Facial feature Extraction and Emotional Analysis

Group Member Names

18BCE0022- L Ruthvik Raj

18BCE0026- Bitra HarshaVardhan

18BCE0011- M.Manikanta Sahith

Faculty Name

Dr. CHIRANJI LAL CHOWDHARY

Associate Professor, CITE, VIT Vellore

Winter Semester

December 2019 - June 2020

This the video link of our project presentation and execution –

https://drive.google.com/file/d/10Bfe_LiM34etZJsplbYhjFeT_RxaQOJh/view?usp=sharing

SYNOPSIS

In the pattern recognition domain, deep learning algorithms are currently widely used and they have achieved fine results. However, these deep learning algorithms make particular demands, especially in terms of their requirement for big datasets and GPU. Aiming to gain better results without deep learning, we propose a simplified algorithm using fusion features extracted from the salient areas of faces. Furthermore, the proposed algorithm has achieved a better result than some deep learning algorithms.

Through coherent facial expressions, humans share a universal and basic set of feelings. An algorithm that detects, extracts and evaluates these facial expressions will enable the automatic recognition of human emotions in pictures and videos. Here is a hybrid feature extraction and facial expression identification technique that uses Viola-Jones cascade object detectors and Harris corner key points to obtain faces and facial characteristics from pictures and uses main component analysis, linear discriminating analysis, histogram-of-oriented gradients (HOG) extraction. The hybrid method enables for rapid original classification through the projection of a test picture onto a calculated individual vector, a basis specifically calculated to emphasize the separation of a particular emotion from others. This original phase operates well for five of the seven feelings that are more distinguishable. If further prediction is required, the computationally slower extraction of the HOG function will be carried out and a class prediction will be created using a qualified SVM. The predictor achieves reasonable precision, depending on the test set and test feelings. Accuracy with disdain is 81 percent, a very difficult-to-distinguish emotion, included as a target emotion, and the runtime of the hybrid strategy is 20 percent quicker than using the HOG strategy alone.

The most widely used classification algorithms are fisher’s algorithm and PCA. But classification algorithms like Principal Component Analysis and Fisher’s linear discriminant analysis suffer from illumination effects, the amount of light in an image severely affects their performance. While the more advanced algorithms like HOG take some initial time to compute gradients from the image making the overall performance of the software very slow. Using HOG algorithm is the most accurate way of classification but fisher algorithm is faster.

These basic emotions are anger, contempt, disgust, fear, happiness, sorrow, and surprise. The universality of these expressions implies that recognition of facial emotions is a job that computers can also accomplish. In addition, computers, like many other significant duties, can provide benefits in analyzing and problemsolving over humans. Computers that can acknowledge facial expressions can discover applications where effectiveness and automation can be helpful, including in entertainment, social media, content analysis, criminal justice, and health care.

Keywords: Support Vector Machines Classification; Histogram of Oriented Gradients; Facial expression recognition; Viola-Jones Cascade object detectors Image Processing; Machine Learning; Computer Vision

REVIEW 0

In review 0 we submitted our projected title and the members involved in the project. We gave a brief description our of project. Our project is **Facial feature extraction and emotional analysis,** we explained what all we are going to implement in this project and what the is basic aim of our project. We also gave a brief idea about what all we are going to use to make the project run what our output will be and how it will be better than the previous version and the improvements which we are going to bring in it to get a better working of the project .

REVIEW- 1(2)

In our Review 1 we gave the background of our project and gave an explanation of our methods that we are going to use and what output we are expecting, we also presented it with our literature survey. This is how we presented our review-

(a) <u>Background:</u>

Through facial expressions, people share an all-inclusive and fundamental arrangement of sentiments and feelings. An algorithm that identifies, removes and assesses these outward appearances will empower the programmed acknowledgment of human feelings in pictures and videos. Here is a hybrid feature extraction and facial expression identification technique that uses Viola-Jones cascade object indicators and Harris corner key focuses to acquire faces and facial attributes from pictures and uses main component analysis, linear discrimination analysis, histogram-of-oriented gradients (HOG) extraction. The most widely used classification algorithms are fisher's algorithm and PCA

(b) <u>Objective:</u>

We chose SVM to acknowledge faces as its outstanding performance in solving inseparable linear problem. SVM may discover an optional separating hyperplane, which makes the training sample range close to maximization. The objective of SVM is to minimize empirical danger and confidence intervals in order to obtain excellent statistical sample laws and enhance the generalization capacity of machine learning. For linear inseparable problems, SVM maps input into a higher dimensional feature space

(a) <u>Expected results:</u>

In the first step using the Voila Jones algorithm we detect the faces in the dataset provided and crop the images to size 100x100 for Principal Component analysis. Once we have the cropped dataset, we run the principal component analysis on the images to obtain the top ten Eigen faces or in other words we project the images data to a 10-dimensional space using 10 feature vectors. Once we have the data points from all the images, we now must find a linear separation between the different classes which is achieved with the help of Fisher's LDA. Once both the models have been trained, we run the test script to test the images and we record the time taken by each of the algorithms including our own algorithm

Finally, we also gave literature survey for 7 research papers and explained them properly in our review along with the above contents.

REVIEW-3

TABLE OF CONTENTS

INTRODUCTION

Motivation

Facial expression plays an important role in our daily communication with other people. For the development of intelligent robots, especially indoor mobile robots, emotional interactions between robots and humans are the foundational functions of these intelligent robots. With automated facial expression recognition technology, these home service robots can talk to children and take care of older generations. Also, this technology can help doctors to monitor patients, which will save hospitals much time and money. In addition, facial expression technology can be applied in a car to identify whether the driver has fatigue, and this can save many lives. Facial expression recognition is needed because many situations need this technology.

Due to its diverse facial expressions, poses and illumination, face recognition is a difficult area in computer vision and pattern recognition. Face recognition and identification of facial expression is very important in fields such as access management, human-computer communication, manufacturing control, elearning, fatigue driving recognition and emotional robot. Facial expression technique is classified to sequence video or static picture based on the input information.

Contribution

The project is aimed to extract the features from the offline facial image dataset and perform Support Vector Machines Classification and Histogram of Oriented Gradients, to extract the emotions. Presented in this project is hybrid feature extraction and facial expression recognition method with a dual classifier approach that uses Viola-Jones Cascade object detectors and Harris key-points to extract features from the images. We are interested in seven fundamental emotions such as Anger, fear, Happiness, Disgust, Sadness, Surprise and Contempt. The project covers disciplines such as Image Processing, Machine Learning and Computer Vision.

LITERATURE SURVEY

Reference	Methods Used	Evaluation	Merits and Demerits
Cai, J., Meng, Z., Khan, A. S., Li, Z., O'Reilly, J., & Tong, Y. (2018, May). Island loss for learning discriminative features in facial expression recognition. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 302-309). IEEE.	Experimental findings on four benchmark expression databases have shown that the CNN with the suggested island loss (IL-CNN) outperforms either traditional SoftMax loss or center loss model CNN and achieves similar or better efficiency relative to state-of-the-art facial expression identification techniques.	Finally, to generate the distribution over the target phrases, a SoftMax loss is calculated on the choice layer. In the CNN training, the island loss and the SoftMax loss are minimized jointly to drive the finetuning process.	For the recognition of facial expression, a CNN with the island loss (IL-CNN) is created to show the efficacy of the suggested island loss. The architecture of IL-CNN involves three convolution layers, each followed by a layer of PReLU and a layer of batch standardization (BN). After each of the first two BN layers, a max pooling layer is used.

Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 115, 101-106.	Deep Neural Networks (DNNs) outperform traditional models in numerous optical recognition missions that include Facial Expression Recognition (FER), an imperative process for clinical practice and behavioural description in the next generation Human-Machine Interaction (HMI).	The proposed network architecture comprises of Convolution layers followed by Recurrent Neural Network (RNN), which extracts the relationships within facial pictures and can be regarded during the classification by using the recurring network of the temporal dependencies that occur in the pictures.	Existing FER techniques are not highly accurate and in real-time applications are not practical enough. Thus, propose a technique for FER in Images for a Hybrid Convolution-Recurrent Neural Network.

Tsai, H. H., & Chang, Y. C. (2018). Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Computing, 22(13), 43894405.	**Facial expression recognition (FER)** is a technique based on support vector machine (SVM) for the FER. Three schemes, the **Angular radial transform (ART)**, the **Discrete cosine transform** (DCT) and the **Gabor filter** (GF), are simultaneously employed in the design of the feature extraction for facial expression in the FERS technique	As a result, the FERS technique possesses better detection rate because the face detection method gets more accurate in locating face regions of an image. The main reason is that the SQI filter can overcome the insufficient light and shade light	More specifically, they are employed in constructing a set of training patterns for the training of an SVM. The FERS technique then exploits the trained SVM to recognize the facial expression for a query face image.
Xie, S., & Hu, H. (2018). Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Transactions on Multimedia, 21(1), 211220.	A deep-based framework, consisting primarily of two CNN branches. One branch extract local characteristic from picture patches while the other extracts holistic characteristics from the entire expressive picture.	These two kinds of hierarchical characteristics constitute different-scale expressions. Compared with most current methods with single type of feature, the model can represent expressions more comprehensively.	Additionally, in the training stage, a novel pooling strategy named expressional transformation-invariant pooling is proposed for handling nuisance variations, such as rotations, noises, etc.
Xue, M., Mian, A., Duan, X., & Liu, W. (2019, May). Learning Interpretable	Different facial elements contain	Firstly, **spatialtemporal**	The expression-sensitive features from the

Expression-sensitive Features for 3D Dynamic Facial Expression Recognition. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-7). IEEE.	distinct amounts of data to be transmitted for recognition of 3D dynamic expression. It is therefore essential for the recognition of discriminative facial expression to identify facial elements that are extremely important to modifications in expression	**features (HOG3D)** are extracted from local depth patchsequences to represent facial expression dynamics. A twophase feature selection process is then proposed to determine the facial components that can best distinguish the expressions.	corresponding area are fed into a hierarchical classifier for facial expression recognition. Additionally, the resulting HOG3D features after feature selection can be used to generate semantic interpretation of the expression dynamics.
Revina, I. M., & Emmanuel, W. S. (2018). Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/j.jksuci.2018.03.015.	This paper uses the EMDBUTMF method to remove the noise in the face image. The given face expression acknowledgment method is tolerant to the distinctive kind of face pictures.	According to the matching recognition analysis, Face expression recognition analysis, FER time analysis, and average time analysis, the proposed method is better than the existing methods.	In future, the face expression recognition ratio can be improved by another face descriptor Fast Representation using a Double Orientation Histogram (FRDOH). Another similarity method using Convolution Neural Network (CNN) can be combined with the existing Support Vector Machine classifier.
Utami, P., Hartanto, R., & Soesanti, I. (2019). A Study on Facial Expression Recognition in Assessing Teaching Skills: Datasets and Methods. *Procedia Computer Science*, *161*, 544–552. doi: 10.1016/j.procs.2019.11.154	A dataset that has an enormous possibility of being utilized as practically identical information with a current calculation is CK +.	Technique related issues that can be taken from a few past investigations: highlights change is utilized to defeat the low goals picture, a restrictive probabilistic figuring out how	The utilization of profound learning requires high preparing gadget determinations. Along these lines, further investigations are required with respect to the arrangement so its usage can be gotten to by clients utilizing a low determination preparing gadget

		to conquer head present variety, neighbourhood and worldwide weighting highlights to beat impediment, contrasting enlightenment levels and redundant learning with smother lighting condition variety.	

BACKGROUND OF OUR PROJECT

In the project we have used both fisher algorithm and HOG in a two-step classification process. Normally when we consider emotions then we can easily distinguish emotions like happiness, surprise and sadness because they involve visible changes in facial features while if we try to distinguish fear and contempt then we might not get accurate results because the change in facial features is not much. Hence, we use fisher algorithm to identify emotions for widely distinguished emotions like happiness, sadness and surprised and if confidence level is less then we try to find the emotion using HOG algorithm. This in turn helps our software to achieve a better overall time than any other software using HOG without compromising the accuracy.

Pose variation - The ideal scenario for face detection would be one in which only frontal images were involved. But, as stated, this is very unlikely in general uncontrolled conditions. Moreover, the performance of face detection algorithms drops severely when there are large pose variations. It’s a major research issue. Pose variation can happen due to subject’s movements or camera’s angle.

Illumination Effects - Most algorithms work by considering the pixel intensity values and fail in cases where the illumination comes into play, the illuminated pixels give binary classifiers a hard time to classify information and to come up with a plane that divides the points representing the expressions.

To overcome most of these drawbacks of the previous models, we have used the most efficient algorithms in the field and joined them up with a two-stage classification process. The major difference between our algorithm is, the previous ones are we are considering the confidence level as a parameter for choosing which algorithm to use. When the images have marked feature differences an algorithm like fisher's can work well and provide accuracy up to 81%, in such cases there is no need to use more computationally expensive and slow algorithms like HOG, they have to be used as a last resort in cases where the confidence level from the fisher's algorithm is not enough to provide an accurate result.

