ISSN: 2583-3286(Online)

IPEC Journal of Science & Technology
Vol. 01 (01), Feb. 2022, pp. 1-6

A Comparative Analysis of Various Methods for Human Facial Emotions Recognition

Shardul Singh Chauhan[1], Ajay Kumar[2], Rajan Rajarshi[3]

[1,2]*Assistant Professor, CSE Department, Inderprastha Engineering College, India*

[3]*Information Security Consultant, Department of Cyber Security, Nihilent Limited, India*

Accepted on 21 February 2022

Abstract: Our life in many ways is guided by our emotions, so knowing more about human emotions will helps us to know more about any human's nature. Since human emotions carry information so, it becomes very useful for man machine interaction. Since emotion recognition has attracted a lot of research, it has found its applications in various domains such as market analysis, customer satisfaction, monitoring, e-learning, medical, etc. An Emotion recognition system (ERS) make use of the facial images of humans captured for the detection of emotions. The study starts with the working of ERS system which helps the user to detect different kinds of human emotions. The aim of this work is to show a brief comparison of some of the existing methods which are used for detecting emotions based on various parameters like dataset used, accuracy achieved, etc. and to recommend the technique which recognizes emotions with a higher classification rate.

Keywords* — *computer vision, emotion recognition system (ERS), feature extraction, image preprocessing, image processing.

I. INTRODUCTION

An emotion [1] can be defined as various forms of the positions or motions of the muscles that lie below the face skin. With reference to the set of various theories, these forms of motions express the state of emotion of any individual to the observers. Emotions represent a form of non-verbal communication ([2] and [3]).

Humans can naturally adapt to an emotion either voluntarily or in an involuntarily way. Although the neural mechanisms which are responsible for controlling of the expression differs in each case. Emotion recognition is often an experience of emotion for the amygdala and brain is highly involved in the process of recognition.

The eyes of the humans are often viewed as paramount features of facial emotions. Aspects such as blinking rate can possibly be acclimated to designate whether a person is nervous or whether he or she is being mendacious. However, there are a lot of cultural differences which deals with the convivial propriety of maintaining eye contact or not.

Beyond the appurtenant nature of face emotions in verbalized type of communication between the people, they play a paramount role in communication with dactylology. Many phrases in dactylology include emotions in the exhibit.

There is an argument circumventing the question of whether emotions are ecumenical and macrocosmic exhibits among humans. Adherents of the Universality Hypothesis claim that many countenances are innate and have roots in evolutionary predecessors. The contrary of this view, question the precision of the studies used to test this claim and believe that the face emotions are conditioned, and that view of the people and understand the emotions in immensely colossal part from the convivial situations around them.

The work begins with a brief introduction to the ERS. Then the three fundamental steps of emotion classification which is followed by comparison of some popular techniques will be discussed. To identify a better method, a comparison table will be shown, which will compare some existing approaches which are based on the dataset considered and the accuracy achieved. The final goal is to put forward that method which possess a high classification rate and to improvise the efficiency of the suggested method.

II. STEPS INVOLVED IN ERS

An ERS process can be divided into three fundamental steps which are as follows:

A. Emotion Image Preprocessing
B. Emotion Image Feature Extraction
C. Emotion Image Classification

Date of Submission: 15 Jan 2022
Corresponding Author: Shardul Singh Chauhan (e-mail: shardul.chauhan@ipec.org.in).

Published online : 25 February 2022

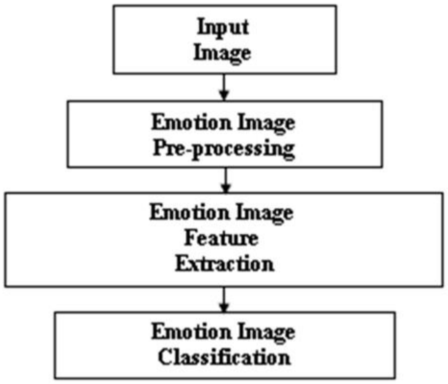

Figure 1 Emotion Recognition Steps

c

ases all the fundamental steps which are required to classify emotions of humans. A detailed description of all these steps is discussed as under:

A. Emotion Image Preprocessing

Here capturing of image is done at initial level by using an electronic device. The image captured is in RGB color model. For processing of this model further, it can be converted to some image color models also like: HSI model, YCbCr model, etc. The motive of this phase is to generate an improvised image. Here noise and unwanted distortion is removed to generate an image which contains more enhanced details. The obtained image is then used for identifying those regions where the face of the human being can be in an easy way.

