Facial Expression Recognition and Analysis Techniques: A Survey

[1]Victor Mokaya

Department of Computer science, School of Engineering & Technology

Suresh Gyan Vihar University, Jaipur, India

[2]Dr. (Commodore) H.P Singh

Pro-president Academics, Suresh Gyan Vihar University,Jaipur,India

Abstract: *Facial expression recognition FER has emerged as a promising and active field of research in machine learning, pattern recognition and computer vision. Recognition of facial features in digital image processing continues to be a challenging task in the highly dynamic environment. In this survey paper two conditions that accurately classify different facial emotions and ambiguities that make it difficult to distinguish between sadness and anger have been considered. We provide a pipeline for the implementation of FER with decisions taken at each stage. At each stage different methods are applied to provide a comprehensive comparison for selecting the best method with high accuracy. The three steps applied are feature extraction; feature detection and feature classification. Feature detection has three techniques geometric based, template based and feature invariant. Geometric based methods were found to be the most appropriate for feature detection. Feature extraction utilizes geometric and appearance based techniques also. Based on the classification of the expressions, by different classifiers are based on the feature detection and feature extraction technique applied.*

Keywords: FER, Feature detection, Feature Extraction, Feature classification, Geometric methods, Template methods, Feature Invariant.

I. INTRODUCTION

The progressive research in FER and machine learning has enabled various challenges in this field and those relating to the medical, industrial, finance, Human Computer Interaction trade and technology to be solved. FER uses non-verbal communication cues from facial components i.e eyes; mouth, cheeks, forehead and eye brows to enhance the conversation, these cues are expressed in the form of emotions. Essentially there are six types of emotions that are portrayed in facial expression recognition as discussed by (Ekman.p 1971); they include sad, anger, happy, disgust, surprise and contempt. Neutral has been excluded.

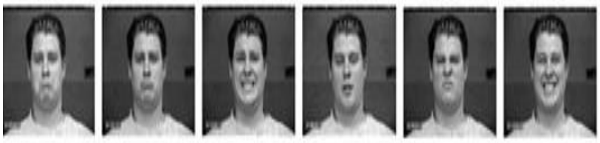


Fig1.0: facial expression images from the Cohan Kanade dataset

All of them convey different messages which influence our communication pattern. In the realm of this dynamic area are challenges associated with capturing of accurate and correct expressions, they include low quality of vision library, lack of multiple camera angles to capture different facial scenes, illumination conditions, partial occlusion and pose variation. Hence to overcome this challenges suitable algorithms and approaches are necessary as discussed below.FER is applied in several applications which are not less than psychology, mental state identification, automatic counseling systems, facial expression synthesis, lie detection, Human Computer

Interactions (HCI) for animating avatars, computer graphics (emoji/emoticon creation) and transport security (detecting driver fatigue).

The process of facial expression recognition takes the following steps.

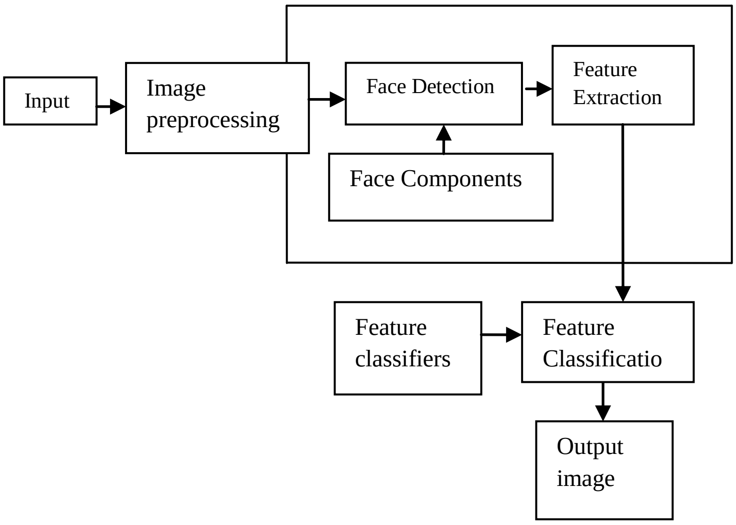


Fig1.2 Face Expression Recognition Flow Diagram

When an input image taken is from an existing dataset or directly captured, it undergoes an initial process known as pre-processing. This process is essential for removing of unwanted distortions and enhancing the required features. In order to improve the image data scaling, rotation, congruence and resizing as discussed by [S. Monica *et al* 1993] are performed. Feature detection is the process of finding the salient and non salient features useful in the determination of expression. Facial components like eyes, nose, eyebrows, cheeks and mouth are applied depending on the approach used. The facial extraction step applies both the output from image preprocessing and faces detection as input to identify the important facial features by matching the most preferred locations with the discriminating locations. Finally feature classification is applied to group feature points into class membership for desired recognition. Feature classifiers are algorithms that map the data into categories and enhance the accuracy of the prediction such as Perceptron, Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbor, Artificial Neural Networks and Support Vector Machine. The final classified class members are merged to form a single set that is obtained as the final classified output image.