PROPOSED WORK/ALGORITHM

The basic workflow of our algorithm is –

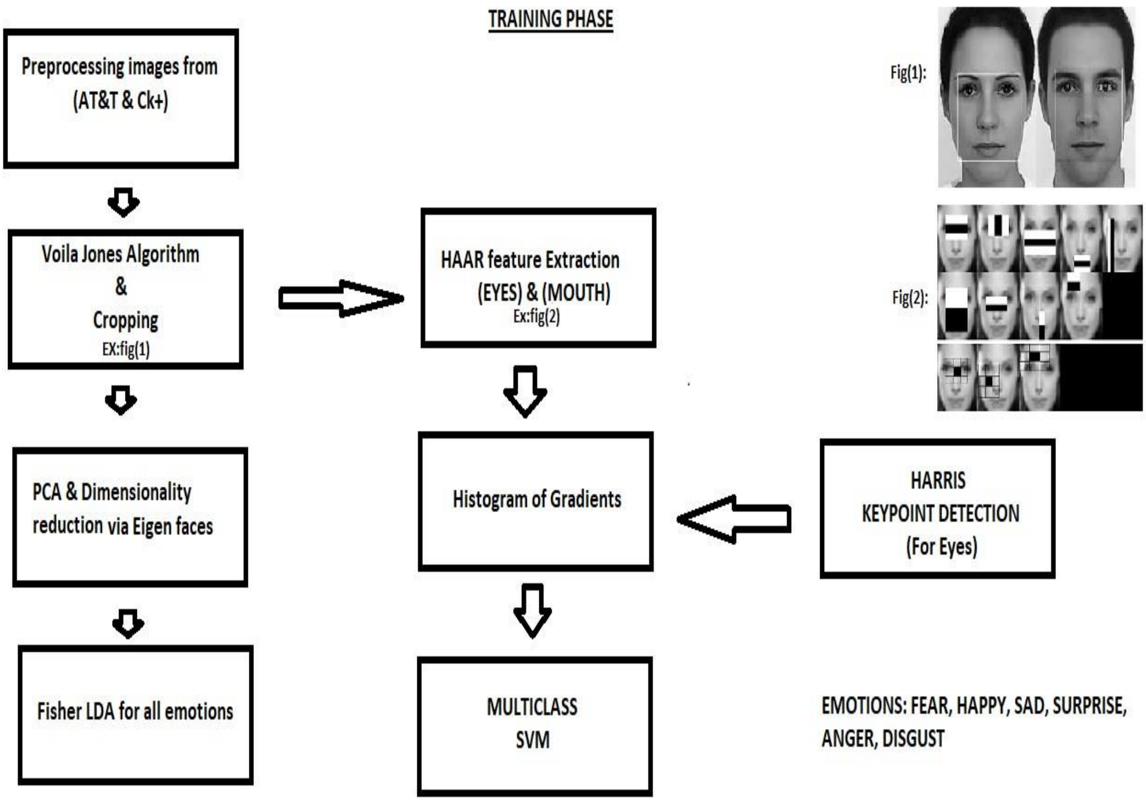


FIGURE 1

Cohn Kanade Dataset-

Like every other machine learning algorithm, we too need an extensive dataset to work upon and increase the accuracy of our system. In this project we have used the Cohn Kanade extended dataset which facial behavior of 210 adults, it was recorded using two hardware synchronized Panasonic AG-7500 cameras. Participants were 18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups. Participants were instructed by an experimenter to perform a series of 23 facial displays; these included single action units and combinations of action units, which means that they were asked to slowly change their facial expressions from one to another so that each frame may have a slight variation in the muscles around the mouth. Each display began and ended in a neutral face with any exceptions noted. Image sequences for frontal views and 30-degree views were digitized into either 640x490 or 640x480 pixel arrays with 8- bit gray-scale or 24-bit color values.

FIGURE 2

FIGURE 3

To train the designed model we took the pictures from the dataset and organized them into specific expression folders. Along with these folders we took some images from the dataset and used them as test cases to test the algorithm once it was trained it was a necessary move to make sure that the algorithm has been correctly.
As per the workflow diagram now the acquired images are processed using the Voila Jones algorithm.

Voila Jones Algorithm-

The Viola-Jones algorithm is a widely used mechanism for object detection. The main property of this algorithm is that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications.

The Viola-Jones algorithm uses Haar-like features, that is, a scalar product between the image and some Haar-like templates. More precisely, let I and P denote an image and a pattern, both of the same size N x N .The feature associated with pattern P of image I is defined by-

$$\sum_{1 \le i \le N} \sum_{1 \le j \le N} I(i,j) 1_{P(i,j) \text{ is white}} - \sum_{1 \le i \le N} \sum_{1 \le j \le N} I(i,j) 1_{P(i,j) \text{ is black}} .$$

FORMULA 1

To compensate the effect of different lighting conditions, all the images should be mean and variance normalized beforehand. Those images with variance lower than one, having little information of interest in the first place, are left out of consideration.

The efficiency of the Viola-Jones algorithm can be significantly increased by first generating the integral image. The Voila Jones algorithm is a strong binary Classifier and can be used to perform different sorts of object detection with reasonable accuracy.

Detection happens inside a detection window. A minimum and maximum window size is chosen, and for each size a sliding step size is chosen. Then the detection window is moved across the image as follows:

1. Set the minimum window size, and sliding step corresponding to that size.
2. For the chosen window size, slide the window vertically and horizontally with the same step. At each step, a set of N face recognition filters is applied. If one filter gives a positive answer, the face is detected in the current widow.
3. If the size of the window is the maximum size stop the procedure. Otherwise increase the size of the window and corresponding sliding stepto the next chosen size and go to the step 2.

Each face recognition filter (from the set of N filters) contains a set of cascadeconnected classifiers. Each classifier looks at a rectangular subset of the detection window and determines if it looks like a face. If it does, the next classifier is applied. If all classifiers give a positive answer, the filter gives a positive answer and the face is recognized. Otherwise the next filter in the set of N filters is run.

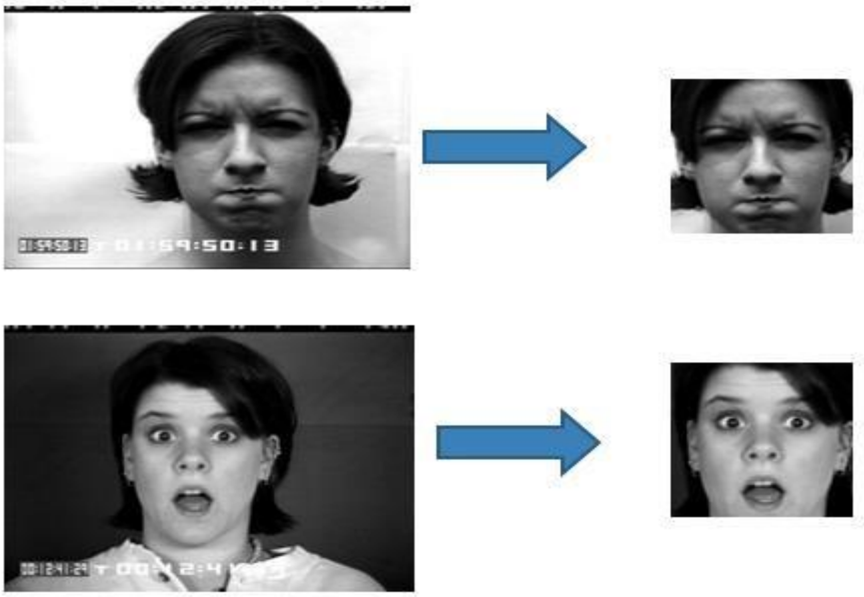


FIGURE 4

Applying the Voila Jones algorithm was necessary because the principal component analysis requires that the images be of the same size. Here we have the images of size 100x100 to perform PCA, since the voila Jones algorithm is considered to be very fast in detection detecting faces will not take much time in real case scenarios.

Principal Component Analysis

In simple terms it can be termed as transformation of set of possibly correlated axes to another set of uncorrelated axes. The uncorrelated axes are called the principal components and can be used for identifying patterns with the given input image by comparing how the input basis vectors are correlated with the given data.

The orthogonal projection of data onto lower dimension linear space is performed such that the variance of the projected data is maximized and the mean squared distance between the data points and projections are minimized. The basic idea is given data points in a d-dimensional space project into lower dimensional space while preserving as much information as possible.

PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a projection of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

The algorithm is as follows:

Following are steps involve;

Step 1: Column or row vector of size N2 represents the set of M images (B1, B2, B3…BM) with size N*N

Step 2: The training set image average (μ) is described as

$$\mu = \frac{1}{m}\sum_{n=1}^{M} Bn \qquad (1)$$

Step 3: the average image by vector (W) is different for each trainee image

$$Wi = Bi - \mu \qquad (2)$$

Step 4: Total Scatter Matrix or Covariance Matrix is calculated from Φ as shown below:

$$C = \sum_{n=1}^{M} wnwnt = AAT , \qquad (3)$$

where A= [W1W2W3…Wn]

Step 5: Measure the eigenvectors UL and eigenvalues λL of the covariance matrix C.

