1),	8.A8(q)=L14(2)-L8(2),
3.A3(q)=L6(1)-L16(1),	9.A9(q)=L15(2)-L9(2),
4. A4(q) = L11(1)-L4(1);	10.A10(q)=L13(2)-L9(2),
5.A5(q)=L11(1)-L1(1),	11.A11(q)=L8(2)-L14(2),
6. A6(q) = L18(1)-L16(1)	12.A12(q)=L9(2) - L15(2)

3.6 Classification

Classification is the process of classifying the unknown input data into respective classes. We used multiclass support vector machine (SVM) and K-NN for classification of facial expression into 7 groups such as anger, disgust, fear, happy, normal, sad and surprised.

3.6.1 SVM

SVM is first developed by Vapnik [12] and the learner. The input image to the SVM classifier is the feature vectors which are generated during the training phase. The classification of this trained data in the given feature space is done by using hyperplane concept defined by the type of kernel function used. Here the SVM classifier uses the RBF (Radial Basis Function) type of kernel which provides better accuracy as compared to linear and polynomial based classifiers.

The classification of data in high dimensional feature space is based on algebra and geometry with its non-linear rules in the input space. For this purpose the learning algorithm is formulated by using kernel function (RBF) which allows efficient computation of the trained vectors. Using the nonlinear mapping Φ that embeds input vectors into feature space, kernel has the form-

$$K(x, z) = \langle \Phi(x) \cdot \Phi(z) \rangle \tag{1}$$

The separation of trained database is based on hyperplane defined by the type of kernel used. The hyperplane of maximal margin is defined as the sum of the distances of the hyperplane defined by the type of kernel used. The hyperplane of the maximal margin is defined as the sum of the distances of the hyperplane from the nearest data point of each of the two classes. The SVM methodology learns non-linear functions of the form-

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i y_i K(x_i \cdot x) + b) \tag{2}$$

Where, α_i is the Lagrange multipliers of a dual optimization problem, it is to show that α_i are non-zero in the optimal solution which are arising from the training points nearest to the hyperplane called as support vectors.

3.6.2 KNN

KNN algorithm (K-Nearest Neighbor) is a non-parametric tool generally used for pattern recognition. It is used for both the purpose i.e. classification and regression respectively; in which the input contains the nearest value (weight) of K training examples in the feature space.

The classification is based on majority vote of its neighbors, with the object being assigned to the class most common among its K nearest neighbor, where K is considered as positive integer. This choice of K being positive is based upon the data; it is considered that larger the value of K lesser is the interference of noise while classification. The algorithm for K-NN is shown below in the Fig-4.

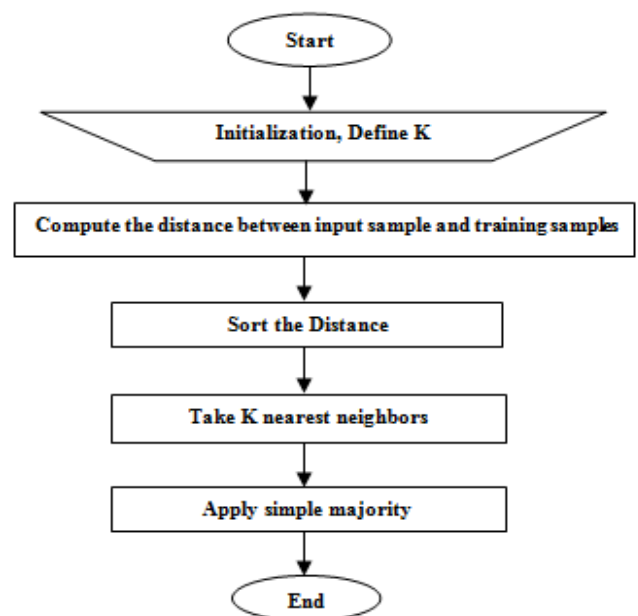


Fig-4: K-NN Flowchart

4. PERFORMANCE ANALYSIS

In the proposed approach, 18 landmark points were localized i.e. L1-L18 using Viola-Jones algorithm with HAAR cascade like features with the input as 63 facial images of different expressions from JAFFE database and 54 facial images from CK database. While training, by using those 18 landmark points 12 features (A1-A12) were extracted in which 6 (A1-A6) are horizontal features and rest 6 (A7-A12) are vertical. As explained previously in this paper the feature extraction phase is carried out by using landmark detection algorithm which is based on Euclidean distance method. The selection of features is purely based on practical analysis. Later for testing 35 images (5 images for each expression) from both the databases were taken and classification was done by using two classifiers i.e. SVM and K-NN. Hence this section includes comparison of performance of both the classifiers and databases by considering the parameters like accuracy and execution time.

