Emotion Detection through Facial Feature Recognition

Wasim Ahmed Syed[1], Arun Kumar Uppalapti[2], Amruth Vadakattu[3], Akhil Vasireddy[4], Purna Chandra Reddy[5]

[1,2,3,4]Under Graduate Student, Dept. of Electronics and communication engineering, VVIT, JNTU Kakinada, Andhra Pradesh, India
[5]Asst.Prof, Dept. of Electronics and communication engineering, VVIT, JNTU Kakinada, Andhra Pradesh, India

***

Abstract – *People share a common and fundamental set of emotions which exhibited through consistent facial expressions. Detection, extraction, and evaluation of these facial expressions are being performed by an algorithm which will allow for automatic recognition of human emotion in images. Extraction of faces and facial features from images are done by using Viola-jones cascade object detectors and Harris corner key points and uses principal component analysis(PCA), linear discriminant analysis(LDA), histogram-of-oriented-gradients (HOG) feature extraction, and support vector machines (SVM) in order to train a multi-class predictor for classifying the seven-basic human facial expressions. This approach allows for initial classification via projection of a testing image onto a calculated eigenvector, of a basis that has been specifically calculated to focus attention on the separation of a specific emotion from others. This initial step works well for five of the seven emotions which are easier to distinguish. For further prediction computationally, slower HOG feature extraction is performed and a class prediction is made with a trained SVM. Depending on the testing set and test emotions reasonable accuracy is achieved with the predictor. Contempt a very difficult-to-distinguish emotion, is achieved included as a target emotion and the run-time of the hybrid approach is 20% faster than using the HOG approach exclusively.*

***Key Words*: Viola-Jones cascade object detector, Harris corner key points, Eigen vector, SVM, LDA, HOG.**

1. INTRODUCTION

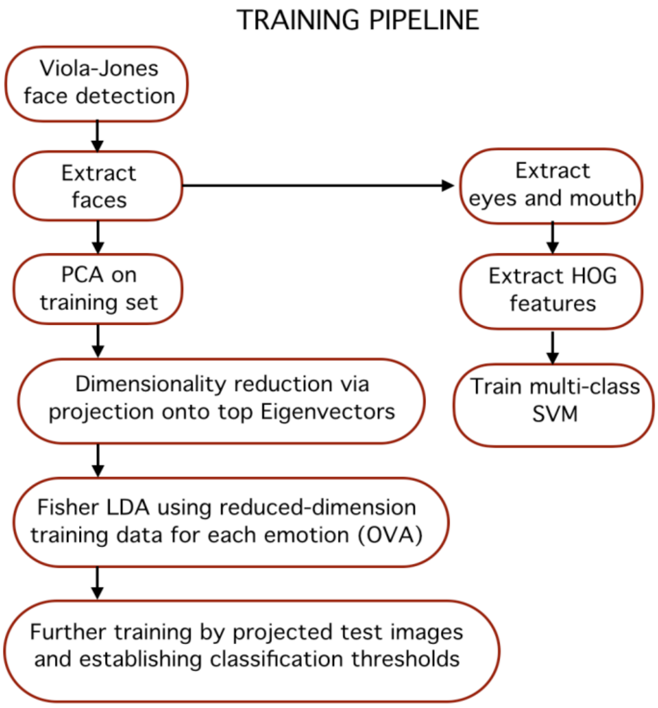
Interpersonal interaction success is often predicated upon a variety of factors as it is intricate. These factors include the context, mood, and timing of the interaction which range widely, as well as the expectations of the participants. To be a successful participant, one must perceive a counterpart's disposition as the interaction progresses and adjust accordingly. For humans this ability is largely innate, with varying levels of proficiency. Humans can quickly and even subconsciously assess a multitude of indicators such as body language, voice inflections, word choices to know the sentiments of others. This analytical significance is from the fact that universal set of fundamental emotions shared by humans. The facial expressions exhibited through emotions are correspondent. Irrespective of culture and language barriers, fundamental set of facial expressions which are shared by people and communicate with. Extensive research, says that seven fundamental facial expressions shared by humans reflects the experiencing of fundamental emotions. Anger, contempt, disgust, fear, happiness, sadness, and surprise are fundamental emotions [1][2]. Unless a person wontedly suppresses their expressions, examining a face of a person can be effective method of discerning their genuine mood and reactions. The recognition of emotions is universal task that can be accomplished by computers also. Similar to other important computers are advantageous over humans in problem-solving and analysis. Computers recognize facial expressions can find application in entertainment, social, criminal justice, content analysis, healthcare where efficiency and automation can be useful. By determining the reactions of a consumer, content providers will adjust their future offerings accordingly. For a detection approach one should note, it is performed by a human or a computer, to have a taxonomic reference for identifying the seven-target emotion. Facial Action Coding System (FACS) a popular facial coding system used by renowned psychologists and computer scientists namely Ekman [1] and Cohn Kanade [3] group. To classify emotions certain facial muscles and muscle group movements are used by this system. Facial movement specifics such as the inner or the outer brow raising, or nostrils dilating, or the lips pulling or puckering, as well as as optional intensity information for those movements are detailed in these action units. FACS indicate discrete and discernible facial movements and changes in accordance to the emotions of interest, digital image processing and visual facial features analysis can allow for successful facial expression predictors to be trained.