1. FEATURE DETECTION

. Different features are responsible for different actions in analyzing the expression expected, this provides a good chance for obtaining accurate results. The different approaches that are applied for this task include geometrical, template and feature invariant.

I. Geometry based Approach.

The geometry based feature detection approach takes into account the basic shape of the facial structure; eyes, mouth, and nose and use them as input. The variation in size, shape and location of the features in the face geometry are taken into account as useful input. The geometrical based approach includes

i).Discrete Cosine Transform

The discrete cosine transform (DCT) is a lossy image compression with strong energy contraction. It's a viable tool for reducing the size of the data without sacrificing the key attributes. In FER we fetch an image information in terms of pixels in the range 0-255 and divide it into a block of 8x8 matrix.DCT reduces the size of the data significantly by transforming an image from a spatial representation into a frequency domain where lower frequencies obtain large amplitudes and the higher frequencies have smaller magnitudes. It is therefore argued that DCT coefficients of lower frequency modes, capture the most dominant and relevant information of the facial expressions.

ii).Point distribution Model (PDM)

A Point Distribution Model (PDM) is a parametric model based on the statistics. It applies the model of deformable shape, which is an essential part. An instance of a given object can be denoted with L landmark points as a 2L× 1 vector.

$$S = [x1, y1, x2, y2, \ldots\ldots xL, yL]^T$$

Consists of the Cartesian coordinates of the points the construction of a PDM commonly involves the following steps:

1. Generalized Procrustes Analysis is applied to align a set of training shapes. This will eliminate the similarity component such as rotation and rotation from the shapes.

2. Principal Component Analysis (PCA) is applied on the aligned shapes by aligning the center by removing the mean shape $\bar{s}$ followed by computing the eigenvectors $U_s \in \mathbb{R}^{2L \times N-1}$

3. Augumenting of the vectors obtained is done to control the similarity component of the object, [1] and retain n eigenvectors. A linear shape model of the kind $\{\bar{s}, U_s\}$ where $U_s \in \mathbb{R}^{2L \times n}$ becomes an orthonormal function of n eigen vectors and $\bar{s}$ as the mean shape vector.

When applying the model operations such as projection or reconstruction, the subspace applied is eliminated. Although the last captured components are removed to minimize noise. This is achieved by setting the number of components applied explicitly.

iii).Gabor Filters

Gabor filters are facial extractors that are used to extract sensitive facial key points such as eyes, mouth and nose. It is a biological representation of visual neurons that captures the receptive field in multi-scale representation; it also relies on the spatial and frequency properties of optical localization which are highly used in various facial expressions. In FER Gabor filters each image is convolved with multiple spatial resolutions which provide the channel for extraction of facial express ions. Gabor kernels are produced when the spatial and distinct orientations are combined together [Jun Ou *et* al 2010]. If there are n spatial frequencies and m distinct orientations then the p is given by (n × m = p). p is therefore, the number of filters that are present in the image which are referred to as the gabor wavelet kernel filter. This method is more appropriate in extraction of images that require classification, though it's less competitive to convolutional neural networks

iv). Discrete Cosine transforms

DCT is expressed as a sequence of finite data points from a sum of cosine functions that are within different frequencies. It's a lossy compression method that is applied to critical compression. Often referred as "type-II DCT" as first described by [ahmed *et* al] or simply "the DCT". In FER DCT is applied to reduce the size of the image data across different facial parts. An 8 x 8 array of two-dimensional DCT-II are computed and appropriate formula applied to all rows and columns. The results obtained are a coefficient array which is quantized, and entropy coded.

v).Optical Flow Method

Optical flow is a detection method that is based on the Horn-schunck optical flow algorithm used to measure the motion of facial expressions based on the dense optical motion field, it uses the assumption of grayscale consistency on an image to measure the dense optical field. The optical calculations accuracy are based when the brightness is invariant and no motion object exists in deformation. Facial motion is atypical non-rigid motion which leads to inaccurate facial recognition rate when used with traditional optical flow method due to inconsistent optical flow field. [[D.Bereziat et al]] came up with an advanced adaptive method capable of computing the optical flow field of motion using the continuity of motion equation, which is robust to changes in deformation and brightness. This method is more appropriate than discrete cosine transform because geometric features are more obvious which results in greater accuracy in recognition.