4.1 Qualitative Analysis

Qualitative measurement focuses on collecting information that is not numerical. The output of the proposed system at different stages are shown below in Fig-5

4.1.1 Localization of 18 Landmark Points L1-L8

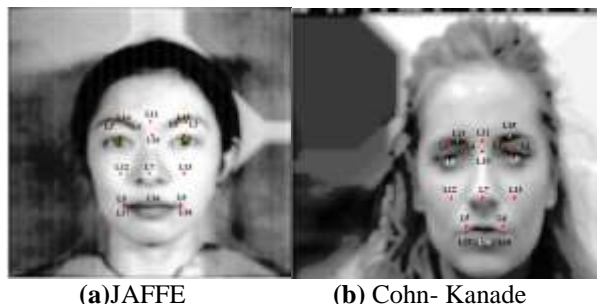
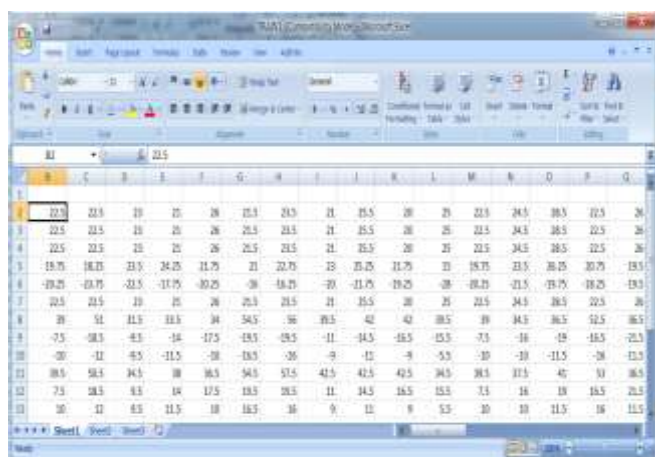
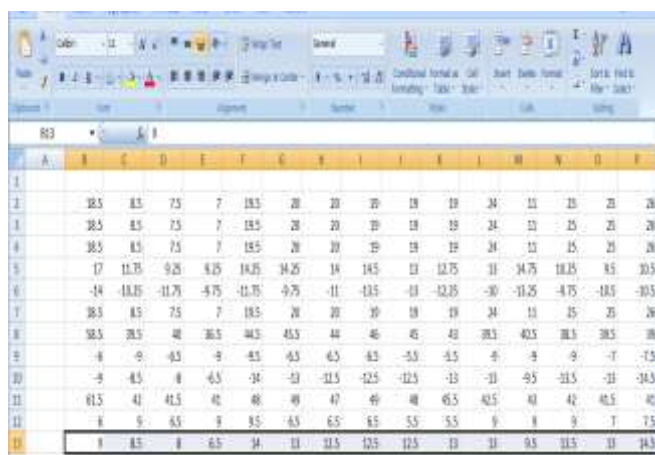


Fig-5: (a) JAFFE and (b) Cohn-Kanade Database Labeled Facial Landmark Points L1-L18.

4.1.2 Extraction of Features for both the Databases and Storing the Feature Vector in Excel (.xls) File



(a)



