A SURVEY ON FACIAL EXPRESSION RECOGNITION SYSTEM

[1]Priyanshi, [2] Dr. Neelu Jain
[1]Student , [2]Professsor
[1]Electronics and Communication Engineering Department,
[1]Punjab Engineering College(Deemed to be University), Chandigarh,India

Abstract : Facial expression carries information about the emotional and physical state of a human being. Facial expression recognition has been an active research area in the past ten years, with growing application areas such avatar animation, neuromarketing and sociable robots. As computers have started becoming a part of our living spaces and workspaces, and began to interact more and more with humans, the systems are supposed to be more accurate in understanding the emotional states and moods of humans. Having an intelligent facial expression recognition system makes the creation of good visual interfaces easier and helpful for human and computer interactions. In this paper the geometric feature extraction methods with the combination of different classifiers has been discussed . The Facial Expression Recognition system also captures the dynamic variation of facial expression along with the feature extraction method. The system learns and extract the features from the training image databases using deep Neural Network. The survey on facial expression recognition system has been done for various technique based on the average recognition rate for different image databases.

***IndexTerms* -Facial Expression Recognition, FACS, Feature Extraction, classification, Deep Networks**

I. INTRODUCTION

Human face to face communication is an ideal model for designing a multimedia human-computer interaction(HCI). Facial expression is one of the most powerful, natural and universal signal for human being to convey their emotions. For example, a smile may indicate a positive and optimistic mood while fear, sadness and disgust indicate negative mental state. In HCI the face plays an essential role in interpersonal communication[1].The two major areas where HCI research is going towards are on biometric data and emotion detection. Various advanced application areas are security, lie detection, mental state identification, face expression fusion, automated tutoring systems, operator exhaustion detection etc. Research into automatic Facial Expression Recognition(FER) is very imperative in this era of technology. Since, expression of face is one of the non-verbal communication method by which one can apprehend the mood or mental state of a person. Facial expression is the obvious transformation of the human face due to the reflex responses to the emotional instability. In most of the situations it is unprompted and uncontrollable. Mainly, facial expressions are the seven typical ones, namely; anger, fear, surprise, sad, disgust, happy, and neutral.[2]

Facial expressions recognition is very active and open research area in the domain of machine learning. FER have many real time and fully automatic applications like behavioral research, human computer interaction, video-calling and vision systems helps in interpreting non-verbal facial gestures. But the major challenge nowadays is to develop a system to perform natural interactions between man and a machine. The communication among humans is effective as they can give responses according to other person's expression, so for interacting effectively with the humans, the computers are also supposed to gain this ability. Human Computer Interfaces and robotics are not the only applications of facial expressions recognition systems, it rather finds its applications in several distinct areas like Video Games, Animations, Psychiatry, Educational Software, Sensitive Music, Medical science, Forensics, Criminal Interview etc. As the facial expressions recognition systems are becoming robust and real time, many other innovative applications and uses are yet to be seen.

In this paper, different techniques for Facial Expression Recognition are reviewed .In FER system the image is processed to extract such information which can help to recognize the six universal expressions. The processing is done in several phases including image acquisition ,feature extraction and finally expression recognition and classification, as shown in figure 1 . A face acquisition stage involves an automatic face detector that permits to locate the face in complex scenes with muddled backgrounds. Certain methods capture just the facial feature while other methods capture the face as a whole. Face analysis is convoluted by the variation in face appearance caused due to pose and illumination changes. Pose variations may occur due to scale changes as well as in-plane and out-plane variations. Therefore, it is suggested to do normalization before proceeding for the further analysis. But then again, it is not compulsory to do normalization with the understanding that features are normalized prior to classification. Feature extraction process is the next stage of FER system. Feature extraction is finding and depicting of positive features of concern within an image for further processing. In image processing computer vision feature extraction is a significant stage, whereas it spots the move from graphic to implicit data depiction. Then these data depiction can be used as an input to the classification.

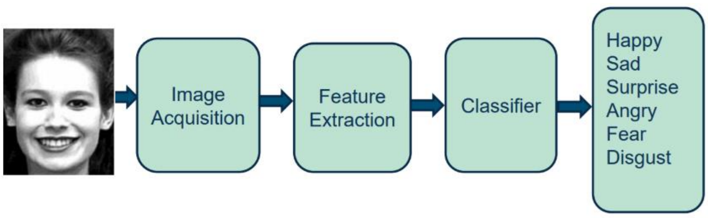

Figure 1: Facial expression recognition system

FER systems can be divided into two main categories according to the feature representations: static image FER and dynamic sequence FER. In static-based methods the feature representation is encoded with only spatial information from the current single image, whereas dynamic-based methods consider the temporal relation among contiguous frames in the input facial expression sequence. The majority of the traditional methods have used handcrafted features like LBP(Local Binary Patterns) , Gabor filters, feature tracking. For training and recognition of expression features K-nearest neighbor (KNN),support vector machine(SVM) has been considered in FER system. Recently, deep learning methods like Deep Neural Network(DNN),Deep Belief Network(DBN),Convolution Neural Network(CNN) have been attracting many artificial intelligence and machine learning applications. To extract the temporal features the images are given directly as input to deep spatio-temporal network models.

