International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-08, Issue-02, MAY 2022

Analysis of Human's Emotional Expression: A Review

*Vaishnavi Borgaonkar, #Prof. Sarita Sanap

*PG Student, #Assistant Professor, Maharashtra Institute of Technology college of engineering, Aurangabad, India, *vaishnavi910deshmukh@gmail.com, #sarita.sanap@mit.asia

Abstract: Facial emotions are important factors in human communication that help us understand the intentions of others. In general, people infer the emotional states of other people, such as joy, sadness, and anger, using facial expressions and vocal tone. According to different surveys verbal components convey one-third of human communication, and nonverbal components convey two-thirds. Among several nonverbal components, by carrying emotional meaning, facial expressions are one of the main information channels in interpersonal communication. Therefore, it is natural that research of facial emotion has been gaining lot of attention over the past decades with applications not only in the perceptual and cognitive sciences, but also in affective computing and computer animations. The most effective way to recognize human's emotion is to notice human's facial expressions. The facial expressions have rich information about human's emotion or mood. For that reason, if computer animated agents or robots can automatically recognize the facial expressions, those artificial systems are easily able to understand or estimate human's emotion or mood. This recognition technique can be also used as a component of human-robot interaction (HRI).

***Keywords* —review, human expression, RNN, CNN,**

I. Introduction

Human face is a very special and it has been used to express emotions by the human being in order to express their feelings to others. But when talking about Human Robot Interaction (HRI), it becomes difficult to identify the exact emotion. So far, there are seven recognized emotions: Natural, Happy, Sad, Anger, Surprise, Fear, and Disgust. However, apart from these emotions, human being can express combination of these emotions. Therefore, recognition of the expression and what emotional expression should be given to the person as a response becomes more difficult.

II. Literature Review

A. Existing scenario

Panti, et al. [1] has done survey on work done in automating facial expression analysis in facial images and image sequences identifies three basic problems related to facial expression analysis. The characteristics of an ideal automated system for facial expression analysis has been described in this paper. These problems are: face detection in a facial image or image sequence, facial expression data extraction, and facial expression classification. The capability of the human visual system with respect to these problems also has been described. Under facial expression analysis, very first, face from the image has been detected. After which facial expression analyzation has been done. In preceding step, facial expression information from the observed facial image or sequence has been extracted.

Essa, et al. [2] fitted a 3D mesh model of face geometry to a 2D face image and classified five facial expressions using the peak value of facial muscle movement.

Lanitis, et al. [3] used Active Appearance Models (AAM) to interpret the face, and Huang and Huang used a gradient-based method to classify face shape using a Point Distribution Model(PDM) [4].

Zhang, et al. [5] fitted the face with sparsely distributed fiducial feature points and distinguished facial expressions. Generally, these model-based methods are robust to occlusions. However, they are inadequate for a real-time system because they require much time to fit the model to the face image and need high resolution input images to analyze facial expressions.

Kim, et al. [6] introduced an artificial facial expression mimic system which can recognize facial expressions of human and also imitate the recognized facial expressions. They proposed a classifier that is based on weak classifiers obtained by using modified rectangular features to recognize human facial expression in real-time.

Breazeal, et al. [7] developed a sociable robot Leonardo, which has an expressive face capable of near human-level expression based on an active binocular vision

 DOI : 10.35291/2454-9150.2022.0192

system to recognize human facial features.

To achieve affective Human-Robot Interaction, mostly primary focus is on presenting different methodologies commonly used for facial expression recognition and imitation. The automatic recognition of emotions is necessary multimodal, that is, it requires of verbal and non-verbal channels (face, gesture, and body language), physiological signals or midterm activity modeling, among others [9], [10]. One of the most significant works used by the scientific community in facial expression recognition using visual information is based on Paul Ekman's study [11]. This author identifies and classifies the facial expressions through the study of different facial muscles in each expression, giving rise to so-called Facial Action Coding System (FACS). The recognition of facial expressions is a very diversified field in its classification or detection methods, ranging from the use of active Appearance Models (AAM) [11], Support Vector Machines (SVM) [12], Gabor filter bank [13] and Dynamic Bayesian Network (DBN).

