2013 International Conference on Advances in ICT for Emerging Regions (ICTer): 051 - 055

Facial Expression Recognition using Active Shape Models and Support Vector Machines

K. Samarawickrame#1, S. Mindya#2

#Department of Computer Science, Informatics Institute of Technology
Sri Lanka

1k.samarawickrema@gmail.com

2s.mindya@gmail.com

***Abstract*— Facial Expression Recognition is the subsequent step after Face Detection and Real time recognition of facial expressions is a challenging task. Various technologies of Facial Expression Recognition has been experimented by researchers over the past few years. In this paper, it has been observed the accuracy and effectiveness of employing Active Shape Models and Support Vector Machines to achieve higher recognition rates. Active Shape Model is used to locate the facial feature deformations of a face detected by using Haar classifiers. These facial coordinates are fed into a Support Vector Machine and the trained system classifies the expressions into seven categories, namely happy, sad, anger, disgust, fear, surprise and neutral. The system was tested on JAFFE Database and Cross Validation had been used as a mechanism for analysing the results of the experiment.**

***Keywords*— Feature Extraction, Machine Learning, Active Shape Models, Support Vector Machines**

I. INTRODUCTION

Facial expressions convey non-verbal cues, which play an important role in interpersonal relations. Facial Expression Recognition is a challenging task since it involves face detection, facial landmark points detection and also performing machine learning to identify variations in facial expressions. According to [2], although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge.

Facial expressions play a vital role in human communication. Thus identifying facial expressions has an utmost importance. Referring [7], Mehrabian has reported that facial expressions have a considerable effort on a listening interlocutor; the facial expression of a speaker account for about 55% of the effect, 38% of the latter is conveyed by voice intonation and 7% by the spoken words." This implies that the facial expressions form the major modality in human communication.

Face Detection is the preliminary step in expression recognition. According to [12], face detection has being classified into 3 groups, namely

- Knowledge Based Methods
- Feature Invariant Approaches
- Template Matching Methods

However, [10] discusses three initial mechanisms used in face detection

- Novel image representation named Integral image [8]
- Creating a classifier by selecting a small number of important features using AdaBoost [3]
- Combining more complex classifiers in a cascade structure by focusing attention on promising regions of the image [1]

for the development of a real time face detection classifier. It could be seen this is a more efficient method for face detection since it was implemented adopting three renowned contributions in Computer Vision. Furthermore the statement in [9], "Of all the face detectors currently in use, the one introduced by Viola and Jones is probably the best known and most widely used" further confirms this opinion.

Hence, it was decided to employ Viola Jones face detection Haar classifier for face detection.

Active Shape Models enable users to mark landmark points thus a model of the image could be created. Active Shape Models manipulate a shape model to describe the location of the structures in a target image.

Support Vector Machines (SVMs) are based on the results of statistical learning theory carried out by Vapnik. SVM maps feature vectors into a higher dimensional space and classify data using linear algebra by employing a kernel function. Then an optimal hyper plane that fits into the training data is created. In a linear classification the margin between the separating hyper plane and the nearest feature vectors from both classes is maximal. The feature vectors closest to the hyper plane are called "support vectors".

SVM has evolved from sound theory to implementation and experiments while Neural Networks has followed a more heuristic path, from application and extensive experimentation to theory [11]. Moreover it states that SVM has achieved practical learning benchmarks in digit recognition, computer vision and text categorization.

Furthermore SVM has been adopted as the classification technique by many researchers in computer vision. None withstanding this, research conducted in [5] also has adopted SVM for classifying facial expressions in real time. They have used the coordinates of feature points as the input to a multi class SVM as discussed in the feature extraction section. Assuming the training data is $(g_1,l_1),\ldots,(g_N, l_N)$ where $g_j \in \Re^F$ $j = 1,\ldots,N$ the deformation feature vectors and $l_j \in \{1,\ldots,6\}$ $j=1,\ldots,N$ are the facial expression labels of the feature vector. It constructs 6 (six facial expressions) two-class rules where the k-th function, $w_k^T \emptyset(g_i) + b_k$ separates training vectors of the class k from the rest of the vectors. Hence, there are 6 decision functions, all obtained by solving one SVM problem. 93.7% accuracy rate has been achieved with the Cohn-Kanade database. Moreover it indicates that the given accuracy is the highest reported in literature for Cohn-Kanade database up to 2005 according to their knowledge. Thus it proves that using SVM for classification was a driving factor for achieving a high accuracy rate as given.

