International Journal of Multimedia and Ubiquitous Engineering
Vol.12, No.11 (2017), pp.21-30
http://dx.doi.org/10.14257/ijmue.2017.12.11.03

Emotion Detection through Facial Feature Recognition

Maliha Asad [1], Syed Omer Gilani[1] and Mohsin Jamil[2,3]

[1] Department of Biomedical Engineering and Sciences
[2] Department of Robotics and Artificial Intelligence
School of Mechanical and Manufacture Engineering (SMME),
National University of Sciences and Technology (NUST), Islamabad, Pakistan
[3]Department of Electrical Engineering, Faculty of Engineering, Islamic University
Madinah, Saudia Arabia
maokhan7@gmail.com, omer@smme.nust.edu.pk and mohsin@smme.nust. edu.pk

Abstract

Humans share a universal and fundamental set of emotions which are exhibited through consistent facial expressions. An algorithm that performs detection, extraction, and evaluation of these facial expressions will allow for automatic recognition of human emotion in images and videos. Presented here is a hybrid feature extraction and facial expression recognition method that utilizes Viola-Jones cascade object detectors and Harris corner key-points to extract faces and facial features from images and uses principal component analysis, linear discriminant analysis, histogram-of-oriented-gradients (HOG) feature extraction, and support vector machines (SVM) to train a multi-class predictor for classifying the seven fundamental human facial expressions. The hybrid approach allows for quick initial classification via projection of a testing image onto a calculated eigenvector, of a basis that has been specifically calculated to emphasize the separation of a specific emotion from others. This initial step works well for five of the seven emotions which are easier to distinguish. If further prediction is needed, then the computationally slower HOG feature extraction is performed and a class prediction is made with a trained SVM. Reasonable accuracy is achieved with the predictor, dependent on the testing set and test emotions. Accuracy is 81% with contempt, a very difficult-to-distinguish emotion, included as a target emotion and the run-time of the hybrid approach is 20% faster than using the HOG approach exclusively.

1. Introduction

For humans, understanding and identifying emotions can be extremely interesting and useful, as genuine emotions are at most only partially controllable and often display their presence through facial expressions of the person experiencing them. A person's emotions can sometimes be very distinct and obvious and at other times may very transient and difficult to notice; however, as long as their cues are visually present, it ostensibly possible for a computer to perform image processing and classification of that expression. There are many applications, ranging from entertainment, social media, criminal justice, to healthcare where the automated ability to process and detect emotion of a person can have functional benefits. For example, content providers can judge a person's authentic and immediate emotional response and tune their product accordingly, or health tracking apps that would monitor emotional stability and fluctuation of a user.

Humans can quickly and even subconsciously assess a multitude of indicators such as word choices, voice inflections, and body language to discern the sentiments of others. This analytical ability likely stems from the fact that humans share a universal set of fundamental emotions. Significantly, these emotions are exhibited through facial

Received (July 5, 2017), Review Result (October 30, 2017), Accepted (November 20, 2017)

ISSN: 1975-0080 IJMUE

expressions that are consistently correspondent. This means that regardless of language and cultural barriers, there will always be a set of fundamental facial expressions that people assess and communicate with. After extensive research, it is now generally agreed that humans share seven facial expressions that reflect the experiencing of fundamental emotions. These fundamental emotions are anger, contempt, disgust, fear, happiness, sadness, and surprise [1] [2].

It is important for a detection approach, whether performed by a human or a computer, to have a taxonomic reference for identifying the seven target emotions. Figure 1 shows a computer based emotion detection by facial feature recognition.