B. Deep-Learning Based FER Approaches

Kahou et al. [14] proposed a hybrid RNN-CNN framework for propagating information over a sequence using a continuously valued hidden layer representation. In this work, the authors presented a complete system for the 2015 emotion recognition in the wild (EmotiW) challenge and proved that hybrid CNN-RNN architecture for a facial expression analysis can outperform a previously applied CNN approach using temporal averaging for aggregation.

Kim et al. [15] utilized representative expression-states (e.g., the onset, apex, and offset of expressions), which can be specified in facial sequences regardless of the expression intensity. The spatial image characteristics of the representative expression-state frames are learned using a CNN. In the second part, temporal characteristics of the spatial feature representation in the first part are learned using an LSTM of the facial expression.

Breuer and Kimmel [16] employed CNN visualization techniques to understand a model learned using various FER datasets, and demonstrated the capability of networks trained on emotion detection, across both datasets and various FER-related tasks. Jung et al. [17] used two different types of CNN: the first extracts temporal appearance features from the image sequences, whereas the second extracts temporal geometry features from temporal facial landmark points. These two models are combined using a new integration method to boost the performance of facial expression recognition.

Zhao et al. [18] proposed deep region and multi-label learning (DRML), which is a unified deep Network. DRML is a region layer that uses feed-forward functions to induce important facial regions, and forces the learned weights to capture structural information of the face. The complete network is end-to-end trainable, and automatically learns representations robust to variations inherent within a local region.

Therefore, it is necessary to reduce the computational burden at inference time of deep learning algorithm. Among the many approaches based on a standalone CNN or combination of LSTM and CNN, some representative works are shown in Table 2.

Reference	**Emotions Analyzed**	**Recognition Algorithm**
Hybrid CNN-RNN [14]	23 basic and compound emotions	Hybrid RNN-CNN framework for propagating information over a sequence Using temporal averaging for aggregation
Kim et al. [15]	Six emotions	Spatial image characteristics of the representative expression-state frames are learned using a CNN Temporal characteristics of the spatial feature representation in the first part are learned using an LSTM
Breuer and Kimmel [16]	Eight emotions, 50 AU detection	CNN-based feature extraction and inference
DRML [18]	12 AUs for BP4D, eight AUs for DISFA	Feed-forward functions to induce important facial regions Learning of weights to capture structural information of the face
Joint Fine-Tunning [19]	Seven emotions	• Two different models • CNN for temporal appearance features • CNN for temporal geometry features from temporal facial landmark points
Multi-level AU [20]	12 U detection	Spatial representations are extracted by a CNN LSTMs for temporal dependencies
3D Inception-ResNet [21]	23 basic and compound emotions	LSTM unit that together extracts the spatial relations and temporal relations within facial images Facial landmark points are also used as inputs to this network
Candide-3 [22]	Six emotions	Conjunction with a learned objective function for face model fitting Using a recurrent network for temporal dependencies present in the image sequences during classification
Multi-angle FER [23]	Six emotions	Extraction of the texture patterns and the relevant key features of the facial points. Employment of LSTM-CNN to predict the required label for facial expression

Reference	**Emotions Analyzed**	**Visual Features**	**Decision Methods**
EmotioNet [24]	23 basic and compound emotions	Euclidean distances between normalized landmarks Angles between landmarks Gabor filters centered at of the	Kernel subclass discriminant analysis

 DOI : 10.35291/2454-9150.2022.0192

		landmark points	
InfraFace [25]	Seven emotions, 17 AUs detected	Histogram of gradients (HoG)	A linear SVM
Stepwise approach[26]	Six prototypical Emotions	Stepwise linear discriminant analysis (SWLDA) used to select the localized features from the expression.	Hidden conditional random fields (HCRFs)
Global Feature [27]	Six emotions	Local binary pattern (LBP) histogram of a face image	Principle compound analysis (PCA)

III. METHODOLOGY

In the preceding step data will be passes to emotion state recognizer. Internal block diagram of emotion state recognizer is as shown in fig.8 which consist of three steps

- Face Detection
- Features Extraction.
- Face Recognition

Face Detection is one of the essentials and first step to all facial analysis.

