International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 11
ISSN: 2454-4248
510 – 519

An Analysis of Facial Expression Recognition Techniques

Vandana Patidar(PG Student)[1*], Dr. Devang Pandya[2], Asst. Prof. Dharmesh Tank[3]
[1*]Computer Engineering, Silver Oak College Of Engineering and Technology, Gujarat Technical University, Ahmedabad, India
[1]*patidar.vandana@gmail.com*, [2]*devang1979@gmail.com*, [3]*dharmesh.tank1@gmail.com*

***Abstract*—** In present era of technology , we need applications which could be easy to use and are user-friendly , that even people with specific disabilities use them easily. Facial Expression Recognition has vital role and challenges in communities of computer vision, pattern recognition which provide much more attention due to potential application in many areas such as human machine interaction, surveillance , robotics , driver safety, non- verbal communication, entertainment, health- care and psychology study. Facial Expression Recognition has major importance ration in face recognition for significant image applications understanding and analysis. There are many algorithms have been implemented on different static (uniform background, identical poses, similar illuminations) and dynamic (position variation, partial occlusion orientation, varying lighting)conditions. In general way face expression recognition consist of three main steps first is face detection then feature Extraction and at last classification. In this survey paper we discussed different types of facial expression recognition techniques and various methods which is used by them and their performance measures.

Keywords-*Facial Expression recognition, Face detection,Feature Extraction, Facial Action Coding System, classification*

******

I. INTRODUCTION

Facial Expression Recognition(FER) is progressive area of research. It is interesting part of image processing. Now- a-days image processing techniques give significant emphasis on two regions one is upgrade of pictorial data for human elucidation and second is processing of image for storage transmission and portrayal for automatic machine perception. Emotion recognition can be done by different modalities such as speech, body gesture, facial expression. From all the modalities facial expression has given much attention because of its potential value for practical applications and its theoretical challenges. The challenges associated with face expression recognition is varying faces from one person to another person due to different ages, pose variations, occlusion, illumination conditions etc. Instead of this challenges facial expression has many applications. It plays vital role in non verbal communication. Nonverbal communication is communication between people by sending and receiving wordless clues, in other words the people with disabilities to understand other people emotions. Automatic facial expression recognition has applied in various areas for example ad-liking prediction, student automatic e-learning, detection of unlawful activities, face expression synthesis, surveillance, lie detection, music for mood, mental state identification etc. It has been very active over past decades in communicable interests(behaviour science, human computer interaction, health-care and security).Many researches covers the design and development of facial expression surveillance system based on the application domain of security surveillance systems and facial expression recognition. Although there are recent applications of facial expression recognition in human machine interaction(HCI)-where human machine interaction is ability to automate the capture of emotion personality traits and cognitive intention from facial expression could be a significant advancement.

We begin with overview of facial expression recognition(FER). Facial expression is one or more motion and positions of muscles beneath the skin of the face. According to Chin and Kim(at 2009) "Facial expression is formed by relaxing and contracting different muscles of human face. " By Ekman and Frisen[1] (at 2003) " Facial expressions are rapid signals which differs with change in facial features like open mouth, raising eyebrow, lips, eyes, cheeks etc and these features movement affects the performance of system. " In 1971 Ekman and Frisen discovers six different expressions such as happy, sad, disgust, fear, surprise, anger along with their neural face. These facial emotions are universally accepted as core emotions. Paul Kman and Wallance (in 2002) described a new expression named as contempt. This ex

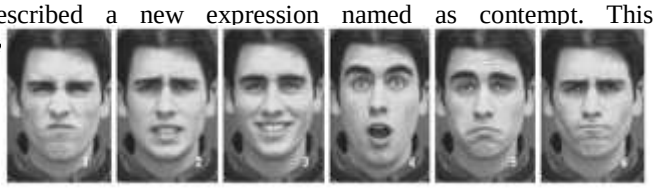

Figure 1: Emotion specified Facial Expressions(1-Disgust, 2 Fear, 3-Happyness, 4-Surprice, 5-Sadness, 6-Anger)[44]

FACS is widely used in both psychology and computer science fields. This system described visually distinguished facial movement(face action activities)[2].In regards of FACS every facial expression can be represented or decompose into one action unit or combination of action units. In psychology research study any one can express his feeling and attitude by speaking of message contribute up to 7%, vocal part contribution up to 38% ,while facial expression contribute 55%[3].

II. FACIAL EXPRESSION RECOGNOTION

Facial expression recognition process organize in five basic phases which is shown in figure 2. In first phase take the images as input .Input may be static images or sequences of image(video frames) in expression recognition. The second phase performs various pre-processing techniques such as noise reduction ,image enhancement on the input image. Nose reduction is process of removing noise from image. Noise can be random or white noise with no coherence, or coherent noise introduced by the device's mechanism or processing algorithm. Normalization is one of the pre-processing technique to

 ISSN: 2454-4248
510 – 519

remove illumination problem. Normalization is the process that changes the range of pixel intensity values. Normalization against variation of pixel position or brightness is done.