B. Emotion Image Feature extraction

This phase deals with the segmentation of those details in images which are required to perform emotion classification. Initially that region is located where the face exists in an image. For figuring out the region where the face exists, an image mask is used. This image mask works by moving from pixel to pixel and then finds out the face region. Then image segmentation is performed, which involves segmenting out the region which has face region available in an image. Here rest of the details is discarded because facial emotions can be recognized by using faces of humans only. After locating the face regions, some features are discovered which is necessary for an emotion like eyes, eyebrows, ears, nose, lips, gap between nose and lips and so on. These features extracted are used for further classifying human emotions.

C. Emotion Image Classification

This includes the classification of emotions by using those features which are extracted in emotion image feature extraction phase. Each emotion possesses certain characteristics like the gap between nose and lips, eye movement, etc. based upon these extracted features and emotion is classified. The classification process implements an algorithm and starts with the training for some labeled input and output samples. This is done to train the system for identifying the various types of human facial emotions using a training dataset. After training the system, it is finally checked for its accuracy using test dataset and this shows that up to what extent the system has learned.

III. LITERATURE STUDY

Extracting faces using AdaBoost as in [4], which is followed by selection of the most appropriate models AAM1 & AAM2 based on degree of facial similarity and face direction. A frontal view of the image is generated by using a Vector Equation. Further SVM is used for classifying the emotions.

Neha Kulkarni et. al. [5] has used VCF and snakes for emotion recognition. Here the faces are divided in to 2 halves (upper and lower) and scrutinizes each one of them separately. It uses horizontal and vertical axes to figure out the facial features. Then extraction of the contours of the eyes, mouth and the eyebrows are performed by using active contour technique. This is followed by Facial features extraction.

Deep network framework is used in [6]. The concept of successive frames is employed to detect facial regions and the features are extracted from each detected region. The features are segmented using those algorithms that are suitable for segmentation. Automatic Extraction of Regional LBP Features uses the information given by equations of gray scale value of pixel and thresholding binary operator to generate histograms of facial regions. Each of these feature vectors is then concatenated and a concatenated histogram can be defined eventually. Data Fusion Using the method of Autoencoders involves extensive learning & implementation of mathematical functions to learn a better representation of a feature set in a compressed form. A feature vector x is used that corresponds to a facial image and the output of the classifier y represents the emotion class that has 6 emotions in it.

The work in [7] consists of Pose alignment and Shape Normalization. In Pose Alignment, 4 facial landmarks are considered as rigid points and are applied for the alignment of the tilted face image. In Shape Normalization, a technique for shape normalization makes use of an affine transformation to accommodate variability in shape. Here Bayesian type of Network is employed for representation of probability distribution and utilizes graph theoretic algorithms for learning and inference.

Making use of the concept of deep belief network and local direction based robust features for emotion identification as discussed in [8]. In Feature extraction, a LDDP code which is an eight-bit code is given to each pixel of an input depth face. Once locally salient features are obtained, it results in increase in the dimension of features. Hence PDA is adopted for dimension reduction.

K-means Approach [9] for detection of emotions. The Pre-Processing phase starts with several image processing stages which is required to for enhancement of the image quality and to minimize noise level. The mouth as well as eye region present in the image are detected using the edge detection method. Canny method for edge detection is applied as it has low error rate. Feature extraction step is performed, and the features of mouth along with eyes are stored into a dataset for use in classification. Three equal zones are created, the smile mouth expression falls in upper and middle zones and the neutral expression in middle and lower zones.

Feature extraction in [10] is done by splitting face in 80 regions and optical flows are calculated for each region based on the movements in vertical directions. Here discrete type of Hopfield neural networks is implemented for emotion recognition. The optical flow data is converted in discrete form for the neural networks. Furthermore, personal learning is used for neural networks.

Contourlet transform as in [11]. Decomposition of the image is done at different levels to obtain band pass image, which is further processed by the DFB method. Image preparation and face detection is done using Viola jones method which is tested on Caltech Image Database. The feature area is extracted by cropping image according to the predefined measurements. Further, gaussian smoothing filter is applied to resize the image and then histogram equalization is performed to overcome illumination differences. Feature selection phase involves entropy of different directional sub bands at different levels.