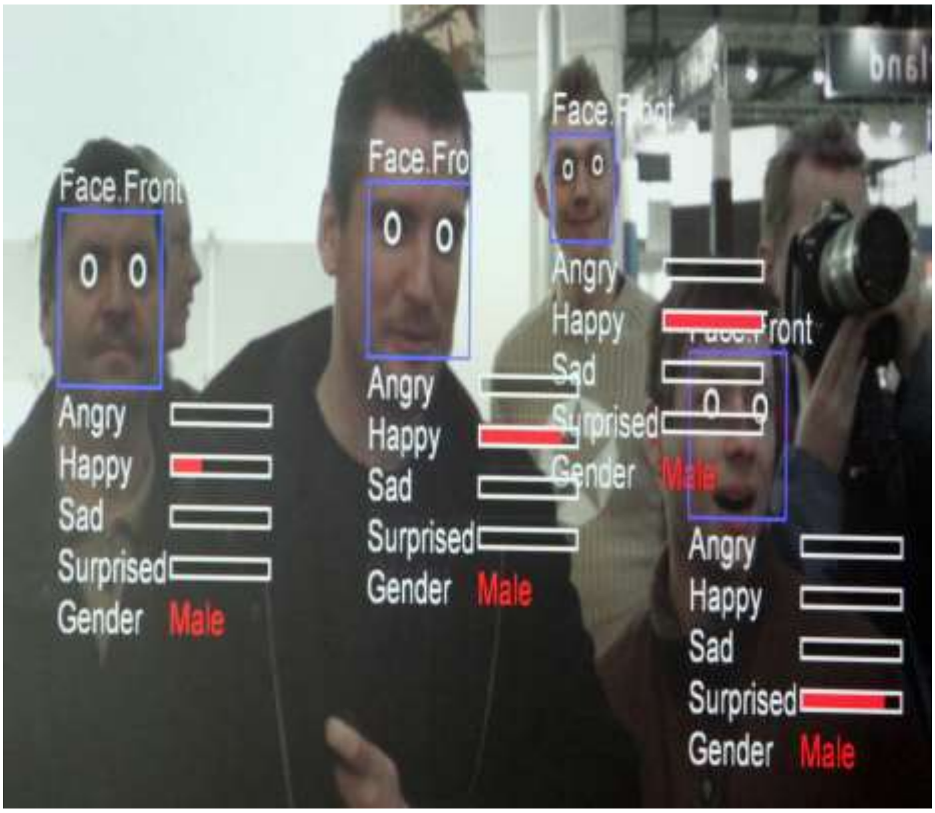

Figure 1. Computer Based Emotion Detection Shows Percentage of Each Emotion It Detects

A popular facial coding system, used both by noteworthy psychologists and computer scientists such as Ekman [1] and the Cohn- Kanade [3] group, respectively, is the Facial Action Coding System (FACS).

The system uses Action Units that describe movements of certain facial muscles and muscle groups to classify emotions. Action Units detail facial movement specifics such as the inner or the outer brow raising, or nostrils dilating, or the lips pulling or puckering, as well as optional intensity information for those movements. As FACS indicates discrete and discernible facial movements and manipulations in accordance to the emotions of interest, digital image processing and analysis of visual facial features can allow for successful facial expression predictors to be trained.

2. Methodology

The detection and recognition implementation proposed here is a supervised learning model that will use the oneversus- all (OVA) approach to train and predict the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise). The overall face extraction from the image is done first using a Viola-Jones cascade object face detector. The Viola- Jones detection framework seeks to identify faces or features of a face (or other objects) by using simple features known as Haar-like features. The process entails passing feature boxes over an image and computing the difference of summed pixel values between adjacent regions. The difference is then compared with a threshold which indicates whether an object is considered to be detected or not. This requires thresholds that have been trained in advance for different feature boxes and features. Specific feature boxes for facial features are used, with expectation that most faces and the features within it will meet general conditions. Essentially, in a feature-region of interest on the face it will generally hold that some areas will be lighter or darker than surrounding area. For example, it is likely that the nose is more illuminated than sides of the face directly adjacent, or brighter than the upper lip and nose bridge area. Then if an appropriate Haar-like feature, such as those shown in Figure 2, is used and the difference in pixel sum for the nose and the adjacent regions surpasses the threshold, a nose is identified. It is to be noted that Haar-like features are very simple and are therefore weak classifiers, requiring multiple passes.

