

# Facial Expression Analysis and Estimation Based on Facial Salient Points and Action Unit (AUs)

**Shashank M Gowda<sup>1</sup> and H N Suresh<sup>2</sup>**

<sup>1</sup>AP, DoECE, YIT, India

<sup>2</sup>Prof, DoEIE, BIT., Bengaluru, Karnataka, India

\*Correspondence: Shashank M Gowda; Email: shashank.m.gowda.91@gmail.com

**ABSTRACT:** Humans use their facial expressions as one of the most effective, quick, and natural ways to convey their feelings and intentions to others. In this research, presents the analyses of human facial structure along with its components using Facial Action Units (AUs) and Geometric structures for identifying human facial expressions. The approach considers facial components such as Nose, Mouth, eyes and eye brows for FER. Nostril contours such as left lower tip, right lower tip, and centre tip are considered as salient points of Nose. Various salient points for Mouth are extracted from the left and right end point, upper and lower lip mid points along with curve. These salient points are extracted for all facial expression of the same subject considering neutral face as reference. The Geometric structure for neutral face is mapped along with other facial expression faces. The deformation is estimated using the Euclidean distance. The classification algorithms such as LibSVM, MLP, RF has achieved classification accuracy of 86.56% on an average. The findings of the experiments show that the extraction of picture characteristics is more efficient in terms of computing and gives promising outcomes.

**Keywords:** Facial Expression, Action Unit, Salient points, Random Forest, Pre-processing, Facial Landmark.

## ARTICLE INFORMATION

**Author(s):** Shashank M Gowda and H N Suresh

**Received:** Dec 15, 2021 **Accepted:** Feb 20, 2022 **Published:** Mar 30, 2022;

**e-ISSN:** 2347-470X;

**Paper Id:** 0122SI-IJEER-2022-06;

**Citation:** doi.org/10.37391/IJEER.100102

**Webpage-link:**

https://ijeer.forexjournal.co.in/archive/volume-10/ijeer-100102.html

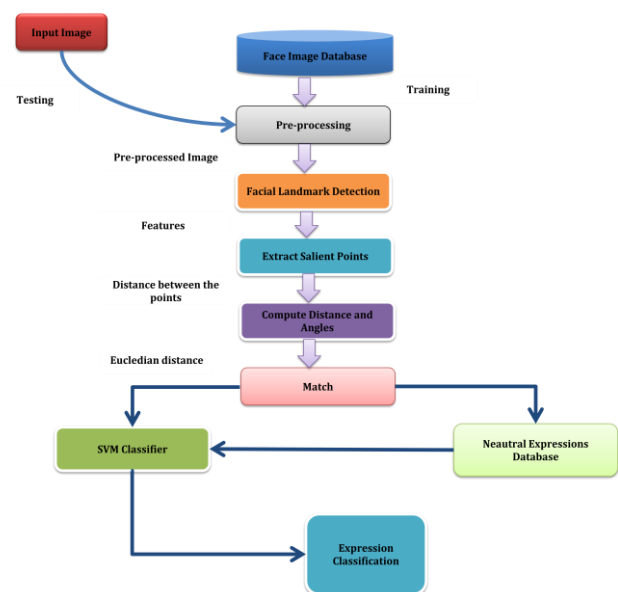


**Publisher's Note:** FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

## 1. INTRODUCTION

Building automatic FER system is a complex task, since the human face varies from person to person due to various factors such as age, culture, ethnicity, etc. Apart from these, other factors like facial hair, glasses, translation, scaling, rotation, illumination make this task more complex. Thus, there is a need for an effective approach which can handle the above mentioned factors. Further, accurate localization of facial points is a key step for extracting feature to categorize facial expressions. However, in many real time applications, the performance of feature extraction approach also depends on the environmental factors such as illumination, brightness, etc. If the illumination is non-uniform, extracted feature will be inaccurate and results in poor recognition rate. Hence, it is necessary proposed effective FER approach, which may use methods like DCT Normalization (Chen and Er, 2006 [1]), Histogram Equalization (Gangolli et al, 2019 [2]) etc. before facial feature extraction. The suggested FER system is shown in Figure 1. In this paper, the proposed approach analyses and estimate the facial expressions using facial landmarks and its salient points. The proposed approach considers the JAFEE Dataset for analysis. The dataset comprises of 256 x 256-pixel grayscale photos of ten Japanese women. There are a total of 213 photos in this collection, with seven different face emotions shown, including neutral, surprise, happy, sad, and

disgusted. There are 30 photos of rage, 29 photographs of disgust, 31 images of joy, 32 images of fear, 31 images of grief, 30 images of surprise, and 30 images that are neutral. The Dataset is cross validated into training and testing set. The proposed approach detects the face using Viola Jones (Jain and Choudhary, 2018 [3]). The detected face is cropped using background subtraction and rescaled for 132x132 pixels. The normalization and filtering techniques are used to handle illumination invariant, occlusions, noise etc., and using morphological Sobel edge detection (Singh and Singh, 2016 [4]).



**Figure 1.** Schematic Diagram of the Proposed FER System

The next phase is feature extraction phase. The approach considers eyes, eyebrows, nose and mouth as facial components. The rescaled sample images are processed for detecting facial landmarks using Cascaded CNN algorithm.

The face region is segmented using Region based segmentation technique using landmark. The approach identifies the salient points for each of these facial components using landmarks and computes the distances and angles between these salient points for anger, disgust, fear, happiness, sadness and surprise. The distances and angles between the neutral face and face under consideration is calculated. The Euclidian distance is used for comparing the distance and angle and the degree of difference is calculated.

## 2. LITERATURE SURVEY

A person's emotional and mental condition are represented by their facial expressions, which are very important. Based on the Psychological survey, Language conveys 7% of overall information. 38% by language auxiliary such as speech rhythm, tone, etc. and 55% by facial expression. Hence, FER is considered as an important research topic in terms of theoretical and practical research and the life application values. An emotional feeling, say *happy*, *sad*, *fear*, etc. are expressed in terms facial expression and these are caused due to occurrence of various situations. The author Ekman, (1992 [5]) determined 6 basic emotions from the human face such as *anger*, *happiness*, *surprise*, *sad*, *disgust* and *fear* that are expressed frequently through facial expressions. Automatic FER attracted a large number of researchers in the last few decades due to the rise of numerous applications in the field of computer vision, robotics, medical, HCI, etc. However, there exist certain limitations in FER system such as traditional approaches may not handle this real-time huge data. Apart from this, lighting conditions, noise, occlusion and head position fluctuations are some of the elements that impact the FER systems resulting in uncertainty and ambiguity. Many solutions employed handmade characteristics for FER and hence need additional efforts in terms of programming and calculation expense. In this paper, an architecture specification of FER system is proposed to handle the above mentioned research issues. The proposed approach considers six basic emotions mentioned by Ekman, 1992 [5]. The approach builds the FER model by considering facial components such as eyes, eyebrows, nose and mouth features. The approach proposes geometric-based feature extraction for these facial components considering *neutral* face as reference image. The geometric structure and their properties are exploited for each of the facial components. For instance, the lips are mapped to parabola and its properties; nose is mapped to tetrahedron/pyramid and its properties, etc. The similarity between the geometric properties of *neutral* face and geometric properties of other expression faces are analyzed. The degree of deformation for the facial expressions is calculated.

