Face Expression Recognition

Sushirdeep Narayana
Department of Electrical and Computer Engineering
Texas A&M University, College Station
Texas, United States of America
sushir_369@tamu.edu

Ashok Shanker
Department of Electrical and Computer Engineering
Texas A&M University, College Station
Texas, United States of America
ashokshanker@tamu.edu

Abstract—**Ones face directly links to his/her emotions. Our project aims to recognize the facial expression from the frontal face image. Recognizing Facial Expressions has many obstacles and challenges starting from correctly identifying a face in the image and keeping track of the key features on the face to then associating these features to the correct emotional expression. In our project we explore various features and their combinations useful for expression recognition from static face images. Various features extracted from the face images were combined and used for classification by generating models related to the combinations of these features. The result of the proposed classification was compared with other popular existing techniques.**

Keywords—Expression Recognition; Emotion Classification; Face Detection; Feature Extraction; Feature Combination

I. INTRODUCTION

The study of Facial Expression analysis goes well back into the 19th century when Charles Darwin demonstrated the universality of facial expressions and their continuity in man and animals. The next milestone in the study of Facial Expression Analysis came in 1970s when Ekman and Friesen postulated six primary emotions that possess each a distinctive content together with a unique facial expression. The six basic emotions comprise happiness, sadness, fear, disgust, surprise and anger. Traditionally, facial expressions have been studied by psychologists, medical practitioners and artists. However, in the last quarter of the 20th century, with the advances in the fields of robotics, computer vision and image processing researchers have started showing interest in the study of facial expressions.

Facial Expression is an important communication channel other than language. As robots begin to interact more and more with humans and start becoming part of our living spaces and work spaces, they need to become more intelligent in terms of understanding the human emotions. This is precisely what the Human- Computer Interaction research community is focusing on. Facial Expression Recognition plays a significant role in building responsive Human –Computer Interaction (HCI) interfaces. Other Practical real-time applications have also been demonstrated. Bartlett et al. developed an animation character that mirrors the expressions of the user called the CU Animate [11]. Another real-time application includes 'EmotiChat' where users can chat based on their Facial Expressions. Other uses of Facial Expression Recognition are in Behavioral Science and Video Games.

Facial behaviors and motions were parameterized based on muscle actions by Ekman and Friesen when they developed the Facial Action Coding System (FACS). The process of Facial Action Coding involves identifying the various facial muscles individually or in a group causing changes in facial behaviors. These changes in the face and the underlying muscles that cause these changes are called Action Units. The FACS system is very comprehensive in its measurement and a lot of parameterization is involved when coding the FACS. The FACS system is very popular in the Computer Animation and Computer Graphics Research community.

The remainder of the Project Report will be organized as follows: Section 2 gives the workflow of the Face Expression Recognition, Section 3(a) covers Face Detection used in the project, Section 3(b) describes the various feature extraction methods employed in our project, Section 3(c) gives a description of the combinations of the feature models that were created, Section 3(d) covers the Classification employed for Face Expression Recognition, Section 4 describes how the procedure was implemented on the Cohn-Kanade (CK+) database and the results obtained from implementing the proposed method, Section 5 gives the Conclusion and Section 6 mentions the References.

II. FACE EXPRESSION RECOGNITION WORKFLOW OUTLINE

A. *Face Expression Recognition System Outline*

Figure 1 shows the system outline of our project. The first module of our system consists of Face Detection from the raw image taken from the dataset. The second module consists of extracting feature points from the detected face region of the image. The third module combines the various features extracted in the second module. The classification for Expression Recognition is performed by the fourth module.

III. MODULES OF FACE EXPRESSION RECOGNITION

The detailed explanation of the various modules of the Face Expression Recognition system outlined above is given in this section.

A. *Face Detection*

The first step in Face Expression Recognition is to detect the face in the given image or video sequence. Face detection is the process of locating the face within an image whereas the process of locating and tracking the face across different frames of a video sequence is termed as Face Tracking. In

2001, Paul Viola and Michael Jones developed a method that was very fast and could rapidly detect frontal view faces [8].