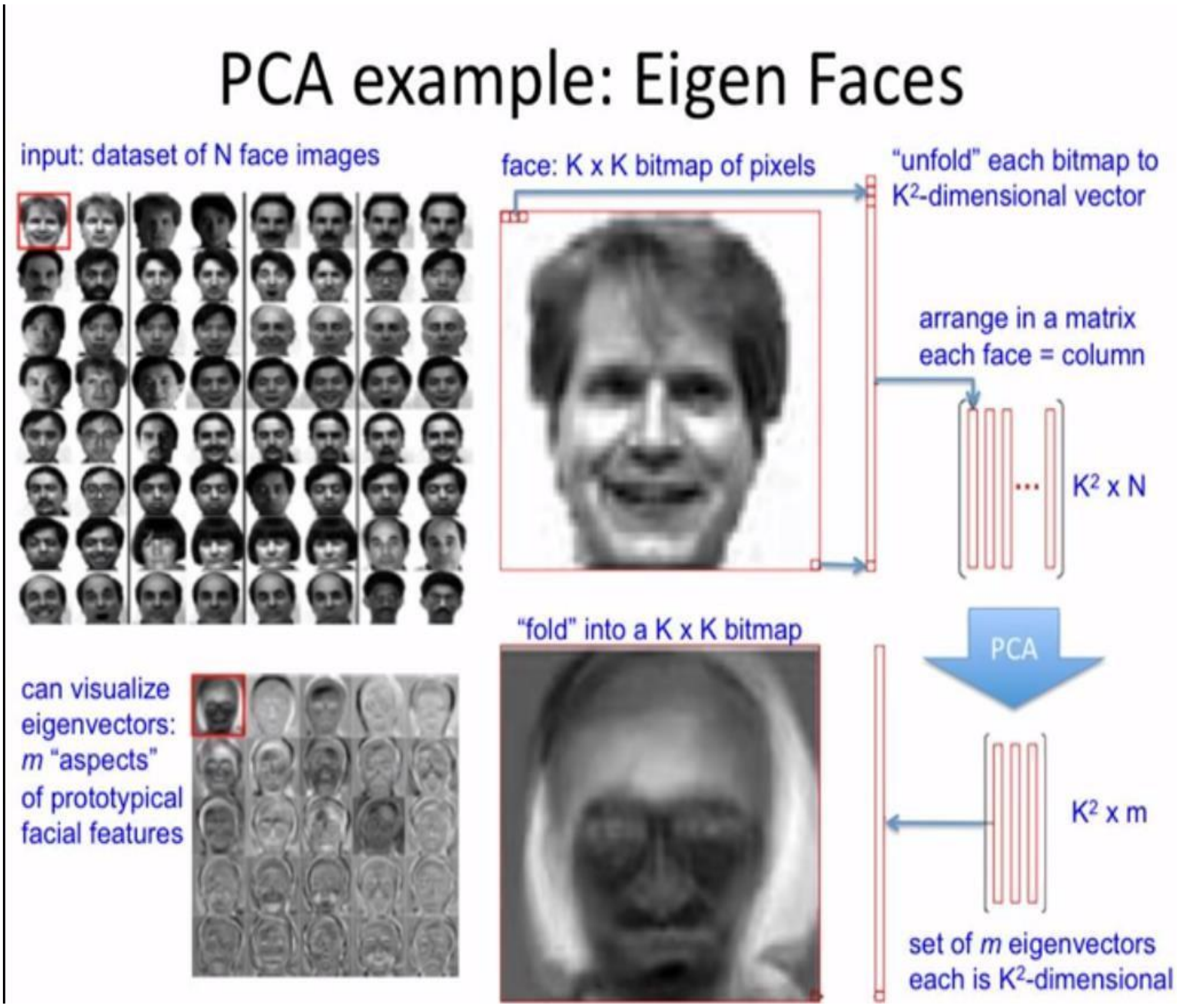
Step6: For image classification, this feature space can be utilized. Measure the vectors of weights

$$\Omega T = [w1, w2, \ldots, wM'], \qquad (4)$$

whereby,

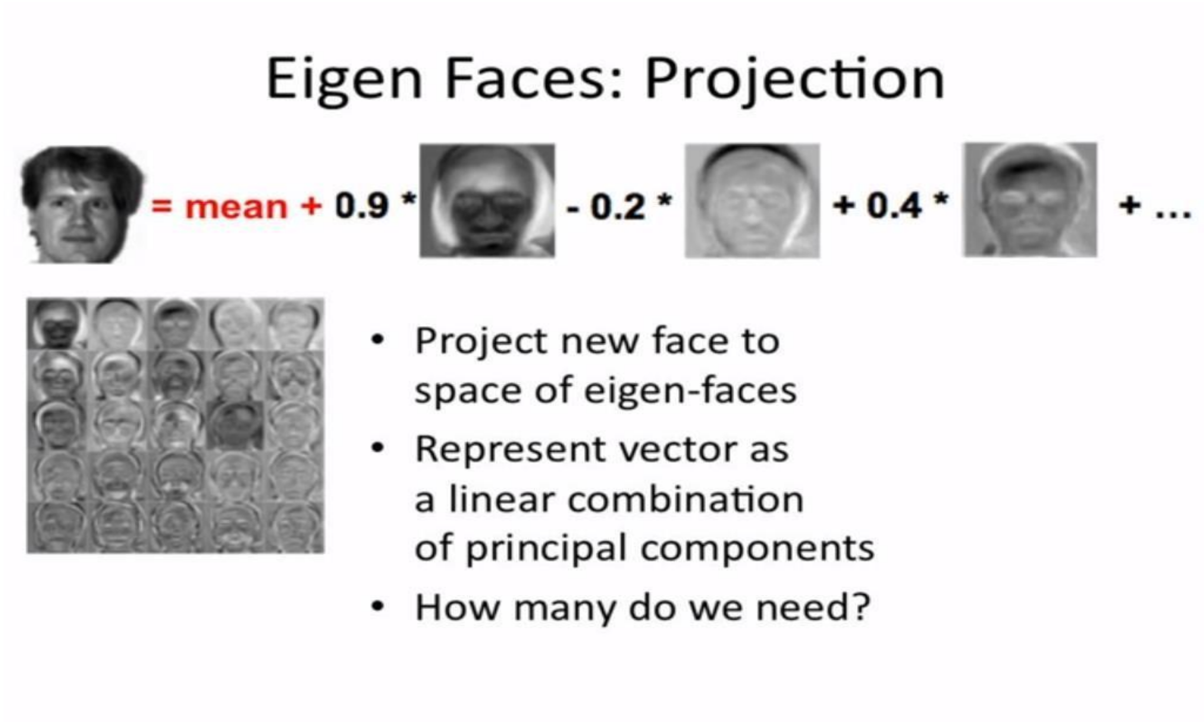
$$Hk = UkT (B - \mu), k = 1, 2, \ldots, M' \qquad (5)$$

ALGORITHM 1


PCA example: Eigen Faces
input: dataset of N face images
face: K x K bitmap of pixels
"unfold" each bitmap to K²-dimensional vector
arrange in a matrix
each face = column
K² x N
"fold" into a K x K bitmap
PCA
K² x m
can visualize eigenvectors: m "aspects" of prototypical facial features
set of m eigenvectors each is K²-dimensional

EIGEN FACES


Anger
Fear
Happy
Sad
Surprise



Once the dimensionality of the images has been reduced then we can use the fisher's linear discriminant analysis to perform the classification of the points involved.