### 2.1 Reviews on Pre-Processing Techniques in FER System

The image pre-processing step improves the FER system's performance. This process deals with input samples data and carried out before feature extraction process. Image pre-processing includes various techniques in handling the raw data such as image detection, cropping, distortion removal,

image scaling, normalization, segmentation, additional enhancement processes, etc. to improve the image frames. Below section represents literatures of various pre-processing techniques.

#### 2.1.1 Reviews on Face Detection and Tracking

It is a technique for extracting the face from a picture by using Face Detection. It's important to do this step since each picture has a distinct size and orientation. Due to the observed person's motions, the pictures captured by the cameras will be of varying sizes and angles. As a result, searching for a certain pattern in the picture becomes time-consuming. Detecting faces in photos with noise and occlusion may be challenging. Face tracking will be difficult if the backdrop and lighting circumstances of the facial picture are complicated. If the test photograph is taken in a different lighting environment, facial expression recognition is more likely to fail. Facial landmark points cannot be detected accurately because of Illumination factors. To avoid these factors, accurate face detection in pre-processing is required. This section discusses various approaches related to detecting, cropping and scaling the facial images. The authors Li and Lyu, (2016) [6] depth image feature-based face detection approach has been suggested. HOG feature is used as main feature to train the image for classification. The detected image is divided into cells and gradient is calculated pixel by pixel to produce the Histogram of Oriented Gradient. The gradient feature vector is normalized block by block. All the HOG feature vectors are collected using zigzag scan sliding window. The detection rate is further improved by combining LBP feature with HOG feature and the features are classified using Advanced SVM classifier. The authors SumaLakshmi and Vasuki, (2020) [7] have classified the models into edge, dimensional and middle for facial detection. A black-and-white picture is utilised for the design of Viola and Jones models. Upon importing it onto the image, the pixels inside the black and white portion of rectangle is removed to detect the facial features. Viola Jones is used to identify the face, then the LBP histogram from three sections of the face is analysed and the final histogram from the three regions is returned. The authors Narayan and Deshpande, (2017) [8] detected the face using Viola Jones algorithm, and then later the Fusion of PCA is used to identify images and FFNN. The authors Wagh et al., (2015) [9] captured the image picture captured with a high definition camera; the image is preprocessed using Histogram equalisation. Face detection is accomplished via the use of the Viola-Jones algorithm. Skin classification technique is done to improve the face detection algorithm.

#### 2.1.2 Reviews on Handling Head posture, Illumination, Occlusion, Scale and Translation Factors

Image cropping is an essential step in the FER system as it is important to crop the relevant section from the image for feature extraction. Images are generally dominated by noisy data, occlusions, disturbing backgrounds, and illumination conditions. Cropping is a process where it removes the image background and presents the actual image useful for feature extraction and classification. The author Kumar et al., (2012) [10] has proposed an approach where an image is given as an

input to detect some faces. It is suggested to use a face detection algorithm in conjunction with an occlusion detection method to identify all of the occluded faces. This experiment is conducted in a classroom with thirty students and the proposed algorithm detected the faces well without occlusions. The author Kavinmathi et al., (2016) [11] has proposed an approach to detect the face for automatic online attendance. The author has used background subtraction method initially on input image and then later the face is cropped and detected using Eigen value method to recognize the faces.

The authors Zeng and Veldhuis, (2020) [12] has explored the occlusion problem and presented various face recognition techniques which deals with occlusion robust feature extraction and occlusion recovery techniques. The author Kahatapitiya et al., (2019) [13] has proposed a method to remove occlusions in an image automatically. This method identifies the occlusions in an image by subtracting the foreground and background classes using vector embedding and discards them through in painting. Face Restoration is an important process to locate occluded regions in the image. The authors Srinivasan and Balamurugan, (2015) [14] used Gappy Principal Component Analysis (GPCA) method for 3D image occlusion in which the image is masked initially by computing the Distance from Feature Space (DFFS). The output image is refined to remove the regions detected from the Masking.

The authors Ma and Mohamed, (2015) [15] have proposed a new concept called as image processing pipeline to improve the effects of illumination in the image. BU-4DFE Dataset images are used as training Dataset and 3D images of different facial expressions are tested by adding various lighting conditions. An image processing pipeline is proposed which includes gamma correction, selective filtering and contrast equalization. SIFT key point's detector (Guo et al., 2018) [16] is used to compare the training and testing illuminations of the images. Fisher face classifier is used to evaluate the results of preprocessing pipeline on a classifier.

The authors Pal and Porwal, (2015) [17] have proposed a Local Brightness Normalization (LBN) algorithm to correct global and local vertical stripes in images. In order to accomplish this task, they had utilized noise fraction-transformation-based filtering to remove noise from the image. Then based on the output values, global and local stripes are corrected. It is observed that though detecting human facial emotions using geometrical structure is popular technique in recent times, the complexity lies in geometric variability in both emotion expression and the neutral face as it directly affects geometric FER methods.

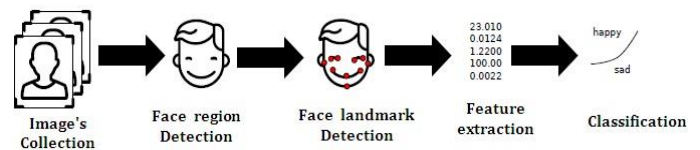
### 2.1.3 Reviews on Facial Segmentation and Landmark Detection

Image segmentation is a critical stage in the FER preprocessing process. Here, the facial image is divided into multiple segments to extract data from the attributes of facial image. This supports in locate facial components and boundaries in the facial image. In FER, ROI segmentation technique is used in general as it segments those interested

face regions which are actively participated in expression recognition. The authors used Interface's 49-point facial landmark detector has been used to localize the points of Eyebrows, Lips and Nose. The upper lip and eyebrows edges are detected by using edge detection algorithms such as Sobel, Canny, Prewitt and Roberts (Zhang and Yan, 2009) [18]. After performing the edge detection, any false rates are removed by using Otsu's Thresholding method (Feng and Zhao, 2017) [19]. Finally, emotion classification is done using MLP classifier. By extracting the principal curvatures and form index of the eyes and nose, it is possible to recognise facial emotions more rapidly by analysing the characteristics of the face. (Vezzetti et al., 2017) [20].

## 2.2 Conventional approaches

In addition to aforementioned distinction between static FER and FER on sequences, approaches to FER can also be categorized by the features used for classification. Conventional FER approaches use handcrafted features inferred from face in the facial extraction step of FER process. Deep-learning approaches often use CNN to extract features directly from images during training process. Approaches in this category usually adhere to following FER process schema depicted in Figure.2.