II). Template based Approach

The image is represented as a two-dimensional (2-D) array of intensity values which is compared with suitable metric (like Euclidean distance)[Roberto Brunelli *et* al 1993]. A face is stored in a set of distinct templates which represent different viewpoints of eyes, mouth, nose and eye brows. The Point Deformation Model (PDM) is heuristically built into a metric that can be used by the matching measure; this technique is known as elastic template [Nidhi N. Khatri *et* al 2014]. The Viola and Jones (V & J) [Roberto Brunelli *et* al 1993] face detector is the widely used image dataset and algorithm for face detection.

i).Active Shape Model (ASM)

[Lu H *et* al 2014] Active Shape Model (ASM) is a model definition method that mimics what the expected image is supposed to look like; it tries to attempt to model a match between the actual image and the expected image. It is widely used in image recognition applications. In FER is represents all the deformations of the face to form a Point Distribution Model (PDM) respectively. In all the multi-resolution of a training image, a gray-level alignment is created for every landmark after the PCA. In order to improve the extraction level increasing the width of search profile is performed in order to reduce the effect of noise. Grouping of special landmark features to avoid distortion is done as well.

iii).Eigen Vector

Eigenvectors and Eigen values are dependent on the concept of orthogonal linear transformation. An Eigenvector is basically a non-zero vector. The Prominent Eigenvector within a matrix is one corresponding to the largest Eigen value of that matrix. This dominant Eigenvector is important for many real world applications. In his paper, an Eigenvector based system has been presented to recognize facial expressions from digital facial images. [Jeemoni Kalita *et* al 2013]In the approach, firstly the images were acquired, and cropping of five significant portions from the image was performed to extract and store the Eigenvectors specific to the expressions. The Eigenvectors for the test images were also computed, and finally the input facial image was recognized when similarity was obtained by calculating the minimum Euclidean distance between the test image and the different expressions

.

III. Feature invariant based Approach

This approach depends upon the concept of structural representation of features like eyes and mouth where a structural classifier classifies the regions into two types' facial region and non- facial region [T.C. Chang *et* al 1994]

i).Space Gray-level Dependence (SGLD)

[Dai, Ying et al 1994] Facial images can be regarded as a form of texture plane with special textural characteristics. The texture of facial images cannot be orientated and duplicated; rather it only appears as a whole. The symmetric of the face allows expressions to be varied with respect to medial axis of the head. Gray level variation of these expressions applied on the local region are unbalanced in regard to the vertical and horizontal axis, because of the influence features such as the eyes, mouths, and other facial gradient. This homogenous creation of the local region enables the accuracy in the extraction process.

ii).Matrix of face pattern

This approach involves the extraction of common properties of images which individual classes in the training set are determined by subtracting the differences in images. A matrix common to all classes is created and used in face recognition. Two methods are used in order to find a common matrix for every individual class from the training set. First one is Gram-Schmidth orthogonalization method while the other uses the scatter matrix of each class.[Turhal, Ü *et* al 2005]. Projection of corresponding zero eigen values of any matrix into eigen matrices produces a common matrix for that class. Since finding solution for high dimensional data are infeasible, PCA and LDA are applied to minimize the dimension in these vectors. Reduction process, discards important information due to lack of dimensionality reduction.

S. No	Facial Detection Approach	Features Used	Methods
1	Geometry	eyes, mouth, and nose	Point distribution Model (PDM) Gabor Filters Discrete Cosine transform (DCT) Optical Flow Method (OFM)
2	Template	eyes, mouth, nose and eye brows	Active Shape Model (ASM) Shape template Eigen Vector
3	Feature Invariant	eyes and/or mouth	Space Gray-level Dependence (SGLD) Matrix of face pattern(MFP) Mixture of Gaussian (MG)

Fig 2.0 Feature detection approaches

2. FEATURE EXTRACTION

. There are three basic methods for the extraction of facial features which include. Appearance based, geometric based and fusion based methods.

I). Geometrical based methods

Features extracted are relative to position and size of eyes, nose, mouth and other important components of the face. Consideration is made upon two components, Firstly the detection of edges which are made up of direction of region images and important components. Afterwards feature vectors are built from these edges and directions. Finally geometric methods are used based on grayscale difference of important components and unimportant components. Feature blocks and Haar-like feature block [Saranya R Benedict *et* al 2016] in Adaboost method are used to change the grayscales values in the feature vector.

i).Geometric Shaped Facial Feature Extraction for Face Recognition system (GSF2EFR)

GSF2EFR is used for identification and verification of the extraction features. The extraction method uses two modules detection and localization module. [Saranya R Benedict *et* al 2016]The detection model performs mapping of the image dimensions; the edges are detected and then filtered using gabor filters. Support vector machine (SVM) classifier extracts additional facial features which assist to remove the false negatives and true positives that emerge during feature extraction. The localization module is finally applied to the output of the detection module by refining the location of the extracted feature from the rest of the face.

ii)PHOG(Pyramid Histogram Of Orientation Gradients)

Pyramid Histogram of Oriented Gradients (PHOG) for feature extraction is widely applied especially in smile recognition. It is a spatial shape descriptor that represents an image by its local and spatial layout properties. PHOG features are extracted from the area of interest i.e eye region increases the weight of recognizing a surprise. PHOG method results in a much lower number of features extracted compared to other methods such as Gabor filter. The procedure of feature selection by using the AdaBoost method and Support Vector Machine (SVM) classifier results in best recognition performance. [Yang Bai, Lihua Guo *et* al 2009]

The PHOG feature applies the following steps.