Fisher's linear discriminant Analysis:

In the previous step we applied the PCA to reduce dimensionality of the images but now we need to find a linear separation between the different class of image though we have the data points obtained in the previous analysis but we require a function that can maximize the linear separation between different classes .The work of fisher's algorithm is to maximize a function that will give a large separation between the projected class means while also giving a small variance within each class, thereby minimizing the class overlap.

In other words, FLD selects a projection that maximizes the class separation. To do that, it maximizes the ratio between the between-class variance to the withinclass variance.

In short, to project the data to a smaller dimension and to avoid class overlapping, FLD maintains 2 properties.

1. A large variance among the dataset classes.

2. A small variance within each of the dataset classes.

A large between-class variance means that the projected class averages should be as far apart as possible. On the contrary, a small within-class variance has the effect of keeping the projected data points closer to one another.

The algorithm for multi-dimensional space is-

$$S_W = \sum_{k=1}^{K} S_k \tag{5}$$

$$S_k = \sum_{n \in C_k} (x_n - m_k)(x_n - m_k)^T \tag{6}$$

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T \tag{7}$$

$$\boldsymbol{W} = max_{D'}(eig(S_W^{-1} S_B)) \tag{8}$$

Within-class covariance

Between-class covariance

FORMULA 2

Here, we need generalization forms for the within-class and between-class covariance matrices.For the within-class covariance matrix S_W, for each class, take the sum of the matrix-multiplication between the centralized input values and their transpose. Equations 5 and 6 denote the operation.

For estimating the between-class covariance S_B, for each class k=1,2,3,…,K, take the outer product of the local class mean m_k and global mean m. Then, scale it by the number of records in class k-equation 7.

The maximization of the FLD criterion is solved via an Eigen decomposition of the matrix-multiplication between the inverse of SW and S_B. Thus, to find the weight vector W, we take the D' eigenvectors that correspond to their largest eigenvalues (equation 8).

In other words, if we want to reduce our input dimension from D=784 to D'=2, the weight vector W is composed of the 2 eigenvectors that correspond to the D'=2 largest eigenvalues.
This gives a final shape of W = (N,D'), where N is the number of input records and D' the reduced feature dimensions.

The example can be shown with a three-dimensional model-

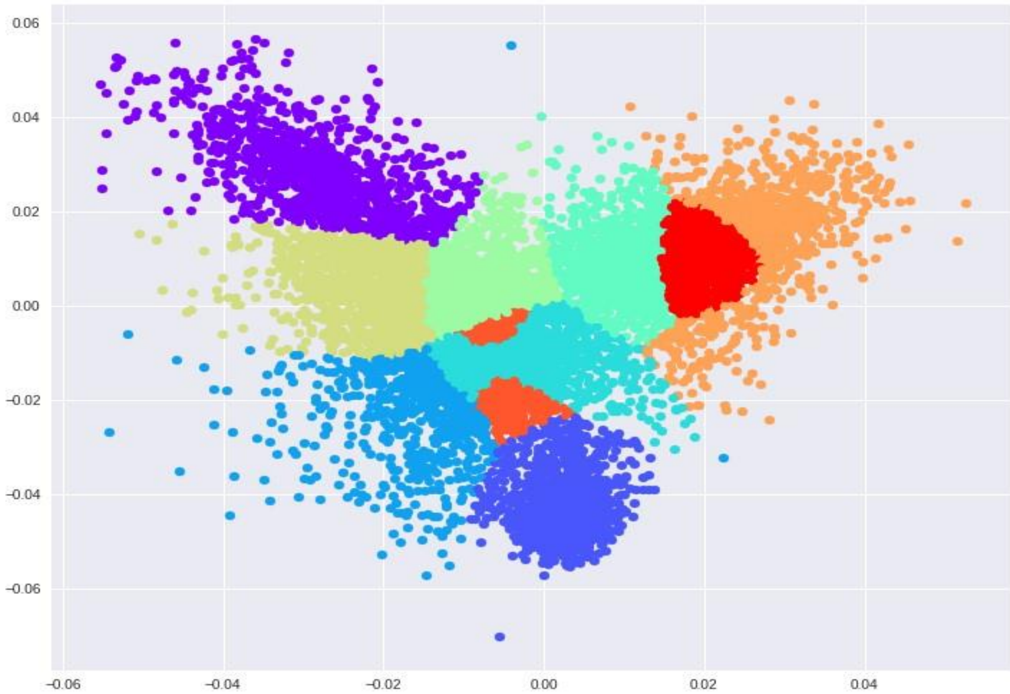


FIGURE 6

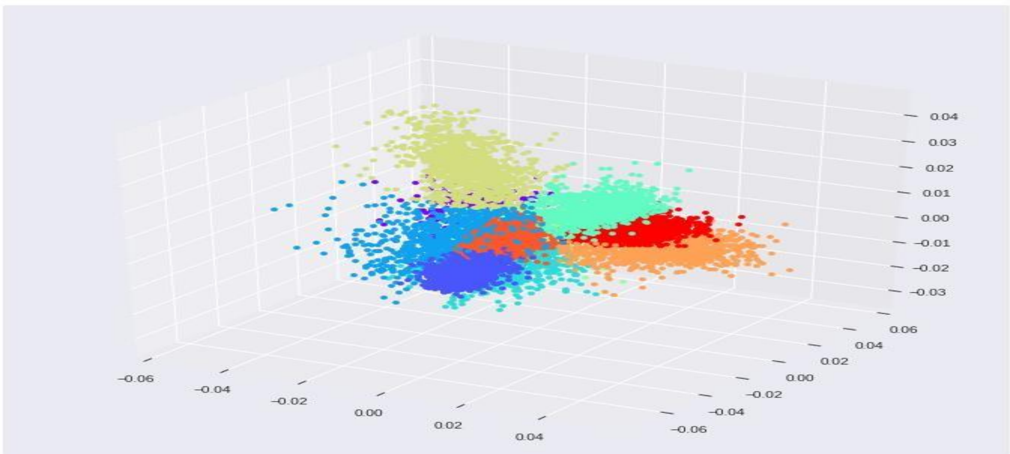
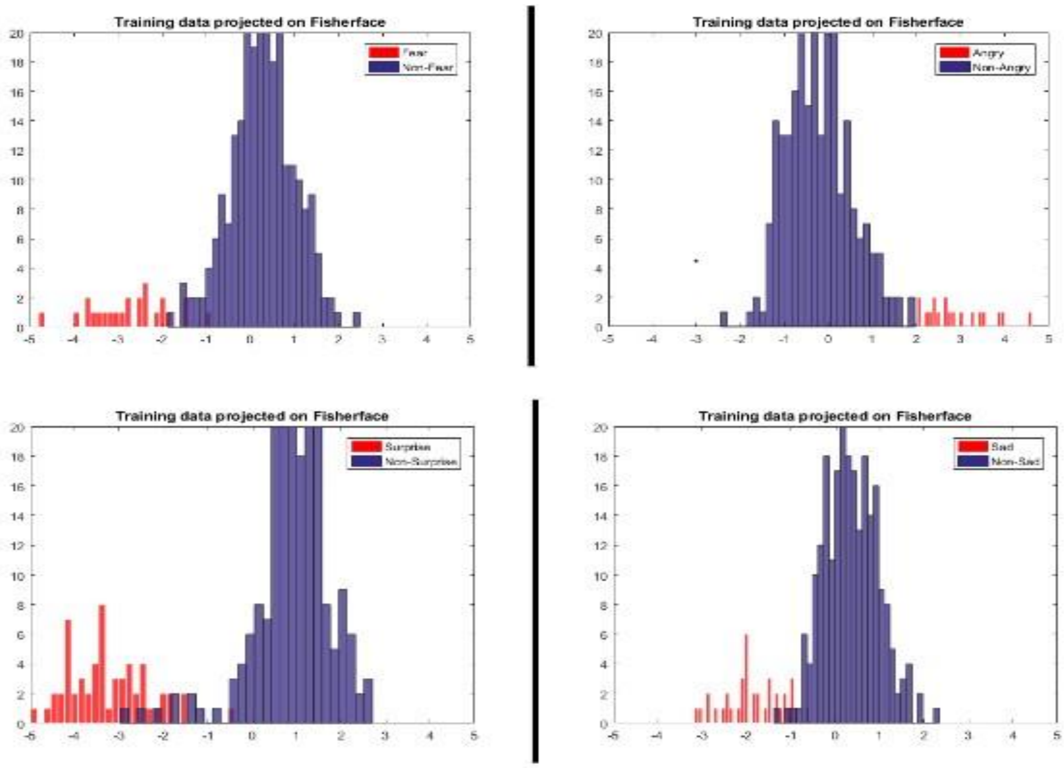