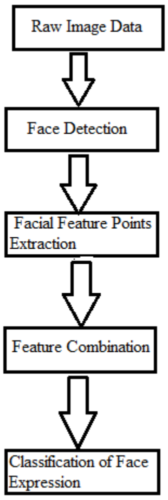

Figure 1: Face Expression Recognition workflow outline

The Viola-Jones algorithm builds a cascade object detector to detect people's faces. The training for the Viola-Jones algorithm is very slow, but the detection is very fast. The characteristics that make it a good algorithm are that it is robust (has a very high detection rate) and real time (can be used for practical applications). The Algorithm has four stages:

(1) Haar Feature Selection

(2) Creating an Integral Image

(3) Adaboost Training

(4) Cascading Classifiers

In the Viola-Jones algorithm each model is trained to detect a specific type of object by extracting features from a set of known images. Cascade classifiers are used to efficiently process image regions for the presence of a target object. In each stage of the algorithm, the cascade classifier rapidly rejects regions that do not contain the target object as shown in Figure 2. A typical output obtained from the Viola Jones algorithm is shown in Figure 3.

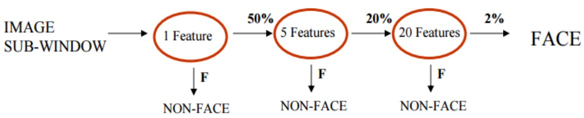

Figure 2: Cascade Classifier of Viola- Jones algorithm

B. Feature Extraction from Face Image

This is a very crucial step in the processing chain, as the facial feature points extracted for this application need to be accurate enough to be useful for expression recognition. The detected face region from the raw image is cropped and the feature extraction is performed only on this cropped face image.

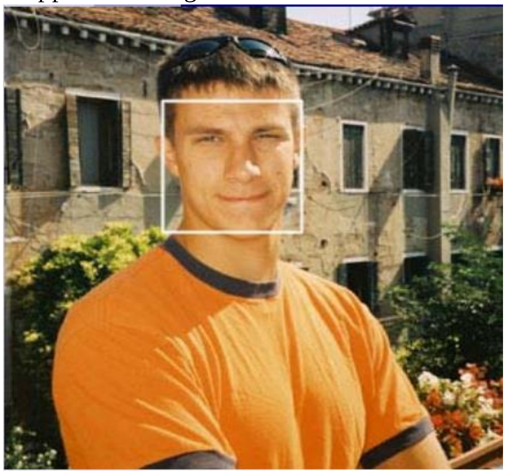

Figure 3: Face Detection using Viola Jones algorithm

The following features were computed:

(a) **Facial Feature points using Morphological Operations**

The boundary of eyes, nose, mouth and the lower jaw differ for different emotional expressions. This is important information for classifying different emotional expressions. The boundary of these features is extracted by Morphological operations.

First, histogram equalization is performed to the face image. Histogram equalization increases the global contrast of the image. Through this process the intensities are better distributed on the histogram. After, histogram equalization the pixels covering the cheeks, forehead have approximately same intensities. The histogram equalized image is now segmented to a binary image using Otsu's segmentation. This process of image segmentation groups together similar-looking pixels for efficiency of further processing. Hence, similar looking pixels of cheeks, forehead are clustered as one group whereas the pixels of the eyes, mouth and the lower jaw are clustered as another group. This process involves finding the optimum global threshold using Otsu's method. The histogram of the equalized image is computed. Otsu's method chooses the threshold value k such that the between–class variance $\sigma^2_B(k)$ is maximized. The between-class variance is defined as

$$\sigma_B{}^2(k) = P_1(k)[m_1(k) - m_G]^2 + P_2(k)[m_2(k) - m_G]^2$$

where, $P_1(k)$ is the probability of set C_1 (set of pixels with levels [0, 1, 2, …,k]) occurring:

$$P_1(k) = \sum_{i=0}^{k} p_i$$

$P_2(k)$ is the probability of set C_2 (set of pixels with levels [k+1, …, L-1]) occurring:

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i$$

$m_1(k)$ is the mean intensities of the pixels in set C_1
$m_2(k)$ is the mean intensities of the pixels in set C_2
m_G is the global mean intensity of the entire image

$$m_G = \sum_{i=0}^{L-1} i p_i$$

After obtaining the value of the optimum threshold k, the image is segmented to a binary image. The boundary of the binary image was extracted using the below equation

Boundary = A – (A erode B)

where, B is a structural element.