Step 1: Extracting edge contours from a given sample image

Step 2: Dividing the image into partitions of several pyramidal levels.

Step 3: Computing the HOG of each grid at each resolution of pyramid level

Step 4: The final PHOG descriptor for an image is a combination of all the HOG vectors at every level of pyramid resolution. The merging of all the HOG vectors introduces spatial information of the image which are then normalized to the sum of all the pyramid levels.

iii) Gabor Feature Based Boosted Classifiers

According to [Vukadinovic *et* al 2005] the method adopts a fast and robust face detection algorithm, which represents an adapted version of the original Viola-Jones face detector. The detected face region is divided into 20 points of interest; each point is examined to predict the location of the facial feature points using a feature patch template. Feature models called GentleBoost templates built from both gray level intensities and Gabor wavelet features are used for this purpose. When trained and tested using the Cohn-Kanade database, the method achieves an average recognition rate of 93%.

iv). Enhanced Active shape model

According to [Mahoor *et* al 2006] active Shape Model (ASM) for facial feature extraction highly relies on the initialization and the representation of the local structure of the facial features in the image. Color information is used to improve the ASM approach for facial feature extraction and to localize the centers of the features of interest to assist the initialization step. The feature points in the RGB color space are modeled and a 2D affine transformation aligned to the facial features that are bothered by head pose variations. The 2D affine transformation compensates for the effects of head pose variations and the projection of 3D data to 2D. Experiments on a face database of 50 subjects show that this approach outperforms the ASM and is successful in facial feature extraction.

v). Principal Component Analysis

Principal Component analysis PCA is a method of statistics that demands an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.PCA is used for two main reasons. Firstly it focuses on reducing the dimensions of data to computable feasible size, secondly it extract the most relevant features out of the input data which as a result reduces the input size. The main features are retained to represent the original data [Tanaya Mandal *et* al 2007]. A covariance matrix from the feature image matrix is obtained firstly, followed by the Eigen vectors of covariance transformation. Eigen vectors are those vectors that are invariant to direction during a transformation, which can also be used as a representative set of the whole big dataset [N.G.Chitaliyaa *et* al 2010].

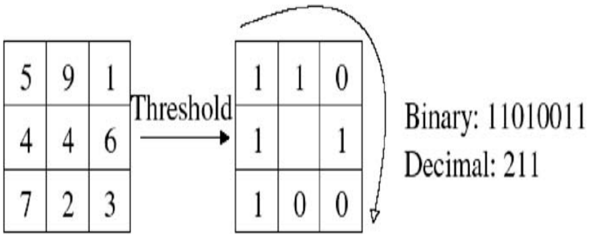
vi). Point distribution Model:

PDM basically represents the average landmark points of a feature along with additional statistical models of variation obtained from the training set. It is achieved by combining the local edge feature detection and a model based approach. This produces a robust and simple method of representing an object and how its structure can deform. PDM depends upon major landmarks available at the locus of every shape instance within the training set of a given dataset.

II. Appearance based methods

i). Local Binary Pattern

LBP operator was first introduced by [T. Ojala *et* al 1996]. It’s been proved as means of texture description. Operators are used in pixel labeling of an image by thresholding 3 × 3 matrix neighborhoods of each pixel with a center value. Afterwards the results are finally considered as a binary number.



5	9	1
4	4	6
7	2	3

1	1	0
1		1
1	0	0

Binary: 11010011

Decimal: 211

Fig.2 [T. Ahonen *et* al 2004]

The binary numbers obtained are coded to local primitives like spots and curved edges. Every LBP code is known as a micro-texton [stéphane G. Mallat 1999]. An LBP operator suffers limitation due to its small 3 × 3 neighborhoods which doesn’t obtain dominant features with large-scale structures. Hence the operator is later extended to use neighborhood of different sizes [T. Ahonen *et* al 2004]. This is derived by using circular neighborhoods and bi-linearly interpolating the pixel values. A uniform Local Binary Pattern contains at most two bitwise movement patterns from 0 to 1 or vice versa when the binary string is said to be circular. Accumulating the patterns which have more than 2 transitions into a single bin yields an LBP operator, denoted LBP$u2_{P,R}$.Taking an instance whereby 256 labels are used by a neighborhood of 8 pixels for the standard LBP and 59 for LBP^{u2}. After labeling an LBP operator to an image, a histogram of the labeled image $f_l(a,b)$ can be defined as

$$H_i = \sum_{a,b} I(f_l(a,b) = i), i = 0, \dots, x - 1$$

Where *x* corresponds to the number of different labels obtained from LBP operator and