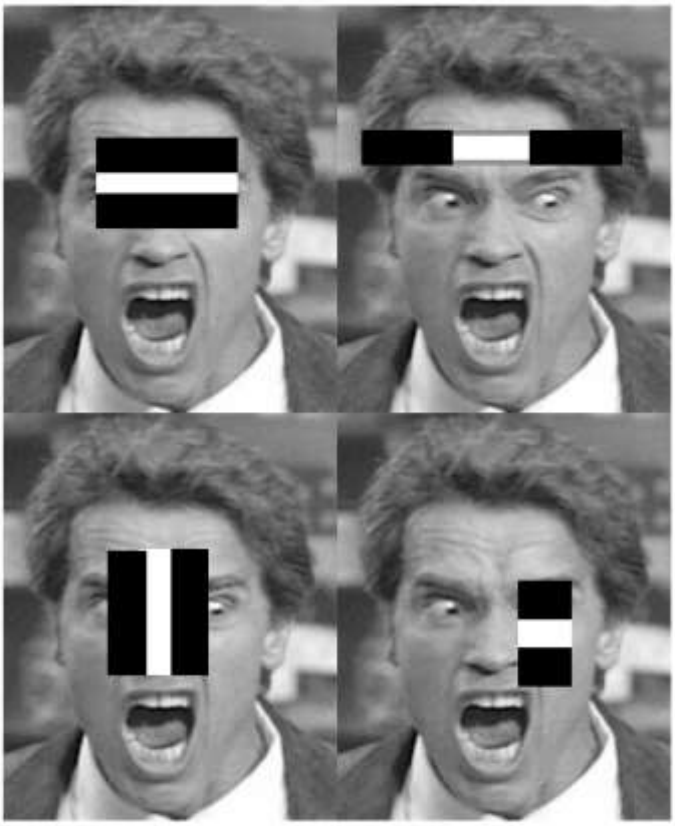

Figure 2. Computation of Haar Like Features on Example Face

However, the Haar-like feature approach is extremely fast, as it can compute the integral image of the image in question in a single pass and create a summed area table. Then, the summed values of the pixels in any rectangle in the original image can be determined using a total of just four values. This allows for the multiple passes of different features to be done quickly. For the face detection, a variety of features will be passed to detect certain parts of a face, if it were there. If enough thresholds are met, the face is detected.

Once the faces are detected, they are extracted and resized to a predetermined dimensional standard. As Zhang has shown that lower resolution (64x64) is adequate, we will resize the extracted faces to 100x100 pixels. This will reduce computational demand in performing the further analysis. Next, the mean image for all training faces will be calculated. The entire training set is comprised of faces from the Extended Cohn-Kanade [3] dataset, and comprises faces that express the basic emotions. The mean image is then subtracted from all images in the training set. Then using the mean-subtracted training set the scatter matrix **S** is formed. The intention is to determine a change in basis that will allow us to express our face data in a more optimized dimensionality. Doing so will allow the retention of most of the data as a linear combination of the much smaller dimension set. PCA accomplishes this by seeking to maximize the variance of the original data in the new basis. We perform PCA on the using the Sirovich and Kirby method, where the eigenvalues and eigenvectors of the matrix **SHS** are first computed to avoid computational difficulties. The eigenvectors of the scatter matrix, defined as **SS**H, can then be recovered by multiplying

the eigenvector matrix by **S.** Retaining the top eigenvectors, also known in this context as eigenfaces, allows us to project our training data onto the top eigenfaces, in this case the 100 associated with the top eigenvalues, in order to reduce dimensionality while successfully retaining most of the information. This allows us to proceed to the Fisher linear discriminant analysis (LDA) in a reduced dimensionality. For each emotion that we wish to train a predictor for, we will perform Fisher LDA, in which the goal is to optimize the objective function that minimizes within class variance and maximizes between class variance to gain clear class separation between the class of interest and the other classes. We then project all the training data used to calculate the Fisherface for each emotion onto that particular Fisherface. Binning the projection values into histograms to examine the distribution allows us to determine thresholds for each Fisherface's projection values. The Fisherfaces do reasonably in separating the classes for each emotion, as shown in Figure 3.