1.1 Related Work

For detection, extraction, and recognition of human facial features and expressions there have been many approaches in using computers, as this topic is of heed in many fields covering both social sciences and engineering. For instance, geometric positions of facial fiducial points as well as Gabor wavelet coefficients at the same points are detailed by Zhang [4] to perform recognition based on a two-layer perceptron. Because of low frequency nature of expression information Facial expression detection is achievable with low resolution and it is shown by Zhang significantly. Most of the useful expression information which is encoded within the inner facial features is also shown by Zhang. This allows facial expression recognition to perform with relatively less computational requirements and that too in a successful way. Congregation of methods are utilized to perform the feature extraction task, and the subsequent characterization. Gabor transforms coupled with neural networks is a general approach as well as popular which is similar to Zhang's approach. Shan's [6] local binary patterns, Carcagni's[7] histogram of oriented gradients, Lucey's [3] facial landmarks with Active Appearance Modelling are some other extraction methods that have been used. Learning models are often used to perform classification such as support vector machines.

2. METHODOLOGY

This is a supervised learning model that will use the one versus- all (OVA) approach to train and predict the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise) for detection and recognition. Viola-Jones cascade object face detector is used for identifying Face from the image which is used to detect the features from the face by using Haar-like features. The process involves 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. For different feature boxes and features threshold have been trained in advance. Most faces and the features within special feature box used for facial features 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,



that the nose is more highlighted than sides of the face adjacent, or brighter than the upper lip and nose bridge area. Then if an appropriate Haar-like feature, such as those shown in Figure 1, is used and the difference in pixel sum for the nose and the adjacent regions surpasses the threshold, nose is identified. Haar-like features are very simple and weak classifiers, which requires multiple passes.

FIG 1: Sample Haar-like features for detecting face features

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, which allows for the multiple passes of different features to be done quickly. A variety of features will be passed to detect certain parts of a face, if it were there. The face is detected after enough thresholds are met then face is detected and resized to predetermined level. As Zhang has shown that lower resolution (64x64) is adequate, we will resize the extracted faces to 100x100 pixels which reduces computational demand in performing the further analysis. Later, the mean image for 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. 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 SSH, 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 100associated with the top eigenvalues, in order to reduce dimensionality while successfully retaining most of the information.

For each emotion, we will proceed with one versus all for performing LDA where all nontarget emotion training samples will be grouped. Then, we perform PCA once again on *SW-1SB.* The eigenvector corresponding to the largest eigenvalue is the known as the Fisher face for the emotion in-training, some of which are shown in Figure 2 . Fisher face for each emotion is calculated by projecting all the training dataonto that particular Fisher face. binning the projection values into histograms to examine the distribution allows us to determine thresholds for each Fisher face's projection values. The Fisher faces do reasonably in separating the classes for each emotion, as shown in Figure3.

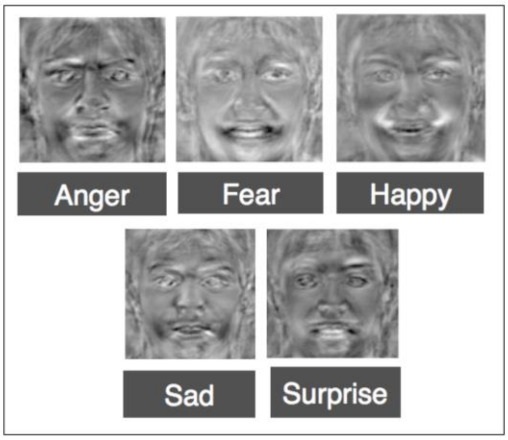


Fig. 2 Top eigenvectors reshaped to 100x100 images (Fisher faces) after Fisher LDA for Anger, Fear, Happy, Sad and Surprise.