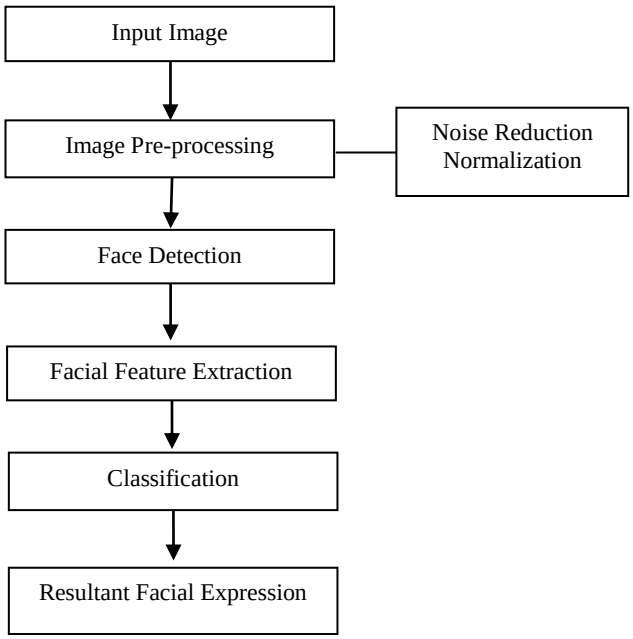

Figure:2 Facial Expression Recognition Block Diagram

Next phase described face detection and various approaches of face detection

A. *Face Detection*

Face detection is essential phase of emotion recognition in light of facial expression. The objective of face detection is utilized to recognize and find any face present in the image , in straightforward words we say that it is process identifying the face from input image. For identification and understanding the mood of people, it is very important that computer recognize face very effectively. Amid the face detection process different issues experience to be broke down, for example, recognizing geometry(front and non-front) illumination (colour, shadow and self shadow) the imaging procedure, for example, resolution , occlusion, focus and their related effects.[4]. Face detection isn't simple since scale and orientation of the face can change starting with one face then onto the next face. Facial feature extraction progress toward becoming fall flat if the test image has distinctive lighting condition than their preparation images. There are four diverse methodologies of face detection is given as[5]:

1) Knowledge based approach: This approach detect the face using set of rules(like human has 2 eyes ,nose, mouth, lips, along with their certain distance and position co-relate to each other) based on human knowledge.[4-5]

2) Template based approach: The image which is represented as a bi-dimensional (2-D) array of intensity values is compared with suitable metric (like Euclidean distance)[6]. In this approach a face is stored in a set of distinct smaller templates represented by multiple view points. The deformation Model (DPM) is heuristically built into the metric used by the matching measure this technique is known as elastic template[45]. The Viola and Jones (V & J)[6] face detector is widely used in face detection from image.

3) Feature invariant approach: This approach is depends on the concept of structural features of face in which structural classifier is used to classify facial region and non- facial region[7].

4) Appearance based approach: This approach is based on set of empower training face images to design face models. Face detection techniques are generally used with eigen values for face detection. According to research appearance based approach has better performance than other once and following table shows summarization of different approaches and methods[8].

TABLE I: ABSTRACTION OF DIFFERENT FACE DETECTION APPROACHES [8]

Sr. No.	Face Detection Approaches	Methods
1	Knowledge based	Multi-resolution rule based method Grouping of edges
2	Feature invariant	
	-Facial feature	Space Gray-level Dependence
	-Texture	Matrix of face pattern(SGLD)
	-Skin Color	Mixture of Gaussian (GMM)
	-Multiple Feature	Integration of skin color, size
3	Template matching-	
	Predefined face templates	Shape template Active Shape Model (ASM) Eigen Vector decomposition and clustering
	-Deformable Templates	Gaussian distribution and multilayer Perceptron
4	Appearance based	
	-Eigen face	Ensemble of neural networks
	-Distribution based	Arbitration scheme

B. *Facial Feature Extraction*

Feature extraction transform the pixel image data into higher level representation of motion, appearance of inner structures. Feature extraction methods extract holistic features from the initial representation features of face. They map an input representation onto a lower dimensional space to discover a delegate structure from the representation. This is important phase The main goal is reducing the dimensionality of the input space and to minimize the variance in the data caused by sudden conditions such as lighting, alignment errors or (motion) blur, and to reduce the sensitivity to contextual effects such as identity and head pose. Although several techniques have been proposed, segmentation of features still has wide fields of research. According to various research study feature extraction has many methods. In this survey paper we categorize feature extraction methods in three categories.

 ISSN: 2454-4248
510 – 519

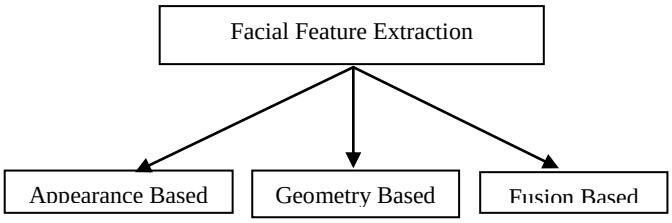

Figure 3: Categorization of Feature Extraction Methods

1. *Appearance Based Facial Feature Extraction* - Nowadays appearance based feature extraction methods are most commonly used. This methods describe the color and texture of a facial region. This method is depends on the pixel intensities in the texture calculations, and tries to build a model utilizing the learning database of faces. Appearance Features narrate the variations in face texture when specific movement is performed such as wrinkles, bulges, forefront, regions surrounding the mouth and eyes. Image filters are used, applied to either the entire face or specific regions in a face image to extract a feature vector[9]. The following section gives a brief description of some of the appearance methods widely used for feature extraction.