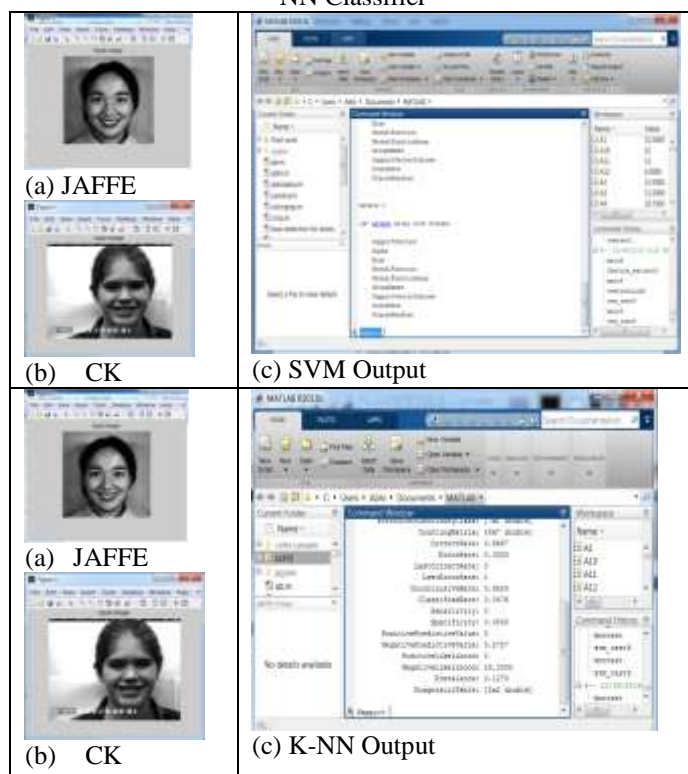
(b)

Fig-6: (a) Feature Vectors of 63 images from JAFFE Database, (b) Feature Vectors of 54 images from CK Database

4.1.3 Expression Classification

Below Table-3 shows the final expression classification done by using the SVM and K-NN classifier. The extracted feature vectors during testing are compared with the previously stored vectors during training phase. Comparison is done using SVM's hyper plane and K-nearest neighbor concept on the basis of which image is been classified and hence we get the final result.

Table-3: Expression Classification by using SVM and K-NN Classifier



4.2 Quantitative Analysis

For quantitative analysis 35 images from JAFFE and CK database 5 images from each class, were given as input to SVM and K-NN classifier to measure the accuracy and execution time of classification. Table-4 shows the comparison made between two classifiers SVM and K-NN using CK and JAFFE databases. The comparison is based on expression recognition accuracy which is plotted in the below graph i.e. represented in the Chart-1. The comparison made by considering execution time is also shown in Table-5 and Chart-2 respectively.

Table-4: Comparison based on expression recognition accuracy

		Classifier	
		SVM	K-NN
Database	JAFFE	82.85%	65.71%
	CK	93.30%	80%

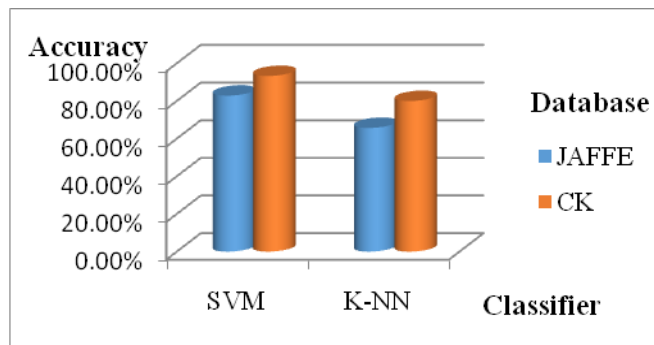


Chart-1: Comparison graph plotted, based on accuracy

Table.5. Comparison based on accuracy and time

	Accuracy (%)	Time (sec)
JAFFE+SVM	82.86	11.963
JAFFE+K-NN	65.71	12.209
CK+SVM	88.57	14.101
CK+K-NN	80	12.547



Chart-2: Comparison graph plotted, based on accuracy and time

In the above analysis 35 images from both the databases were tested separately which includes 5 images from each class i.e. anger, disgust, fear, happy, neutral, sad and surprise using SVM and K-NN classifier respectively. The accuracy of expression recognition for SVM+JAFFE is 82.86% and execution time required by it is 11.963sec; likewise for SVM+CK is 65.71% and time required is 12.209sec, CK+JAFFE is 88.57% and time required is 14.101sec and lastly CK+K-NN is 80% and time required by it is 12.547sec respectively.