$I\ (A) = \{^{1\ A\ is\ true}_{0\ A\ is\ false}$

The LBP information about distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image are contained and can be used to describe image characteristics statistically. Images of the face are essentially considered as micro-patterns that can be efficiently described using LBP histograms. These characteristics provide a useful tool to represent facial features using LBP [T. Ojala *et* al 1996, stéphane G. Mallat 1999]. A LBP histogram calculated upon the whole face image decodes only the appearance of the micro-patterns without any indication of their positions.

ii). Gabor Filters

Gabor filters are applied in facial image recognition for feature extraction. They are applied because of their optimal localization in spatial and frequency domain. Gabor filter is represented by the following equation:

$$\Psi_{x,y} = \frac{\parallel k_{x,y} \parallel^2}{\sigma^2} \exp\left\{-\left(\frac{\parallel x, y \parallel^2 \parallel z \parallel^2}{2\sigma^2}\right)\right\}\left[exp\left(izk_{xy}\right) - \ exp\left(-\left(\frac{\sigma^2}{2}\right)\right)\right]$$

where

$$k_{x,y} = \left(k_y cos\phi_x / k_y sin\phi_x\right), \phi_x = \frac{\pi u}{k}, -\ _{k_y} = \ 2^{-\left(\frac{y+2}{2}\right)}{}_{\pi}$$

Where z=(x, y) belongs to the location of pixels in the spatial domain. k_y and ϕ_x are frequency and orientation modulating. x is the filter orientation while v is the filter scale v used to come up with the wavelength.

$$\left(exp\left\{-\left(\sigma^2/_2\right)\right\}\right)$$

The second equation of the Gabor filter is used to compensate for direct current component value because the cosine component contains nonzero mean and the sine component zero mean. Gabor filter provides a good resolution for both frequency and spatial field [J. Ou *et* al 2010]

iii). Haar Wavelets

As the name suggests DWT was invented by a Hungarian mathematician by the name of Alfréd Haar. Discrete wavelet transform (DWT) is a wavelet transform whereby wavelets are discretely sampled using numerical analysis and functional analysis methods. Compared to other fourier wavelet transform, (DWT) has a key advantage as it uses temporal resolution. It contains both frequency and location information of the data. Given an input is represented by a list of 2*n* numbers, Haar wavelet transformation can be achieved by pairing up input values which stores the difference between the input values and passing the sum. This process is recursively iterated by pairing up the sums to establish the next available level. The final step thus results in $2n - 1$ difference and one final sum. Haar DWT portrays the necessary features of wavelets in general. The first step performed uses *O* (*n*) operations; the next intermediate step involves capturing not only a notion of the input frequency content through different levels of examining, but also by temporal content like the number of frequency occurrences. These characteristics make the Fast Fourier Transform (FFT) an advanced alternative to Fast wavelet transform (FWT). The discrete wavelet transform has a huge number of applications most notably in signal coding, to elaborate a discrete signal in a much more redundant way is conditioned for data compression [stéphane G. Mallat 1999]. According to [Chin-Chen Chang *et* al 2004] the Haar transformation technique is applied in wavelet formation because of its simplicity and wide scalability. Haar wavelet uses a set of low-pass and high-pass filters for image decomposition. These filters are first applied in the image columns and then to the image rows independently. The output of the first level Haar wavelet produces four sub-bands. These four sub-bands are PP, MP, PM, and MM. The low-frequency band PP can be further decomposed into four sub-bands *pp*, *mp*, *pm*, and *mm*.

pp is a reduced resolution corresponding to low frequency level of the image. The other three sub-bands are high frequency level in the vertical, horizontal, and diagonal directions, respectively [N.G.Chitaliyaa *et* al 2010].Lastly the Haar transform is finally applied to the cross-multiplies function against the Haar wavelet with has within it various shifts and stretches, including Fourier transform. [Mohammed Alwakeel *et* al 2010]

iv). Facial Action Coding System

FACS is a system that uses the human anatomy for the selection of special features of interest. The face is the classified into two separate regions, the higher and lower part where the facial motions are classified into action units. The action units are the visible and different movements of the facial muscles that make it possible to create emotions. The facial expressions are initially identified using the underlying muscles movement that produces the expressions. Using two frames 8 feature points are marked bounding the eye brows regions in the first frame of each image using a computer mouse pointer. Each point of the frame consists of vertical and horizontal flows due to large facial feature motion displacement caused by sudden rising of the eyebrows. Using pyramid optical flow approach large feature point movements that are sensitive to precise facial motion such as eyebrow movement are used. The facial feature points are tracked automatically in the same way in the remaining frames of the image sequence.

