American Journal of Science and Learning for Development

ISSN 2835-2157
Volume 2 | No 6 | Jun -2023

A Study on Face Expression Observation Systems

Sandeep Kumar [1], Rishabh [2], Kirti Bhatia [3], Rohini Sharma [4]

[1] P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

[2, 3] Assistant Professor, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

[4] Assistant Professor, Department of CS, GPGCW, Rohtak, Haryana, India

Abstract: *Facial expression perception (FEO) is a new and important area of study in the field of pattern recognition. In surveillance films, expression assessment, recognition of gestures, smart homes, video games, stress treatment, tracking patients, treatment for depression, the psychoanalytic paralinguistic interactions, operator tiredness detection, and robotics, facial emotion analysis is effectively applied. In this essay, we provide a thorough analysis of FEP. This overview is based on several deep learning (DL) techniques as well as traditional machine learning (ML) methods. The goal of this study is to emphasize the potential research gap in this field for new researchers by providing a comprehensive evaluation of FEO utilizing conventional ML and DL approaches.*

Keywords: *Facial Expression Perception, Machine Learning, Deep Learning.*

I. Introduction

Analysis of facial expressions of emotion is crucial in non-verbal communication. It improves the effectiveness of oral communication and facilitates conceptual understanding [1,2]. Additionally, it is useful for identifying human attention, including behaviour, mental health, personality, criminal propensity, and lying. Most people have a good understanding of facial expressions regardless of their gender, country of origin, culture, or ethnicity. Automation of facial expression recognition and classification, however, is a difficult task. A few fundamental emotions are used in study, including fear, aggression, frustrated, and happiness. Yet it can be quite difficult for robots to distinguish between various emotions [3,4]. Computers need to be properly trained in order to comprehend their environment, particularly the intents of other people. Robots and computers are included in the phrase "machines" when it is used. One distinction is that while their design incorporates some level of autonomy, robots integrate interpersonal skills to a more inventive amount [5,6]. The fundamental issue in categorising feelings of individuals is the diversity in image or video quality, age, race, gender, and nationality. It is essential to equip a system that can identify face expressions with knowledge comparable to that of humans. Especially in the previous two decades, FEO has just emerged as a new area of study. Reliable computerised facial recognition systems are being expanded for safety and medical purposes using computer vision (CV) approaches, AI, image processing, and machine learning [7].

The initial phase in the FEO process for finding or identifying faces in a movie or single image is face identification. The pictures don't just show faces; they also include elaborate backdrops. While machines without great training find it challenging to anticipate facial emotions and other facial traits from an image, humans can do so very simply [8]. To distinguish face images from background (non-faces) images is the main goal of face detection. Gesture recognition, video surveillance, autonomous cameras, gender identification, facial feature identification, face recognition, tagging, and teleconferencing are a few face detection fields [9]. First, these systems must be able to recognise faces as inputs. Initially these systems must be able to recognise faces as inputs. Everywhere there is a colour sensor for image acquisition that takes colour pictures. As a result, most face recognition methods used today rely primarily on grayscale data, and there are very few methods that can work with colour photos. These systems use either window-based or pixel-based approaches, which are the two main types of facial recognition methodologies, to work better. The face is more difficult to distinguish from the hands of another skin region of the individual using the pixel-based methodology [10].

II. Related work

A review by authors in [11] focused primarily on traditional ML methods. While authors in [12] published a state-of-the-art evaluation of facial emotion detection approaches utilising visual data, authors in [13] proposed a brief review of DL in FEO. The only distinctions they made were between DL approaches and traditional ML techniques. Fig. 1 shows different machine language-based emotions classifiers.