The Facial Action Coding System (FACS) is based on facial expressions by action units (AUs). Action units are defined in terms of the amount of muscular effort is invested with respect to one's neutral pose. FACS is thus defined independently of a person's facial structure and characteristics. As such, the same FACS coding can be made on individuals with completely different appearance . Out of 44 FACS AUs that they defined, 30 AUs are anatomically related to the contractions of specific facial muscles: 12 for upper face, and 18 are for lower face. AUs can occur either singly or in combination. Some of the AUs are shown in Figure 2 .When AUs occur in combination they may be additive, in which the combination does not change the appearance of the constituent AUs, or nonadditive, in which the appearance of the constituents does change. Although the number of atomic action units is relatively small, more than 7,000 different AU combinations have been observed .

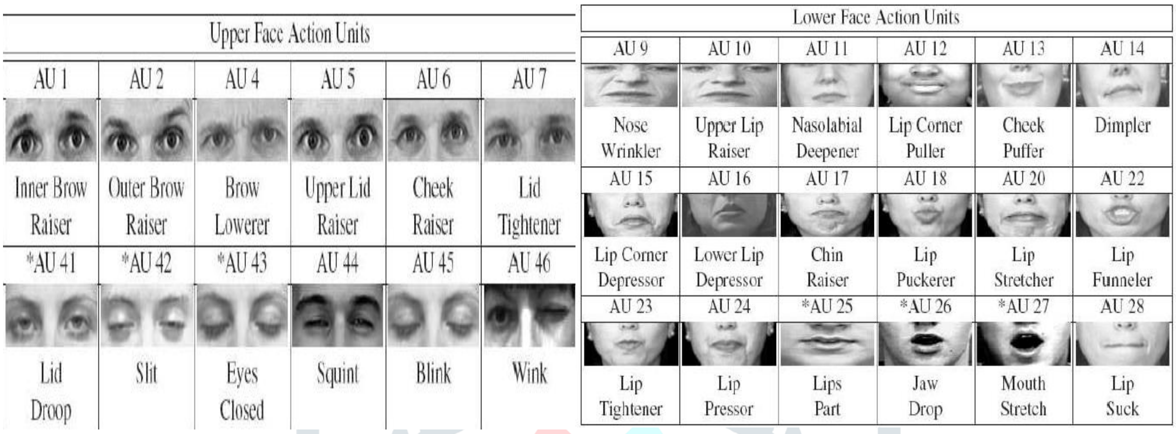

Figure 2: Facial Action Coding System action units

FACS provides the descriptive power necessary to describe the details of facial expression. Multistate face and facial component models are proposed for tracking and modelling the various facial features, including lips, eyes, brows, cheeks, and furrows. During tracking, detailed parametric descriptions of the facial features are extracted. With these parameters as the inputs, a group of action units (neutral expression, six upper face AUs and 10 lower face AUs) are recognized whether they occur alone or in combinations. [2]

Having sufficient labelled training data that include as many variations of the populations and environments as possible is important for the design of a deep expression recognition system. In this section, we discuss the publicly available databases that contain basic expressions and that are widely used in our reviewed papers for deep learning algorithms evaluation. We also introduce newly released databases that contain a large amount of affective images collected from the real world to benefit the training of deep neural networks. Table 1 provides an overview of datasets, including the number of subjects, number of image or video samples and expression distribution.

Table 1: Overview of Various Databases

Database	Samples	Subjects	Resolution	Expression Distribution
CK	486 image sequences	97	640 × 490	6 basic expressions plus and neutral
CK+	593 image sequences	123	640 × 490	6 basic expressions plus and neutral
MMI	740 images and 2,900 videos	25	720 × 576	6 basic expressions plus neutral
JAFFE	213 images	10	256 × 256	6 basic expressions plus neutral
Oulu-CASIA	2,880 image sequences	806	320×240	6 basic expressions
KDEF	4,900 images	70	762 × 562	6 basic expressions plus neutral
BU-3DFE	2,500 images	100	512 × 512	6 basic expressions plus neutral