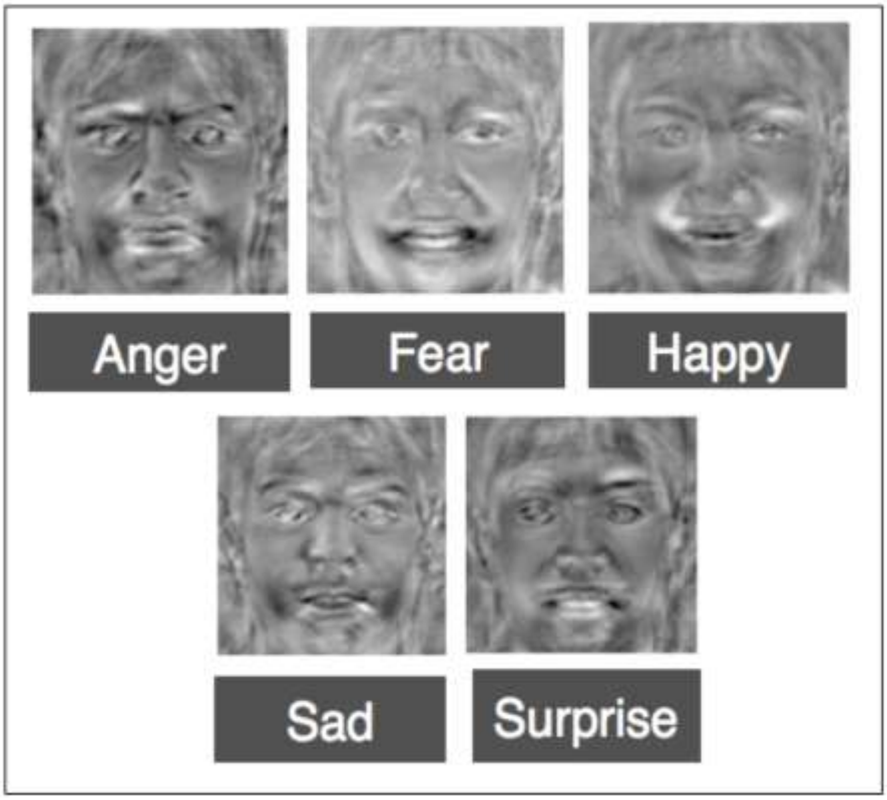

Figure 3. Top Eigen Faces (Fisher Faces) for Each Emotion (Resized to 100x100 Pixels)