The structural element used was

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

The boundary of the face image obtained using the above procedure is shown in Figure 4.

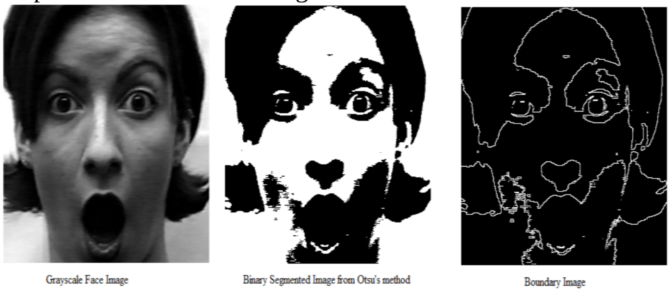

Figure 4: Illustration of Boundary Features extraction

(b) **Facial features using Landmark points**

A landmark point is a point in a shape object where in which correspondences between and within the populations of the object are preserved. Landmark points are widely used in biometrics for solving detection and recognition problems.

We used the Landmark points already extracted by using the Active Appearance Model (AAM). This makes the process of Landmark Localization faster. The AAM model is very efficient and robust model as it uses a set of parameters to characterize identity, pose, expression, lighting etc. The Landmark points extracted from the AAM model enclose the face in the form of a wire-frame mesh as shown in Figure 5. 68 Landmark points were extracted from each face image as features.

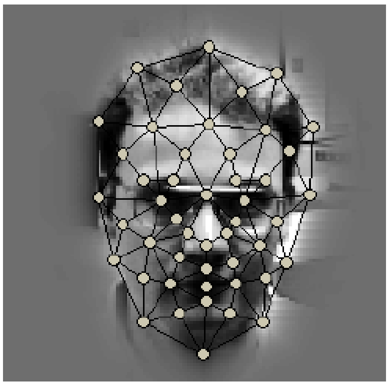

Figure 5: Landmark Points enclosing the face as a wireframe mesh

(c) **Eigenfaces**

Eigenfaces refers to the features of a set of eigenvectors of the face images when used for image processing problems. The eigenvectors are derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images. The idea of Eigenfaces was developed by Mathew Turk and A. Pentland for Face Recognition. Each face image can be made up of "proportions" of K "Eigen faces". Each face image can be expressed as the equation shown below

Face Image _1 = 23 % of Eigenface1 + 30% of Eigenface2 + 3% of Eigenface3 + …..

Figure 6 shows a picture illustrating a set of Eigenfaces.

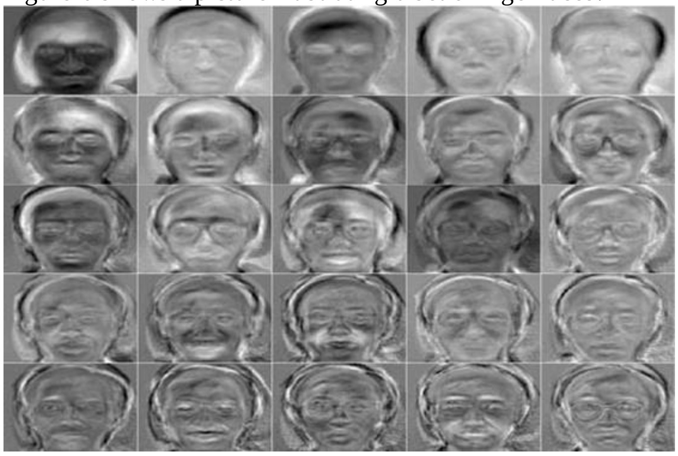

Figure 6: A set of Eigenfaces

The procedure for calculating the features related to the Eigenfaces is described below:

1. The mean input face image is calculated by taking the mean of all the input face images in the database.
2. The mean shifted images are calculated by subtracting the input images of the database from the mean face image.
3. The eigen-vectors and eigenvalues of mean-shifted images are calculated.
4. The eigen-vectors are ordered by the order of decreasing corresponding eigenvalues.
5. Eigen-vectors with only significantly large eigenvalues are retained.
6. The features of Eigenfaces are computed by projecting the mean shifted images to the eigenspace. This is done by multiplying the ordered eigenvectors with the mean shifted face images.

(d) **Gabor Features**

Gabor Features are obtained by processing the face image with a Gabor Filter Bank. The Gabor Filter is widely used in image processing and is named after Dennis Gabor. The Gabor features capture frequency components and pattern orientations of the objects in an image. Image analysis with Gabor Filters is similar to the visual cortex of the mammalian brains. In the spatial domain, a 2D Gabor filter can be modelled as a Gaussian kernel function modulated by a sinusoidal plane wave. The basis functions of the Gabor filters are non-orthogonal. Figure 7 shows an example of a 2 dimensional Gabor Filter.

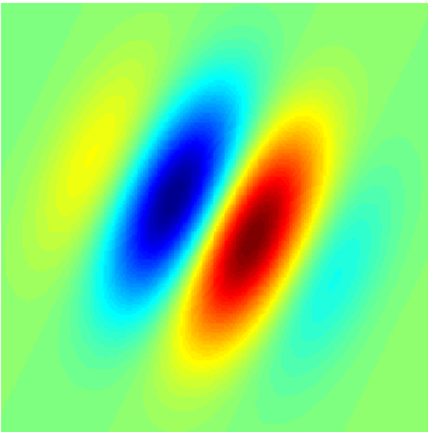

Figure 7: Typical 2D Gabor Filter

The Gabor filter has a property of minimizing the product of standard deviation in time and frequency domain. In our project we used Gabor Filter Bank containing 5 scales and 8 orientations. Figure 8 illustrates the Gabor Filter Bank used and Figure 9 illustrates the Gabor feature one obtains by applying a face image to the Gabor filters.

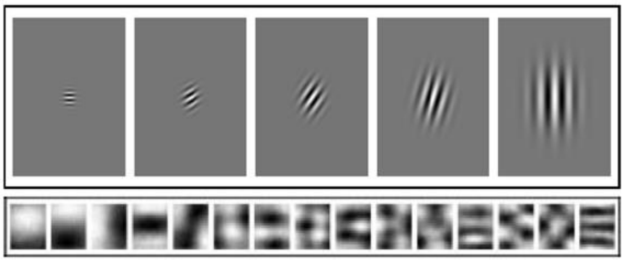

Figure 8: Illustration of Gabor kernels containing 5 scales and 8 orientations

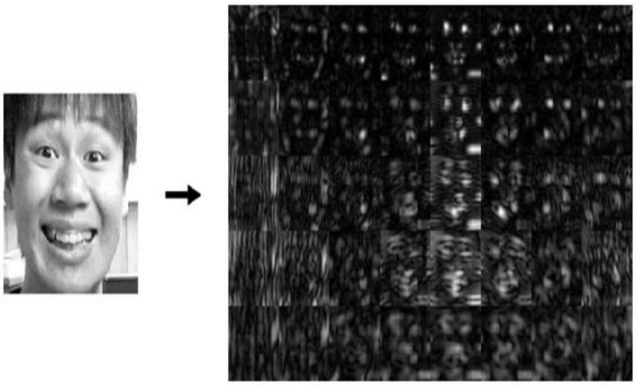

Figure 9: Gabor Response matrices

The Gabor Response matrices are converted to features by taking the absolute value of the response matrices.