Similarly reason in [6] is that SVM's ability to outperform ANN in a variety of applications was a factor considered in selecting SVM for their Real Time Facial Expression Recognition system. They have achieved an 86% accuracy rate for still images and 71.8% for person independent classification.

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

II. System Description

As shown in figure 1 the components of the system were broken down into three sub components namely, facial feature extraction, machine learning and human intervention.

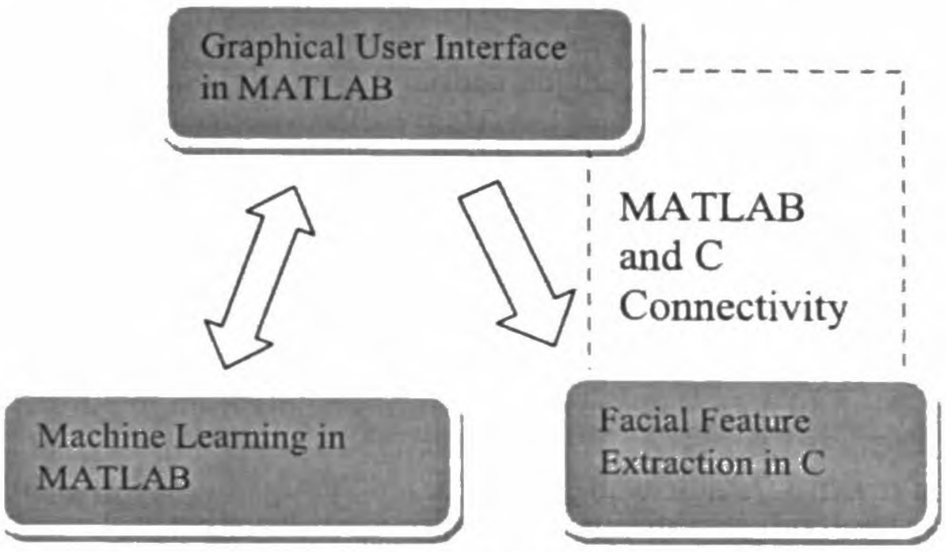


Fig. 1 System Description

A. Facial Feature Extraction Module

i) User Input

User should input an image that needs to extract facial feature points. The image will be sent to the facial feature extraction module. As described in the diagram above, this module is used to localize the facial features and identify the values of their coordinates. GUI has been developed using MATLAB.

ii) Face Detection

Face detection module is used to detect faces in a given image. In order to locate the facial feature points, detection of the face in the given image is essential.
In [10] it was identified that using face detection classifier is an optimal solution for face detection in this research. This decision was taken since OpenCV providés Haar Classifiers for face detection which could be integrated in the project.
Haarcascade_frontalface_alt2.xml was used as the haar classifier for detecting the face. OpenCV also provides the following Haar classifiers for face detection

- Haarcascae_frontalface_alt.xml
- Haarcascade_frontalface_alt_tree.xml
- Haarcascade_frontalface_default.xml
-

