A Review on Facial Expression Recognition Techniques

Divya R. Kushwah[1]
M.E. (Computer)
Silver Oak College of Engineering &Technology,
Ahmedabad, India
Divyakushwah2026@gmail.com

Dr. Devang Pandya [2]
Prof. Computer department
Silver Oak College of Engineering &Technology,
Ahmedabad, India
Devang1979@gmail.com

Abstract—Facial expression is in the topic of active research over the past few decades. Recognition and extracting various emotions and validating those emotions from the facial expression become very important in human computer interaction. Interpreting such human expression remains and much of the research is required about the way they relate to human affect. Apart from H-I interfaces other applications include awareness system, medical diagnosis, surveillance, law enforcement, automated tutoring system and many more. In the recent year different technique have been put forward for developing automated facial expression recognition system. This paper present quick survey on some of the facial expression recognition techniques. A comparative study is carried out using various feature extraction techniques. We define taxonomy of the field and cover all the steps from face detection to facial expression classification.

***Keywords—**facial expression,FACS, multimodal, HMM,*

*****

I. INTRODUCTION

Facial expression plays vital role in indicating person's intention, feeling and other internal state. Research shows that the facial expression contributed to about 55% effect of overall emotion expression during interaction [1]. In the year of 1872 Charles Darwin was the one of the first scientist to recognize that facial expression is most powerful and essential mean for human to communicate.

In the year of 1972, Ekman, Friesen and Ellsworth[2] works on the idea of Darwin and found that according to psychology perspective facial expression were cultural specific like any culture has its own language, emotion also have a language of facial expression. Facial expression shows the mood or emotional state of an individual how he is feeling at a particular moment like sad, happy, anger etc. In additional it provide information about the cognitive state such as interest, boredom, confusion and states

Fig. 1: Primary emotions expressed on the face. From left to right: disgust, fear, joy, surprise, sadness, anger. [4].

Paul Ekman [3] give six universal emotion namely sad, happiness, anger, fear, disgust and surprise (see figure 1).

Facial expression recognition (FER) is an important and challenging problem in researcher communities of computer vision as it attracts due to its potential application. Areas such as human computer interaction, robotics, driving safety, surveillance. Different modalities are combined with speech, gaze and standard interaction like movement, mouse movement and key strokes [5]. Similarly one can build E-learning platforms and implement FER in order to understand the frustration level of the students. FER are implemented in robots with social skills [6]. Authors in [7] proposed a method to detect the pain level of patient from the video, frame by frame. Depression recognition from FE is very important application in analysis of psychological disorder. It is implemented in system that is used to communicate with the children who are suffering from autism or down syndromes.

II. INFERING FROM FER

FER consists of steps which are face detection, feature selection, feature extraction and feature classification. It is depicted by the following flow chart (see figure 2).

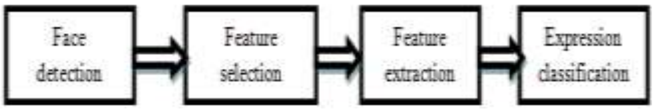


Fig.2: Block diagram of FER

In this process, the image or video are the input. They are loaded first and the basic step is face detection followed by feature selection where key point or face is parameterized. This process is very essential as it extracts the significant feature and help in extracting more relevant feature. Feature extraction is next step after this is classification and recognition approaches that infer emotions or mental states based on the extracted FE features. The vast majority of techniques use a multi-class classification model where a set of emotions mental states are to be detected.

III. TEXONOMY OF FER

Face detection is first step in the process of recognizing the expression. The public availability of model (eg. Open CV and matlab) its reliability for frontal image. Another open library dib make face detection easy. Many works have adopted the DPM i.e. deformable part model [8] to perform face detection by using feature such as face edge, texture, skin color And motion. HAAR classifier is one of the algorithms which are used for face detection. This feature can be easily scaled by increasing and decreasing the pixel to be examined. These features are suitable as it get selected in training phase which leads to achieving high detection accuracy.