**Figure 2.** Conventional FER process

Facial images are first collected and preprocessed (histogram equalization, noise reduction, etc.). FER is usually performed on grayscale images as color does not carry significant information about the expression. Next step is face region detection. It is important to regionalize face in the image before attempting to localize facial landmarks to avoid false-positives. Multiple approaches to face region detection have been proposed over past few decades. Haar cascade classifier is one of the more popular approaches. Localization is performed via AdaBoost method using Haar-like features (descriptors of contrast change between adjacent rectangular groups of pixels). Detected face region is then used as a region of interest for face landmark estimation (face alignment). Many face alignment approaches use cascade of regressors. Each regressor is improving on landmark position estimate based on image features relative to the previous landmark position estimate. In [25] author's use ensemble of regression trees learned by gradient boosting to achieve super-real-time performance while maintaining state-of-the-art accuracy on face alignment problem. Feature extraction step uses face landmarks to produce feature vector for training. Temporal and appearance features are also often extracted in addition to geometric landmark features. SVMs are dominant classification method in conventional FER approaches. Radial Basis Function (RBF) kernel SVM seems to usually outperform linear SVM in FER.



In [26] author's present a real-time mobile application for FER using a set of SVMs to recognize 7 basic emotions. Active Shape Model (ASM) [14] is used to locate 77 face landmarks which are then used to generate 13 high-level distance features. The model performs classification based on displacements relative to the neutral feature set. During classification process each frame is first classified by binary classifier detecting neutral emotion state. Extracted features from neutral frames are then used to update the current neutral feature set. A CK+ dataset was used for training. Reported accuracy on CK+ dataset is 87.9%.

In [27] author's used Elastic Bunch Graph (EBG) [16] to initialize 52 landmark positions which are then tracked in rest of the frames in sequence using Gabor jets. The classification is performed by SVM using features of two types. First type is  $x$  and  $y$  displacement of 52 landmarks relative to neutral features. Second type is Euclidean distance and angle change between all pairs of landmarks relative to the distances and angles in neutral features. Neutral frames are not being recognized in-process but rather an assumption is made that neutral frame is always the first frame in sequence. Final feature vector is selected from a feature pool consisting of the two aforementioned types of features using AdaBoost with Dynamic Time Warping (DTW) similarity. CK+ dataset was used for training and reported accuracy on this dataset is 97.2%. In [28] author's present real-time FER system using multi-block Local Binary Pattern (LBP) appearance features and Principal Component Analysis (PCA) to classify 6 basic emotions (neutral emotion is not being classified). In proposed model Haar cascade is first used to detect face region in source image. Face region is then divided into small subsections and the LBP histogram is calculated for each block. The resulting feature vector is formed by joining together separate LBP histograms. The classification is done using PCA Eigen values for each emotion. Reported accuracy on custom dataset is 97%. The author, in contrast to [29], who recovered appearance features from the global face area, derived region specific appearance LBP features by partitioning the face region into 29 sector particular local regions. Incremental search approach was employed to localize important local regions in order to reduce dimensionality. In addition to appearance LBP features, geometrical landmark features were also extracted using implementation of [12]. Final feature vector is presented to linear SVM classifier. Model was validated against CK+ dataset with reported accuracy of 91.8% when classifying 7 basic emotions which are summarized in the table 1.

### 2.3 Deep-learning approaches

Deep-learning approaches to FER often use CNN to either perform classification directly or to extract latent features. In order to capture temporal aspect of expressions Recurrent Neural Networks (RNN) are sometimes used as well.

In their submission to the 2015 Emotion Recognition in the Wild contest, author's in [30] examined effectiveness of transfer-learned CNN on FER problem with small available dataset. They used pre-trained CNN model of VGG-CNN-M-2048 [20] architecture which was trained on generic image

recognition task using images from ImageNet. This base model was then transfer-learned in two fine-tuning phases using EmotiW and FER-2013 dataset. Resulting model achieved 55.6% accuracy on the test set.

In [31] the authors introduced a combination deep temporal appearance convolutional network (DTAN) and deep temporal geometry network (DTGN) model. Softmax outputs of these two networks is connected by element-wise addition with softmax applied to produce the final output. DTA is a 3D convolutional network where convolutional filters are shared along the time axis. This network captures temporal difference in appearance of the input images. Sequence of facial landmarks is used as input for the DTGN. Each landmark point is centered around a nose point and normalized using division by standard deviation of according dimension. Horizontal flipping and rotation were applied to input image sequences in order to increase the amount of data available for training. Model was trained using MMI dataset. Accuracy of 97.25% on CK+ dataset is reported.

Breuer and Kimmel employed deep CNN visualization methods to examine the relation between CNN-learned features and AUs in [22]. They used architecture of three convolutional blocks (consisting of a convolutional layer of 5x5 filters, activation by ReLu and max pooling layer with 2x2 window) and two fully-connected layers to perform emotion classification. This architecture achieved 98.5% accuracy on CK+ dataset measured by 10-fold cross validation. After examining the neuron activation in individual layers they found high correlation between learned features and FACS AUs. They then performed transfer learning on the same architecture to detect individual AUs and found high accuracy of 97.5% in AU presence detection and 96.1% in AU intensity prediction. This work demonstrates viability of deep CNN networks in FER related tasks.

Submission to the 2015 Emotion recognition in the Wild challenge (EmotiW) by authors of [32] proposes using hybrid CNN-RNN network for video classification. CNN network is used to extract high-level representation of input frames. Multiple architectures of CNN network with various depths were tried. Since the data provided as part of the challenge contained only videos labelled with single emotion per video, other static datasets were used for training of the CNN network. It was observed that deeper architectures tend to overfit on the static datasets and therefore a 3 convolutional block (consisting of convolutional layer of 9x9 filters, ReLu activation and max-pooling) was chosen as the best contender. The features extracted by CNN were used as input for IRNN network (RNN of ReLu using initialization trick as described in [24]). In addition to appearance features extracted by CNN authors also used geometrical landmark features and audio features to enhance the performance of the final model. To combat different lightning conditions between datasets histogram equalization was applied to the images. Best reported accuracy on the test dataset provided as part of the challenge was 52.875% and showed an improvement over pure-CNN approaches, as seen in the Table 2 below.

**Table 1. Conventional FER approaches**

Reference	Emotions classified	Classification method	Validation dataset	Reported accuracy
[13]	7 basic	RBF SVM	CK+	87.9%
[15]	7 basic	RBF SVM	CK+	97.2%
[17]	6 basic	PCA	Custom	97%
[18]	7 basic	SVM	CK+	91.8%

**Table 2. Deep learning FER approaches**

Reference	Emotions classified	Classification method	Validation dataset	Reported accuracy
[19]	7 basic	VGG CNN	FER2013	55.6%
[21]	7 basic	DTAN & DTGN	CK+	97.25%
[22]	7 basic	CNN	CK+	98.5%
[23]	7 basic	RNN-CNN	EmotiW	52.875%

## 2.4 Gaps in the Existing Research

In this research, the issues related to pre-processing, feature extraction and classification are considered. Here, the issues related to the face detection, illumination, noise, occlusion, facial landmarks, segmentation, localization of facial regions, geometrical mapping, feature vector construction and classification are discussed. The proposed approach uses various aspects such as distance metric, degree of deformation and degree of direction, geometrical properties along with a suitable distance measure to address these issues. The proposed approach has addressed these issues by superimposing the grid over features and extracting the feature vectors by using the degree of distortion, the direction, and the magnitude of choosing best classifier for classifying emotions with highest classification accuracy.