iii) Fitting a Face Model

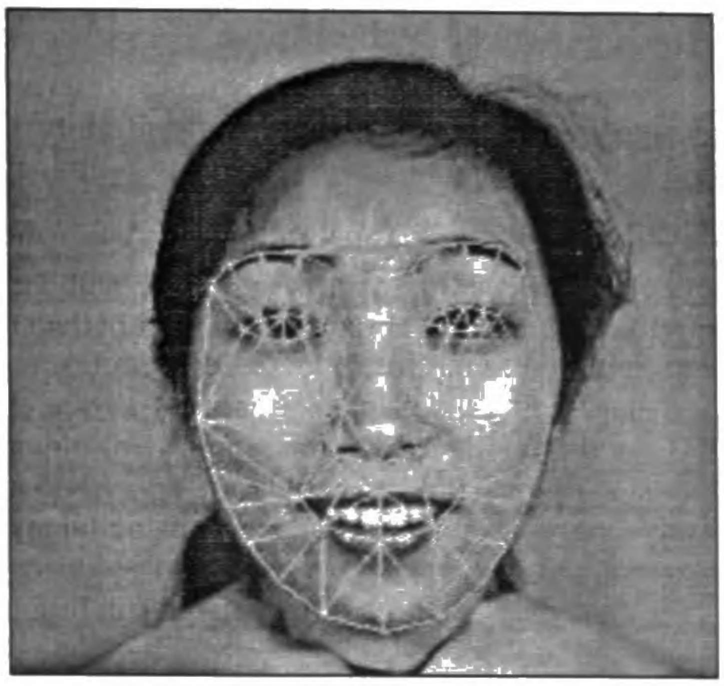


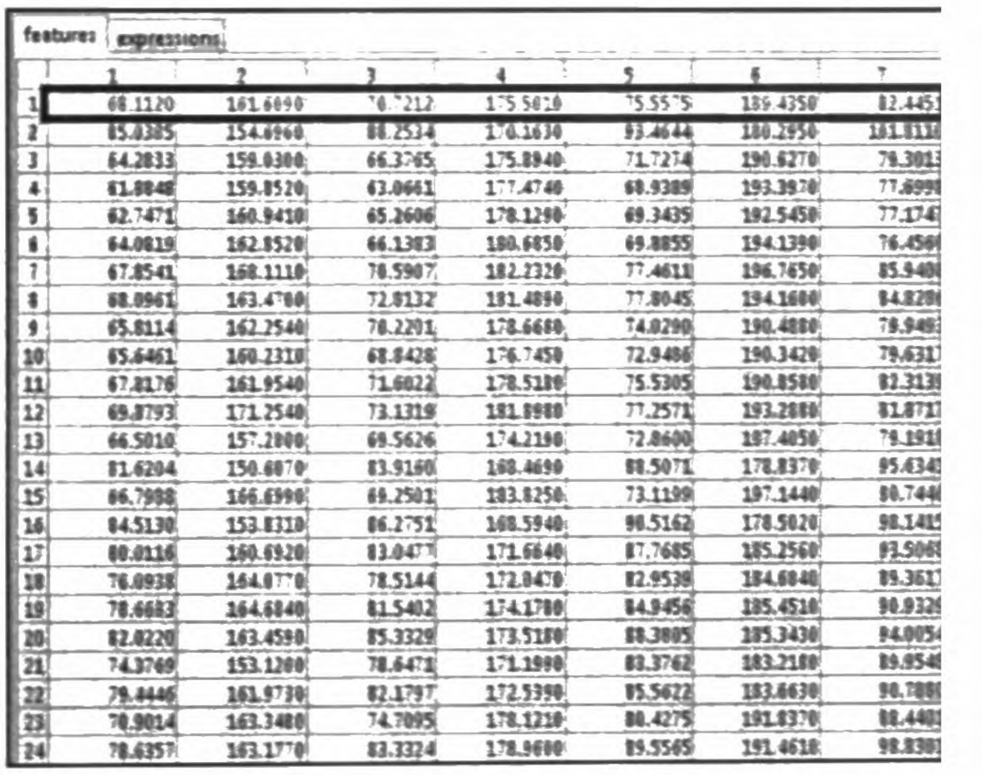
Fig. 2 – An image in Jaffe database after fitting the face model Image in Jaffe database (Michel et al, 1998)

iv) Detecting Feature Points

After fitting the face model on to the given image, the facial landmark points can be extracted and written to a file.

B. Machine Learning Module

This module is designed to train the Support Vector Machine with labelled data for the expressions and classify expressions. Thus it should identify an expression of a given image. Figure given below, shows a snapshot of the Matlab matrix file containing the training data (140*117) of all the feature points denoted by a variable called 'features' and the labels corresponding to each row denoted by a variable called 'expressions'. In the figure, tab named features is highlighted in order to display the two matrices clearly.