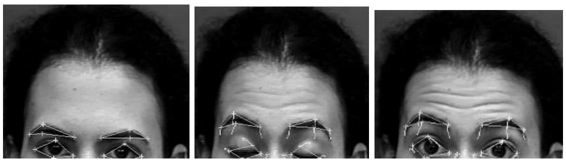


Fig 4

3. FEATURE CLASSIFICATION

1) Learning Vector Quantization (LVQ)

LVQ was developed by Kohonen as one of the most frequently used unsupervised clustering algorithms based on the winner-takes-all philosophy. Several versions of LVQ exist [Kohonen, T, 2001] and LVQ-I is most frequently used. LVQ-I has two layers, the competitive and output layer. The neurons in the competitive layer are also called sub-classes. Each sub-class contains a weight vector which is very similar to the input vector. When an input vector is applied to an LVQ network, the best match is searched in the competitive layer and the best match is called the winning neuron. When a particular neuron in the competitive layer wins, the particular output belonging to the class of the neuron is set high. Multiple neurons in the competitive layer may correspond to the same class in the output layer, but a neuron in the competitive layer is associated only with a particular class. It is for this reason that the neurons in the competitive layer are called sub-classes.

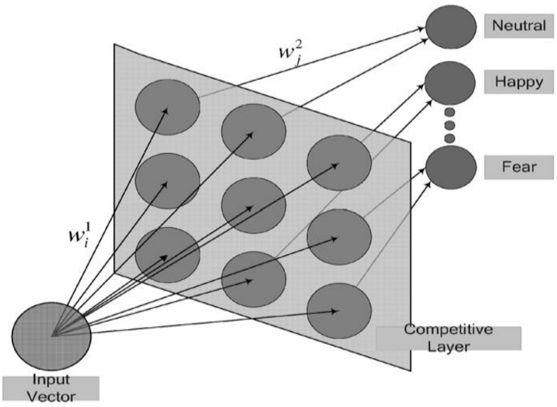


Fig.4 Linear vector Quantization diagram

The learning method commonly used with LVQ is the competitive learning rule in which for each training pattern, the competitive layer neuron that is the closest to the input is determined, the corresponding output neuron is called the winner neuron. The weights of the connections to this neuron are then adapted using the following equation

$$T_i^1(q) = \frac{t_i^1(q-1)+ \alpha\left(p - t_i^1(q-1)\right)}{t_i^1(q-1)- \alpha(p- t_i^1(q-1))}$$

In Eq. above t_i is the input layer weight while s is the input vector and α becomes the learning rate. The direction of the weight adaptation depends on firstly whether the class of the training pattern and that of the reference vector are same or not. If they are same, the reference vector is moved closer to the training pattern; otherwise, it is moved farther away. This movement of the reference vector is controlled by the learning rate. The learning rate is represented by a fraction of the distance moved by the training pattern. The learning rate decreases over time such that the initial changes finally obtained become greater than those obtained during the initial training of the process. [S. Bashyal 2008 *et* al] conducted an experiment by excluding the two expressers from the data set using JFFE (Japanese Female Facial Expression) dataset and obtained a recognition rate from 90.22%.

ii). Support Vector Machine

Support vector Machine is a supervised classification algorithm that classifies data through a set of support vectors. The support vectors are training inputs that provide the outline of a hyper plane in a feature space. The dimensional hyper plane which represents the number of features of input vectors sets the boundary between different classes. The training data used are classified into a set of binary classes. The kernel function is used to provide a generic mechanism that fits the hyper-plane into the data. The user provides a polynomial, sigmoid curve or a line which then selects surfaces along the surface of the function. This allows wide range of samples to be classified. The SVM uses the LIBSVM package for experimentation with wide features like parameterized kernel functions and multiclass classification. [Anvita Bajpai *et* al 2010] a mean accuracy level of 88.1% with the POFA dataset using the linear kernel with C - SVC formulation was attained.