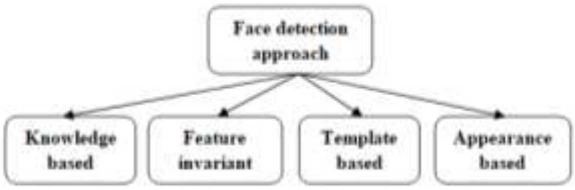


Fig.3: classification of face detection approach

There are four approaches for detecting face which represented in the above figure. **Knowledge based approach** detect the face using a set of rule (like human face has two eyes, nose, mouth and lip with the certain distance and position related to each other). **Feature invariant approach** is based upon structural feature of the face in which classified is trained and used to differentiate between facial and non- region. **Template matching approach** also known as image matching uses predefined face template to locate and detect the face also determine the face between the template and input image. **In Appearance based approach**, face detection done with Eigen value and show superior performance then the other approaches.

A. Parameterization

This section deals with the defining the coding scheme for the description of FEs. This scheme is further categorized into to as shown by the following diagram:

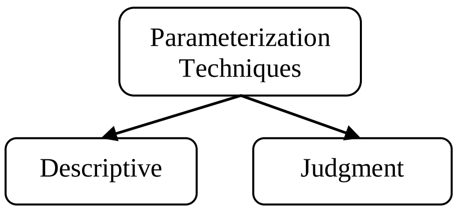


Fig.4: techniques of parameterization

Descriptive coding scheme do the parameterization of the FE in terms of surface property. This scheme focus on what face can do. The most known example is facial expression coding system (FACS) and face animation parameter (FAP).

1. *FACS(facial action coding system)*

The facial expression coding system, a total of 46 action unit which are related to contraction and relaxation of one or more set of muscle is defined. There is total 17 facial muscles that are connected to each other or with skin. The FACS manual is a comprehensive descriptive of facial behavior. FACS labels each observable component of facial movement as an action unit or AU. The FACS manual describes the criteria for observing and coding each AU. It also describes how Aus appear in combinations. (see figure 5)

1. *(FAP) Facial animation parameter*

A facial expression is about the deformation of facial features and muscles. Different combination of facial features and muscles produces different type of facial expressions. For instance, how to differentiate between the true smile and polite smile since both types of smile share the same action unit (AU) deformation which is the lip corner puller. In this case, the cheek raiser AU needs to be checked; if it's in action, the true smile is performed and vice versa.

FAPs are a set of parameters, used in animating MPEG-4 models that define the reproduction emotions, expressions and speech pronunciation. It gives the measurement of muscular action relevant to the AU and it provides temporal information that is needed in order to have a life-like facial expression effect [10]. There are many applications that can use information contained in FAPs, including audio-visual automatic speech recognition (ASR) [11], audiovisual speaker recognition, facial expression recognition [12-13], etc. FAPs can be used to concisely represent evolution of facial expressions. The FAPs that contain significant information about facial expressions are those that control eyebrow and mouth movements [14].

FACS itself is purely descriptive and includes no inferential labels. By converting FACS codes to EMFACS or similar systems, face images may be coded for emotion-specified expressions (e.g., joy or anger) as well as for more molar categories of positive or negative emotion.

Fig. 5: AU and their corresponding action [9]

Judgment coding scheme

This scheme on the other hand describes Fes in term of latent emotion or effect that is believed to generate them. Because the single emotion or affect may result in the multiple expression. The two examples that best explain this scheme are EMFACS and AFFEX. This are described briefly below.

1. EMFACS(emotion facial action coding system)

EMFACS gives us additive directive to interact emotion from FACS code. EMFACS only deal with AUs that correspond to emotion and other AUs are not discussed. The below table gives the sample rule and have discussed only few rule the reason is that there are many rule for each emotion. Which means different AU combination and varying intensities may result in same emotion with different degree.

Table I

SR.NO	SAMPLE AU	
	EMOTION	***COMBINATION***
1	ANGER	4+3+7+23
2	CONTENT	124A+14A
3	DISGUST	9+15+16
4	FEAR	1+2+4+5+20+26
5	HAPPINESS	6+12
6	SADNESS	1+4+15

A hybrid approach is to define emotion labels interms of specific signs rather than latent emotions or affects. Examples are EMFACS and AFFEX [15]. In each, expressions related to each emotion are defined descriptively. As an example, enjoyment may be defined by an expression displaying an oblique lip-corner pull co-occurring with cheek raise. Hybrid systems are similar to judgment-based systems in that there is an assumed 1:1 correspondence between emotion labels and signs that describe them. For this reason, we group hybrid approaches with judgment based systems.