Feature Extraction is a simultaneous process, sometimes face detection suit comparatively difficult and require 3D Head Pose, facial expression, face relighting, Gender, age and lots of other features.

Face Recognition is less reliable and its accuracy rate is still not up to the mark.

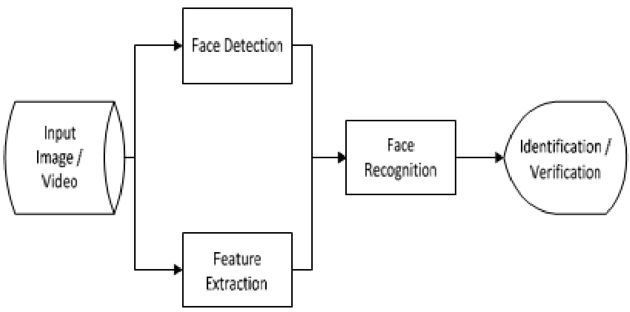

Fig.8. Block diagram of emotion state recognizer

The modern hybrid technique for face recognition will be used Face recognition by means of using the nose tip for the main attribute of feature extraction phase. Then a hybrid 3D model is used for the recognition purpose. A research work is done on the face recognition with the help of the Gabor filter approach and the normalization approach. With the combination of holistic and feature based, a hybrid method can be proposed using Markov Random Field, in which facial images were sub divided into patches. The IDs were allocated and compared using BP algorithm.

To model this responses the important will be needed which is synthesizer. Different approaches can be used to design the synthesizer. Fuzzy kohonen clustering network can be one of the best approach to design the synthesizer.

There are different approaches of face recognition are discussed and it is concluded that the hybrid approach is comparatively best approach as it uses two approaches so hybrid approach is considered as best approach. Therefore hybrid approach can be used for recent work.

IV. CONCLUSION

This paper summarizes existing research about emotion recognition based of Deep-Learning Based FER Approaches. The modern hybrid technique for face recognition will be used Face recognition by means of using the nose tip for the main attribute of feature extraction phase. Then a hybrid 3D model is used for the recognition purpose. A research work is done on the face recognition with the help of the Gabor filter approach and the normalization approach. With the combination of holistic and feature based, a hybrid method can be proposed using Markov Random Field, in which facial images were sub divided into patches. The IDs were allocated and compared using BP algorithm.

To model this responses the important will be needed which is synthesizer. Different approaches can be used to design the synthesizer. Fuzzy kohonen clustering network can be one of the best approach to design the synthesizer.

There are different approaches of face recognition are discussed and it is concluded that the hybrid approach is comparatively best approach as it uses two approaches so hybrid approach is considered as best approach. Therefore hybrid approach can be used for recent work.

REFERENCES

[1] M. Pantic and L. J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. Pattern Recognition and Machine Intellignece, vol. 21, no. 12, pp. 1357-1362, 1999.

[2] I.A. Essa and A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, 1997.

[3] .A. Lanitis, C. J. Taylor and T. F. Cootes, "Automatic Interpretation and Coding of Face Images using Flexible Models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol 19, no, 7, pp. 743-756, 1997.

[4] C. L. Huang and Y. M. Huang. "Facial Expression Recognition using Model-Based Feature Extraction and Action Parameters Calssification," Journal of Visual Communication and Image Representation, vol. 8, no. 3, pp. 278-290, 1997.

[5] Z. Zhang M. Lyons, M. Schuster and S. Akamatsu, "Comparision between Geometry-Based and Gabor-

 DOI : 10.35291/2454-9150.2022.0192

Wavelets-Based Facial Expression Recognition using Multi-Layer Perceptron," In Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, pp 454-459, 1998.

[6] Do Hyoung Kim, Sung U k Jungt, Kwang Ho An, Hui Sung Lee and Myung Jin Chung, Development of a Facial Expression Imitation System, Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9 - 15, 2006, Beijing, China.