The work in [12] shows a 3D Average Face and Ameliorated AdaBoost for identification of emotions. If after a facial detection, the features are not extractable, an average face is reconstructed in 3-D. The rotation and coordinate algorithms for the same are implemented. A weak classifier based on Hair-Like feature is constructed to train emotional features by multi-class AdaBoost.

Emotions are recognized using Hessian Regularized Support Vector Machine in [13]. In this technique the Face images are normalized. Facial features are extracted. To avoid dimension dilemma, only the Gabor coefficients are extracted on some fiducially face points. By making use of the extracted Gabor coefficients, Hessian matrix was computed. The Hessian matrix and the extracted features were provided as an input to the HesSVM classifier for training.

An image is inputted in [14] for emotion categorization. Multi-pose AAM templates are fitted that includes view based AAM templates training, orientation assignment, expression key point descriptor and optimal AAM template matching. This will help in facial landmark tracking. A SIFT descriptor is run that gives AAM_SIFT feature descriptions, feature points which are later grouped giving a more detailed description of the face. The centers are clustered, and the regions are weighted based on FCM. The expressions are classified with the help of SVM classifier.

Multi -Scale Filter [15] begins with image acquisition which can be done with a camera. The image is preprocessed. Gabor Wavelet can be used for the same as it gives detailed and appealing information about the image. The filters are implemented for emotion feature extraction because of its optimal localization properties in both spatial as well as frequency domain.

An auto encoder is a symmetrical three-layer neural network [16], where encoding is done in the input layer and decoding in the output layer and there is a layer between them which is the hidden layer. In these layers, the input features are mapped into the hidden representation which is further mapped into the output layer. Deep sparse auto encoders are a neural network that is built by stacked sparse auto encoders are used for improvising the results.

The work cited in [17] has come up with Boosted NNE Collections for Multicultural Facial Expression Recognition. Here the probabilities of the existence of the expression are calculated and the output of the neural network is combined which indicates the existence of a particular face expression. Inter Expression Correlation method is used to quantify the similarities between various expressions and values obtained. System Verification involves the performance of system using test data is evaluated.

A framework which is based on the concept of local Zernike moment and motion history image [18] for emotion identification. Feature extraction consists of 8 steps which begin with computation of Motion History Image, followed by Optical Flow Algorithm where the dynamic information is captured, then comes the Optical Flow Based MHI. Next there is entropy where the probabilities of intensity values occurring in the face image are examined. The entropy values of discrete intensities and enMHLOF are examined. The local spatial textural information is incorporated to achieve a higher classification rate. Dimension reduction is done using PCA technique.

A 3D emotion recognition [19] by making use of kernel method on Riemannian manifold is discussed. The data is acquired in the beginning, then preprocessing of the image where the image undergoes a lot of processes like smoothing etc. to essentially refine it. After that feature points are extracted, and feature vector is computed which encodes the local geometric and spatial properties of the point and each facial patch is characterized in a covariance matrix. When the covariance matrices are extracted and geodesic distances have been defined, a Gaussian Kernel representation can be done, and its expression is shown in the paper. A multi-class SVM classification algorithm is used for emotion categorization.

An automated technique to discover the pixels present in a face image is discussed in [20]. The process begins with Preprocessing, where the image undergoes several operations such as rotating and reshaping the face dimensions, normalizing face size to 50*50 and histogram

Published online : 25 February 2022

equalization, wherein the features are detected as well as the orientations etc. are corrected and the width and height values are computed.

A fuzzy based classification approach is discussed in [21]. For the face segmentation, Viola-Jones method is used to segment the face from the image background. The irrelevant contents of the face are removed by splitting the face image into its 3 color components, subtracting the green component and then the resulting image is binarized. Adjustments of the face dimension wherein the moments of the image are estimated as well as the edges of the faces are estimated and finally the face image is segmented. Then the mouth and forehead/eye segmentation are done. Feature extraction is done with the help of Gabor functions. PCA is performed for dimensionality reduction. The classifier operates in a supervised form to generate several clusters for each class.

A PCA based dictionary method [22] for building an efficient emotion categorization system via a sparse representation. A class label is allocated to test sample by making use of the coefficients of the vector estimated. Different images are used to create a dictionary. Difference in images is studied with different expressions.

An experimental method to showcase the influence of facial expressions on recognition performance has been implemented [23]. The experimental design consists of 2 factors. The very first factor over here is represented by emotional display of the person targeted. The other factor is determined by an emotional display of identification of previously seen person. Each factor has emotional displays of positive and negative facial expressions.