B. Feature selection

This step is implemented to reduce the dimensionality and to extract the relevant feature. Many techniques have such proposed.PCA (principal component analysis) have been widely used for feature reduction [16]. Though it has less discriminating power. Y.Zang and et.al. [17] has manually selected feature, they extract the total of 24 representative features i.e. Δd Distance change using 22 key facial points out of 121. Another feature selection technique includes fisher linear discriminant (FLD).

FLD provide more class discriminative feature. It maximizing the mean between the classes and minimizing the variation with classes [18][19].moreover, the lajevardi and hussain [20] invested feature selection by adopting the GI, MI and mrMR combining it with Gabor filter. mrMR opted by author is greedy approach and heuristic based that can lead to optimal solution that would affect discriminative feature selection.

1. PCA(PRINCIPLE COMPONENT ANALYSIS)

Principle Components Analysis (PCA) is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, PCA is a powerful tool for analyzing data. It is a statistical pattern recognition and signal processing for data reduction and feature extraction. It is based on desired number of principal component of multi-dimensional data. The purpose of PCA is to reduce the large dimensionality of the data space to smaller intrinsic dimension of feature space that would be independent variable.

PCA also known as Eigen face method which defines a feature space and it work as follow. The image of rectangular matrix is converted into column vector which contain mean of the each row. Normalization is calculated by determining the difference between the column vector of an image and the mean vector that is calculated from all column vectors.

PCA also known as Eigen face method which defines a feature space and it work as follow. The image of rectangular matrix is converted into column vector which contain mean of the each row. Normalization is calculated by determining the difference between the column vector of an image and the mean vector that is calculated from all column vectors.

$$\mu = \left(\frac{1}{n}\right) \sum_{n=1}^{m} x_n \quad (1)$$

$$c = \left(\frac{1}{m}\right) \sum_{n=1}^{m} (x_n - \mu)(x_n - \mu)^t \quad (2)$$

The mean and covariance matrix is calculated by using equation (2). This normalization vector is given as an input to the principal component analysis which returns score in the image. The rows of scores correspond to the observations and columns to the components.

Principal component returns latent *i.e.*, a vector containing the Eigen values of the covariance matrix of the image. Rows in the matrix X, represents the observations and the columns correspond to the variables. 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 large computation also.

Principal component returns latent *i.e.*, a vector containing the Eigen values of the covariance matrix of the image. Rows in the matrix X, represents the observations and the columns correspond to the variables. 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 large computation also.