II. LITRATURE SURVEY:

In this section various techniques that are dealing with the Facial expression are reviewed. Kotsia and Pitas[3] focused on the effect of occlusion when classifying six facial emotion expressions. In order to achieve this, several feature extraction techniques and classification models were combined achieving the maximum accuracy rate of 99.7% on CK database. Shan et al [4] implemented a recognition system based on texture features called Local Binary Pattern (LBP). This feature extraction technique has achieved average recognition rate of 93.8% on JAFFE database. Niu and Qiu [5] proposed a method that used an improvised version of Principal Component Analysis (PCA) and SVM for expression recognition. In this paper the authors mainly focused on the feature extraction process and improving its effectiveness. Song and Chen[6] has put forward a person dependent approach for facial expression recognition using a active appearance model (AAM) for feature extraction. The design mainly focused on robotic applications such as improving human-robot interaction with recognition rate of 98.7% on BU-3DFE database. Oliveira et al [7] has introduced a feature-selection method which is applied on the feature matrix extracted by the 2DPCA method .To show proposed approach's efficiency the two classifiers k-NN and SVM are used. Rivera et al [8] proposed a novel local feature descriptor, local directional number pattern (LDN) for face and expression recognition. This method produces more discriminative codes then other methods. Liu et al [9] proposed a AU-aware Deep Networks (AUDN) for facial expression recognition by elaborately utilizing the prior knowledge that the appearance variations caused by expression can be decomposed into a batch of local facial Action Units (AUs). Experiments on three expression databases CK+, MMI and SFEW with accuracy rate of 92.05%, 74.76% and 26.14% respectively. The recognition rate is poor when simulated on real-world conditions .Luo et al [10] , introduced a hybrid method of PCA and LBP based features extraction of global grayscale features which are environmentally sensitive. The SVM is used for facial expression recognition achieving the maximum accuracy rate of 89.64 % on RGB camera.

Table 2: A survey on some Facial Expression Recognition Systems

Author, Year	Feature Extraction	Classifier	Database (Average Recognition Rate)
Kotsia and Pitas, 2007.[3]	tracking of Geometric deformation	SVM	CK(99.7%)
Shan et al ,2009[4]	LBP	SVM	MMI (86.9%) JAFFE(93.8%)
Niu and Qiu, 2010[5]	Weighted PCA	SVM	CK (88.25%)
Song and Chen,2011[6]	Active Appearance Model	BPNN	BU-3DFE(98.7%)
Oliveira et al.2011[7]	2DPCA	Modified k nearest neighbor	CK(91.51%) JAFFE(89.67%)
Rivera et al., 2013[8]	LDN	SVM	CK+ (89.3%)
Liu et al,2013[9]	AU-aware deep networks	RBM	CK+(92.05%) MMI(74.76%) SFEW(26.14%)
Luo et al ,2013[10]	PCA +LBP	SVM	RGB camera(89.64%)
Dahmane and Meunier (2014)[11]	HOG	SVM	JAFFE (98.14%)
Shojaeilangari et al 2015[12]	Spatial temporal descriptor based on optical flow	Extreme Sparse Learning	CK+ (94.48%)
Hamster et al 2015[13]	Convolution autoencoder+CNN	Fully connected CNN	JAFFE(95.8%)
Happy and Rautray 2015[14]	LDA+PCA	SVM	CK (94.04%) JAFFE (92.22%)
Ryu et al. (2017)[15]	LDTP	SVM	CK+ (94.2%) JAFFE(93.2%)
Kuo et al 2018[16]	CNN	CNN	CK+ (98.47%)
Li and Deng, 2019[17]	DLP-CNN	DCNN	CK+(95.78%) MMI(78.46%) SFEW 2.0(51.05%)

Beside the facial representation problem, the same displayed facial expression may vary differently across humans. This can be true even for the same person in different contexts. To cope with these variable factors, the concept of prototype-based model is introduced by Dahmane and Meunier [11] as anchor modeling through a SIFT-flow registration. To characterize the facial expression appearance, histograms of oriented gradients (HOG) are processed on each image. The method obtained the best results 87% with the person-independent evaluation strategy on JAFFE and 83% on the complex setting of the GEMEP-FERA database. To robustly recognize the facial emotions in real-world natural situations, Shojaeilangari et al [12] proposed an approach called extreme sparse learning, which has the ability to jointly learn a dictionary and a nonlinear classification model. A pose-invariant Optical flow -based spatio-temporal descriptor is able to robustly represent facial emotions even when there are head movements while expressing an emotion with 94.48% average recognition rate. Hamster et al [13] proposed a new architecture based on the Multi-channel convolutional neural network (CNN) for recognizing facial expressions and 95.8% accuracy was achieved using fully connected CNN for classification. Happy and Rautray [14] used LBP and PCA for feature extraction. The Active facial landmarks are used to improve the localization of salient feature patches which are estimated during the training stage. The method is teste on CK and JAFFE database with recognition rate of 94.04% and 92.22 % respectively. Ryu et al. [15] uses a coarse grid for stable codes and a finer one for active. This multi-level approach enables the system to do a finer grain description of facial motions while still characterizing the coarse features of the expression. The accuracy rate measured on CK+ (94.2%)and JAFFE(93.2%) database. Kuo et al [16] proposed a framework extended to a frame-to-sequence approach by exploiting temporal information with gated recurrent units. In addition, an illumination augmentation scheme is developed to alleviate the overfitting problem when training the deep networks with hybrid data sources In Li and Deng [17] Deep Locality-preserving CNN(DLP-CNN) is proposed to preserve the locality closeness which enhance the discriminative power of deep features and has maximum accuracy rate of 95.78% on CK+ database.