(e) **Haar Wavelet features**

When digital images are to be viewed at multiple resolutions the Discrete Wavelet Transform (DWT) is very efficient. The Haar Wavelet Transform comes under the family of various Discrete Wavelet Transforms. According to the Haar Wavelet Analysis, an image can be analyzed as a linear decomposition of the wavelet coefficients multiplied by their wavelet basis functions as shown by the equation below

$$f(x,y) = \sum_j \sum_k a_{j,k} * \psi_{j,k}(x,y)$$

where, $f(x,y)$ is the image, $a_{j,k}$ are the real valued Wavelet coefficients and $\psi_{j,k}(x,y)$ are the wavelet expansion functions.

The Haar Wavelets has certain special properties which are listed as follows:

1. The Haar Wavelet Transformation of an image gives a very good space –frequency localization.
2. The Haar Wavelet basis functions are orthogonal to each other.
3. The Haar Wavelet coefficients contain more information than the signal directly.
4. A typical Haar Wavelet basis function has the following form

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k) \quad where\ t \in R$$

Figure 10 illustrates a Haar Wavelet transform on a face image.

Figure 10: Haar Wavelet Transform of a face image

The Haar Wavelet Coefficients were chosen as the features for the face image.

The various coefficients extracted were now combined to three different combinations.

s

C. Combination of Feature Models

The various features extracted from the face images were then combined into three different combinations. The three different combinations are

(1) Combination of Gabor features with Haar Wavelets and the feature points obtained by Morphological operations.

(2) Combination of Eigenfaces with Haar Wavelets and Gabor Features.

(3) Combination of Facial Landmark Detector features with Eigenfaces and Haar Wavelets.

The three feature combinations were selected assuming that each feature would contribute complementary information with respect to the other features in the same combination set. Since the dimensions of the Gabor features, Haar Wavelet coefficients and the boundary features from Morphological operations were very large, the dimensions of these features were reduced using Principal Component Analysis (PCA) before combining them. The feature models were normalized to be in the same range before performing the combination. The feature models were then cascaded to obtain three different feature combination models.

D. Classification for Expression Recognition

Multiclass Support Vector Machine (SVM) was implemented to classify the face expressions. Multiclass Support Vector Machine is a supervised Machine Learning algorithm. Multiclass classification refers to a classification task with more than two classes. Our project requires multiclass classification because 7 classes of facial expressions are involved namely, anger, contempt, disgust, fear, happy, sadness and surprise. In Multiclass classification each data point belongs to one of N different classes. The goal is to construct a function which, given a new data point, will correctly predict the class to which the new point belongs.

The type of Multiclass classification implemented was One-vs-All classification. In this technique N different binary classifiers are built. For the i^{th} classifier, the positive examples are considered to be all the data points in class i and the negative examples are all the data points not in class i.

The Support Vector Machines were introduced by Boser, Guyon and Vapnik in 1992 and has become rather popular since. It has been known to give an empirically good performance in many applications in the field of bioinformatics, text recognition and image recognition. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. For this reason SVMs are known as optimal margin classifier. An SVM classifies the data points by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. These concepts are best illustrated by Figure 11. The mathematical formulation of an SVM is viewed as an optimization problem for finding the optimal margin classifier such that the equation given below is satisfied.

$$min_{w,b}\frac{1}{2}||w||^2 \quad s.t.\ y^{(i)}(w^T x^{(i)} + b) \geq 1,$$

$i = 1, 2, \ldots, m$

where, m is the number of training examples.

The kernel function used in the SVM algorithm we implemented is Linear Kernel.