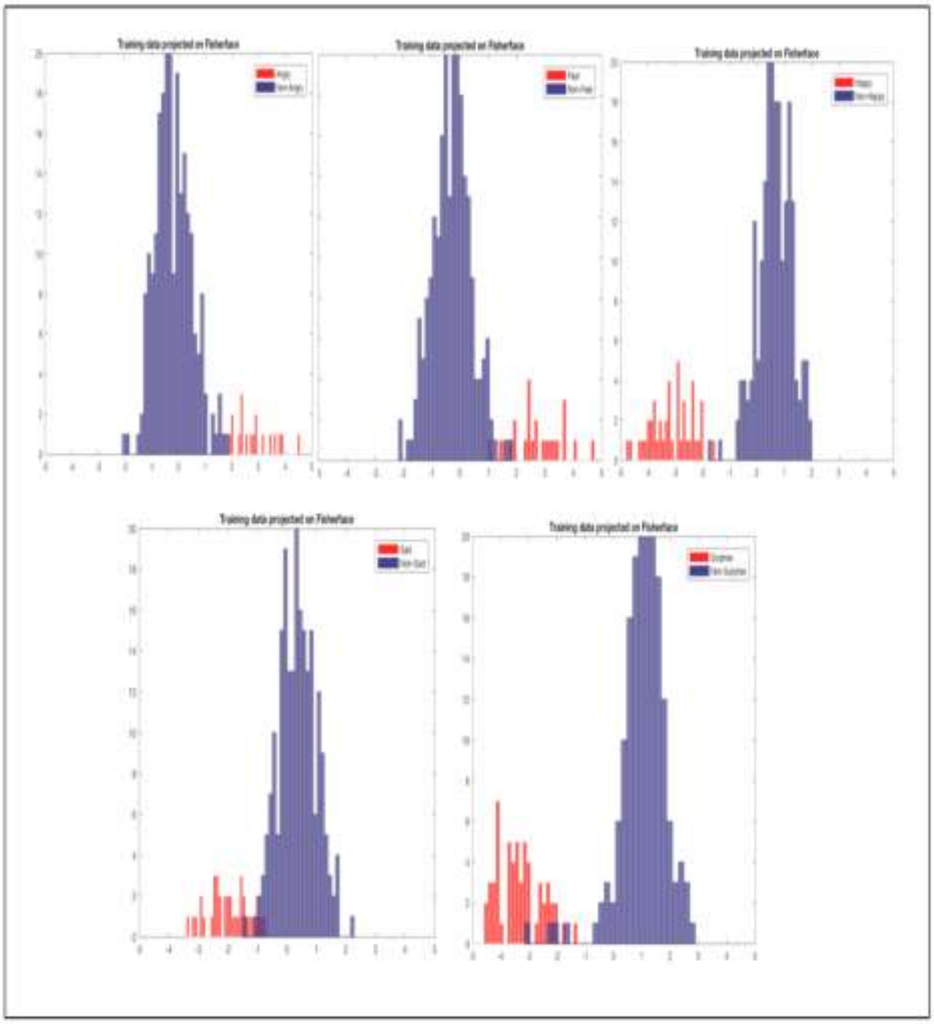

Figure 4. Distributions of Training Data Projected Back Onto Calculated Fisherfaces For Anger, Fear, Happy, Sad and Surprise. Distributions of Within-Class Shown In Red and Outside-Class Shown In Blue Are Relatively Well Separated

These Fisherfaces thresholds can then be used to classify test data that we have. We will detect and crop the test images in the same manner in which we did for the training images, and then project the test image onto each Fisherface. Then a classification prediction can be made based on the projection coefficient and the threshold we have established.

3. Results

Top 10 Eigen faces for all images

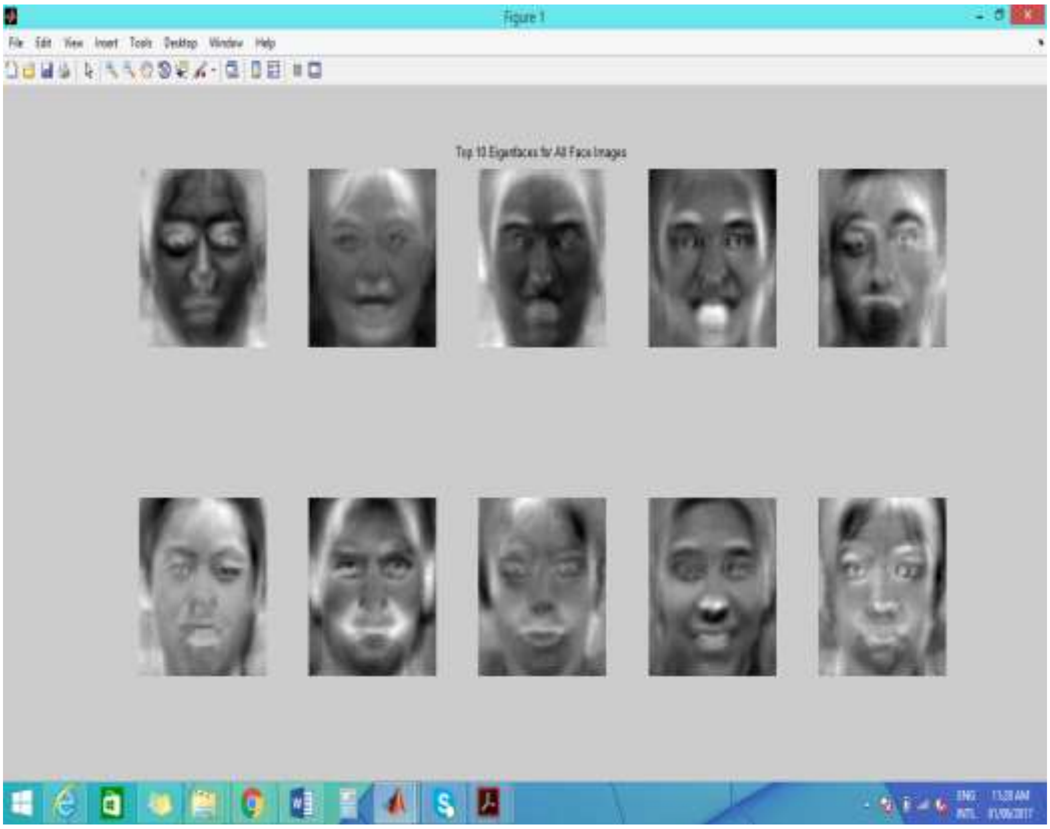
Figure 1
File Edit View Insert Tools Desktop Window Help
Top 10 Eigenfaces for All Face Images