## 3. EMOTION AND ITS EXPRESSION

This section aims to summarize psychological background of emotion and its manifestation in facial expression. Even though the science of emotion is an active field there is no universal definition of what emotion is and how to distinguish it from other psychological states. According to Dr. Paul Ekman emotion has following characteristics: There is a particular universal expression connected with given state which works as a signal, even though it might be very subtle. Presence of expression is not a sufficient evidence of presence of an emotion, as the expression can be simulated. While this characteristic does not help to clarify boundaries of emotion it is important to note as an explanation of universality. Emotional expressions involve multiple signals. There are limits on the duration of emotion. The duration of emotional expression is correlated to the strength of emotional experience (possibly modified by an attempt to manage said expression). Based on his cross-cultural experiments Ekman

identified seven basic emotions which are surprise, anger, fear, contempt, sadness, disgust and enjoyment.

### 3.1 Universality of Emotion Expression

The question of universality of expression is an important one in order to establish whether emotion recognition based on facial expression is a viable general method. Until the second half of 20th century most academics believed that expressions of emotion are culturally bound and that only members of same or similar culture express emotions in same way. Charles Darwin, however, thought otherwise.

In [33] author's argued that expressions of emotion were universal as they were a product of evolution. To support this claim he proposed three principles.

Principle of serviceable habits describes some expression habits as helpful and therefore reinforced by natural selection. An example would be raising the eyebrows to increase field of view in an event of danger (correlated with fear emotion). Antithesis principle states that some expressions, such as shoulder shrugging, exist merely because of their opposite nature to a serviceable habit. Some expressions, as proposed by the expressive habit principle, are a result of discharge of excitement in the nervous system. Vocal roar of anger would be an example of such expression [34].



**Figure 3.** AU examples. AU1: inner brow raiser, AU25: lips part, AU9: nose wrinkler, AU12: lip corner puller, AU6: cheek raiser.

In mid-1960s Paul Ekman took an interest in this issue. Based on hundreds of hours of film capturing isolated cultures in New Guinea highlands, taken by Carleton Gajdusek and Richard Sorenson, Ekman found that in response to given stimuli the face expressions observed were in accordance with his expectations. No culturally unique expressions were observed either. Even though he was leaned towards the culturally relativistic viewpoint at first, this experience swayed him that Darwin might be right and inspired him to travel to the New Guinea highlands.

After conducting his own experiments and collecting supporting evidence for universality of expression, Ekman came up with the idea of "display rules" (a set of socially learned, culturally unique behaviors that are used to mask, exaggerate, diminish or exhibit expressions in specific cultural contexts) that would explain culture-based differences in expressions. In late 1960s he gathered evidence supporting this explanation by conducting a study of students in Tokyo, Japan and Berkeley, California. Author found that both Japanese and American students reacted the same way to emotion inducing clips as long as they were filmed alone by a hidden camera. However, when a scientist entered the room, Japanese students masked negative expressions with positive ones [27]. Some of the AU examples are presented in Figure.3.

Expression universality has been widely accepted as a number of cross-cultural studies yielded supporting results. In [18] authors argue that only a few of these studies were truly cross-cultural. They claim that cultures that have been exposed to the western culture have adapted their emotion expressions and concepts. Furthermore, in the studies that were truly cross-cultural (such as Ekman's experiments in New Guinea), an emotion conceptual context was included in the experimental method by asking the subjects to assign facial expression to word or description. Their free label experiment with participants from Himba ethnic group and America did not find supporting evidence for expression universality and authors are suggesting that emotion expressions are actually culture based to some degree. Whether the emotion expression is truly universal or not, the findings of universality between cultures exposed to western culture is sufficient for vast majority of potential FER applications.

### 3.2 Expression Measurement

In order to be able to measure and describe facial expression Dr. Paul Ekman and Dr. Wallace Friesen developed an anatomically-based system designed to measure human facial movements called Facial Action Coding System (FACS). Action Units (AUs) are used to characterize face movements that result in changes in appearance. Action unit is a numeric code that represents muscle activity of certain facial muscles or muscle groups. FACS distinguishes 46 different AUs (e.g. AU1 - Inner brow raiser, AU23 - Lip tightener).

Resulting FACS code is a string of present AUs. Presence of emotion is decided based on rules of presence of certain AUs. Even though FACS was primarily developed to help describing facial expressions while studying emotion it is a robust system that can be used in other areas as well [19].

## 4. PROPOSED METHOD

The proposed approach is divided into three phases: Pre-processing, Feature extraction and Classification phases and are discussed below.

### 4.1 Pre-processing

The Viola Jones algorithm (Jain and Choudhary, 2018) [21] is used to recognize faces in pictures during the pre-processing phase. On the greyscale picture, the Viola-Jones method is used to identify the face. Using Haar-like properties, it defines a box and searches for a face inside it. The  $M \times N$  image is taken as input and a window scale multiplier is used as a parameter to compute the windowed images mean and standard deviation. A cascade of false negative and positive rates is validated. The positive group of windows declared positive by the cascade is obtained as output and the face is detected. The detected face is cropped using background subtraction method where the foreground image is subtracted from the background image and the facial part is cropped effectively and the image is rescaled for  $132 \times 132$  pixels. The rescaling is necessary pre-processing step as the image with higher pixel values appear to be fuzzier and pixelated. Also to bring multiple image files to a common scaling in the Dataset,

rescaling is important. Scaling down the image makes the feature extraction process easier; hence, the face images are rescaled for  $132 \times 132$  pixels. Further, the approach use normalization and filtering techniques to handle illumination invariant, transition and occlusions. To detect edges, the proposed FER system makes use of the Sobel edge detection technique (Singh and Singh, 2016) [22]. The Sobel edge detection technique is technically a discrete differential operator, and it is used to compute an approximate estimation of the gradient of the image brightness function in the proposed FER system. In mathematics, the Sobel operator is a typical edge detection operator that is based on the first derivative of a function. When using a similar local average operation as a result of the operator's introduction of the operation, the noise has a smooth impact and may effectively limit the influence of noise.

### 4.2 Feature Extraction

The feature extraction phase is the second phase in proposed FER system. The proposed approach considers the six basic emotions along with *neutral* expression as reference. Figure 4 shows these basic emotions. The approach considers eyes, eyebrows, nose and mouth as facial components.