features | expressions

	1	2	3	4	5	6	7
1	68.1120	161.6090	70.7212	175.5020	75.5575	189.4350	82.445
2	85.0385	154.6960	88.2534	170.1630	93.4644	180.2950	101.811
3	64.2833	159.0300	66.3765	175.8940	71.7274	190.6270	78.301
4	61.8848	159.8520	63.0661	177.4740	68.9389	193.3970	77.699
5	62.7471	160.9410	65.2606	178.1290	69.3435	192.5450	77.174
6	64.0819	162.8520	66.1383	180.6850	69.8855	194.1390	76.456
7	67.8541	168.1110	70.5907	182.2320	77.4611	196.7650	85.940
8	68.0961	163.4700	72.8132	181.4890	77.8045	194.1600	84.820
9	65.8114	162.2540	70.2201	178.6680	74.0290	190.4880	78.949
10	65.6461	160.2310	68.8428	176.7450	72.9486	190.3420	79.631
11	67.8176	161.9540	71.6022	178.5180	75.5305	190.8580	82.313
12	69.8793	171.2540	73.1319	181.8980	77.2571	193.2880	81.871
13	66.5010	157.2000	69.5626	174.2190	72.8600	187.4050	79.191
14	81.6204	150.6070	83.9160	168.4690	88.5071	178.8370	95.634
15	66.7988	166.6990	69.2501	183.8250	73.1199	197.1440	80.744
16	84.5130	153.8310	86.2751	168.5940	90.5162	178.5020	98.141
17	80.0116	160.6920	83.0477	171.6640	87.7685	185.2560	93.506
18	76.0938	164.0770	78.5144	172.0470	82.9539	184.6840	89.361
19	78.6683	164.6840	81.5402	174.1780	84.9456	185.4510	90.932
20	82.0220	163.4590	85.3329	173.5180	88.3805	185.3430	94.005
21	74.3769	153.1200	78.6471	171.1990	83.3762	183.2180	89.954
22	79.4446	161.9730	82.1797	172.5390	85.5622	183.6630	90.788
23	70.9014	163.3480	74.7095	178.1210	80.4275	191.8370	88.440
24	78.6357	163.1770	83.3324	178.9600	89.5565	191.4618	98.838

Fig. 3 – Matlab Matrix with the training features

The figure given below displays Matlab matrix file containing the expression corresponding to each row from the 'features' matrix, which is one image with a particular expression. The tab is highlighted in order to display it clearly.