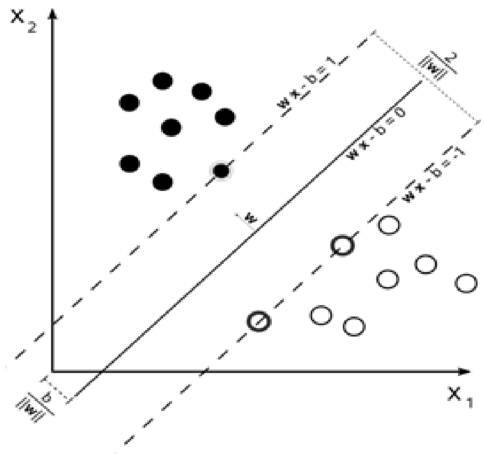

Figure 11: Illustration of Maximum margin hyperplane for an SVM

IV. EXPERIMENT AND RESULTS

All the programs were written and composed in Matlab 2015a (The Mathworks Inc. United States). The database used was Extended Cohn-Kanade (CK+) database obtained from the Affect Analysis Group of the University of Pittsburg. There are in total 593 sequences across 123 subjects. But, only 327 of the 593 sequences have been labelled with emotional facial expressions. All sequences are from the neutral face to the peak expression. The seven classes of Emotional Face Expression are given below

1 = Anger, 2 = Contempt, 3 = Disgust, 4 = Fear, 5 = Happy, 6 = Sadness, 7 = Surprise and it is assumed 0 = Neutral.

We have used only the 327 sequences which have been labelled with Emotions. In each sequence only the image containing the peak expression was considered for feature extraction.

A. Intermediate Results

The Face Detection was implemented using vision.CascadeObjectDetector of Matlab 20015a Computer Vision Toolbox. This function uses the Viola-Jones Algorithm to implement Face Detection. The face was detected successfully in all the face images of the database. The output of Face Detection for one of the face images in the database is shown in Figure 12.

Feature Extraction was performed once the face detection was completed. Figure 13 illustrates the output obtained from extracting the boundary pixels from morphological operations. Figure 14 depicts the landmark points chosen to enclose the face structure in the form of a wireframe mesh. Figure 15 shows an Eigenface obtained by projecting a mean shifted face image to the eigenspace. Figure 16 shows the response matrices of a cropped face image with the Gabor Filter Bank used. Figure 17 illustrates the Haar Wavelet pyramid on a face image. For extracting Haar Wavelet coefficients the Wavelet Toolbox of Matlab 2015a was used.

Figure 12: Output of Face detection on a sample image

Figure 13: Illustration of Boundary pixel extraction from Morphological operations

Figure 14: Landmark points on a Face image

Figure 15: Eigenface of a mean shifted face image

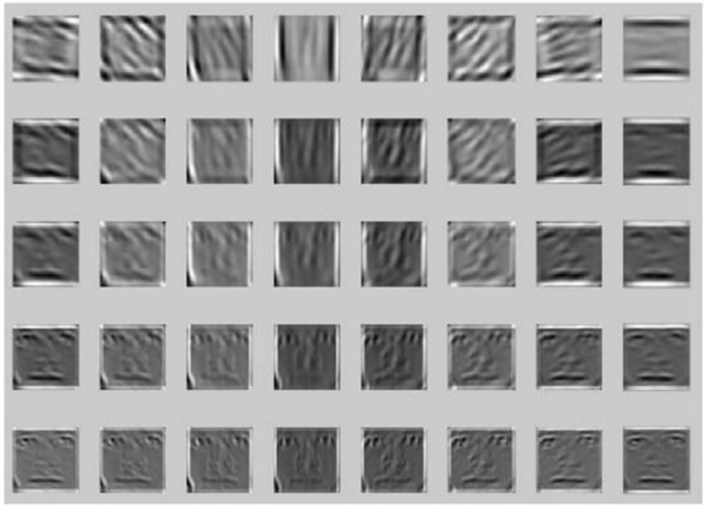

Figure 16: Gabor response matrices for a face image

Figure 17: Haar Wavelet Coefficient pyramid of a face image

B. Results of Classification

Single trial Multiclass SVM classification was performed using the Statistics and Machine Learning Toolbox of Matlab 2015a. Three SVM Classification models were obtained based on the three feature combination models that were computed. Each classification model consisted of 7 binary learners as there are 7 classes of emotional Face expressions involved. The 10 fold Cross Validation Error was computed as a measure of accuracy for the classification models constructed. Table 1 summarizes the results of 10 fold cross validation.