5. CONCLUSION

The proposed system presents the computationally efficient FER system which accurately classifies the 7 universal facial expression classes. All the majorly active regions on the face which are responsible during the formation of expressions are extracted using landmark detection technique. These 12 features are extracted using previously located 18 landmark points. These 18 points are determined from the salient areas of the face where the features are discriminative for different expressions. Following with the

classification using SVM and K-NN classifier, comparison using the parameters like expression recognition accuracy and execution time is done. This shows that the computation is better in the SVM classifier than the traditional K-NN algorithm, and also the performance is stable for any type of database used.

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to my guide, **Mr. D. D. Dighe**, for his constant encouragement, wonderful technical guidance and support throughout my work. I sincerely thank **Prof. S. D. Pable** of Electronics & Telecommunication Department for his advice and support during course of this work. I express my thanks to all teaching & non-teaching staff of Electronics & Telecommunication Department for their kind cooperation and guidance for preparing and presenting this paper. I take this opportunity to express my gratitude towards my parents and my friends without whom it would have not been possible.

REFERENCES

- [1] M. Pantic and L.J.M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", *Image and Vision Computing J.*, vol. 18, no. 11, pp. 881-905, 2000.
- [2] Z. Hammal, "Segmentation des traits du visage, analyse et Reconnaissance d'expressions faciales par le Modèle de Croissance Transferable", PhD Thesis in Cognitive Science, University of Joseph Fourier, 2006.
- [3] M. J. Black and Y. Yacoob, "Recognizing facial expression in image sequences using local parameterized models of image motion", *Computer Vision*, 1995.
- [4] Y. Chahir, Y. Zinbi and K. Aziz, "Categorization des expressions faciales par marches lavatoires sur graphe", *CORESA*, 2007.
- [5] L. Sirovich and M. Kirby, "Low Dimensional Procedure for Characterization of Human Faces", *J. Optical Soc. Am.*, vol. 4, pp. 519-524, 1987.
- [6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, Jul. 1997.
- [7] Y. Tayal, P. Pandey and D. B. V. Singh, "Face Recognition using Eigenface", *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)* 3 (1), pp. 50-55, Dec.12-Feb.2013.
- [8] Z. Zhang, X. Mu and L. Gao, "Recognizing Facial Expressions Based on Gabor Filter Selection", 4th International Congress on Image and Signal Processing, 2011.
- [9] C. Shan, S. Gong and P. W. Mcowan, "Facial expression recognition based on Local Binary Patterns: A comprehensive Study", *Image and Vision Computing* 27, 803-816, 2009.

-
- [10] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches", IEEE transaction on affective computing, vol.6, no.1, January-march 2015.
- [11] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Syst., ManCybern. vol.9, no.1, pp. 62–66, Jan.1979.
- [12] V. N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [13] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", In Proceedings of ECML-98, 10th European Conference on Machine Learning, pages 137–142, Heidelberg, DE, Springer Verlag., 1998.
- [14] A. Shukla and D. D. Dighe, "A Review on Automatic Facial Expression Recognition System", IJESRT, ISSN 2277-9655, pg. no. 272-279, April 2016.
- [15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in Proc. IEEE Conf. Comput. Vis. Pattern Recog., pp. 511–518, 2001.
- [16] M. Philipp and R. Kaliouby, "Real Time Facial Expression Recognition in Video using Support Vector Machines", ICMI'03, 2003.
- [17] M. Song, D. Tao, Z. Liu, X. Li, and M. Zhou, "Image ratio features for facial expression recognition application", IEEE Trans. Syst., Man, Cybern., Part B: Cybern., vol. 40, no. 3, pp. 779–788, Jun. 2010.
- [18] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis", in Proc. IEEE Conf. Comput. Vis. Pattern Recog., pp. 2562–2569, 2012.
- [19] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Pearson Education, New Jersey, US, 2008.
- [20] T. Kanade, Y. Tian and J. Cohn, "Comprehensive database for facial expression analysis", in International Conference on Face and Gesture Recognition, vol. 4. IEEE Computer Society, France, pp. 46–53, 2000.
- [21] X. Wang, X. Liu, L. Lu and Z. Shen, "A New Facial Expression Recognition Method Based on Geometric Alignment and LBP Features", 2014 IEEE 17th International Conference on Computational Science and Engineering (CSE), pp. 1734-1737, 19-21 Dec. 2014.