Figure 4. Sample Face Expressions

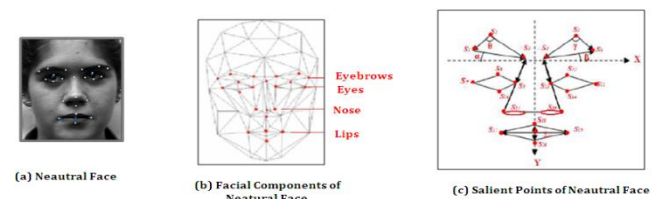


Figure 5. Salient Points of Facial Components in Neutral Face

In this phase, the pre-processed training set images are processed for extracting feature. The proposed approach initiates with image detecting facial landmarks using ensemble of regression tree approach (Kazemi & Sullivan, 2014) [23]. Considered here is a series of tagged images of face landmarks for training purposes. This data is used to train the regression trees in an ensemble to predict face landmark locations directly from the pixel intensities.

Thus, the facial landmarks of eyes, eyebrows, nose and mouth components are detected accurately for further processing. The image is segmented based on Region of Interest (RoI) such as

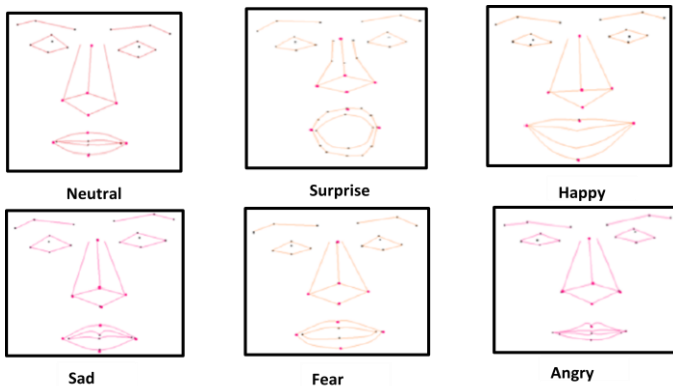


eyes, mouth, nose etc. by using Threshold segmentation. Using the grey values of several targets, this method separates the picture into distinct sections. Using the local threshold technique (Kohler, 1981), a large number of segmentation thresholds are determined, and the image is separated into several target regions and backgrounds by a large number of segmentation thresholds [24].

The salient points are identified and extracted using the facial landmarks from the segmented RoI. The salient points that are extracted from the RoI include facial components such as eyes, eyebrows, nose and mouth and proceeded to analyse the facial expressions. The Figure.5 (a), (b) and (c) depicts the salient points of facial components. For instance, mouth component has five salient points represented as  $S_{17}, S_{18}, S_{19}, S_{20}, S_{21}$ . Nose component has two salient points  $S_{15}, S_{16}$ . Left eyes and right eyes together have eight salient points  $S_7, S_8, S_9, S_{10}$  and  $S_{11}, S_{12}, S_{13}, S_{14}$ . The left and right eyebrows together have six salient points  $S_1, S_2, S_3$  and  $S_4, S_5, S_6$ . The movement of eyebrows are represented in terms of angles  $\alpha, \beta, \gamma$  and  $\theta$  and based on Geometric structures the expressions are analysed as explained in detail in below sub-section.

### 4.3 Geometric Structure Analysis

This sub-section presents the salient points and geometric structure of facial components for *neutral* and other facial expressions. This is depicted in Figure 6. The proposed FER system considers all facial expressions along with *neutral* expression for geometric structure analysis. For want of space and clarity, the sub section shows the *neutral* and *happy* facial expressions alone. Also, the proposed FER framework analyze the geometric structure of all facial components, for simplicity, nose and mouth alone are presented.

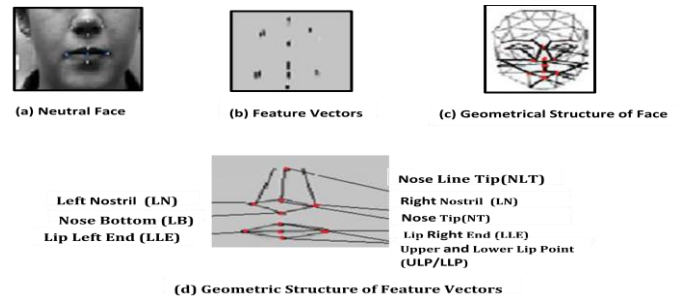


**Figure 6.** Salient Points and Geometric Structure of Facial Expressions

#### 4.3.1 Geometric Structure Analysis of Neutral Face

The proposed approach considers *neutral* face as reference face. Hence, the geometric structure analysis of *neutral* face is necessary and is captured. Based on the facial landmarks, the salient points are extracted that locate two nostrils. This includes sharp edge of the left nostril and right nostril along with lower tip of the both nostrils. The mouth component extracts salient points of the lips which represents left end point, right end point, upper and lower lip mid points along with curve. The salient points that are extracted from the nose

and mouth components supports in building the geometric structure for these in the *neutral* face. Figure 7 shows the salient points and its geometric structure for nose component in *neutral* face. The distances are computed between the salient points for each component. Similarly, the salient points that are extracted for eyes and eyebrows are linearly combined and considered as geometric structure.



**Figure 7.** Salient Points and Geometric Structure for Nose and Mouth components in Neutral face

Table 3 Represents nose and mouth components feature vectors and their salient points that represents the pixel information (x, y). The pixel information is used to compute the distances.

The proposed approach analyses the facial expressions in terms of distance and angle measurement using distance metric.

**Table 3. Nose and Mouth Component: Feature Vectors and their Salient Points for Neutral face**

Facial Components	Feature Vectors	Salient Points (X, Y)
Nose	Nose Tip (NT)	(327, 356)
	Nose Bottom (NB)	(326, 409)
	Left Nostril (LN)	(253, 381)
	Right Nostril (RN)	(390, 373)
	Nose Line Tip (NLT)	(317, 242)
Mouth	Lip Left End (LLE)	(230, 477)
	Lip Right End (LRE)	(415, 474)
	Upper Lip Point (UL)	(327, 452)
	Lower Lip Point (LL)	(323, 504)

**Table 4: Nose and Mouth Component: Distance Measure for Neutral face**

Facial Components	Feature Vectors	Distance Measure
Nose	Distance(NLT & NT)	114
	Distance (LN & RN)	137
	Distance (NB & UL)	43
Mouth	Distance (LLE & LRE)	185
	Distance (UL & LL)	52
	Distance (LRE & LL)	97

#### 4.3.2 Euclidian Distance Measure

The Euclidean distance metric, which reflects the normal distance between two locations and can be calculated using the Pythagorean formula, is defined as follows: The salient points of facial components for *neutral* expressions are computed using distance metric. The approach uses Euclidian distance metric for computing the distance between the salient points within the component and across the components. This is depicted in Eq. (1). For instance,  $d_1$ ,  $d_2$ , and  $d_3$  represents three Euclidean distances between salient points within the component and  $d_4$ ,  $d_5$ , and  $d_6$  represents three Euclidean distances (ED) between salient points across the component and is shown below:

$d_1$  is the ED between inner corners of left and right eyebrows (LEB-IC & REB-IC),  $d_2$  is the ED between left and right nostrils (LN & RN),  $d_3$  is the ED between upper and lower Lip Points (UL & LL),  $d_4$  represents top of the nose and Outer corner of the right eye,  $d_5$  represents the top of the nose and Inner corner of the left eye,  $d_6$  is represents the top of the nose and Left mouth corner.