An Emotion Recognition System along with its application which is based on the concept of PSO-SVM and the curvelet transform is implemented in [24]. Here, the images are first normalized to 100*100.The curvet transform of the image is done. Dimension reduction of the curvelet coefficients is performed because the size of curvelet coefficients was extremely large. For this problem 2 types of SVM are used i.e., LibSVM and PSO-SVM. In the first, 2 matrices are used for classification which are label and attribute matrices. The PSO-SVM is a parallel evolutionary computation technique; this is where optimization of SVM parameters takes place.

Sparse local fisher discriminant analysis [25] is employed, which helps in extraction of facial features, enhancing discriminant power of LFDA and work well for multimodal problem and to control weights of original variables or features. Moreover, the algorithm for sparse LFDA is formulated with the help of concepts of matrices. Experiments have been performed on multiple databases.

Table 1 shows a comparative analysis of some of the work done for recognizing human emotions based on some aspects like dataset considered, accuracy achieved, etc.

Paper	**Algorithm Used**	**Dataset Used**	**Rate of Accuracy of the System**
[4]	Support Vector Machine	ATR Facial Expression Image	82%
[5]	Support Vector Machine	Own Dataset	Approximately 90%
[6]	Self-Organized Maps based classifier	MMI and extended Cohn-Kanade (CK+)	97.55% and 98.95%
[7]	Bayesian Network	Own Dataset	95.7%
[8]	Deep belief Network	Own Dataset	96.67%
[9]	K- Means	Own dataset	76.5%
[10]	Discrete Hopfield Neural Networks	Own dataset	92.2%
[11]	Support Vector Machine	JAFFE and Cohn-Kanade	98.67-99.33% And 95.33%-99.67%
[12]	AdaBoost	JAFFE	85%-95%
[13]	Hessian regularized Support Vector Machine	JAFFE	Around 95%
[14]	Fuzzy C Means Clustering Algorithm	BU-3DFE	81.4%
[15]	Gabor filter	Cohn-Kanade AU-Coded Facial Expression	89.04%
[16]	Deep Sparse Autoencoders	Cohn-Kanade (CK+)	95.79%
[17]	Boosted NNE Collection	JAFFE, TFEID, and RadBoud	52.76%, 55.85%,68.57%
[18]	Support Vector Machine	CK+ and MMI	88.3%,79.8%
[19]	Support	BU-3DFE	92.62%, 86.14%

	Vector Machine	and Bosphorus	
[20]	Multi-Layer Perceptron	GENKI, JAFFE and FERET	92%,82%,91%
[21]	Fuzzy classification	KDEF	97.5% approximately
[22]	Principal Component Analysis based dictionary building	CK+, MMI, and Bosphorus	95.67%, 78.51%, 72.41%
[23]	Own method	Own dataset	Approximately 95%
[24]	Particle Swarm Optimization with Support Vector Machine	JAFFE	94.94%
[25]	Sparse Local Fisher Discriminant Analysis	JAFFE and C-K	Around 90%, around 97%

IV. CONCLUSIONS

The work cited some of the popular techniques which have been used for identifying the facial emotions. An analysis of these methods along with their accuracy rates has been carried out for emotion recognition. For figuring out a better algorithm various parameters were taken into consideration like accuracy rate, dataset used, and the type of features used. Based on this work, it can be finally concluded that Support Vector Machine technique produced the best results when compared to some of the other methods. It is obvious that always there are chances of improvement of the SVM algorithm too. So, the efficiency of this technique can be enhanced by combining it with other classifier along with the combination of appearance and geometric based features .

REFERENCES

[1] A.Freitas-Magalhães, “Facial Expression of Emotion from Theory to Application”, Porto: FEELab Science Books, 2011.

[2] A. J. Fridlund, “Human facial expression”, Academic Press, 1994.

[3] J. R. Dolls and J. Fernandez, “The psychology of facial expression”, Cambridge University Press, 1997.

[4] T. T. Y. A. Tatsu Okada, "Pose robust and person independent facial expressions recognition using AAM selection", IEEE 13th International Symposium on Consumer Electronics, Kyoto, 2009.

[5] N. Kulkarni, S. Kulkarni, M. Pardeshi, and P. Sanap, "Facial expression recognition using VFC and snakes", International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, 2015.