1) Principle Component Analysis(PCA): PCA is widely used statistical technique for dimensionally reduction in facial extraction . PCA is also known as Eigen space projection because it is based on linearly projection the image space to a low dimension feature space that is known as eigen space. There are many PCA-based face-recognition systems have been developed in the last decade. A 2-D facial image can be represented as 1-D vector by concatenating each row (or column) into a long component vector . The PCA algorithm[10] work as follows for feature extraction: The image of rectangular matrix is converted into column vector which contains the mean value of the each row. Normalization vector is calculated by determining the difference between column vector of an image and the mean vector calculated from all column vectors.

$$\mu = (1/m)\sum_{n=1}^{m} X_n$$

$$c = (1/m)\sum_{n=1}^{m} (X_n - \mu)(X_n - \mu)^T \qquad (1)$$

The mean and covariance matrix is calculated by using equation (1)[10]. This normalization vector is given as an input to the principal component analysis which returns score in the image. The economy of principal component returns only the elements of latent that are not necessarily zero. There is also some drawback of PCA like poor discriminating power within the class and it has taken large computation time.

2) *Eigen Faces:* Eigen faces[12] are a set of eigenvectors used in the human face recognition. Specifically, the Eigen faces are the primary components of a distribution of faces, or equivalently the eigenvectors of the covariance matrix of the set of face images, where an image with N x N Pixels is reviewed as a point (or vector) in 2-dimensional space. Mathematically, it is observed that the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, considered an image as a point or a vector in a very high dimensional space. These eigenvectors can be imagined as a set of features that together characterize the variation between face images. Each image locations contribute more or less to each eigenvector, so that we can display the eigenvector as a sort if "shadowy" face which we call an eigen face. Eigen face algorithm falls under two stages. In first step initialize the eigenvectors by acquiring the set of face images(training set) and calculate the corresponding distribution in M-dimensional weight space. In second step recognize the eigen faces with respect to known and unknown faces[12].

3) *Discrete cosine Transformation(DCT)*: The Discrete cosine transform is used to extract the global features. The large area illumination variations are also alleviated by discarding the first few low frequency DCT coefficients. DCT uses cosine functions[13]; the resulting matrix depends on the horizontal, diagonal, and vertical frequencies. DCT features encode texture frequency using predefined filters that depend on the patch size. DCTs are not sensitive to alignment errors, and their dimensionality is the same as the original image[5][19]. The DCT has many advantages such as energy compaction, orthogonality and seperability, image compression However, higher frequency coefficients are usually ignored, therefore potentially losing sensitivity to fine image structures such as wrinkles and bulges.

4) *Single Value Decomposition (SVD) :* The singular-value decomposition (SVD) is a factorization of a real or complex matrix in linear algebra. It is the generalization of the eigen decomposition of a positive definite normal matrix (for example, a symmetric matrix with positive eigen values) to any matrix via an extension of the polar decomposition. It has many useful applications in signal processing and statistics. Applications that employ the SVD include computing the persudoinverse, least squares fitting of data, multivariable control, matrix approximation, and determining the range and null space of a matrix The singular value decomposition of a matrix A of m x n matrix is given in the form,

$$A = U\sum V^T$$

Where U is an m x m orthogonal matrix; V an n x n orthogonal matrix, and Σ is an m x n matrix containing the singular values of A along its main diagonal[11].

5) *Local Gabor Filter Bank:* It can be seen that the Gabor representations are very similar using the filters with the same orientation, especially using the filters with the two neighboring frequencies. These features result from convolving every location of a region with a set of filters. While they have strong expressive power, they lack some robustness to affine transformations and illumination changes[14].

6) Local Discriminate Analysis (LDA): Linear discriminate analysis (LDA) is a generalization of Fisher's linear discriminate, a method used in statistics pattern

 ISSN: 2454-4248
510 – 519

recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. According to research work discriminate analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA [8]is a supervised learning method, which utilizes the category information associated with each sample. The goal of LDA is to maximize the between-class scatter while minimizing the within-class scatter[15].

7) *Local Binary Pattern:* Local Binary Pattern (LBP)[16] is a simple yet very efficient texture operator which labels the pixels of an image by threasholding the neighbourhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. it has further been determined that when LBP is combined with the Histogram of oriented gradients (HOG) descriptor. Based on the operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region.[16] .It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis.

8) *Gaber wavelet:* Gabor wavelet can be applied to images to extract features aligned at particular angles (orientations)[19]. .A Gabor filter is a function obtained by modulating the amplitude of a sinusoid with a Gaussian function. Gabor filters are thought to mimic the functions of simple cells in the visual cortex. The various two-dimensional receptive-field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of two-dimensional Gabor filters [17][12]. Gabor wavelet possess optimal localization properties in both spatial and frequency domains, and they have been successfully used in many pattern recognition applications.

2. Geometric Based Facial Feature Extraction

This approach extract the features by using relative positions and sizes of the important facial components such as eyes, nose, mouth and other important face components. It is also detect the track changes of facial components in near frontal face images. Tian etal researcher developed multi-state models to extract the geometric facial features[18].This model describes three state model for lips such as open ,close, tightly closed. Two state model is used for eyes such as open and closed. similarly for cheek and brow has one state model. some other features are also represented for this. The following methods are used for geometric feature extraction.