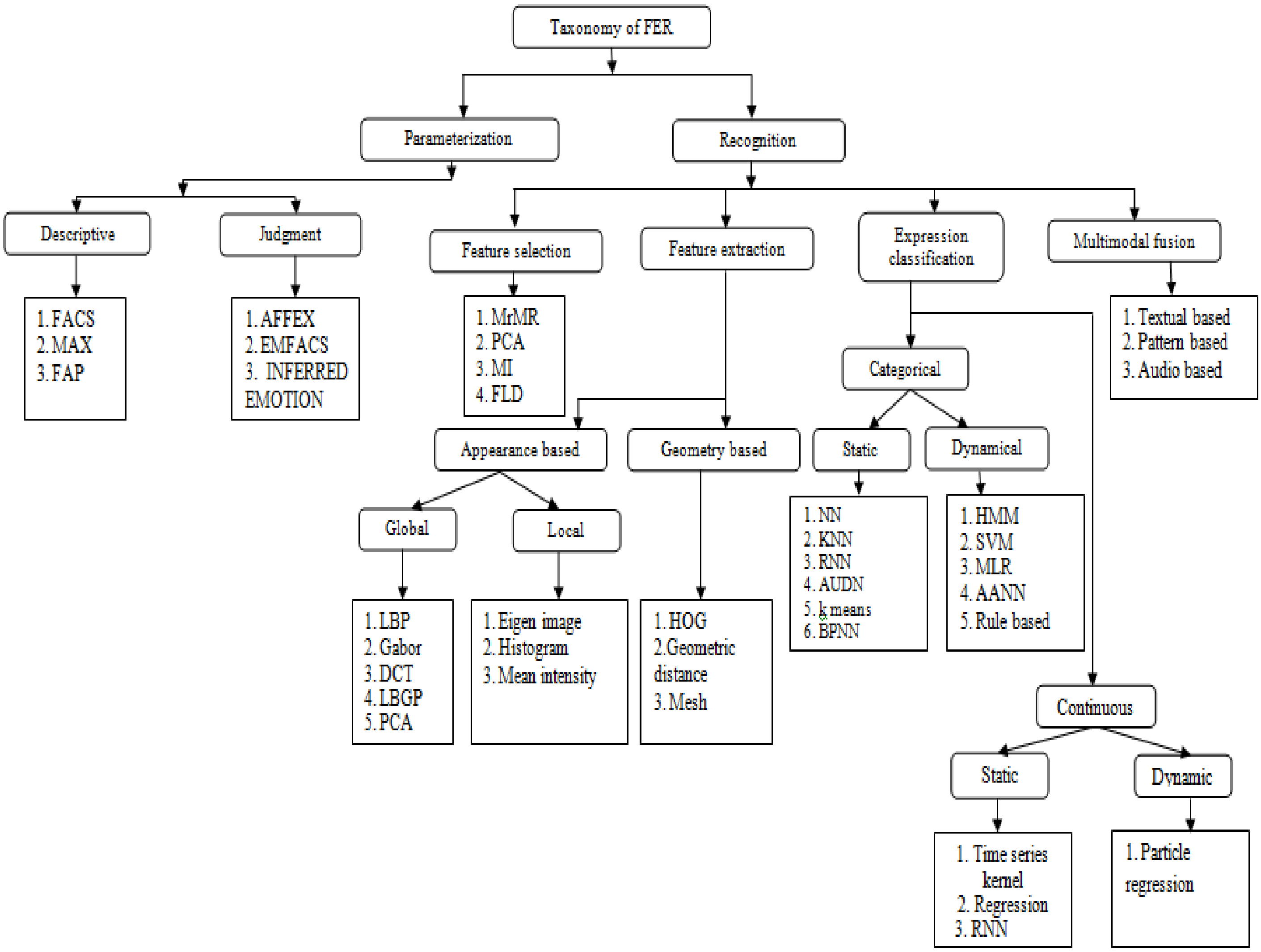
1. MI (mutual information)

Authors in [23] perform feature selection in the following manner. For each frame, the mutual information between the current frame and the initial frame is calculated, and the frame with the minimum mutual information is selected as the frame that represents an emotion with the maximum intensity. It is an information theoretic feature selection method that is not limited to linear dependencies, and is able to maximize the information in the class.

Mutual Information (MI) is a basic concept in information theory. It estimates the quantity of information shared between random variables. For two random variables U and V, their mutual information I(U;V) is defined as follows:

$$I(U;V) = H(U)H(V|U) \quad (3)$$

ISSN: 2454-4248
560 – 567



Where H() is the entropy of the random variable. The entropy H(U) quantifies the uncertainty of U. For a discrete random variable U, H(U) is defined as

$$H(U) = -\sum u \in Up(u)logp(u) \quad (4)$$

Here p(u) represents the marginal probability distribution of U. The conditional entropy H(U|V) quantifies the remaining uncertainty of U, when V is known. The research on the performance improvement has brought the development of MrMR (Minimal redundancy maximal relevance) which is variant of MI.

1. mrMR (Minimal redundancy maximal relevance)

Feature selection is one of the basic problems in pattern recognition and machine learning, identifies subsets of data that are relevant to the parameters used and is normally called *Maximum Relevance*. These subsets often contain material which is relevant but redundant and mRMR attempts to address this problem by removing those redundant subsets. mRMR has a variety of applications in many areas such as cancer diagnosis and speech recognition.