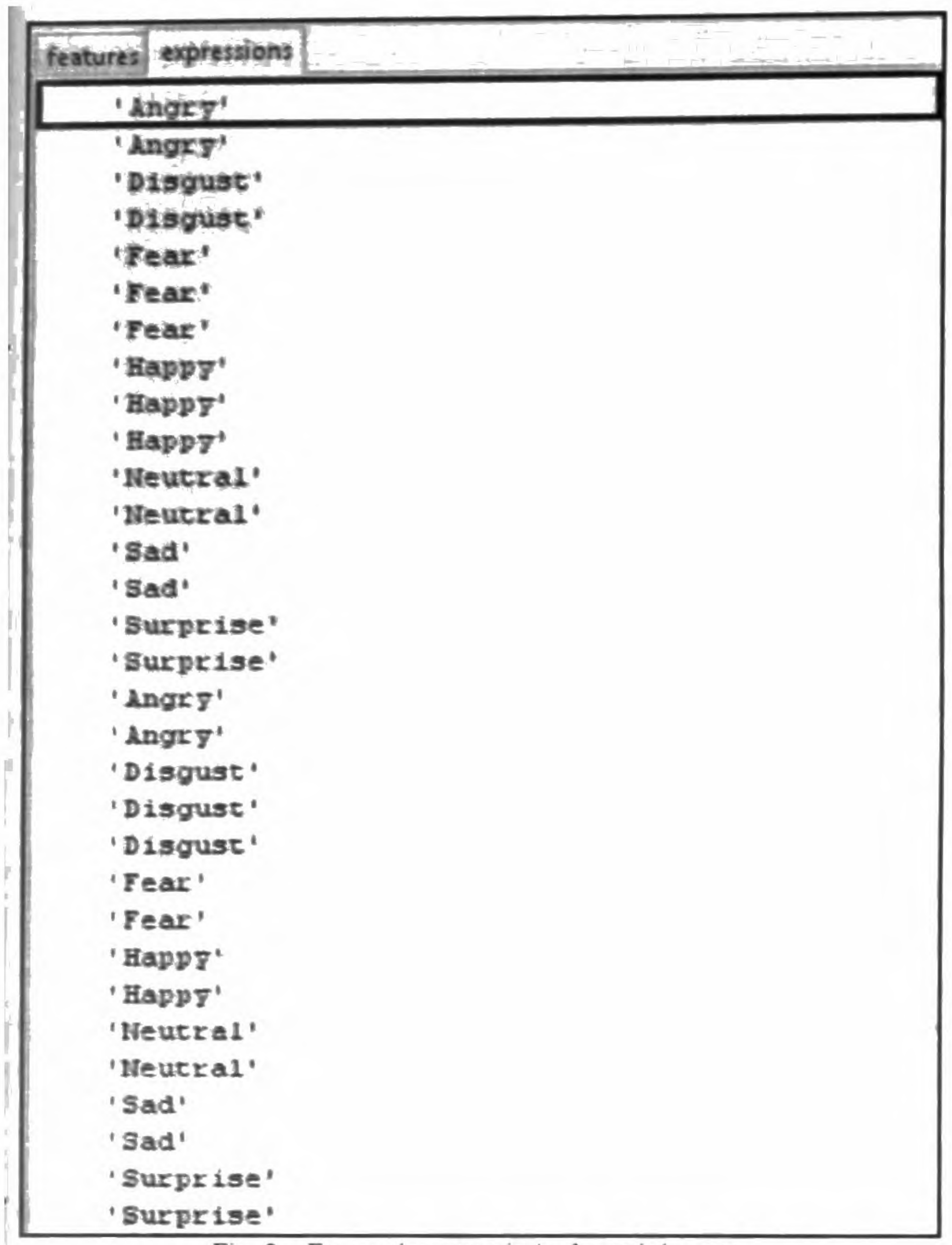


Fig. 3 – Expressions matrix in the training set

B. SVM Training

SVM Training module is used in order to train the support vector machine about the differentiations in landmark points for different facial expressions. User could select the feature file that need to be fed into the SVM and train the system with the labelled data. The basic architecture of the SVM can be displayed in figure 4.

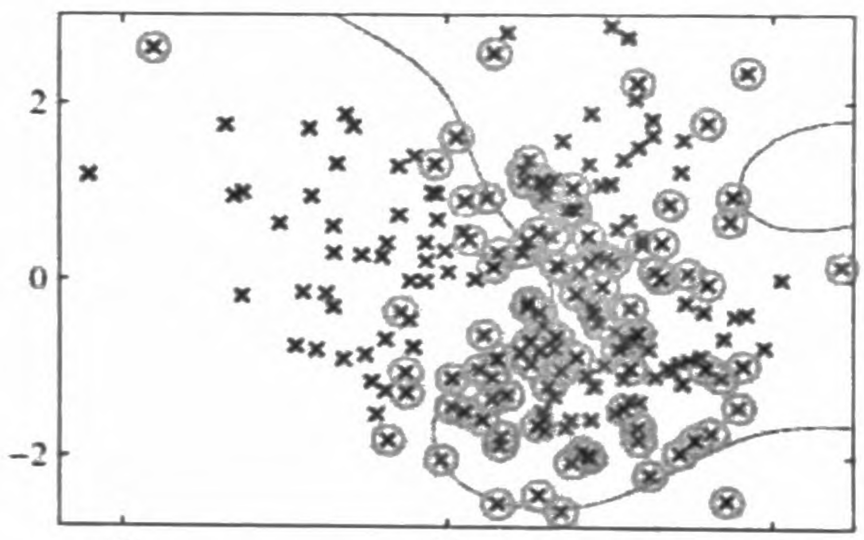


Fig. 4 - SVM Classification of data (Bishop, 2006)

SVM applied to a non-separable data set in two dimensions and the Support Vectors are indicated by circles.

C. SVM Classify

This module is designed to recognize the expression of a given image and output the expression to the user. As a user inputs an image to recognize expression, the feature extraction process described above will be performed. Afterwards, the user is required to select the feature file corresponding to his image and the SVMClassify module is implemented in such a way to output the expression to the user.
Although RBF kernel had been used for separating the training data into feature space, better accuracy rates could have been achieved, if the system was tested for a set of kernel functions when they are training the SVM.

III. Results and Discussion

Cross Validation, sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. This method has been used as the Testing Metric for evaluating the given system.

JAFFE database was used for evaluation of the project since it contains face images of a variety of expressions such as Happy, Sad, Neutral, Angry, Disgust, Fear, Surprise and Neutral. Author tried to carry out testing using Yale Face Database as well. But for some of the images, a problem occurred in detecting the faces in the images using OpenCV. Thus the author carried out testing using the JAFFE face database.

In this section author has presented the average correct rates and error rates for cross validation performed 10 times. Each time random data will be selected for training and testing.