1) *Point distribution Model:* The Point Distribution Model (PDM) is a shape description technique that is used in locating new instances of shapes in images. It is also referred to as Statistical Shape Model. It basically tries to "understand" the shape, as opposed to just building a rigid model. It is very useful for describing features that have well understood general shape, but which cannot be easily described by a rigid model. The PDM has seen enormous application in a short period of time[20]. PDM basically represents the mean geometry of a shape, along with statistical modes of geometric variation inferred from a training set of shapes. It is formulated by combining local edge feature detection and a model based approach. This gives a fast and simple method of representing an object and how its structure can deform. We need to understand that PDM relies on landmark points. A landmark is basically an prominent point on a given locus for every shape instance across the training set.

2) *Gabor Filters:* Gabor Filter is named as linear filter used for texture analysis which means that it basically analysis whether there are any specific frequency content in the image in specific directions in a localized design and the point and region of analysis frequency and orientation represented of Gabor filter are similar to those of human visual system and they have been found to be particularly appropriate for texture representation discrimination. In the spatial domain, a 2-D filters a Gaussian kernel function modulated by sinusoidal plane wave[21]. Gabor filters with different frequencies and with orientations in different directions have been used to localize and extract text-only regions from complex document images (both gray and color). It has also been applied for facial expression recognition. The Gabor space is very useful in image processing applications optical character recognition, iris recognition and fingerprint recognition. Multi-resolution, different dimensional Gabor filter represent at each facial landmark are now used to track the landmarks in the consequence frames. A feature vector corresponding to each landmark point, as well as feature vectors from each pair of landmarks, are now created, and considered as a feature pool[22].

3) *Landmark Initialization and Tracking Using Elastic Bunch Graph:*-The elastic graph matching (EGM) method was first proposed by Lade set al.[23], and applied to face recognition. It is extricated more than one feature on one landmark point, called it EBG, and applied it to face recognition In this landmark initialization and tracking process locating a landmark in a novel has two steps. First, the location of the landmark is represented based on the known locations of other landmarks in the image, and second, that estimate is refined by extracting a Gabor jet from that image on the near about locations and comparing that jet to one of the models. To make the system fully automatic the approximate locations of at least one or two landmarks are needed at the beginning[24]. This goal is achieved by first localizing the face region in the image

 ISSN: 2454-4248
510 – 519

using the Haar-like feature based face detection method proposed in[24]. The movements of facial landmarks, as a particular expression evolves, are not independent of each other.

4) *Optical Flow Method:-* Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene. Optical flow stimulus for the perception of movement by the observer in the world; perception of the shape, distance and movement of objects in the world; and the control of loco motion. Motion information of facial expression can be analysed by the motion field of optical flow estimated. Horn-Schunck optical flow algorithm [25] is a dense optical flow field based on the assumptions of gray scale consistency, the optical flow calculations would be inaccurate when the brightness is not invariance or a motion object exists deformation. Facial motion is typical of non-rigid motion. Using the traditional optical flow method will inevitably lead to inaccurate of optical flow field and affect the facial expression recognition rate. According to extended optical flow constraint equation, estimating facial expression change detection based on Horn and Schunck[26] is used to calculate optical flow field of facial expression sequences.

3. *Fusion Based Facial Feature Extraction methods*

Feature level fusion has vital role in process of data fusion. Fusion of features is process of merging different features which describes different characteristics of images, to gain the superiority of all the features .According to the researchers it is capable of deriving and gaining the most effective and least-dimensional feature vector sets that benefit the final decision. In other words we say that it is method of extracting effective features from salient features (patches of face)of face. There are various fusion based methods according to requirement of higher level feature. According to this survey paper some of them discussed below .

1) *Fusing of LDA and PCA:* This method propose dimensionality reduction by linear combination of features. In effect, linear methods propose the high dimensional data into a lower dimensional space, we say it feature compression. These two traditional methods are finding effective linear transformations. PCA(Principal Component Analysis) finds a projection that best represents the original data in a least-squares sense, and LDA(Linear Discriminate Analysis) seeks a projection that best separates the data in a least-squares perception.
A PCA [4][10] is an unsupervised learning approach, which handles samples of the different classes in the same way. PCA can significantly reduce the dimensionality of the original feature without loss of much information in the sense of representation, but it may lose important information for discrimination between different classes. The detail description of PCA is given above in section of appearance based feature extraction methods.
A LDA [15] is a method, which utilizes the category information associated with each sample. The goal of LDA is to maximize the between-class scatter while minimizing the within-class scatter. The detail description of LDA is given above in section of appearance based feature extraction methods.
Fusing of PCA and LDA propose that PCA maps the original t-dimensional feature x_i to the f -dimensional feature y_i as an intermediate space, and then LDA projects the PCA output to a new g -dimensional feature vectors z_i. More formally, it is given by [15]

$$z_i = W_{lda}^T W_{pca}^T x_i (i = 1,2,3.....N)$$

Where W_{lda} , W_{pca} is linear transformation matrix for X_i features.
When using PCA+LDA method, the dimensionality drastically reduced to 6 dimensions[21] and the recognition performance is improved several percent compared with PCA. Experiments show that PCA+LDA feature may partially eliminate the sensitivity of illumination.

2) *Fusing LBP and LGC*: The fusing method proposed comparison of the neighboring pixels with central pixel as well as with other neighboring pixels at the same time in the binary form. The decimal values are calculated from the both and the maximum of the two is put as the new value. The matrix obtained from this process is divided into blocks and histogram of each block calculated separately and then concatenated.
LBP (Local Binary Pattern) [16][27] is the most popular texture based feature extraction method. In this method, the central pixel value is compared with its P-neighboring pixel values along radius R. The greater neighboring pixel values are assigned 1, other's 0. This generated binary number is converted into decimal.
LGC (Local Gradient Code) [28] is another method which extracts the local information. This method compares the neighboring pixels horizontally, vertically and diagonally instead of only with central pixel value. The optimized forms of LGC are LGC-VD (except horizontal) and LGC-HD (except vertical)[28].