Features can be selected in many different ways. One scheme is to select features that correlate strongest to the classification variable. This has been called maximum-relevance selection. Many heuristic algorithms can be used, such as the sequential forward, backward, or floating selections.

On the other hand features can be selected to be mutually far away from each other while still having "high" correlation to the classification variable. This scheme, termed as ***Minimum Redundancy Maximum Relevance*** (mRMR) selection has been found to be more powerful than the maximum relevance selection [21]. In [22], an automatic feature selection method based on mrMR is proposed by the authors in order to identify the most discriminative and information feature sets for AUs for intensity estimation.

2. FLD (Fisher linear discriminant)

Fisher linear discriminant (FLD) has recently emerged as a more efficient approach for extracting features for many pattern classification problems as compared to traditional principal component analysis. It has been successfully applied to face recognition, which is based on linear projection from image space to a low dimension space by maximizing the between class scatter and minimizing the within class scatter. The main idea of FLD is that it finds the projection to line so that samples from different classes are well separated. The drawback of FLD is for the problem of small samples of facial expression recognition, within-class scatter matrix is difficult to solve for it may be the singular.

1.1 Mean intensity

The mean intensity μ in the selected region of interest (ROI) is given in (5)

$$\mu = \frac{1}{N}\int_{xy}^{1} ROI(x,y)dxdy \quad (5)$$

The intensity-based features are extracted from the gray-level or color histogram of the image. This type of features does not provide any information about the spatial distribution of the pixels. The intensity histogram in a hand is employed to define features. For the background corrected images, the gray value of each pixel is converted to its corresponding optical density (OD) as given in (6):

$$\mu = \frac{log\,10\;(gray\;\; value\;\; of background\;\;)}{gray\;\; value of pixel} \quad (6)$$

Subsequently, they compute the sum and the mean of the optical densities of the pixels to define its intensity-based features. In addition to the features extracted in each RGB color channel, intensity-based feature is also extracted based on difference between the red and blue components. In [27] the facial region is divided by a grid, applying a bank of Gabor filters at each cell and radially encoding the mean intensity of each feature map. Authors in [28] describe the appearance of gray scale frames by spreading an array of cells across the mouth and extracting the mean intensity.

1.2 Eigen vector and Eigen value.

Eigenvectors and Eigenvalues are dependent on the concept of orthogonal linear transformation. An Eigenvector is basically a non-zero vector. The dominant Eigenvector of a matrix is the one corresponding to the largest Eigenvalue of that matrix. This dominant Eigenvector is important for many real world applications. The motivation behind Eigenfaces is that the previous works ignore the question of which features are important for the classification, and which are not. Eigenfaces approach seeks to answer this by using Principal Component Analysis (PCA) of the facial images. This analysis reduces the dimensionality of the training set, leaving only those features that are critical for face recognition.

Eigenimage representation is a popular technique in facial analysis in both visual [29] and thermal spectrum images [30]. Eigenimages is based on Principal Component Analysis (PCA), where images are projected into a lower dimensional space that spans the significant variations (Eigenimages) among known face images. These variations are represented as the eigenvectors ν_i of the covariance matrix C. The eigenvalue λ_i associated with ν_i is the variance of the data distribution along the ν_i direction. Taking Φ_i as the difference between image I_i and an average image Ψ,

$$C = \frac{1}{R}\sum_{j=1}^{R} \Phi_j \Phi_j = \frac{1}{R} AA^T \quad (7)$$
$$A = [\Phi_1, \Phi_2, \dots \dots \dots \dots \Phi_n] \quad (8)$$

Where R is the number of images in the training set.

Computing the eigenvector of AA^T turns out to be computationally prohibitive, due to its size even for a moderate set of images. Instead of the R eigenvectors and eigenvalues υ_i and κ_i A^TA are computed because the size of the matrix RxR is smaller[29].

1.3 GLCM (Gray Level Co-occurrence Matrices)

The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features. The approach has been used in a number of applications, Third and higher order textures consider the relationships among three or more pixels are theoretically possible but not commonly implemented due to calculation time and interpretation difficulty. A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. (The texture filter functions, described in Texture Analysis cannot provide information about shape, that is, the spatial relationships of pixels in an image.)