FIGURE 7

These algorithms were the first step in our classification process the fisher faces which will be generated from this will be used for primary emotional recognition but if the confidence level is not satisfactory, we proceed to the secondary processing method with HOG and SVM.

Projecting all training faces



Histogram of Oriented Gradients (HOG):

A feature descriptor is a representation of an image or an image patch that simplifies the image by extracting useful information and throwing away extraneous information.

In the HOG feature descriptor, the distribution (histograms) of directions of gradients (oriented gradients) are used as features. Gradients (x and y derivatives

) of an image are useful because the magnitude of gradients is large around edges and corners (regions of abrupt intensity changes) and we know that edges and corners pack in a lot more information about object shape than flat regions.

The feature vectors produced by these algorithms when fed into an image classification algorithm like Support Vector Machine (SVM) produce good results.

For a function (x,y) the gradient vector is (f_X,f_Y) while an image is a discrete function of (x,y) so image gradient can be calculated as well. At each pixel image gradient in the horizontal (x-axis) and in the vertical (y-axis) are calculated. These vectors have a direction $atan \frac{f_y}{f_x}$ and a magnitude of $\sqrt{f_y^2 + f_x^2}$

The detailed algorithm to generate gradients for an image-

1. Using 8X8 cells we compute the gradient vectors at each pixel. This generates 64 gradient vectors which are then represented as a histogram.
2. Now each cell is split is into angular bins which means that we classify the vectors according to their angles which we previously derived .In this project we have used 9 bins that means vectors with angles 0-20 degrees are grouped together in the histogram.
3. This effectively reduces 64 vectors to just 9 values and as it stores gradients magnitudes it is relatively immune to deformations.
4. We then normalize the vectors to ensure invariance to illumination changes. For the normalization process we divide the vectors by the gradient magnitudes.

Example of images with HOG performed –

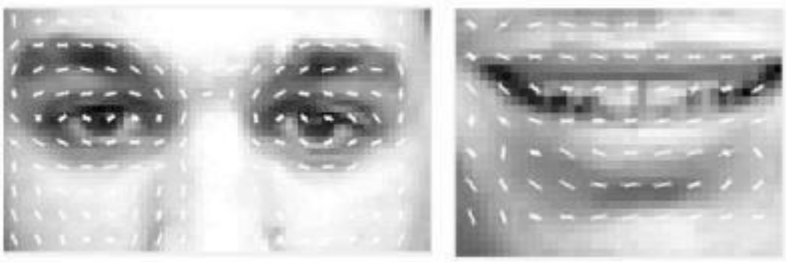
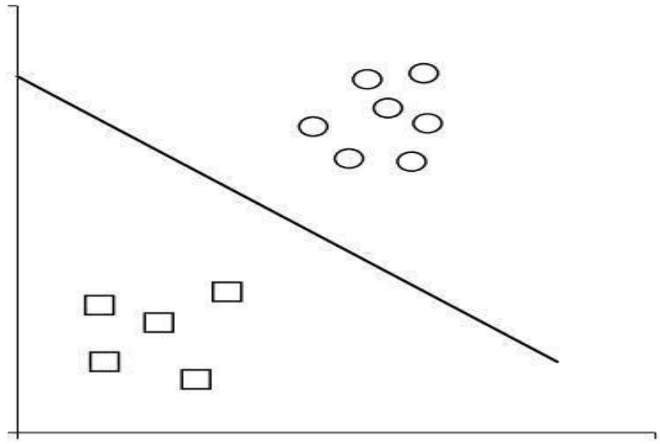
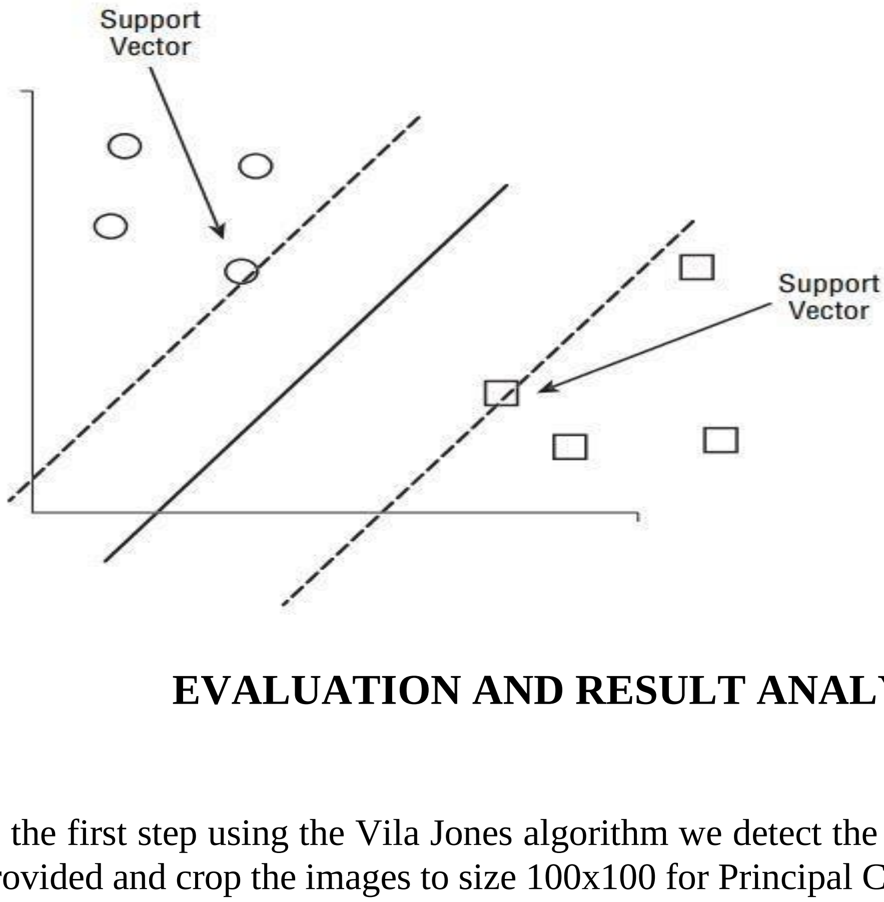