3) *Fusing HOG with LDP:* In this, instead of simply calculating the histogram of the LDP image, the HOG feature extracted from the LDP image is taken the feature vector. This method holds the property of both the methods and recognition rate is promising in comparison to LDP. LBP and HOG[29] descriptors are applied to utilize the texture and orientation information of these expressions. Proper fusion methods are very important factors for the recognition work and uncomfortable methods can make the recognition result worse.LDP is discussed in above section of feature extraction.
HOG(Histogram of oriented gradient) is a feature descriptor which is used in computer vision and image processing [29][28]. The technique counts occurrences of gradient orientation in localized portions of an image. During HOG features extraction, the image is divided into several blocks and the histograms of different edges

 ISSN: 2454-4248
510 – 519

are concatenated as shape descriptor. HOG is invariant to geometric and photometric transformations, except for object orientation.

4) *Fusing HOG and Wavelets:* This is another approach of combining HOG [29]with the wavelets. This approach gives a significant and promising output in comparison to wavelets and HOG. In this, firstly, the wavelet features are extracted and then the HOG of the extracted wavelet feature, results a combined feature vector. This method is also simple. The experiment has been carried out with different number of bins.HOG method is described just above section.
Wavelets [30] is technique of transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time - frequency representation of the signal. It was developed to overcome the short coming of the Short Time Fourier Transform (STFT)[31], which can also be used to analyze non - stationary signals. While STFT gives a constant resolution at all frequencies, the Wavelet Transform uses multi -resolution technique by which different frequencies are analyzed with different resolutions.

Another approach of feature extraction and selection is Facial Action coding System.

Facial Action Coding System
Facial Action Coding System (FACS)[32] is a human-observer-based system designed to detect subtle changes in facial features. Viewing videotaped facial behaviour in slow motion, trained observers can manually FACS code all possible facial displays, which are referred to a section units and may occur individually or in combinations.
Using Action Units: It is an effective method for facial expressions recognition as it divides the face into action units by applying objectivity and flexibility on the image. This approach is described by Ekman et al.[1] (Action Units) on the face. There are total 44 action units at all and out of them 30 are acquired by the contractions of certain muscles. 12 of these 30 action units are of the upper portion of the face while the remaining 18 are for the lower portion of the face. This approach is applicable in those applications in which fine level of changes in the expressions are needed to be identified. There are some methods in which the whole frontal face or all the 44-action units are not used; rather some regions are selected manually from the face and used for the recognition of expressions. Figure 4 shows upper and lower face Action units.
AU combinations: More than 7,000 AU combinations have been observed in everyday life [29]. Co-occurring AUs can be additive, in which the appearance changes of each separate AU are relatively independent, or non-additive, in which one action masks another or a new and distinctive set of appearances is created.

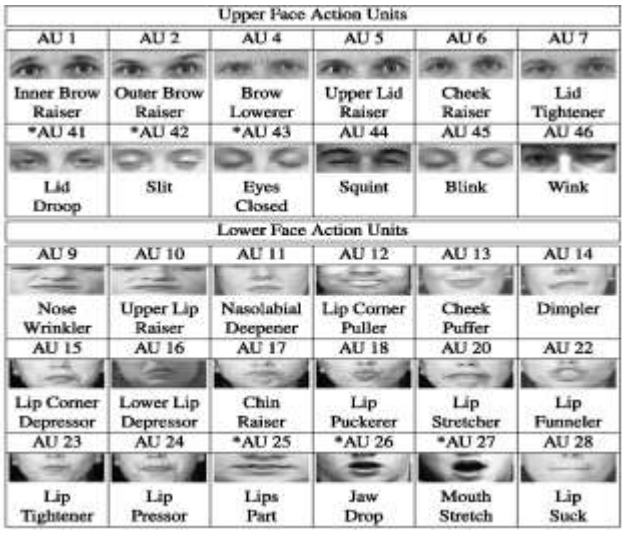

Figure 4: FACS action units for upper and lower face[1]

AUs affecting the same facial area are often non-additive. Furthermore, some AU combinations are more common than others due to latent variables such as emotions. For example, happiness is often expressed as a combination of AU12 and AU6. Table2 shows a number of expressions with their associated AUs.

TABLE II: LIST OF EXPRESSIONS WITH ACTION UNITS (AUs)[17]

FACS	**Upper face:** 1, 2, 4-7, 43, 45, 46; **Lower face:** 9-18, 20, 22-28; other: 21, 31, 38, 39
Expressions:	**AUs:**
Anger	4, 5, 7, 10, 17, 22-26
Disgust	9, 10, 16, 17, 25, 26
Fear	1, 2, 4, 5, 20, 25, 26, 27
Happiness	6, 12, 25
Sadness	1, 4, 6, 11, 15, 17
Surprise	1, 2, 5, 26, 27

C. *Facial Expression Classification*

This is the last and resultant step of facial expression recognition which uses the features extracted from the previous step and tries to classify the features set based on the similarities between the feature data. According to past researches, there are a lot of approaches discovered for the facial expressions classification such as K-Nearest Neighbor(K-NN), Naive Bayesian (NB) classifier, Multilayer Perceptron(MLP), Support vector Machine(SVM), Hidden Markov Model(HMM) etc.