C. Feature classification

FE classifications are divided into categorical and continuous depending on the target expression. In categorical group, there is predefined set of expression. Commonly for each one a classifier is trained. Although other ensemble strategies could be applied. Some work detect the primary expression [36][35].while other detect expression of pain, drowsiness and emotional attachment[7].

The continuous groups, FEs are represented as point in a continuous multidimensional space. The advantage of this is that they have ability to represent subtly different expression, mixture of primary expression. Expression recognition is grouped into static and dynamics. Static model evaluate each frame independently using classification technique such as Bayesian network classifier, support vector machine. On the other side dynamic model take account feature extraction independently from each frame to model the evolution of the expression over time. Some example are hidden markov model (HMM), variable-state latent conditional random field (VSL-CRF).

1. ANN(artificial neural network)

ANN is a machine learning algorithm that has been used for various pattern classification problems such as gender classification, face recognition, and classification of facial expression. ANN classifier has advantages for classification such as incredible generalization and good learning ability. The ANN takes the features vector as input, and trains the network to learn a complex mapping for classification, which will avoid the need for simplifying the classifier. Being able to offer potentially greater generalization through learning, neural networks/learning methods have also been applied to face recognition in [31].

An ANN is based on a collection of connected unites called artificial neuron which is analogous to biological neuron that is fundamental unit of brain. Each connection between consecutive neuron is known as synapse and can transmit a signal to another neuron. The receiving neuron can process the signal downstream neurons connected to it.

2. RNN (recurrent neural network)

Recurrent neural networks are connectionist architecture where one or more of the network layers are connected to itself. The self-connections allow the network to build an internal representation of past events, thereby allowing it to make flexible use of temporal context. In addition, the internal representation tends to be more robust to shifts and distortions in the input sequence (e.g. the same expression enacted at different speeds) than static methods. Since the network is designed for temporal patterns, the sequence of extracted features can be fed to it directly. One refinement to the basic recurrent architecture is the use of long short-term memory (LSTM) [32] cells. LSTM cells use linear units protected by multiplicative gates to store information over long periods of time. This extends the range of temporal context available to the net.

3. KNN (k-nearest neighbor)

K-nearest neighbors is used to classify local patches, performing a dimensionality reduction of the output through PCA and LDA and classify the resulting feature vector. In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. In K-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification.

The K-NN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, it can be useful to weight the contributions of the neighbors', so that the nearer neighbors' contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor. The neighbors' are taken from a set of objects for which the class (for K-NN classification) or the object property value (for K-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required [33].

4. BPNN (Backpropagation neural network)

Y.zang et al. [35] has estimated the intensities of 16 diagnostic AUs using feed forward neural network and support vector regressor. It has following application:

1. It is robust to noise and error involved in training data which may be inevitable in many supervised way.
2. It needs some training cost which depends heavily on sample size, the dimension of the training data, and the accuracy requirement. Once the model trained however it is extreme fast to be applied to be subsequent test instances. This is beneficial to real life application.

The training of a network by backpropagation neural network involves three stages: the feed forward of the input training pattern, the calculation and backpropagation of the associated Error and the adjustment of the weight .the data are feedforward from the input layer, through hidden layer, to output layer without feedback. Then, based on the feed forward error-backpropagation learning algorithm, backpropagation will search the error surface using gradient descent for point. Based on the error, the portion of error correction is computed, and then the weights for all layers are adjusted simultaneously

5. Naïve Bayesian (NB) Classifier

This is a Probabilistic method that has been shown very effective in many classification problems. This method considers that the presence of a particular feature of a class is unrelated to the presence of any other feature. The formula for classification used is

$$c = argmax\{P(C_i)\pi P(F_i|C_i)\} \quad (9)$$

Where $P(F_i \backslash C_i)$ is conditional tables or conditional density learned in training using examples. Naïve Classifier has shown very good classification results for many real datasets.