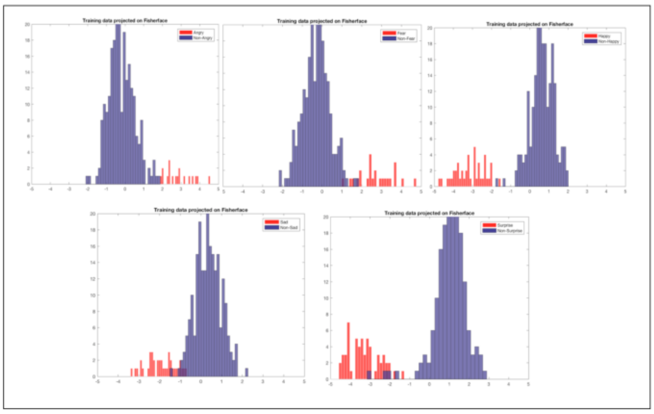


Fig. 3 Distributions of training data projected back onto calculated Fisherfaces for Anger, Fear, Happy, Sad and Surprise. Distributions of within-class shown in red and outside-class shown in blue are relatively well separated

3. RESULTS

Viola-Jones's Haar-like feature cascade detector was used by completed training implementation to detect faces as well as eyes and mouths. PCA was performed after the detected faces were cropped, resized, and mean subtracted. Fisher LDA is performed to extract faces by using the reduced dimensionality training on which we can project test data. Also during training, Haar-like features are used to detect eyes and mouths, or using a Harris corner based approach is Haar-like features fail. Extraction and resizing of the detected eye and mouth regions are performed. HOG features are extracted from each region, and by using a combined eye-mouth HOG vector and training labels a SVM is trained. The foremost reason why we use this dual-classifier approach is refining speed with maintaining accuracy and precision. We can achieve an accuracy of 56% when we use test images from the Extended Cohn-Kanade dataset and by projecting those images onto our Fisher faces for organisation based on our initiated thresholds. It is only minutely better than random guessing, hence it is a poor result. Upon more extreme investigation, this is due to the Fisher face-approach's impotence to explicitly detect the expressions corresponding to abhorrence and contumacy. However, Fisher face approach is more than 90% accurate when only detecting expressions of test images that congruous to anger, fear, happiness, sadness, and surprise.

Fisher face approach's accuracy is 56%, but when using the HOG and SVM classifier only, the accuracy for detection is 81% and it is much better than a Fisher face approach. The accuracy of dual-classifier method is same as the HOG approach and it is 81%, but the testing process of dual-classifier method is 20% faster than HOG approach. This is because only those test images that are not given a prophecy by the much faster Fisher face classifier but not all images must undergo eye and mouth detection, extraction, and then undergo HOG feature extraction.

Performed on test set of 32 images ck+

Algorithm	**ACC**	**Runtime**
Fisher face only	56%	7.40
HOG only	81%	9.87
Fisher face +HOG	81%	7.91

Testing results for classifiers

4. CONCLUSIONS

Face images are used to train a dual classifier predictor which predicts the seven basic human emotions given a test image and it is done by an image processing an classification method. At predicting test data from the same dataset used to train the classifiers the predictor is always comparatively successful. However, at detecting the expression associated with derision the predictor is consistently poor. Contempt, poor pre-training labelling of data, and the intrinsic difficulty at identifying contempt is clearly exhibited by the combination of lacking training and test images. In predicting emotions for test data that have expressions that do not clearly belong exclusively to one of the seven basic expressions the classifier is also not successful, as it has not been trained for other expressions. From different datasets, adding training images investigating more precision methods that still maintain computational coherence, and considering the classification of more nuanced, complex and complicated expressions we will improve the sturdy of the classifiers and future work should demand for improving this robustness of the classifiers.

REFERENCES

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

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

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

[4] Zhang, Z. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. International Journal of Patten Recognition and Artificial Intelligence 13 (6):893-911 (1999).

[5] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998)

[6] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput. 27(6):803–816 (2009)

[7] Carcagnì P, Del Coco M, Leo M, Distante C. Facial expression recognition and histograms of oriented gradients: a comprehensive study. Springer Plus. 4:645. (2015)