1. *K-Nearest Neighbor(K-NN):* K-Nearest Neighbor (K-NN)[33] is a simple algorithm that stores all available cases and classifies new classes based on a similarity measure (e.g. distance functions). K-NN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique [33]. The algorithm works as follows. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbor measured by a distance function. If K = 1, in this case is simply assigned to the class of its nearest neighbor. When there are only two classes, k must be an odd integer.

 ISSN: 2454-4248
510 – 519

However, there can still be times when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance[33]:

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

where x_i is samples to recognize, y_i is training samples.

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross validation is another way to retrospectively determine a good K between 3-10[33].

The shortcoming of K-NN is of great computational complexity, because we must calculate the distance between every sample to classify and all known samples to find out the k nearest neighbours. The often-used solution is to edit the known examples advance to remove the examples which have no great effect on classification.

2. *Naive Bayesian Classification:* This is a Probabilistic strategy that has been demonstrated extremely compelling in numerous classification issues. This technique considers that the nearness of a specific component of a class is irrelevant to the nearness of some other element. The equation for classification utilized is[34]

$$C = \arg\max\{P(C_i)\prod P(f_i/C_i)\}$$

Where $P(f_i/C_i)$ are conditional tables(probability) or conditional density learned in training and C_i is variable set. A Bayesian network having the correct structure and parameters is also optimal for classification because the a-posterior distribution of the class variable is accurately represented. A Bayesian network classifier is a generative classifier when the class variable is an ancestor (e.g., parent) of some or all features. A Bayesian network classifier is diagnostic, when the class variable has none of the features as descendants[34].

The NB classifier makes the assumption that all features are conditionally independent given the class label. Although this assumption is typically violated in practice, NB have been used successfully in many classification applications. One of the reasons for the NB success is attributed to the small number of parameters needed to be learnt.[34].

3. *Multilayer Perceptron(MLP):* MLP is a feed forward artificial neural network that maps sets of input data onto a set of appropriate output. MLP follows supervised learning technique. This supervised learning method also called back propagation for training the network. MLP is an improvisation of the standard linear Perceptron and can distinguish data that are not linearly separable. Hayet Boughrara ·Mohamed Chtourou Chokri Ben Amar and Liming Chen [35] in their study of facial expression recognition based on a MLP neural network using constructive training algorithm works with MLP of three layer. The number of input neurons is equal to the size of related feature vector. Same as, the number of output neurons is equal to the number of facial expressions to be recognized. In the learning phase, the desired output neuron has 1 for the correct input pattern and 0 for all others output neurons. The hidden layer is constructed using the proposed constructive training algorithm. There are two steps on the realization of the facial expression recognition system using the MLP architecture: the training step and the testing step. The learning algorithm used this study is the standard back-propagation [36].MLP requires the network architecture definition before the training. This is known that there is no general answer to the problem of defining neural network architecture for the given problem but MLP works well if the network architecture is properly chosen.

4. *Support Vector Machine(SVM):* Data Classification is one of the major parts in machine learning. SVM is one of the most important classification techniques discovered in 1995[37]. The idea of Support Vector Machine (SVM) is creating a hyper plane in dimensional feature space and separate two classes of data with a maximum margin of hyper plane. The separating hyper plane maximizes the distance between two parallel hyper planes [38]. This optimum hyper plane is produced by maximizing minimum margin between two sets. Therefore, the resulting hyper plane will only be depended on border training patterns called support vectors. More than that, Support vectors are the data points that lie closest to the decision surface . A classification approach separating input data into training and testing sets. Each instance in the training set has one "target value" and more than one "attributes". The aim of SVM is to design a model (based on the training data) which predicts the target values of the test data given only the test data attributes[33].

Given a training set of instance label pairs (x_i, y_i) i= 1,2,..... l where $x_i \in R^n$ & $y_i \in \{1,-1\}^l$ the SVM require the solution of the following optimization problem[33]:

$$\min_{w,b,\xi} \frac{1}{2}W^T W + C\sum_{i=1}^{l}\xi_i$$

$$\text{subject to } y_i(W^T\phi(x_i)+b) \geq 1-\xi_i \qquad (1)$$
$$\xi_i \geq 0$$

Here training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function ϕ. SVM finds a linear separating hyper plane with the maximal margin in this higher dimensional space. C > 0 is the penalty parameter of the error term. The functions defined the kernel in SVM[33]. It is basically four types linear , polynomial, RBF, and sigmoid. Recently several studies have reported that support vector machine (SVM) delivers higher accuracy in terms of data classification compared with other classifier[39]. The SVM classifier views the classification problem as a quadratic optimization problem.

 ISSN: 2454-4248
510 – 519

As the SVM classify the data with the set of support vectors by minimizing the structural risk, the average error between input and their target vectors is reduced. SVM is used in various recognition problems like face recognition, pattern recognition, and emotion recognition and in many more applications.