[6] A. Majumder, L. Behera and V. Subramanian, "Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion", IEEE Transactions on Cybernetics, vol. 48, no. 1, pp. 103-114, 2016.

[7] M. Singh, A. Majumder and L. Behera, "Facial expressions recognition system using Bayesian inference", International Joint Conference on Neural Networks(IJCNN), Beijing, 2014.

[8] M. Z. Uddin, M. M. Hassan, A. S. Almogren, A. Alamri, M. A. AlRubaian, and G. Fortino, "Facial Expression Recognition Utilizing Local Direction-Based Robust Features and Deep Belief Network", IEEE Access, vol. 5, pp. 4525-4536, 2017.

[9] A. M. Zeki, R. b. S. Ali and P. Appalasamy, "K-Means Approach to Facial Expressions", International Conference on Information Technology and e-Services, Sousse, 2012.

[10] M. Yoneyama, A. Ohtake, Y. Iwano and K. Shirai, "Facial Expressions Recognition Using Discrete Hopfield Neural Networks", International Conference on Image Processing, Santa Barbara, 1997.

[11] S. Biswas and J. Sil, "An Efficient Expression Recognition Method using Contourlet Transform", 2nd International Conference on Perception and Machine Intelligence, Kolkata, 2015.

[12] J. Chen, Y. Ariki and T. Takiguchi, "Robust Facial Expressions Recognition Using 3D Average Face and Ameliorated AdaBoost", MM '13 Proceedings of the 21st ACM international conference on Multimedia, Barcelona, 2013.

[13] C. Song, W. Liu and Y. Wang, "Facial expression recognition based on hessian regularized support vector machine", ICIMCS '13 Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, Huangshan, 2013.

[14] F. Ren and Z. Huang, "Facial Expression Recognition Based on AAM–SIFT and Adaptive Regional Weighting", IEEJ TRANSACTIONS ON

Published online : 25 February 2022

ELECTRICAL AND ELECTRONIC ENGINEERING, vol. 10, no. 6, pp. 713-722, 2015.

[15] J. Ou, "Facial Expression Recognition Analysis with Muti-Scale Filter", International Conference on Solid State Devices and Materials Science, Macao, 2012.

[16] N. Zeng, H. Zhang, B. Song, W. Liu, and Y. Yi, "Facial Expression Recognition via Learning Deep Sparse Autoencoders", Journal on Neurocomputing, vol. 273, no. C, pp. 643-649, 2018.

[17] G. Ali, M. A. Iqbal, and T.-S. Choi, "Boosted NNE Collections for Multicultural Facial Expression Recognition", Pattern Recognition, vol. 55, no. C, pp. 14-27, 2016.

[18] X. Fan and T. Tjahjadi, "A Dynamic Framework Based on Local Zernike Moment and Motion History Image for Facial Expression Recognition", Pattern Recognition, vol. 64, no. C, pp. 399-406, 2017.

[19] W. Hariri, H. Tabia, N. Farah, A. Benouareth and D. Declercq, "3D facial expression recognition using kernel methods on Riemannian manifold", Engineering Applications of Artificial Intelligence, vol. 64, no. C, pp. 25-32, 2017.

[20] T. Danisman, I. M. Bilasco, J. Martinet and C. Djeraba, "Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron", Signal Processing, vol. 93, no. 6, pp. 1547-1556, 2013.

[21] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Facial Expression Recognition with Automatic Segmentation of Face Regions using a Fuzzy based Classification Approach", Knowledge-Based Systems, vol. 110, no. C, pp. 1-14, 2016.

[22] M. Mohammadi, E. Fatemizadeh and M. Mahoor, "PCA-based dictionary building for accurate facial expression recognition via sparse representation", Visual Communication and Image Representation, vol. 25, no. 5, pp. 1082-1092, 2014.

[23] F. A. Pavela and E. Iordănescu, "The influence of facial expressions on recognition performance in facial identity", Procedia - Social and Behavioral Sciences, vol. 33, pp. 548-552, 2012.

[24] M. Tang and F. Chen, "Facial expression recognition and its application based on curvelet transform and PSO-SVM", Optik - International Journal for Light and Electron Optics, vol. 124, no. 22, pp. 5401-5406, 2013.

[25] Z. Wang, Q. Ruan and G. Ana, "Facial expression recognition using sparse local Fisher discriminant analysis", Neurocomputing, vol. 174, no. B, pp. 756-766, 2016.