6. HMM (hidden markov model)

HMM is statistical model to describe a markov process with the hidden state. Its state are not visible but visible on the out state, when all the probability distribution of the output of each state is known, the next state is predicted through the current state, or through an associated output sequence. Therefore HMM is widely used in timing pattern recognition.
Authors in [34] have proposed the method in which they have adopted HMM as a classifier. They have calculated minimum and the maximum of each feature corresponding to each feature point and normalization is performed. They have then, initialized the HMM and achieved high accuracy while classifying.

7. SVM (support vector machine)

SVM is one of the most important classification techniques. The SVM classifier views the classification problem as a quadratic optimization problem. 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. The SVM classifier giver better generalization results than traditional neural network classifier. Authors in [36] have proposed method that uses adaboost and SVM for expression recognition. The classification is performed in two stages: support vector machine performed binary decision task. Seven SVM were trained to discriminate the emotion. The emotion category decision then implemented by choosing the classifier with maximum margin.

D. Multi model fusion

Many works have adopted multimodality for recognizing emotion, either by considering different visual modalities. Describing the face or more commonly by other source of information .fusing the multi modalities has the advantage to increase the robustness and provide extra information. Multi model learning for facial expression recognition is proposed by authors in [37]. Two modalities i.e. texture based and landmark based are fused. Texture modalities represent the facial image information, which displays face expression in pixel space. Landmark modalities represents face key points, the corresponding movement of which can help capture the facial expression. The fusion approaches followed by these works can be grouped into three main categories: early, late and sequential fusion depending upon when the fusion is conducted.

IV. CONCLUSION

Development of automated system that accomplishes facial expression recognition is difficult. Many methods have adopted in order to achieve accuracy while recognizing the facial expression. This paper has briefly overviewed each method of facial expression recognition. We have also discussed the taxonomy of FER that would be beneficial for us and other author to conduct more researches. From our study, we have concluded that many traditional approaches are focusing on recognizing a limited set of primary emotion. This is mainly due to lack of diverse datasets. But capturing the depth information factor and can be achieved by modern approaches. These approaches are invariant to pose, illumination and rotation and help to find subtle expression. Efficiency can also be increased by fusing the multiple models as they can enrich the representation space and improve the emotion inference.