Figure 5. Top 10 Results on All Images

We plan to develop another classifier in addition to our Fisherface based classifier since, as we find out experimentally, the Fisherface approach is limited in success by itself. We leverage the fact that most expression information is encoded within the inner facial features, specifically the regions around the eyes, nose, and mouth. As is detailed in FACS, the inner facial features will move in certain distinct combinations with the exhibition of each emotion, as is described by Action Units. Visually, these movements and manipulations should be evidenced in changes of gradients in the areas in the inner facial features. In particular, the brows and mouth, and how they visually warp, are very important the detection of emotions. We will utilize this information to train a classifier which can predict emotions based on the information encoded in the gradients. To begin, we must first extract the eye and mouth regions. We first try to detect these features separately using Haar-like features again. This approach is mostly successful. However, when it is not, perhaps due to illumination issues that affect the Haar-like feature calculations and the thresholding, we need another approach. Here we propose the use of Harris corner detection to detect features such as the eyes in a face image. The Harris corner detection method seeks to find points in an image that are corners by the definition that moving in any direction from that point should provide a gradient change. The approach is to use a sliding window to search for the corner points by examining gradient changes when sliding across that area. We use the fact that the eyes in a face image will be very nonuniform relative to the rest of the face. There white portion of the human’s eye is surrounded by skin that is darker, and the pupil and iris in the center of the eye is almost always darker as well. When viewing a face image with varying pixel intensities, some of the strongest corners are in the eye region. We use this fact to find eyes in a face when the Haar-like feature approach fails. Figure 6 gives an idea of the Harris corner extraction approach. We find the Harris corners on a cropped face image, then keep a number of the strongest corners. We then partition the face into vertical intervals and tally the number of Harris corners that fall in that vertical interval. The interval with the most Harris corners detected "wins the vote" and the eyes are determined to fall in that interval. From that information, the eyes are then extracted.

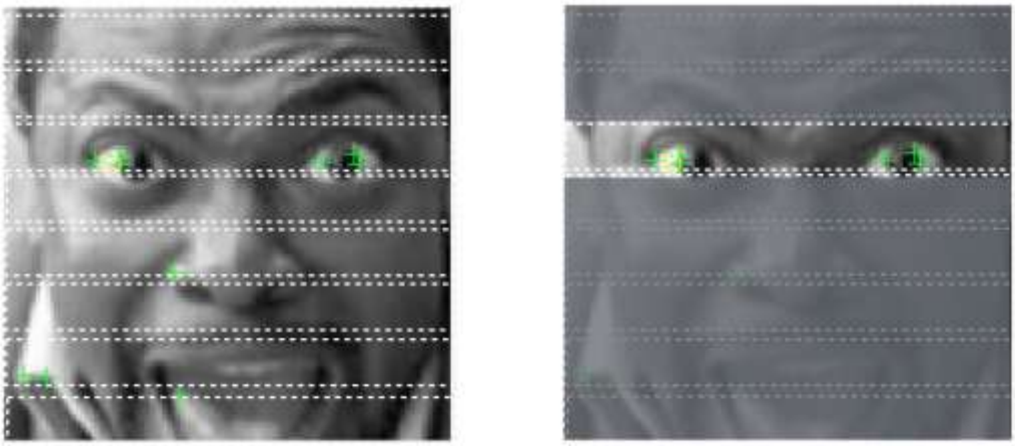

Figure 6. Harris Corner Detector On Faces

Harris corner approach for feature extraction, where the strongest corner points are shown as green crosses. The corner locations are tallied in vertical intervals and the interval in which the eyes reside is determined.