iii). Convolution Neural Networks

CNN is very effective in learning features with a high level of abstraction if used with deeper architectures with new training samples. This type of hierarchical network has alternating types of layers, including convolution layers, sub-sampling layers and fully connected layers. Convolution layers are characterized by the kernel's size and the number of generated maps. [Andre Teixeira Lopes *et* al 2016]The kernel is shifted over the valid region of the input image generating one map. Sub-sampling layers are used to increase the position invariance of the kernels by reducing the map size [D.C. Cirean *et* al 2011]. The main types of sub-sampling layers are maximum-pooling and average pooling [D.C. Cirean *et* al 2011]. Fully connected CNN's layers are likened to those found in general neural networks, its neurons are fully connected with the previous layer (generally: convolution layer, sub-sampling layer or even a fully connected layer). The process of learning by CNNs is made up of identifying the most appropriate synapses' weights. Supervised learning can be performed using a gradient descent method, like the one proposed by [Y. Lecun *et* al 1998]. [Andre Teixeira Lopes *et* al 2016] achieved an accuracy level of 96.76% using the CK+ database for 6 expressions which are state of the earth. During classification the facial expression recognition system performs two learning stages in the classifier. The training phase receives grayscale images with their respective expression Ids and learns a set of weights for the network. Separation of the images is performed for validation to achieve the best set of images that provide accurate learning results. The testing phase uses the same methodology as the training phase: spatial normalization, cropping, down-sampling and intensity normalization. Its output is a single number which is the Id of one of the six basic expressions.

iv). Hidden Markov Model

The HMM finds an implicit time warping in a probabilistic between the hidden states and learns the conditional probabilities of the observations given the state of the model. In the case of emotion expression, the signal is the measurement of the facial motion. This signal is in a state of continuous motion, given expressions are conveyed at varied time intervals, with varying intensities even for the same individual. A HMM is represented as follows.

$$\lambda \quad = (A, B, \pi)$$

$$a_{ij} \; = p(q_{t+1} = Sj/qt = S_i), 1 \leq i, j \leq N$$

$$B \quad = \{b_j(O_t)\} = P(O_t | q_{\,t} = S_j), 1 \leq j \; \leq N$$

$$\pi_j \; = p\;(q_1 = \; S_j)$$

Where A is the state transition probability matrix, B is the observation n probability distribution, and π is the initial state distribution. The number of states is represented by *N.* It should be noted that the observations (O_t) can be either discrete or continuous and can be vectors. In the discrete case, *B* becomes a matrix of probability entries, and in the continuous case, *B* will be given by the parameters of the probability distribution function of the observation.

Basic operations in HMM include

a) How to efficiently compute a probability model

b) How to find the corresponding state sequence given a set of observations and the model.

c) How the parameters learn from the model

These three operations summarize the basic operations that are performed within a Hidden Markov Model. The first operation describes how a set of observations can describe a model; the second operation is an important utilized by the algorithm to recognize the expressions from the input image. The third operation explains how the models of the parameters are learnt. Using 5 subjects and 6 facial expressions the average recognition rate for person-dependent case was 78.49% using single HMM while for multilevel HMM 82.46% accuracy was achieved. [Cohen et al 2000]

v). Bayesian Regularized Recurrent Network

A Recurrent Neural Network (RNN) is a modification to this architecture to allow for temporal classification (Jordan 1986). In this there are closed loop paths from a unit back to itself. i.e. information is sent from later stages to prior stages. A layer of context augmented into the network, is capable of retaining information between each stage. At every time instant, new sources of information are supplied into the RNN. The prior contents of the hidden layer are passed into the context layer. These are then supplied once again to the hidden layer in upcoming steps. In this way, the network keeps a short term memory. As there is feedback, training is fast and accurate.

Elman network is a recurrent network which uses back propagation algorithm. It is invented by Jeff Elman (1990). It has two layers namely, one output layer and one hidden layer. There is a feedback connection from the output of the hidden layer to its input. Hence, the hidden layer is also known as recurrent layer. This feedback path helps the network to learn to recognize and generate temporal as well as spatial patterns. Recurrent network hidden layer neurons are hyperbolic tangent sigmoid neurons. i.e the activation function or transfer function of the neurons in this layer is hyperbolic tangent sigmoid function. The hyperbolic tangent sigmoid activation function is given by.

$$F(n) = \frac{2}{1+\exp(-2^{*}n)} - 1$$

Recurrent network has linear neurons in its output layer. These neurons give the output directly proportional to their input. The hidden layer of this network must have enough number of neurons to get the correct output.

[D. Sivakumar *et* al 2017] conducted an experiment using 6 expressions for testing and the results produced 100% accuracy with an absolute mean error of 29.17%.

Comparison of different accuracy levels

S.No	Method	Accuracy
1	Learning Vector Quantization	90.22%
2	Support Vector Machine	88.1%
3	Convolution Neural Networks	96.76%
4	Hidden Markov Model	82.46%
5	Bayesian Regularized Recurrent Network	100%

CONCLUSION

The various approaches presented have been exhausted reviewed and presented to bring out the most appropriate techniques required. The scope of these facial expression recognition techniques provide an enhanced way of taking care of future scope in this field. Challenges such as partial occlusion, variation in light conditions and different camera angles can take advantage of this approaches provided to improve the accuracy of the obtained results. The above mentioned steps of feature extraction, detection and classification are necessary for the proper and efficient discovery of facial expressions. Hence it’s efficient and sufficient to apply the correct technique to the required extraction feature.