Feature Combination used	10 fold Cross Validation error
Gabor Features+ Haar Wavelets+ Boundary pixels from Morphological operations	8.56 %
Gabor Features + Haar Wavelets + Eigenfaces	20.18 %
Landmark points + Eigenfaces + Haar Wavelets	19. 27 %

Table 1: 10 fold Cross Validation Results

Figures 18, 19 and 20 show the confusion matrix obtained from the classification models of (a) Gabor features, Haar Wavelets and Boundary pixels from Morphological operations, (b) Gabor features, Haar Wavelets and Eigenfaces, (c) Landmark points , Eigenfaces and Haar Wavelets.

Figure 18: Confusion Matrix generated by Multiclass SVM Classification model based on Gabor+ Haar+ Boundary pixels on Morphological operations

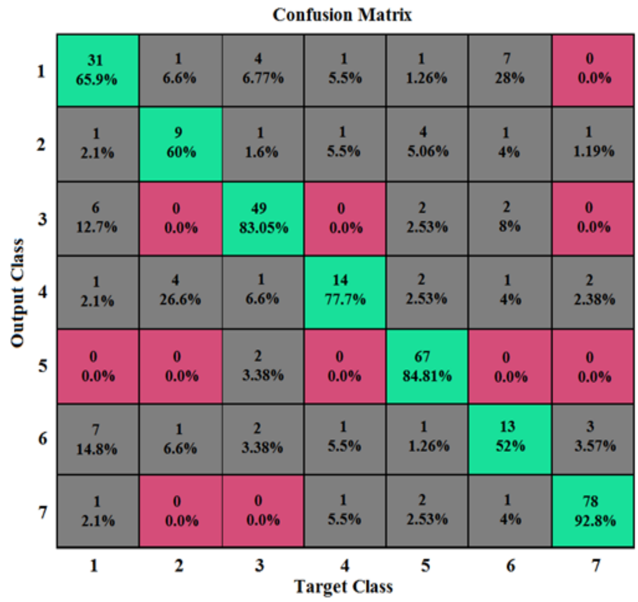

Figure 19: Confusion Matrix generated by Multiclass SVM Classification model based on Gabor+ Haar+ Eigenfaces

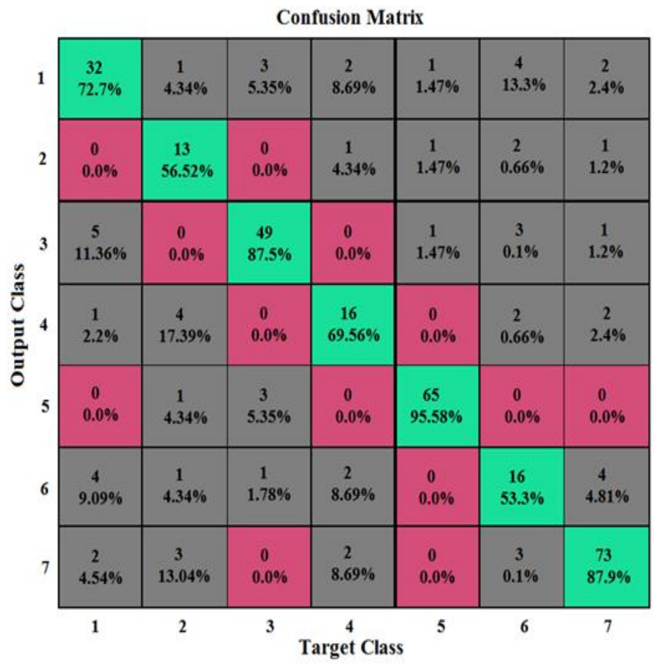

Figure 20: Confusion Matrix generated by Multiclass SVM Classification model based on Landmark points+ Eigenfaces+ Haar Wavelets

CONCLUSION

The objective of our project was to explore and implement new aspects into Face Expression Recognition. We explored the feature extraction of face images and combination of these feature models. The originality of our project comes from the feature combination models we constructed to obtain the multiclass SVM classification models. The classification model obtained from the feature combination of Haar Wavelets, Gabor features and the boundary points obtained from Morphological Operations is the most effective combination for classification of facial expressions as it gives the best performance accuracy of 91.44% compared to other classification models obtained from feature combinations and other classical existing techniques in [3], [4], and [5]. Table 2 gives a comparison of our classification models with other classic techniques of [3], [4], and [5].