FIGURE 8

Support Vector Machines

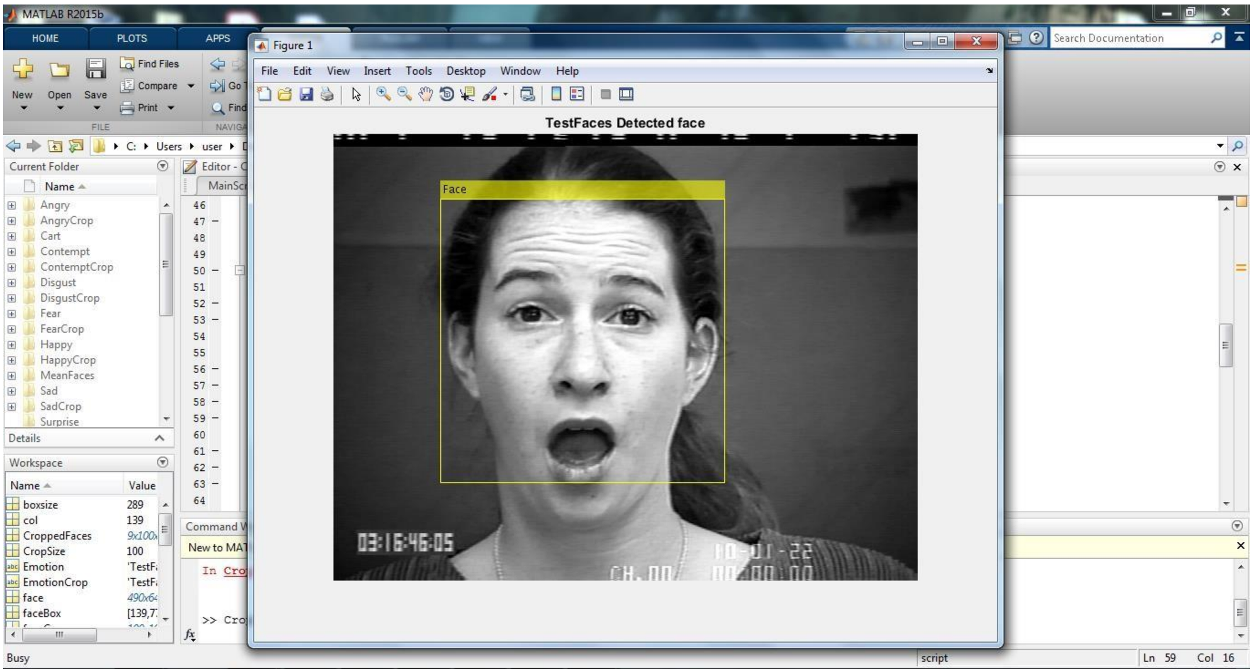
Linear classification is shown below-





EVALUATION AND RESULT ANALYSIS

In the first step using the Vila Jones algorithm we detect the faces in the dataset provided and crop the images to size 100x100 for Principal Component analysis.

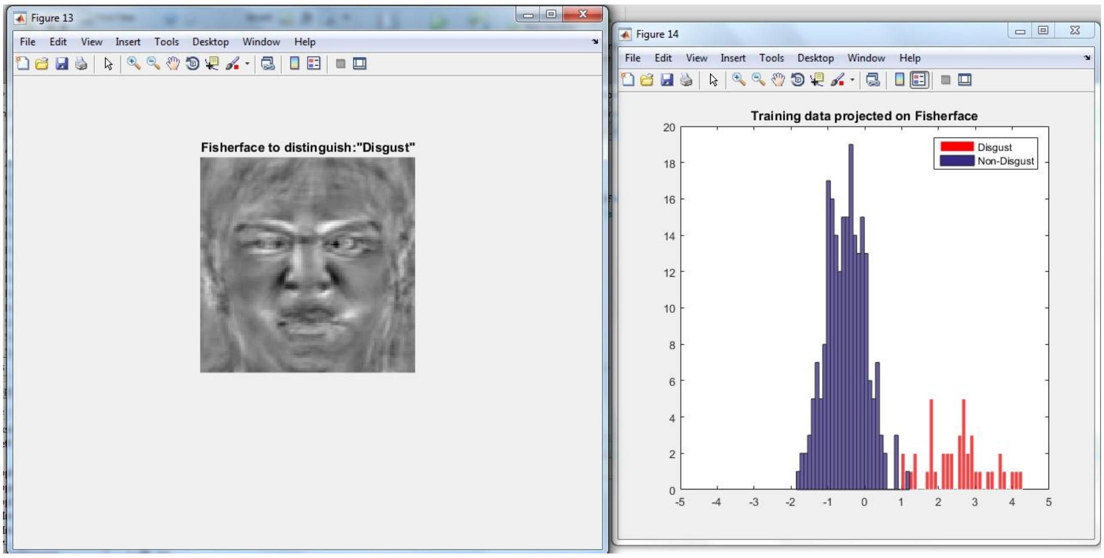


OUTPUT FIGURE 1

Once we have the cropped dataset we run the principal component analysis on the images to obtain the top ten Eigen faces or in other words we project the images data to a 10-dimensional space using 10 feature vectors.

OUTPUT FIGURE 2

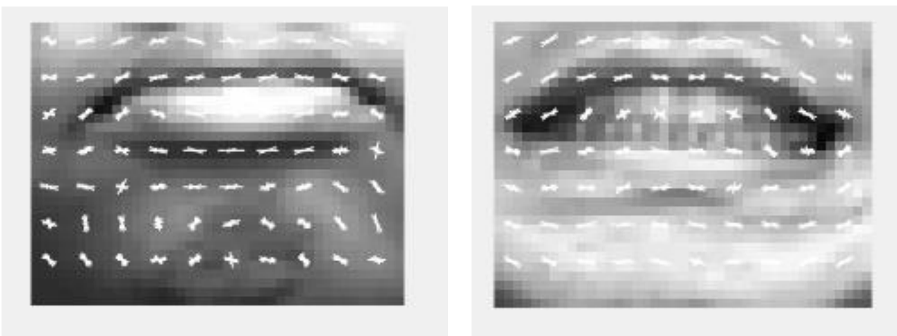
Once we have the data points from all the images, we now have to find a linear separation between the different classes which is achieved with the help of Fisher’s LDA-