Consequently, the distance formula translates Euclidean space to metric space, as shown in the diagram. [PQ] is the length of the line segment that connects points P and Q, and it represents the Euclidean distance between them. If  $P = (p_1, p_2, p_n)$  and  $Q = (q_1, q_2, q_n)$  are two points in Euclidean n-space, then the distance between P and Q is given by:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Where  $(p_1, q_1)$ ,  $(p_2, q_2)$  are coordinates of salient points. When using FER systems, these distances are utilised to compare a neutral expression face with various facial expressions like as happy or sad, among other things. All totally, 160 Euclidean distances were calculated between all pairs that could be made up with the 20 salient points that constitute a vector of 160 elements. This is mathematically represented in Eq. (2). This is depicted in Table 4.

$$D_e = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{160} \end{bmatrix} \quad (2)$$

For the automated facial expression identification system, this vector reflects the neutral expression of the picture sample. Same procedure is carried out for other facial expressions also. This is presented in below sub-section.

#### 4.3.3 Geometric Structure Analysis of Happy Face

The proposed FER system considers other facial expressions such as *happy* expression for the same subject (person) and same procedure is carried out. The salient points with respect to *happy* expression are extracted and its geometric structure is analysed. Figure 8 represents the salient points of mouth and nose components and their geometric structure for *happy* face.

Later, the distances and angles are computed between the salient points within the components and across the components.

Table 5 Represents nose and mouth feature vectors in terms of salient a point that represents the pixel information (x, y) for happy face and is used to compute the distances.

The approach uses Euclidian distance metric for computing the distance and angles between the salient points within components and across components. This is depicted in Table 6. Similarly, the extraction of salient points and geometric structure mapping for other facial expression such as *sad*, *happy*, *surprise*, *disgust*, *angry* and *fear* are performed. Distance and angle computation is performed using distance metric. The vector that represents the basic facial expressions of the same subject is given as input to classification algorithm. The classifiers learn each facial expression and compare them with the *neutral* expression (reference). The classification procedure is discussed in detail in below sub-section.

**Table 5. Nose and Mouth Component: Feature Vectors and their Salient Points for Happy Face**

Facial Components	Feature Vectors	Salient Points (X, Y)
Nose	Nose Tip (NT)	(329,369)
	Nose Bottom (NB)	(332,409)
	Left Nostril (LN)	(253,373)
	Right Nostril (RN)	(401,367)
	Nose Line Tip (NLT)	(324,245)
Mouth	Lip Left End (LLE)	(201,451)
	Lip Right End (LRE)	(451,445)
	Upper Lip Point (UL)	(323,437)
	Lower Lip Point (LL)	(320,528)

**Table 6. Nose and Mouth Component: Distance measure for happy face**

Facial Components	Feature Vectors	Distance Measure
Nose	Distance (NLT & NT)	124.1007
	Distance (LN & RN)	148.1215
	Distance (LN & NT)	76.1051
	Distance (RN & NT)	72.0277
	Distance (NT & NB)	40.1123
	Distance (NB & LN)	86.8158
	Distance (NB & RN)	80.7774
	Distance (NB & UL)	137.5681
Mouth	Distance (LLE & LRE)	250.0719
	Distance (UL & LL)	91.0494
	Distance (LLE & UL)	122.8006
	Distance (LLE & LL)	141.7391
	Distance (LRE & UL)	128.2497
	Distance (LRE & LL)	155.0806



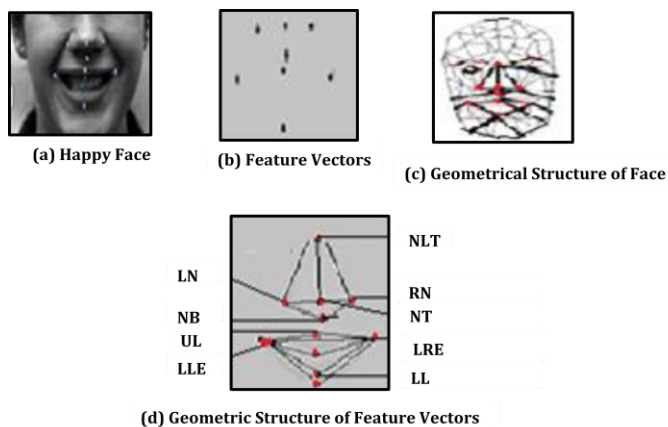
#### 4.3.4 Facial Expression Classification

After computing the distances and angles between these salient points for facial expressions *anger*, *disgust*, *fear*, *sad*, *surprise* and *happy*, the proposed FER system compares these distance measures with the *neutral* face expression. This is given to classification algorithms as hypothetical rules as depicted in Eq. (3)

$$\begin{aligned} \text{if } Dist_{sad} (S_{17}\Delta S_{18}\Delta S_{19}\Delta S_{20}\Delta S_{21}) = \\ = \text{if } Dist_{Neutral} (S_{17}\Delta S_{18}\Delta S_{19}\Delta S_{20}\Delta S_{21}) \end{aligned}$$

then Class: Neutral else Class: Sad (3)

For experimentation, the JAFFE Dataset is divided into training set and testing set using *n*-fold cross validation to train the SVM. The facial images are segregated into different known classes. Based on the hypothesis, the SVM classifier is trained for the predicting the facial expressions. Facial expressions were recognised using a multi-class binary SVM classifier.



**Figure 8.** Salient Points and Geometric Structure of Happy face

For example, the SVM is a supervised classification approach that doesn't have any restrictions on how the data is distributed. Each input ( $S_i, S_j$ ) comprises *k* characteristics that describe the class 'Emotional Expressions,' and this model is trained using the linear kernel function. Once a categorization system for one kind of expression is mastered, it may be applied to all other types. Discrimination of all six possible emotional expressions has been gathered here.

## 5. RESULTS AND DISCUSSION

A number of tests using JAFFE image databases are carried out in order to put the suggested approach to the test and assess its performance. The proposed system considers eyes, eyebrow, nose and mouth facial components of the input images and extracts salient points using facial landmarks. The geometric structures are mapped for each of these components and the distance and angle between the salient points for these components are computed. These were used as input for the SVM classifier, which was trained to learn and categorise facial expressions and obtained an accuracy of 86 percent. Aside from that, the suggested FER system makes use of

multilayer perceptrons and random forests, both of which are extensions of SVM classifiers, for classification. In SVM, the Lib SVM classifier is used in conjunction with the Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid classifiers as kernels, among other things. MLP and RF models are also employed in the training process. The RF tree makes use of the resampling method to choose a random subset of the data from the Dataset.