Feature Extraction and Classifier	Performance/ Accuracy
Haar Wavelets + Gabor features + Boundary pixels from Morphological operations (Multiclass Linear SVM)	91.44 %
Haar Wavelets + Gabor features + Eigenfaces (Multiclass Linear SVM)	79.82 %
Landmark points + Eigenfaces + Haar Wavelets (Multiclass Linear SVM)	80.73%
Vector of feature displacements between neutral and peak expression (SVM) [3]	71.80 %
Gabor Features (Fischer Linear Discriminant Analysis) [4]	74.00 %
Vector of Motion Units using Face Tracker (Bayesian Network) [5]	77.70 %

REFERENCES

[1] K.-E. Ko and K.-B. Sim, "Development of the facial feature extraction and emotion recognition method based on ASM and Bayesian network", in Fuzzy Systems, 2009. FUZZ-IEEE 2009, IEEE International Conference on, Aug. 2009, pp. 2063 2066.

[2] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih,Z. Ambadar, I. Maththews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression",IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010.

[3] P. Michael and R. Kaliouby, "Real Time Facial Expression Recognition in Video Using Support Vector Machines", Proc. 5th Int. Conf. Multimodal Interfaces, Vancouver, BC, Canada, pp.258-264, 2003.

[4] I. Kotsia, I. Buciu and I. Pitas, "An Analysis of Facial Expression Recognition under Practical Facial Image Occlusion", Image and Vision Computing , vol. 26, no. 7, pp.1052-1067, July 2008.

[5] M.N. Dailey, G.W. Cottrell, "PCA = Gabor for Expression Recognition", UCSD Computer Science and Engineering Technical Report CS-629.

[6] I. Cohen, N. Sebe, F. Cozman, M.Cirelo and T.Huang, " Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data" , Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol.1, pp. [-595-]604, 2003.

[7] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art", Tech Report, arXiv:1203.6722, April 2012.

[8] P. Viola and M. Jones "Rapid object detection using a boosted cascade of simple features", CVPR 2001.

[9] M. Uricar, V. Franc and V. Hlavac, "Detector of Facial Landmarks Learned by the Structured Output SVM", VISAPP'12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, 2012

[10] Wang, Sung, "Frontal-view face detection and facial feature extraction using color and morphological operators," Pattern Recognition Letters, 1999.

[11] M.S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan, "Real Time Face Detection and Facial Expression Recognition : Development and Application to Human Computer Interaction", Proc. CVPR Workshop on Computer Vision and Pattern Recognition for human Computer Interaction, vol. 5, 2003.

[12] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," Expert Systems with Applications, vol. 42, no. 21, pp. 7905-7916, 2015.

[13] G. R. S. Murthy, R.S.Jadon "Effectiveness of Eigenspaces for Facial Expressions Recognition" International Journal of Computer Theory and Engineering, Vol. 1, No. 5, December, 2009, pp. 1793-8201.

[14] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol.3, No. 1, pp.71-86, Winter 1991.

[15] Corinna Cortes, Vladimir Vapnik, "Support Vector Networks" , Article, Machine Learning, Springer, Volume 20, Number 3, Page 273, 1995.

[16] G. R. S. Murthy, R.S.Jadon "Effectiveness of Eigenspaces for Facial Expressions Recognition" International Journal of Computer Theory and Engineering, Vol. 1, No. 5, December, 2009, pp. 1793-8201.

[17] I. Kotsia, I. Buciu and I. Pitas, "An Analysis of Facial Expression Recognition under Practical Facial Image Occlusion", Image and Vision Computing , vol. 26, no. 7, pp.1052-1067, July 2008.

[18] T. Kanade, J.F. Cohn, Y. Tian, "Comprehensive database for facial expression analysis", Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on.

[19] Gonzalez and Woods, "Digital Image Processing", 3rd Edition, Prentice Hall.