TABLE I

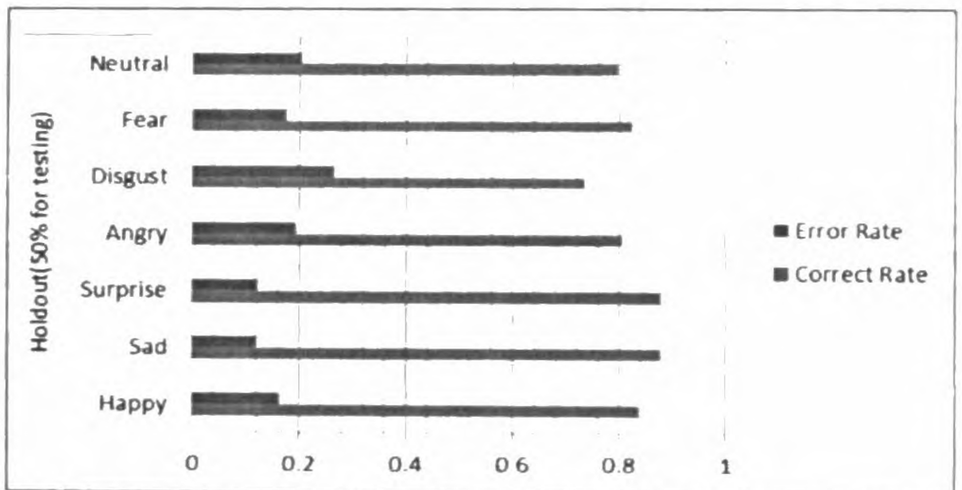


Fig. 5 – Results of using 50% of data for testing and 50% for training

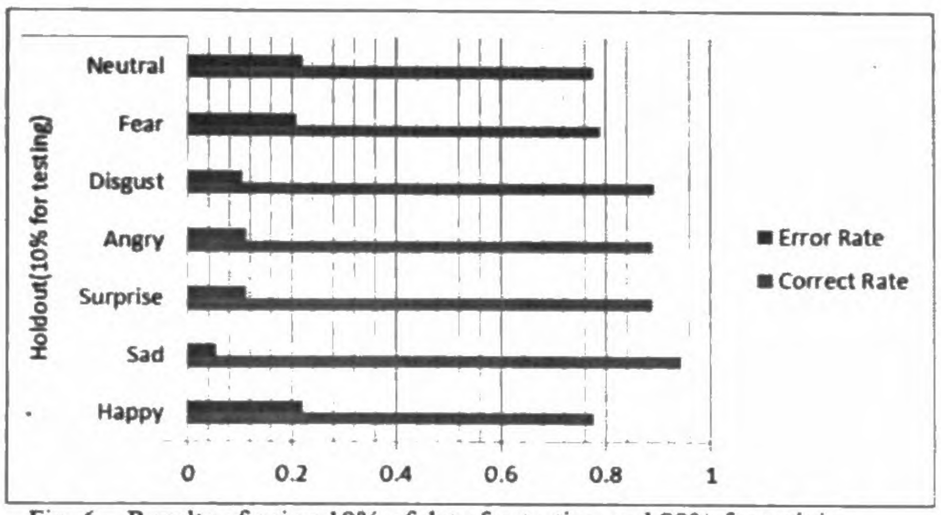


Fig. 6 – Results of using 10% of data for testing and 90% for training

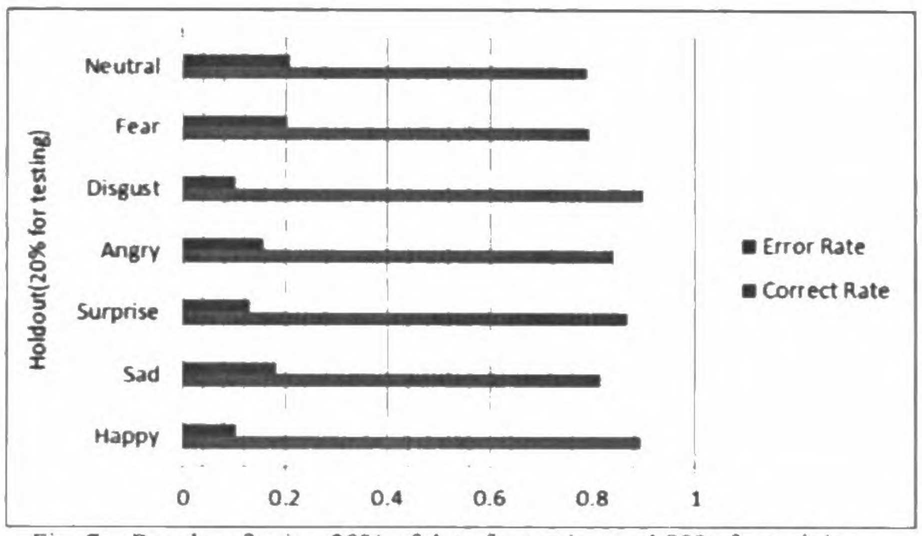


Fig. 7 – Results of using 20% of data for testing and 80% for training

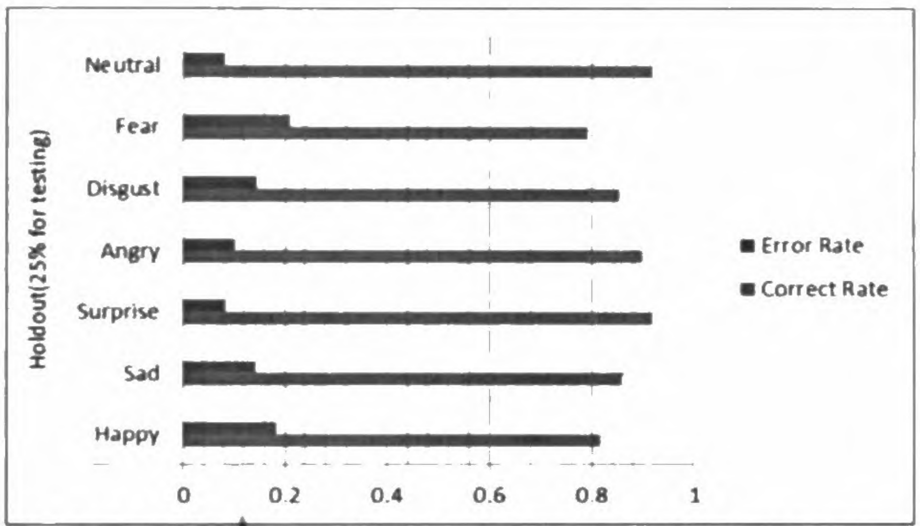


Fig. 8 – Results of using 25% of data for testing and 75% for training

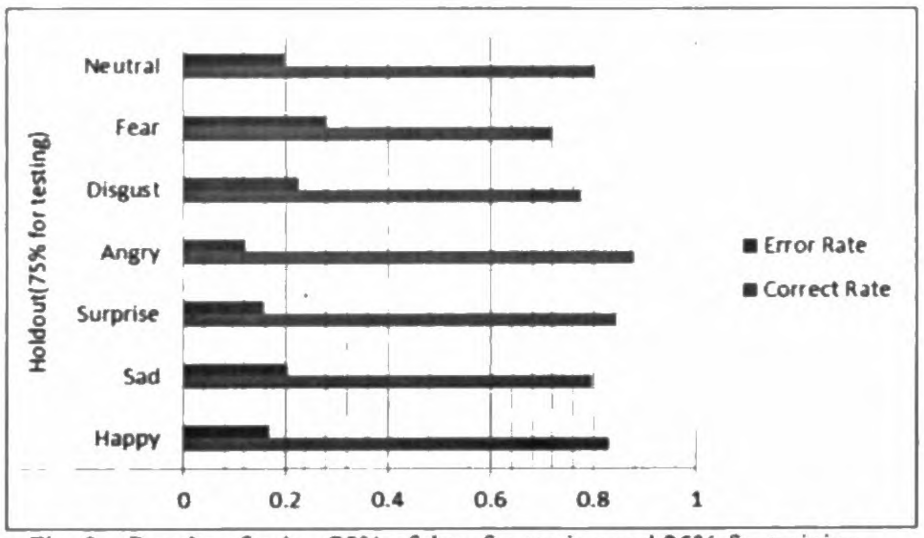


Fig. 9 – Results of using 75% of data for testing and 25% for training

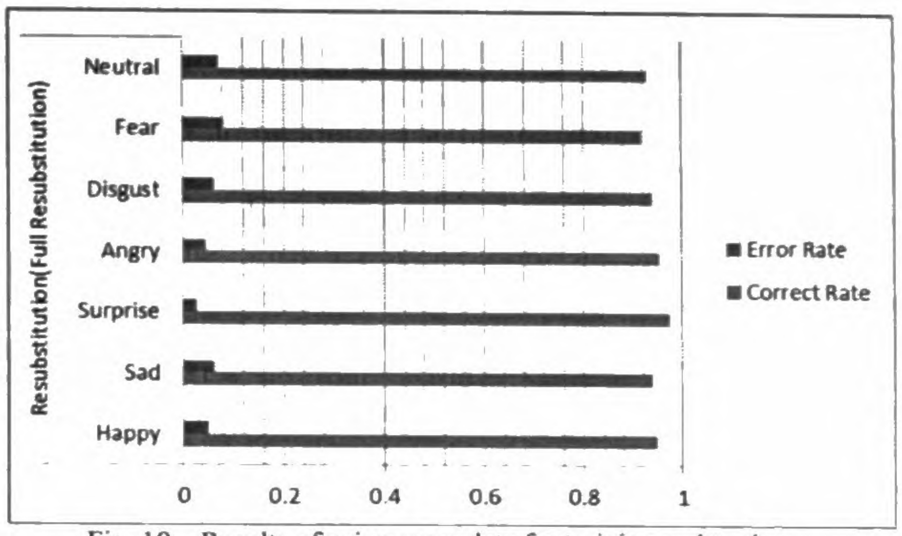


Fig. 10 – Results of using same data for training and testing

TEST RESULTS

From the test results given above, it could be deduced that the highest correct rate is generated for Full Substituion, where the same data will be used for training and testing. Examining the correct rates and error rates it could be inferred that for Holding out 10% out of 196 data for testing, highest correct rate has been achieved. And conversely that lowest correct rate was achieved for holding out 75% of the data for testing. Thus scrutinizing the above findings, it could be inferred that higher accuracy rates could be achieved providing larger data set for training. On the other hand, it could be observed that, sad expression had a higher recognition rate compared to other expressions by obtaining the highest correct rate for 50% holdout and 10% holdout. At the same time, Fear had the lowest correct rate comparatively, by obtaining lowest correct rate for 10% holdout, 75% holdout and Full Resubstitution. Hence it could be inferred that sad expression could be more easily recognized whereas fear expression has a lower recognition rate.