5. *Hidden Markov Model(HMM):* Hidden Markov Model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (i.e. *hidden*) states. The hidden Markov model specifies that simplest dynamic Bayesian network. In Markov models (like a Markov chain), the state is directly visible to the observer, and has only state transition probabilities parameters, while in the hidden Markov model, the state is not directly visualize, but the output, depend on the state, is visible. Individual state has a probability distribution over the possible output reference tokens. Therefore, the sequence of tokens generated by an HMM which specifies the information about the sequence of states. The adjective *hidden* refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a hidden Markov model even if these parameters are known exactly[40]. According to research work HMM describe the observation value series in statistical model, it has exact mathematic result and all the more necessarily mirror the performance attribute of entire observation esteem arrangement.

HMM can be marked by $\lambda = [N, M, \pi, A, B]$

N is number of states in Markov chain, M is number of possible corresponding observation value of every state, π original possibility distribution vector, A possibility matrix of state transfer, B possibility matrix of observation value, for continuous HMM, B is a group of possibility function of observed value[40].

TABLE III COMMON TOOL AND PARADIGMS FOR CLASSIFICATION

S. No.	Models and Techniques	Reference No.	Description
1	K-NN	[33]	This paper compare the performance of machine learning algorithm. K-NN has 98.87% accuracy on CK+ dataset .large amount of data provide high accuracy
2	Naive Bayesian	[34]	This research specifies that structure of Naive Bayesian which used, for un-labelled data improve the classification, otherwise they can actually degrade the performance.
3	Dynamic Bayesian Networks	[41]	This research work showed that DBN models perform better than the state of the art methods for emotion recognition.
4	Support Vector Machine (SVM)	[33][38]	Deep features are extracted Without additional feature engineering steps, using multiple layers of the SVM classifiers with CPONs.
5	Multiclass SVM	[42]	This research work have one-against-one multiclass SVMs for facial Expression classification. This shows that the geometric feature can get good recognition rate for Multiclass SVM.
6	Hidden Markov Model (HMM)	[40]	This introduced the combination of HMM and KNN which reduces training time.
7	Multi observation HMM	[43]	This technique applications in large vocabulary continuous speech recognition.

III CONCLUSION

Facial expressions are very effective way to convey human emotions. It plays a very vital role in human communication irrespective of languages. It is observed that development of an automated system that accomplishes facial expression recognition with good classification accuracy but for limited types of datasets under uncontrolled conditions(like occlusion, posed, illuminations and expression variations). However advance approaches(higher level feature extraction) which involved machine learning statistical techniques improve the performance and accuracy of recognition system. Advanced approaches like K-NN, Dynamic Bayesian Network, Hidden Markov model have succeed to achieve more than 85% accuracy. This paper provides a survey based on timeline view which performs an analysis on different face detection, feature extraction and classification techniques to handle facial expression to recognize face. The fusing of existing strategies reflects effective performance and found robust against the variation of illumination.

REFERENCES

[1] Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel I.,Movellan, J.," Automatic Recognition of Facial Actions in Spontaneous Expressions", Vol. 6, pp. 22-35, 2006

[2] Yingli Tian, Takeo Kanade, and Jeffrey F. Cohn, "Facial Expression Recognition ",IEEE Computer Vision,2011

[3] J. Majumdar and R. Avabhrith , “Human Face Expression Recognition”, International Journal of Emerging Technology and Advanced Engineering(IJETAE), Vol. 4, 2014

[4] M.-H. Yang, D. J. Kriegman and N. Ahuja, “Detecting faces in images: A survey", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol.24, 2002

[5] Neha Bhardwaj and Manish Dixit, " A Review: Facial Expression Detection with its Techniques and Applications ", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.9, pp.149-158, 2016

[6] Roberto Brunelli and Tomaso Poggio," Face Recognition: Features Vs Templates", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol.15, pp.0162-8828, 1993

[7] Priyanka Goel and suneeta Agrawal," An Illumination Invariant Robust and Fast Face Detection, Feature Extraction based Face

Recognition System", IEEE Computer and Communication Technology, Vol 12,2012

[8] Anas Abouyahya, Sanaa El Fkihi, Rachid Oulad Haj Thami, Driss Aboutajdine, "Feature Extraction For Facial Expression Recognition", IEEE Multimedia Computing and Systems, 2016

[9] Wang, Jinghua," A Study of Appearance-based Feature Extraction For Face Recognition", The Hong Kong Polytechnic University,2013

[10] G. Saranya, "Human Expression Recognition using Facial Features", International Journal of computer Applications, 2014

[11] Mandeep Kaur and Rajeev Vashisht, "Comparative Study of Facial Expression Recognition Techniques", International Journal of Computer Applications, Vol 13, 2011

[12] Rupinder Saini, Narinder Rana," Facial Expression Recognition Techniques Database & classifier ", International Journal of Advances in Computer Science and Communication Engineering, vol.2, 2014

[13] S. Dabbaghchian, A. Aghagolzadeh, and M. S. Moin, " Feature Extraction Using Discrete Cosine Transform For Face Recognition ", IEEE Signal Processing and Its Applications, 2007

[14] Wei zha, Jeong·Sun Park, AND Sang Woong L EE," Fully Automatic Face Detection and Facial Feature Point Extraction Using Local Gabor Filter Bank and PCA", IEEE Machine Learning and Cybernetics (ICMLC) , 2011

[15] Hong-Bo Deny, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang ," A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA ", International Journal of Information Technology, Vol. 11, 2005

[16] X. Feng, M. Pietikäinen and A. Hadid ," Facial Expression Recognition with Local Binary Patterns and Linear Programming ",IEEE Pattern Recognition and Image Analysis, Vol. 15, pp. 546–548, 2005