[7] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby and B. Blumberg "Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots," Journal of Artificial Life, Vol. 11, pp.1-32, 2005.

[8] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Rocella, M.C. Carrozza, P. Dario and A. Takanishi, "Effective Emotional Expressions with Emotion Expression Humanoid Robot WE- 4RII," in Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, 2004, pp.2203-2208.

[9] J.A. Prado, C. Simplcio, N.F. Lori and J. Dias, "Visuo-auditory Multimodal Emotional Structure to Improve Human-Robot-Interaction", In International Journal of Social Robotics, vol. 4, no. 1, pp. 29-51, December 2011.

[10] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of Emotion Recognitionusing Facial Expressions, Speech and Multimodal Information", In Proc. of ACM 6th International Conference on Multimodal Interfaces (ICMI 2004), 2004.

[11] K.-E. Ko, K.-B. Sim, "Development of a Facial Emotion Recognition Method based on combining AAM with DBN", in International Conference on Cyberworlds 2010, pp. 87-91 ,2010.

[12] S. S. Ge, C. Wang, C.C. Hang, "Facial Expression Imitation in Human Robot Interaction", In Proc. of the 17 IEEE International Symposium on Robot and Human Interactive Communication, Germany, pp. 213-218, 2008.

[13] P.Ekman, E. Rosenberg, "What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)", 2nd edn. Oxford Press, London.

[14] Kahou, S.E.; Michalski, V.; Konda, K. Recurrent neural networks for emotion recognition in video. In Proceedings of the ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 467–474.

[15] Kim, D.H.; Baddar,W.; Jang, J.; Ro, Y.M.Multi-objective based Spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Trans. Affect. Comput. 2017

[16] Breuer, R.; Kimmel, R. A deep learning perspective on the origin of facial expressions. arXiv 2017, arXiv:17

[17] Jung, H.; Lee, S.; Yim, J.; Park, S.; Kim, J. Joint fine-tuning in deep neural networks for facial expression recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–12 December 2015; pp. 2983–2991.

[18] Zhao, K.; Chu, W.S.; Zhang, H. Deep region and multi-label learning for facial action unit detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3391–3399.05.0184.

[19] Jung, H.; Lee, S.; Yim, J.; Park, S.; Kim, J. Joint fine-tuning in deep neural networks for facial expression recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–12 December 2015; pp. 2983–2991.

[20] Chu, W.S.; Torre, F.D.; Cohn, J.F. Learning spatial and temporal cues for multi-label facial action unit detection. In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 1–8.

[21] Hasani, B.; Mahoor, M.H. Facial expression recognition using enhanced deep 3D convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Hawaii, HI, USA, 21–26 July 2017; pp. 1–11.

[22] Graves, A.; Mayer, C.; Wimmer, M.; Schmidhuber, J.; Radig, B. Facial expression recognition with recurrent neural networks. In Proceedings of the International Workshop on Cognition for Technical Systems, Santorini, Greece, 6–7 October 2008; pp. 1–6.

[23]] Jain, D.K.; Zhang, Z.; Huang, K. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. Pattern Recognit. Lett. 2017, 1, 1–9.

[24] Benitez-Quiroz, C.F.; Srinivasan, R.; Martinez, A.M. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5562–5570.

[25] Gunawan, A.A.S. Face expression detection on Kinect using active appearance model and fuzzy logic. Procedia Comput. Sci. 2015, 59, 268–274.

[26] Zhao, G.; Huang, X.; Taini, M.; Li, S.Z.; Pietikäinen, M. Facial expression recognition from near-infrared videos. Image Vis. Comput. 2011, 29, 607–619.

[27] Bhor H N, Kalla M. TRUST-based features for detecting the intruders in the Internet of Things network using deep learning. Computational Intelligence. 2021;1–25. https://doi.org/10.1111/coin.12473

[28] Happy, S.L.; George, A.; Routray, A. A real time facial expression classification system using local binary patterns. In Proceedings of the 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, 27–29 December 2012; pp. 1–5.

 DOI : 10.35291/2454-9150.2022.0192