REFERENCES

[1] G.U. kharat, S. V. Dudul, "*human recognition system optimally designed with different feature extraction*"vol.7 2008
[2] P.Ekman and W.V. friesen, "*pictures of facial affect*",consulting psycologist press,paco, alto.ca,(1996)
[3] P.Ekman and W.V. friesen, "*constant across culture in the face and emotion*",in the proceeding of journal of personality and social psycology in (1996)
[4] What when how.com
[5] Z. Duric, W. D. Gray, R. Heishman, F. Li, A. Rosenfeld, M. J.Schoelles, C. Schunn, and H. Wechsler, "*Integrating perceptual and cognitive modeling for adaptive and intelligent humancomputer interaction*," Proceedings of the IEEE, vol. 90, no. 7, pp.1272–1289, 2002.
[6] A. vinciarelli, M.Pantic, H. bourald, "*social signal processing:surey og an emerging domain*,"IVC vol.27,2009
[7] Pattrick mleey, jeffery cohn, simon lucey,"*automatically detecting pain using facial action*", in IEEE (2009)
[8] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. *Object detection with discriminatively trained part-based models. Trans. Pattern Anal. and Machine Intel.*, 32(9):1627–1645, 2010
[9] www.pinterest.com
[10] Zhang Y, Ji Q, Zhu Z, Yi B (2008) Dynamic facial expression analysis and synthesis with MPEG-4 facial animation parameters. IEEE Trans. Circuits Syst Video Technol 18(10): 1383–1396
[11] M. Pard`as and A. Bonafonte, "*Facial animation parameters extractionand expression detection using hmm*," in SPIC, 2002, pp.675–688.
[12] P. S. Aleksic and A. K. Katsaggelos, "*Automatic facial expression recognition using facial animation parameters and multistream hmms*," TIFS, vol. 1, no. 1, pp. 3–11, 2006.
[13] J. L. Landabaso, M. Pardàs, and A. Bonafonte, "*HMM recognition of expressions in unrestrained video intervals*," in Proc of ICASSP, pp. 197-200, Hong-Kong, China, 2003.
[14] M. Pardàs and A. Bonafonte, "*Facial animation parameters extraction and expression detection using HMM*," Signal Processing: Image Communication, vol.17, pp. 675-688, 2002.
[15] D. L. M. . H. E. A. Izard, C. E., " *A system for identifying affectexpressions by holistic judgments. Instructional Resources Center,University of Delaware*", 1983.
[16] G. liu, HH wechsier, "*evolutionary persuit and its application to face recognition*",IEEE trans (2000)
[17] Z. zeng, H.Zhang, R. Zhang, Y.Zhang, "*A hybrid feature selection method based on rough conditional mutual information and naïve bayesian classifier*",ISRN, app math (2014)
[18] S.K shah, "*A survey of facial expression recognition method*",ISOR.J.eng 4(4) 2014
[19] S. C. Neoh, li Zhang, Kamlesh mistry, Mohammed alamgir hossain, chee peng lim, naumam aslam, philip kinghorn, " *intelligent facial emotion recognition using a layered encoding cascade optimization model*", ELSEIVER (2014)
[20] S.M. lajevardi,Z.M hussain, "automatic facial expression recognition: feature extraction and selection, signal image video process" IEEE (2012)159-169
[21] https://en.wikipedia.org/wiki/Minimum_redundancy_feature_selection
[22] yang zang and li zang, " *adaptive 3D facial action intensity estimation and emotion recognition*" in Elsevier, pp 1446-1464, 2015
[23] Lajevardi, S. M., Hussain, Z. M., "*Feature selection for facial expression recognition based on optimization algorithm*", Second International Workshop on Nonlinear Dynamics and Synchronization (INDS'09), Klagenfurt, Austria, 2009.
[24] T. Ojala, M. Pietik¨ainen and D. Harwood, "*A comparative study of texture measures with classification based on feature distributions*" Pattern Recognition vol. 29, 1996.
[25] T. Ojala, M. Pietikainen, and T. M aenp "*a. Multiresolution gray- ¨ scale and rotation invariant texture classification with local binary patterns*" IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, July 2002
[26] T. Gehrig and H. K. Ekenel, "*Why is facial expression analysis in the wild challenging?*" in ICMI Workshops, 2013, pp. 9–16.
[27] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "*Facial expression recognition using radial encoding of local gabor*

featuresand classifier synthesis," PR, vol. 45, no. 1, pp. 80–91, 2012.

[28] A. Geetha, V. Ramalingam, S. Palanivel, and B. Palaniappan,"*Facial expression recognition–a real time approach*," Expert Syst.Appl., vol. 36, no. 1, pp. 303–308, 2009.

[29] M. Turk, A. Pentland, "*Eigenfaces for recognition*", Journal of Cognitive Neuroscience, no.3. pp.71-86, 1991.

[30] X. Chen, P. Flynn and K.W. Bowyer, "*PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparitive Studies*", International Workshop on Analysis and Modeling of Faces and Gestures, Oct., 2003.

[31] Omaima N, *"Review of the face detection system based artificial neural network algorithm*".IJMIA,Vol. 6(2014)

[32] M. Wollmer, M. kaiser,F.Eyben, B.schuller and G.Rigoll, "*Lstm-modelling of continous emotions in an audiovisual affect recognition framework*,"IVC vol.31, no.2, pp. 153-163,2013

[33] Alpesh kumar, Dr. nilamani,"expression using KMM and HMM,"IDSR-JECE, (2011)

[34] Shan he, shangfei wang and yanpeg,"*spontaneous facial expression recognition based on facial point tracking*,"IEEE(2011)

[35] Yang zhang, li zhang, M.A. hossain, "*adaptive 3D facial action intensity estimation and emotion recognition*", ELSEIVER (2014)

[36] Marian stewart bartlett, Gwen littlewort,Ian fasel,Javier R. movellan," *real time face detection and facial expression recognition: development and application to human computer interaction*",IEEE(2003)

[37] Wie zhang, Youmei zhang, lin ma, jingwie guan, shiefie, gong, "*multimodal learning for facial expression recognition*",ELSEIVER (2015)