[17] Brais Martinez, Member, Michel F. Valstar, Bihan Jiang, and Maja Pantic, "Automatic Analysis of Facial Actions: A Survey", IEEE Trans ,vol. 13, 2017

[18] Archana Verma &. Lokesh Kumar Sharma," A Comprehensive Survey on Human Facial Expression Detection ",International Journal Of Image Processing, Vol.7, 2013

[19] N. U. Khan," A Comparative Analysis of Facial Expression Recognition Techniques ", IEEE Signal Processing, 2012

[20] TF. Cootes and C. J. Taylor," Combining Point Distribution Models with Shape Models Based on Finite Element Analysis", IEEE computer vision and image processing, 2004

[21] Rosniza Roslan Nursuriati Jamil, " Texture Feature Extraction using 2-D Gabor Filters", IEEE Computer Applications and Industrial Electronics, 2012

[22] Deepak Ghimire and Joon whoan Lee," Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines", Sensors, Vol.13,2013

[23] Lades, M.; Vorbüggen, J.C.; Buhmann, J.; Lange, J.; Malsburg, C.; Würtz, R.P.; Konen, W., "Distortion invariant object recognition in the dynamic link architecture", IEEE Trans. Computer, Vol 42, pp 300–311,1993

[24] Viola, P.; Jones, M.J.," Robust real-time face detection", International Journal of Computer Vision, Vol.57, pp 137–154, 2004.

[25] Ashok Ramadass, Myunghoon Suk, B.Prabhakaran, "Feature Extraction Method for video based Human Action Reconition:Extended Optical Flow Algorithm ",IEEE Acoustic Speech and Signal Processing, Vol 10,2010

[26] E. Meinhararth Hopis, S.Javior And D. Kandermann, "Horn - Schucles Optical Flow with Multiscale Strategy", IEEE Computer Applications, Vol.3, pp 151-172, 2013

[27] Jyoti Kumari, R. Rajesh and Abhinav Kumar," Fusion of Features for the Effective Facial Expression Recognition", Communication and Signal Processing (ICCSP),vol.16, 2016

[28] Jyoti Kumaria, KM.Poojaa," Facial expression recognition: A survey ",Elsevier Journal Computer Vision and the Internet, Vol.5 ,pp 486 – 491,2015

[29] Yanpeng Liu, Yibin Li, Xin Ma, Rui Song, " Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas", preprints, Mathematics and Computer Science ,January 2017

[30] Fan Guojuan, Mu Wanquan, Li Bo," HOGG: Gabor and HOG-based Human Detection ", IEEE Information Technology in Medicine and Education (ITME), 2016

[31] Jung Jun Lee, Sang Min Lee, In Young Kim, Hong Ki Min, Seung Hong Hong," Comparison between Short Time Fourier and Wavelet Transform for Feature Extraction of Heart Sound ", lEEE TENCON, 1999

[32] Maninderjit Singh, Anima Majumder and Laxmidhar Behera," Facial Expressions Recognition system using Bayesian Inference", IEEE Neural Network, Vol.14, 2014

[33] Ratna Astuti Nugrahaeni1, Kusprasapta Mutijarsa," Comparative Analysis of Machine Learning KNN, SVM, and Random Forests Algorithm for Facial Expression Classification ", IEEE Neural Networks,2016

[34] Ira Cohen, Nicu Sebe , Fabio G. Cozman , Marcelo C. Cirelo , Thomas S. Huang," Learning Bayesian Network Classifiers for Facial Expression Recognition using both Labeled and Unlabeled Data", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.3, 2003

[35] H.Boughrara, "Facial Expression Recognition Based on a MLP Neural Network Using contractive Training Algorithm ",springer ,2014

[36] Zaenal Abidin, Agus Harjoko," A Neural Network based Facial Expression Recognition using Fisher face ", International Journal of Computer Applications, vol. 59, 2012

[37] Moulay Smail Bouzakraoui, Abdelalim Sadiq, Nourddine Enneya," A Customer Emotion Recognition through Facial Expression using POEM descriptor and SVM classifier ", Association for Computing Machinery, 2017

[38] Kim, Sangwook, et al. "Deep learning of support vector machines with class probability output networks", International Association of Science and Technology for Development (IASTED),Neural Networks ,2014

[39] Foody, Giles M., and Ajay Mathur. "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM",IEEE Remote Sensing of Environment,2006

[40] Ch. J. Wen,,Y. Zh. Zhan, " HMM+KNN Classifier for Facial Expression Recognition ",IEEE Computer Vision, Vol.8, 2008

[41] Hiranmayi Ranganathan, Shayok Chakraborty and Sethuraman Panchanathan," Multimodal Emotion Recognition using Deep Learning Architectures", IEEE Application of Computer Vision. 2016

[42] Le iGang, Li Xiao-hua, ,Zhou Ji-liu, Gong Xiao-gang," Geometric feature based facial expression recognition using multiclass support vector machines ",IEEE Granular Computing, 2009

ISSN: 2454-4248
510 – 519

[43] Bohumir Jelinek," Review on Training Hidden Markov Models with Multiple Observations", isip.piconepress,2000

[44] http://www.what when how .com/facial expression images

[45] Nidhi N. Khatri, Zankhana H. Shah, Samip A. Patel, "Facial Expression Recognition: A Survey", International Journal of Computer Science and Information Technologies, Vol. 5 , pp.149-152, 2014