Table 7 Displays the JAFFE Dataset's classification accuracy using LibSVM and RBF. Using the confusion matrix, you can see how many variations there are for each face expression. As an example, in Table 3, out of 30 surprise facial expressions, 22 are correctly labelled as surprise, whereas 8 are incorrectly classified as furious (as well as joyful and disgust). Overall, 176 of 213 photos were properly classified by the suggested method, with an overall accuracy of 82.65%. Table 7 shows the categorization accuracy for each facial emotion. Accuracy using Lib SVM (RBF) on JAFFE Dataset is presented in Figure 9. Table 8 shows the accuracy of JAFFE's categorization using MLP. For happy expressions, categorization accuracy is greatest, while for frightened expressions, it is lowest. The total accuracy of categorization is 88.31%. Accuracy using MLP on JAFFE Dataset is presented in Figure 10.

**Table 7. Classification Accuracy using Lib SVM (RBF) on JAFFE Dataset**

Facial Expressions	Ang ry	Disg ust	Fe ar	Hap py	Neut ral	Sa d	Surp rise	Classific ation Accurac y (%)
Angry	25	0	3	0	0	2	0	83.33
Disgust	2	24	1	0	0	2	0	82.75
Fear	1	2	25	0	2	2	0	78.12
Happy	0	0	0	27	2	0	2	87.09
Neutral	0	1	0	0	28	1	0	93.33
Sad	0	1	0	0	3	25	0	80.64
Surprise	2	2	0	4	0	0	22	73.33
Overall Accuracy								82.65

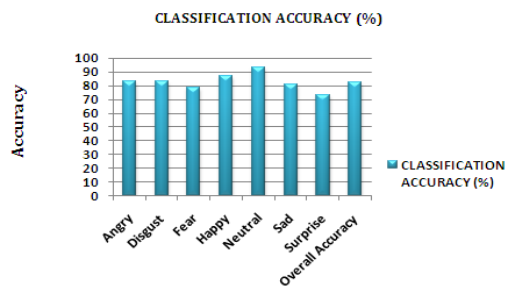
**Table 8. Classification Accuracy using MLP on JAFFE Dataset**

Facial Expressions	Ang ry	Disg ust	Fe ar	Hap py	Neut ral	Sa d	Surp rise	Classific ation Accurac y (%)
Angry	27	0	0	0	2	1	0	90.00
Disgust	2	26	1	0	0	0	0	89.65
Fear	1	2	26	0	2	1	0	81.25
Happy	0	0	0	29	0	0	2	93.54
Neutral	0	0	0	0	27	3	0	90.00
Sad	0	1	0	0	3	27	0	87.09
Surprise	2	0	0	2	0	0	26	86.66
Overall Accuracy								88.31

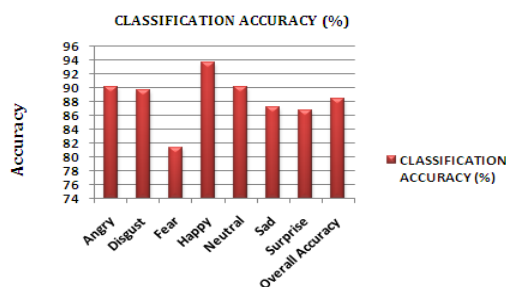
**Table 9. Classification Accuracy using RF on JAFFE Dataset**

Facial Expressions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Classification Accuracy (%)
Angry	26	0	0	0	2	2	0	86.66
Disgust	2	25	1	0	0	1	0	86.20
Fear	1	0	28	0	2	1	0	87.50
Happy	0	0	0	29	2	0	0	93.54
Neutral	0	0	0	0	28	2	0	93.33
Sad	0	2	0	0	3	26	0	83.87
Surprise	0	0	0	3	0	0	27	90.00
Overall Accuracy								88.72

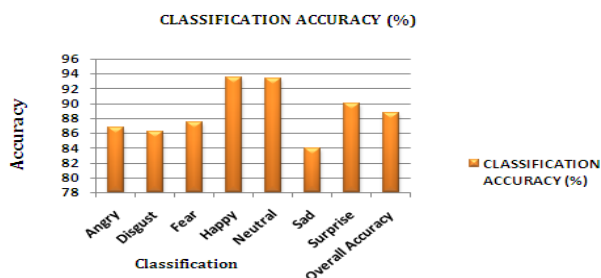
Table. 9 shows the classification accuracy of the RF method. The precision has increased marginally and now stands at 88.72 percent. For happy, neutral, and surprised expressions, the categorization accuracy has improved. Accuracy using RF on JAFFE Dataset presented in Figure 11.



**Figure 9. Accuracy using Lib SVM (RBF) on JAFFE Dataset**

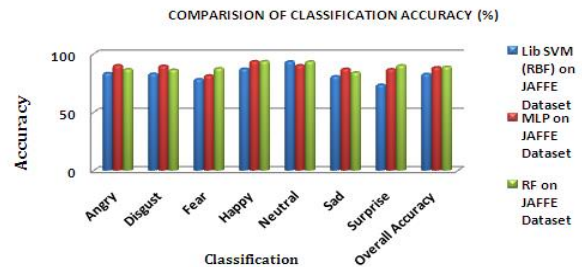


**Figure 10. Accuracy using MLP on JAFFE Dataset**



**Figure 11. Accuracy using RF on JAFFE Dataset**

As can be seen from the aforementioned findings, the degree of precision in determining for happy and neutral expressions is high and for certain expressions, it is low. Thus, it is noticed that the proposed FER system performance on JAFFE Dataset using Euclidian distance between all pairs of salient points within and across facial components have given encouraging results. Comparison of Classification Accuracy is presented in Figure 12.



**Figure 12. Comparison of Classification Accuracy**

The classification algorithms such as LibSVM, MLP, RF has achieved classification accuracy of 86.56% on an average. The findings of the experiments show that the extraction of picture characteristics is more efficient in terms of computing and gives promising outcomes.

## 6. CONCLUSION

The paper presents the analyses of human facial structure along with its components using Action Unit and Geometric structure for identifying human facial expressions. The approach considers facial components such as Nose, Mouth, eyes and eye brows for FER. Nostril contours such as left lower tip, right lower tip, and center tip are considered as salient points of Nose. Various salient points for Mouth are extracted from the left and right end point, upper and lower lip mid points along with curve. These salient points are extracted for all facial expression of the same subject considering *neutral* face as reference. The Geometric structure for *neutral* face is mapped along with other facial expression faces. The deformation is estimated using the Euclidean distance. The classifier is trained for predicting the facial expressions for various basic emotions of input image. The experimental results are encouraging.