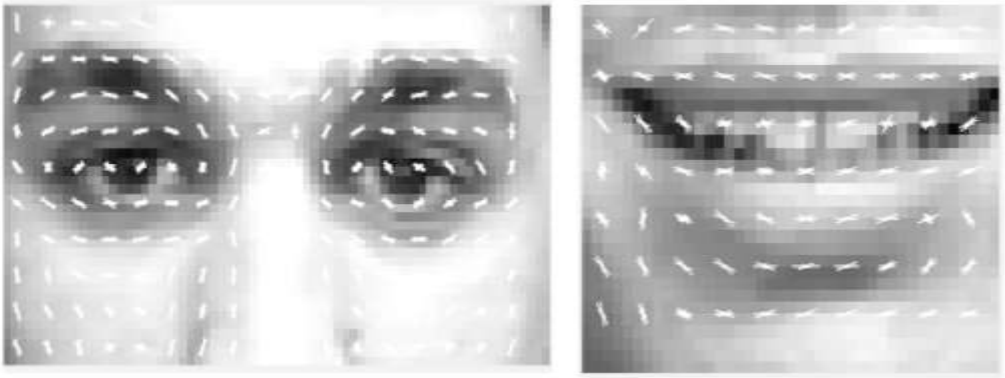

Figure 7. Visualization of HoG Feature

Plotted visualizations of HOG features on extracted eye and mouth regions. It should be expected then that facial expressions that have different muscular manipulations should result in varying HOG features. It should be noted that the extracted and resized eye and mouth regions must be consistent in dimension from image to image so we can extract the same number of HOG features, which is required for our further classifier training. We concatenate the extracted eye HOG vector with the mouth HOG vector for each training image, and assign a corresponding label. This, like the Fisher LDA process, requires us to know the class that each test image belongs to.

Upon completing HOG extraction for each image, we then train a mulit-class support vector machine (SVM) using the concatenated HOG vector.

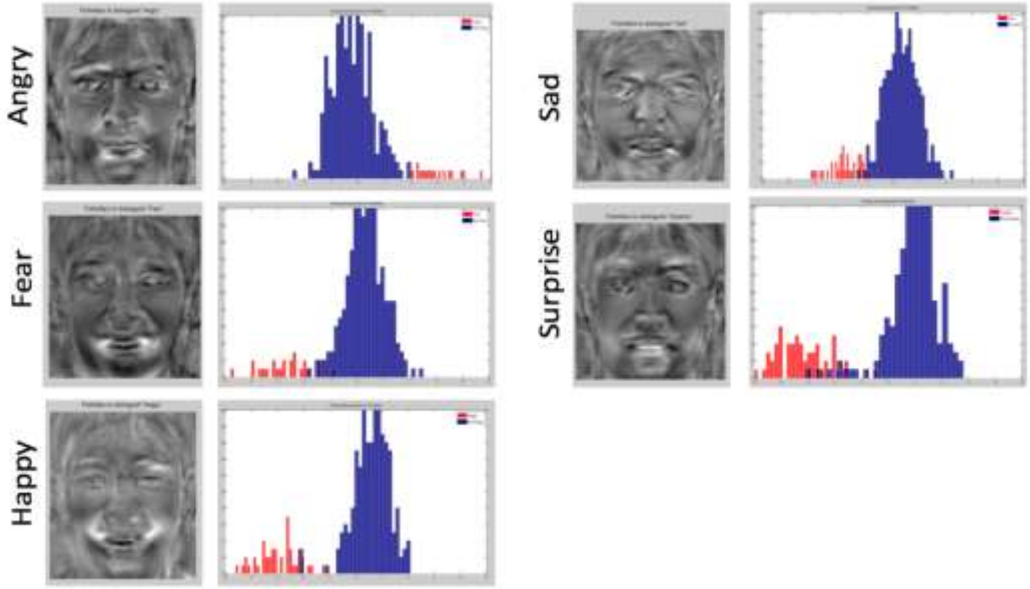

Figure 8. Fisher Faces to Distinguish Different Emotions and Their Projected Distributions

4. Discussion

The completed training implementation uses Viola-Jones's Haar-like feature cascade detector to detect faces as well as eyes and mouths. Detected faces are cropped, resized, and mean subtracted, then PCA is performed. Using the reduced dimensionality training dataset Fisher LDA is performed to extract Fisherfaces on which we can project test data. Also during training, eye and mouths are detected using Haar-like features, or using a Harris corner based approach is Haar-like features fail. The detected eye and mouth regions are then extracted and resized. HOG features are extracted from each region, and a SVM is trained using a combined eye-mouth HOG vector and training labels.