REFERENCES

- “Automatic facial expression recognition using image processing and bayesian regularized recurrent neural network” International Journal of Computer Application (2250-1797) Volume 7– No.2, March - April 2017
- “Real-time Facial Emotion Detection using Support Vector Machines” Anvita Bajpai, Kunal Chadha .(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 1, No.2, 2010
- Ahmed, N., Natarajan, T., and Rao, K. 1974. Discrete cosine transform.*IEEE Trans. on Computers*, 23(1):90–93.
- Andre Teixeira Lopes, Edilson de Aguiar, Alberto F. DeSouza and Thiago Oliveira-Santos, Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order, Pattern Recognition, http://dx.doi.org/10.1016/j.patcog.2016.07.026
- Chin-Chen Chang, Jun-Chou Chuang and Yih-Shin Hu, 2004. "Similar Image Retrieval Based On Wavelet Transformation", *International Journal Of Wavelets, Multi-resolution And Information Processing*, Vol. 2, No. 2, 2004, pp.111–120.
- Chitaliya, N. G., C Engg and A. I. Trivedi. “An Efficient Method for Face Feature Extraction and Recognition based on Contourlet Transform and Principal Component Analysis using Neural Network.” (2010).
- D. C. Cirean, U. Meier, J. Masci, L. M. Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Two, IJCAI’11,AAAI Press, 2011, pp. 1237–1242.
- D.Bereziat, I. Herlin, and L. Younes, A generalized optical flow constraint and its physical interpretation, in Proceedings of CVPR’2000, 2000:487-492
- Dai, Ying, Yasuaki Nakano, and Hidetoshi Miyao. "Extraction of facial images from a complex background using SGLD matrices." *Proceedings of 12th International Conference on Pattern Recognition*. IEEE, 1994

- Ekman, P., 1971. Universals and cultural differences in facial expressions of emotion. In: Nebraska Symposium on Motivation. Lincoln University of Nebraska Press, pp. 207–283.
- IEEE 2016 Saranya R Benedict, J.Satheesh Kumar Geometric Shaped Facial Feature Extraction for Face Recognition
- Jeemoni Kalita, Karen Das (2013) Recognition of Facial Expression Using Eigenvector Based Distributed Features and Euclidean Distance Based Decision Making Technique (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 2
- Jun Ou，Xiao-Bo Bai*，Yun Pei ,Liang Ma, Wei Liu Automatic Facial Expression Recognition Using Gabor Filter And Expression Analysis, 2010 Second International Conference on Computer Modeling and Simulation.
- Lu H., Yang F. (2014) Active Shape Model and Its Application to Face Alignment In: Chen YW., C. Jain L. (eds) Subspace Methods for Pattern Recognition in Intelligent Environment. Studies in Computational Intelligence, vol 552. Springer, Berlin, Heidelberg
- Mahoor, Mohammad & Abdel-Mottaleb, Mohamed. (2006). Facial Features Extraction in Color Images Using Enhanced Active Shape Model.. 144-148. DOI 10.1109/FGR.2006.51.
- 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
- Roberto Brunelli and Tomaso Poggio," Face Recognition: Features Vs Templates", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol.15, pp.0162-8828, 1993
- S. Bashyal, G.K. Venayagamoorthy Engineering Applications of Artificial Intelligence 21 (2008) 1056–1064
- Sonka M., Hlavac V., Boyle R. (1993) Image pre-processing. In: Image Processing, Analysis and Machine Vision. Springer, Boston, MA https://doi.org/10.1007/978-1-4899-3216-7_4
- stéphane G. Mallat, 1999. A wavelet tour of signal processing, *Academic Press*, 1999
- T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European Conference on Computer Vision (ECCV), 2004.
- T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distribution, Pattern Recognition 29 (1)(1996) 51–59.
- Turha Ü. Çiğdem, M. Bilginer Gülmezoğlu, and Atalay Barkana. "Face recognition using common matrix approach." *Signal Processing Conference, 2005 13th European*. IEEE, 2005
- Vukadinovic, Danijela & Pantic, M. (2005). Fully automatic facial feature point detection using Gabor feature based boosted classifiers. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics 2. 1692 - 1698 Vol. 2. 10.1109/ICSMC.2005.1571392.
- Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition 86 (11) (1998) 2278–2324.doi:10.1109/5.726791.
- Yang Bai, Lihua Guo, Lianwen Jin and Qinghua Huang, "A novel feature extraction method using Pyramid Histogram of Orientation Gradients for smile recognition," *2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, 2009, pp. 3305-3308.doi: 10.1109/ICIP.2009.5413938