## 7. ACKNOWLEDGMENTS

We thank our colleagues from Yenepoya Institute of Technology who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

## REFERENCES

- [1] Chen W, Joo EM, Shiqian W (2006) Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE transactions on systems, man, and cybernetics*, 36: 458–66. <https://doi.org/10.1109/TSMCB.2005.857353>.
- [2] Gangolli SH, AJL Fonseca, Sonkusare R (2019). Image Enhancement using Various Histogram Equalization Techniques, In: Global Conference for Advancement in Technology (GCAT), Bangalore, India, pp. 1-5, doi: 10.1109/GCAT47503.2019.8978413.
- [3] Jain U, Choudhary K, Gupta S, Privadarsini MJP, (2018). Analysis of Face Detection and Recognition Algorithms Using Viola Jones

- Algorithm with PCA and LDA, In: *2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 945-950, doi: 10.1109/ICOEI.2018.8553811
- [4] Singh, M. Singh, Singh B (2016). Face detection and eyes extraction using Sobel edge detection and morphological operations, In: *Conference on Advances in Signal Processing (CASP)*, Pune, India, pp. 295-300, doi: 10.1109/CASP.2016.7746183.
- [5] Ekman P, Friesen W (1978) Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists*, Palo Alto, CA, USA, Tech. Rep.
- [6] Li X, Ji Q (2005). Active affective state detection and user assistance with dynamic Bayesian networks, *IEEE Transactions on Systems, Man, and Cybernetics, A, Syst., Humans*, 35(1):93-105.
- [7] SumaLakshmi CH, Vasuki P (2020). Performance Improving of ANN with Preprocessing Stage in Human Face Recognition System. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(4):2278-3075.
- [8] Narayan T, Deshpande, Ravishankar S (2017). Face detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN. *Advances in Computational Sciences and Technology*, 10(5):1173-1189.
- [9] Wagh P, Chaudhari J, Thakare R, Patil S (2015). Attendance System based on Face Recognition using Eigen Face and PCA Algorithms. In: *IEEE International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 303-308.
- [10] Kumar GS, Prasad Reddy PVGD, Gupta S, Kumar RA (2012). Position Determination and Face detection using Image Processing Techniques and SVM Classifier. In: *Proceedings of the International Conference on Information Systems Design and Intelligent Applications*, pp. 227-236.
- [11] Kavinmathi B, Varadharajan E, Dharani R, Jeevitha S, Hemalatha S (2016). Automatic Attendance Management System using Face Detection. In: *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, Coimbatore, India, pp. 1-3, doi: 10.1109/GET.2016.7916753.
- [12] Zeng D, Veldhuis R, Spreuwers L (2020). A survey of face recognition techniques under occlusion, *IET Biometrics*, <https://doi.org/10.1049/bme2.12029>.
- [13] Kahatapitiya K, Tissera D, Rodrigo R (2019). Context-Aware Automatic Occlusion Removal, In: *IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, pp. 1895-1899, doi: 10.1109/ICIP.2019.8803141.
- [14] Srinivasan A, Balamurugan V, (2014), Occlusion Detection and Image Restoration in 3D Face Image, *TENCON 2014 IEEE Region 10 Conference*, Bangkok, Thailand, pp. 1-6, DOI: 10.1109/TENCON.2014.7022477.
- [15] Ma R, Mohamed A (2015), Image Processing Pipeline for Facial Expression Recognition under Variable Lighting, In [https://web.stanford.edu/class/ee368/Ma\\_Mohamed.pdf](https://web.stanford.edu/class/ee368/Ma_Mohamed.pdf).
- [16] Guo F, Yang J, Chen Y, Yao B (2018). Research on Image Detection and matching based on SIFT Features. In: *3rd International Conference on Control and Robotics Engineering (ICCCE)*, Nagoya, Japan, pp. 130-134, doi: 10.1109/ICCCE.2018.8376448.
- [17] Pal MK, Porwal A (2015). A Local Brightness Normalization (LBN) Algorithm for destriping Hyperion images. *International Journal of Remote Sensing*, 36(10): 2674-2696. <https://doi.org/10.1080/01431161.2015.1043761>
- [18] Zhang T, et al.(2009). Multiscale facial structure representation for face recognition under varying illumination. *Pattern Recognition*, 42(2):251-8. <https://doi.org/10.1016/j.patcog.2008.03.017>.
- [19] Feng Y, Zhao H, Li X, Zhang XL, Li HP (2017). A multi-scale 3D Otsu thresholding algorithm for medical image segmentation, *Digital Signal Processing*, 60: 186-199, <https://doi.org/10.1016/j.dsp.2016.08.003>.
- [20] Vezzetti E, Marcolin F, Tornincasa S, Ulrich NDL (2017). 3D geometry-based automatic landmark localization in presence of facial occlusions, *Multimedia Tools and Applications*, 77: 14177-14205, 10.1007/s11042-017-5025-y.
- [21] Jain V, Lamba PS, Singh B, Namboothiri N, Dhall S (2019). Facial Expression Recognition using feature level fusion, *Journal of Discrete Mathematical Sciences and Cryptography*, 22:337-350, <https://doi.org/10.1080/09720529.2019.158266>.
- [22] Singh, M. Singh, Singh B (2016). Face detection and eyes extraction using Sobel edge detection and morphological operations, In: *Conference on Advances in Signal Processing (CASP)*, Pune, India, pp. 295-300, doi: 10.1109/CASP.2016.7746183.
- [23] Kazemi V, Sullivan J, (2014). One millisecond face alignment with an ensemble of regression trees, In: *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1867-1874, doi: 10.1109/CVPR.2014.241.
- [24] Kohler R, (1981). A segmentation system based on thresholding. *Computer Graphics and Image Processing*, 15(4): 319-338.
- [25] EKMAN, P. Introduction. *Annals of the New York Academy of Sciences*, volume 1000, no. 1, 2003: pp. 1-6, ISSN 1749-6632, doi: 10.1196/annals.1280.002.
- [26] Ekman, P. Happy, sad, angry, disgusted. Volume 184, 10 2004: pp. 4-5.
- [27] Darwin, C. *The Expression of the Emotions in Man and Animals*. Cambridge Library Collection - Darwin, Evolution and Genetics, Cambridge University Press, second edition, 2009, doi:10.1017/CBO9780511694110.
- [28] Gendron, M.; Roberson, D.; et al. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, volume 14, no. 2, Apr 2014: pp. 251-262.
- [29] Zhu, Q.; Yeh, M.-C.; et al. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, IEEE, 2006, pp. 1491-1498.
- [30] Kazemi, V.; Josephine, S. One millisecond face alignment with an ensemble of regression trees. In *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*, IEEE Computer Society, 2014, pp. 1867- 1874.
- [31] Ebrahimi Kahou, S.; Michalski, V.; et al. Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ACM, 2015, pp. 467-474.
- [32] Lucey, P.; Cohn, J. F.; et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 *IEEE Computer Society Conference on*, IEEE, 2010, pp. 94-101.
- [33] Szegedy, C.; Vanhoucke, V.; et al. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.
- [34] Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, volume 9, no. 3, 2007: pp. 90-95, doi:10.1109/MCSE.2007.55.
- [35] Stephen Raj. S and SriPriya. P (2021), Eye Drowsiness Tiredness Detection Based on Driver Experience Using Image Mining. *IJEER* 9(1), 1-5. DOI: 10.37391/IJEER.090101.



© 2022 by the Shashank M Gowda and H N Suresh Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).