The primary reason we use this dual-classifier approach is improving speed with maintaining accuracy. When we use test images from the Extended Cohn-Kanade dataset and project those images onto our Fisherfaces for classification based on our established thresholds, we have an accuracy of 56%. This is a poor result, as it is only marginally better than random guessing.

Upon further investigation, this is due to the Fisherface-approach's inability to effectively detect the expressions corresponding to disgust and contempt. However, when only detecting expressions of test images that correspond to anger, fear, happiness, sadness, and surprise, the Fisherface approach is more than 90% accurate.

References

[1] P. Ekman and D. Keltner, "Universal facial expressions of emotion: An old controversy and new findings", In Segerstråle, U. C. & Molnár, P. (Eds.), Nonverbal communication: Where nature meets culture Mahwah, NJ: Lawrence Erlbaum Associates, **(1997)**, pp. 27-46.

[2] D. Matsumoto and C. Kupperbusch, "Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience", Asian Journal of Social Psychology, vol. 4, **(2001)**, pp. 113-131.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy", in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, **(1963)**, pp. 271-350.

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

[5] Z. Zhang, “Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perception”, International Journal of Patten Recognition and Artificial Intelligence, vol. 13, no. 6, **(1999)**, pp. 893-911.

[6] M. J. Lyons, S. Akemastu, M. Kamachi and J. Gyoba, “Coding Facial Expressions with Gabor Wavelets”, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, **(1998)**, pp. 200-205.

[7] C. Shan, S. Gong and PW. McOwan, “Facial expression recognition based on local binary patterns: a comprehensive study”, Image Vis Comput., vol. 27, no. 6, **(2009)**, pp. 803–816.

[8] P. Carcagnì, M. Del Coco, M. Leo and C. Distante, “Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus”, vol. 4, **(2015)**, p. 645.

[9] R. Hari, C. P. Roopesh and M. Wilscy, “Human face based approach for video summarization”, 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS), **(2013)**, pp. 245-250.

[10] B. Chakraborty, “A Novel ANN based Approach for Angle Invariant Face Verification”, Computational Intelligence in Image and Signal Processing 2007. CIISP 2007. IEEE Symposium on, **(2007)**, pp. 72-76.

[11] M. Perera, T. Shiratori, S. Kudoh, A. Nakazawa and K. Ikeuchi, “Multilinear analysis for task recognition and person identification”, Intelligent Robots and Systems 2007. IROS 2007. IEEE/RSJ International Conference on, **(2007)**, pp. 1409-1415.

[12] N. Radji, D. Cherifi and A. Azrar, “Importance of eyes and eyebrows for face recognition system”, Control Engineering & Information Technology (CEIT) 2015 3rd International Conference on, **(2015)**, pp. 1-6.

[13] H. Wang, S. Yan, T. Huang, Ji. Liu and X. Tang, “Misalignment-robust face recognition”, Computer Vision and Pattern Recognition 2008. CVPR 2008. IEEE Conference on, ISSN 1063-6919, **(2008)**, pp. 1-6.

[14] S. M. Sabbir Hossain, A. Yousuf and M. Sheikh Sadi, “Towards an efficient face recognition approach”, Electrical Engineering and Information Communication Technology (ICEEICT) 2015 International Conference on, **(2015)**, pp. 1-5.

[15] L. Torres, J.Y. Reutter and L. Lorente, “The importance of the color information in face recognition”, Image Processing 1999. ICIP 99. Proceedings. 1999 International Conference on, vol. 3, **(1999)**, pp. 627-631.

[16] G. Givens, J.R. Beveridge, B.A. Draper, P. Grother and P.J. Phillips, “How features of the human face affect recognition: a statistical comparison of three face recognition algorithms”, Computer Vision and Pattern Recognition 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, ISSN 1063-6919, vol. 2, **(2004)**, pp. II-II.