Table 2 provide the summary of various techniques for facial expression recognition including database used and the average recognition rate observed. The comparison shows that deep learning methods are superior as compare to the other techniques with accuracy rate of 98.47% on CK+ database. These techniques are more suitable for complex databases.

III. CONCLUSION:

The facial expression recognition is used in many applications ranging from HCI ,surveillance to emotions . The scope is wide and a lot of work can be done in this field. During the survey it is observed that the most of work is limited only on six basic expressions and for a system to be more effective it should be able to detect micro expressions and deal with the different angles of head. The preprocessing of images like face normalization can be done in order to increase the accuracy .All methods only used standard databases like CK+, JAFFE,MMI etc. in which expression were performed by trained subjects. Therefore the future research work can include real time facial expression databases including micro expressions.

REFERENCES:

[1] Pantic, M. and Rothkrantz, L.J.M., 2000. Automatic analysis of facial expressions: The state of the art. IEEE Transactions on pattern analysis and machine intelligence, 22(12), pp.1424-1445.

[2] Tian, Y.I., Kanade, T. and Cohn, J.F., 2001. Recognizing action units for facial expression analysis. IEEE Transactions on pattern analysis and machine intelligence, 23(2), pp.97-115

[3] Kotsia, I. and Pitas, I., 2007. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE transactions on image processing, 16(1), pp.172-187.

[4] Shan, C., Gong, S. and McOwan, P.W., 2009. Facial expression recognition based on local binary patterns: A comprehensive study. Image and vision Computing, 27(6), pp.803-816.

[5] Niu, Z. and Qiu, X., 2010, August. Facial expression recognition based on weighted principal component analysis and support vector machines. In Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on (Vol. 3, pp. V3-174). IEEE.

[6] Song, K.T. and Chen, Y.W., 2011, November. A design for integrated face and facial expression recognition. In IECon 2011-37th Annual Conference on IEEE Industrial Electronics Society (pp. 4306-4311). IEEE.

[7] Oliveira, L., Mansano, M., Koerich, A. and de Souza Britto Jr, A., 2011. 2D Principal Component Analysis for Face and Facial-Expression Recognition. Computing in Science & Engineering, 13(3), pp.9-13.

[8] Rivera, A.R., Castillo, J.R. and Chae, O.O., 2013. Local directional number pattern for face analysis: Face and expression recognition. IEEE transactions on image processing, 22(5), pp.1740-175

[9] Liu, M., Li, S., Shan, S. and Chen, X., 2013, April. AU-aware Deep Networks for facial expression recognition. In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on (pp. 1-6). IEEE.

[10] Luo, Y., Wu, C.M. and Zhang, Y., 2013. Facial expression recognition based on fusion feature of PCA and LBP with SVM. Optik-International Journal for Light and Electron Optics, 124(17), pp.2767-2770.

[11] Dahmane, M. and Meunier, J., 2014. Prototype-based modeling for facial expression analysis. IEEE Transactions on Multimedia, 16(6), pp.1574-1584.

[12] Shojaeilangari, S., Yau, W.Y., Nandakumar, K., Li, J. and Teoh, E.K., 2015. Robust representation and recognition of facial emotions using extreme sparse learning. IEEE Transactions on Image Processing, 24(7), pp.2140-2152.

[13] Hamester, D., Barros, P. and Wermter, S., 2015, July. Face expression recognition with a 2-channel convolutional neural network. In Neural Networks (IJCNN), 2015 International Joint Conference on (pp. 1-8).

[14] Happy, S.L. and Routray, A., 2015. Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing, 6(1), pp.1-12.

[15] Ryu, B., Rivera, A.R., Kim, J. and Chae, O., 2017. Local directional ternary pattern for facial expression recognition. IEEE Transactions on Image Processing, 26(12), pp.6006-6018.

[16] Kuo, C.M., Lai, S.H. and Sarkis, M., 2018, June. A compact deep learning model for robust facial expression recognition. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 2202-22028). IEEE.

[17] Li, S. and Deng, W., 2019. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. IEEE Transactions on Image Processing, 28(1), pp.356-370.