This method has been applied for person invariant facial expression recognition as well. The pictures below display images where expressions were accurately identified as "Happy" although test data contained faces only from Jaffe Database.

Fig. 5 –System used for Person Invariant Recognition

IV. CONCLUSIONS

The paper discussed the important factors in developing a static facial expression recognition system. This sets the groundwork that need to be considered for enhancing the above system and developing a spontaneous facial expression recognition system.

Currently all expressions are not being recognized with the same accuracy. From the given results it could be observed that the sad expression had the highest recognition rate and fear displayed the lowest recognition rate while performing Cross Validation against Jaffe Dataset. This needs to be taken into consideration when enhancing the system.

Identity information is crucial when providing machines with the background knowledge needed to accurately interpret measurements and observations of human expressions. Hence it is a challenge to develop a person independent Facial Expression Recognition System.

One of the challenges in Facial Expression Recognition systems is addressing differences in facial features and facial expressions between cultures (Europeans and Asians) and age groups (adults and children). Hence, the above fact needs to be considered, when developing a robust facial expression recognition system.

Apart from the six prototypic expressions there are a host of other expressions that can be recognized. But capturing and recognizing spontaneous non-basic expressions is even more challenging than capturing and recognizing spontaneous basic expressions. Use of a FACS (Facial Action Coding System) should also be considered in order to increase the accuracy level when developing a robust facial expression recognizer.

REFERENCE

[1] Y. Amit, D. Geman and K. Wilder, "Joint Induction of Shape Features and Tree Classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19(11), pp. 1300 – 1305, 1997

[2] C.C. Claude and F. Bourel, "Facial Expression Recognition : A Brief Tutorial Overview ",Stafford: Staffordshire University, 2002

[3] Y. Freund and R.E. Schapire, "A Decision – Theoretic Generalization of On-Line Learning and an Application to Boosting", Journal of Computer and System Sciences, Vol.55, pp. 119 – 139, 1997

[4] A. Kanaujia and D. Metaxas, "Recognizing Facial Expressions by Tracking Feature Shapes", Proc. ICPR '06, 2006, Vol. 2, pp. 33 – 38

[5] I. Kotisa, and I. Pitas, "Real Time Facial Expression Recognition From Image Sequences using Support Vector Machines", ICIP'05, 2005, Vol. 2, pp. 966-969

[6] P. Michel and R.Kaliouby, Real Time Facial Expression Recognition in Video Using Support Vector Machines, ICMI '03, pp. 258 – 264

[7] 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*, Vol. 22(12), pp. 1424-1444,2000

[8] C. P. Papageorgiou, M. Oren and T. Poggio, "A General Framework for Object Detection", in *Proc. ICCV '98*, 1998, p.555

[9] R. Szelisk, Computer Vision: Algorithms and Applications, Draft, 2009

[10] P. Viola and M. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, Vol. 57(2), pp. 137-154, 2004

[11] L. Wang, Support Vector Machines : Theory and Applications, Berlin, Spinger, 2005

[12] M. H. Yang, D. J. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp34-58, 2002