OUTPUT FIGURE 3

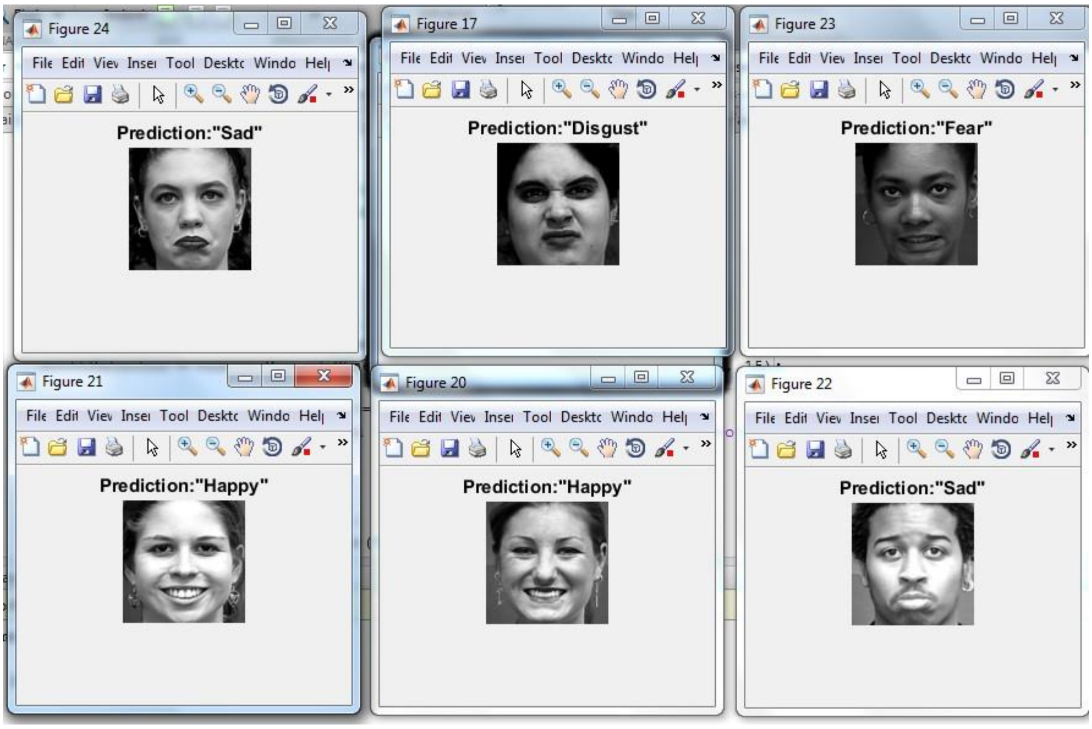
The fisher face generated from this script is used for emotional analysis in test phase, the histogram generated depicts that the values of any particular emotion like Disgust have been completely separated from other classes with no overlapping.

Once the initial phase is done we now train the HOG descriptor on our images. The HOG descriptor does not take any data from any of the previous algorithms it uses the HAAR cascades to detect and extract regions of interest like the mouth and then calculates the gradients for it after which the values are fed into the support vector machine which generates a hyperplane to distinguish the classes of different emotions.



OUTPUT FIGURE 4

Once both the models have been trained, we run the test script to test the images and we record the time taken by each of the algorithms including our own algorithm.



OUTPUT FIGURE 5

For each of the test images we provided the script gave a prediction on their emotions, this was the result from our algorithm and we also checked the total time duration that required to process these images. The script also indicated the user when the Hog prediction was used to detect the facial expression for each of the images.

```
Further predicitive measures needed
Proceeding to HOG extraction
Further predicitive measures needed
Proceeding to HOG extraction
Further predicitive measures needed
Proceeding to HOG extraction
Further predicitive measures needed
Proceeding to HOG extraction
Further predicitive measures needed
Proceeding to HOG extraction
Further predicitive measures needed
Proceeding to HOG extraction

TimeSpent =

   25.3401

fx >>
```

OUTPUT FIGURE 6

TABULAR COMPARISON

The results for each of the algorithms are as such-

Algorithm	Accuracy	Runtime(s)
FisherFace only	56%	7.40
HOG only	81%	9.87
FisherFace+HOG	81%	7.91

OVERALL DISCUSSION

Interpersonal communication is often complex and nuanced, and a range of variables often predict its achievement. These variables vary extensively and may include the background, mood, and timing of the interaction, as well as participants ' expectations. To be a successful participant, one must perceive the disposition of a counterpart as the interaction advances and adjust accordingly. Fortunately, this capacity is mainly intrinsic to animals, with variable skill concentrations. In order to discern the feelings of others, humans can rapidly and even subconsciously evaluate a variety of factors such as word decisions, speech inflections, and body language. This analytical capacity probably arises from the reality that humans share a universal set of basic emotions. Significantly, these feelings are displayed through constantly corresponding facial expressions. This implies that irrespective of language and cultural obstacles, there will always be a set of basic facial expressions that individuals will evaluate and interact with.

The project extracted the features from the offline facial image dataset and perform Support Vector Machines Classification and Histogram of Oriented Gradients, to extract the emotions. Presented in this project is hybrid feature extraction and facial expression recognition method with a dual classifier approach that uses Viola-Jones Cascade object detectors and Harris key-points to extract features from the images. We are interested in seven fundamental emotions such as Anger, fear, Happiness, Disgust, Sadness, Surprise and Contempt. The project covers disciplines such as Image Processing, Machine Learning and Computer Vision.

CONCLUSION

The dual-classifier approach works well when the Fisherface cannot effectively determine a prediction. This happens in two cases. First is if a test image is not one of the "easy-to-distinguish" emotions, and second is if the Fisherface classifier cannot decide between two or more predicted emotions.

When using the HOG and SVM classifier only, the accuracy for detection is 81%, much better than a Fisherface only approach. When using the dual-classifier method, the accuracy is the same as HOG-only at 81%, but the testing process is 20% faster.

With test results we conclude that the two-stage classification approach performed better than the most popular emotional analysis algorithms with a run comparable to fisher's algorithm and accuracy equal to HOG descriptor.

REFERENCES

[1] Cai, J., Meng, Z., Khan, A. S., Li, Z., O'Reilly, J., & Tong, Y. (2018, May). Island loss for learning discriminative features in facial expression recognition. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 302-309). IEEE.

[2] Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 115, 101-106.

[3] Tsai, H. H., & Chang, Y. C. (2018). Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Computing, 22(13), 4389-4405.

[4] Xie, S., & Hu, H. (2018). Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Transactions on Multimedia, 21(1), 211-220. [5] Xue, M., Mian, A., Duan, X., & Liu, W. (2019, May). Learning Interpretable Expression-sensitive Features for 3D Dynamic Facial Expression Recognition. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-7). IEEE.

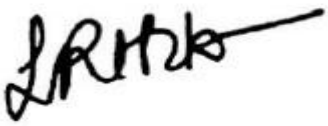
[6] Revina, I. M., & Emmanuel, W. S. (2018). Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/j.jksuci.2018.03.015.

[7] Utami, P., Hartanto, R., & Soesanti, I. (2019). A Study on Facial Expression Recognition in Assessing Teaching Skills: Datasets and Methods. *Procedia Computer Science*, *161*, 544–552. doi: 10.1016/j.procs.2019.11.154

Appendix for Acronyms

DL	Deep Learning
DIP	Digital Image Processing
HOG	Histogram-of-oriented gradients
SVM	Support Vector Machines
PCA	Principle Component Analysis
GPU	Graphical Processing Unit
ML	Machine Learning

Appendix for Individual Contribution Details in Group

Sr No	Reg No and Names	Role and Responsibility	Digital Signature
1	18BCE0026 B Harsha Vardhan	Implementation and coding	
2	18BCE0022 L Ruthvik Raj	Analysis of algorithms	
3	18BCE0011 M Manikanta Sahith	Experimented and studied various research paper