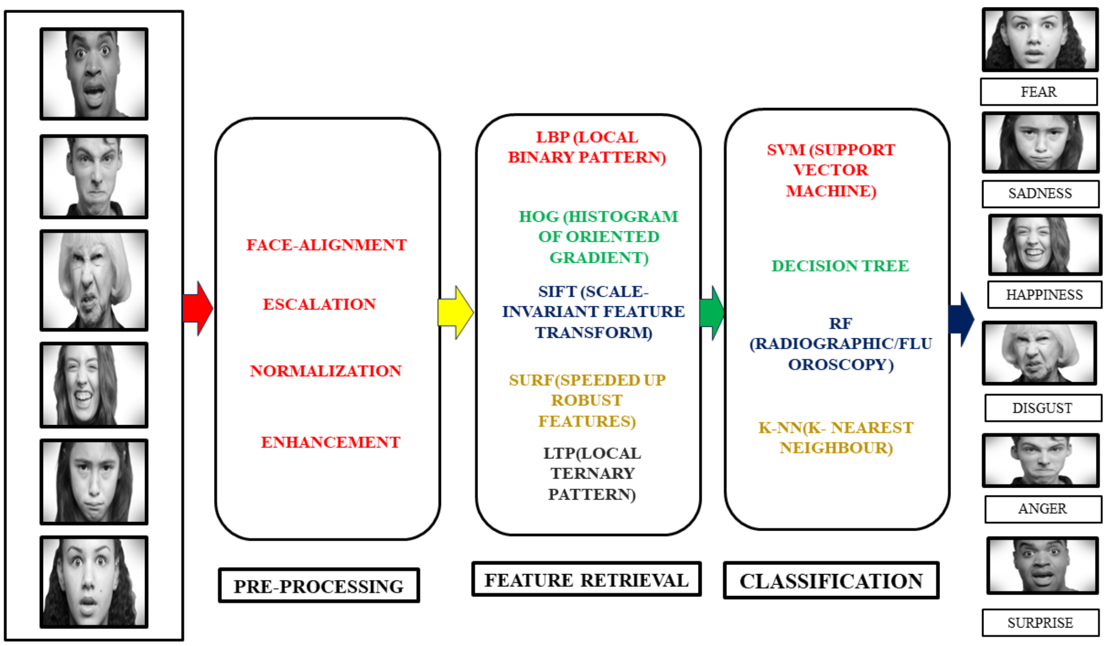

Fig.1: Facial Expression Classification

III. CONVENTIONAL MACHINE LEARNING SYSTEMS

The original method for estimating the strength of facial emotions was based on distance measurement. This method classifies and counts face emotions using high-dimensional rate transformations and local volume differentiation maps. Most systems depict the aspects of face expression in films using Principal Component Analysis (PCA) [14]. The action unit to communicate and define various facial emotions has been recognised using PCA. By mistreating PCA, additional facial expressions are organised and recognised in order to provide a face action unit [15]. Using the active contour framework, authors in [16] identified and extracted the face portion. In order to minimise differences within the face and maximise the distance between it and the setting, scientists used Chan-Vese and Bhattacharyya's energies functions. Additionally, the geometric aspects of emotion in the face and movement of facial features are recovered using optical flow, and noise is minimised using wavelet decomposition. For typical ML approaches like DL methods, considerable computational and memory resources are not required. In order to create embedded devices that do categorization in real time with minimum computing power and deliver appropriate results, greater thought must be given to these algorithms.

IV. DEEP LEARNING SYSTEMS

With RNN and CNN, deep learning (DL) algorithms have completely changed the computer vision industry in the last ten years [17]. Tasks requiring feature extraction, recognition, and classification are accomplished using these DL-based approaches. The main benefit of a DL technique (CNN) is that it can reduce the work needed during the preprocessing and feature extraction phases and eliminate the reliance on physics-based models [18]. Additionally, DL techniques allow for direct learning from input images from beginning to end. For these objectives, DL-based approaches have produced positive outcomes from a number of areas, including FEO, scene awareness, recognition of faces, and recognition of entities [19]. The CNN uses the image or feature maps as input and combines these inputs with a number of filter banks to create feature maps that represent the spatial organisation of the face image. The feature map layer inputs are locally connected, and the convolutional filter weights are connected within the feature map. The second sort of layer, known as subsampling, is in charge of condensing the provided feature maps by adopting one of the most prevalent pooling algorithms, i.e., max pooling, min pooling, or average pooling. A 3D CNN framework was suggested by authors in [20] to identify various moods from films. For the experiment's assessments, they retrieved deep features from three standard datasets: CASME II, CASME, and SMIC. Using a convolutional neural network (CNN), authors in [21] carried out further face cropping and rotation algorithms for feature extraction. The suggested approach was tested using the CK+ and JAFFE databases.

V. APPEARANCE BASED PROCESS

Evident features include the texture, colour, edge, and other properties of face regions associated to expressions. The simplest and most significant of them are histogram and grayscale properties. The number of grey levels in a photograph is specified by the histogram, a statistical description. The image's grayscale feature is more understandable. In experiments on emotion identification and photo texture extraction, gabor and LBP characteristics are frequently used. The Gabor filter, which mimics the sensory stimulus-response of basic cells in the human visual system, is used to recover Gabor features. The Gabor filter's frequency domain settings can be changed to recover features at different scales and orientation.

VI. GEOMETRIC FEATURES BASED PROCESS

Geometrical qualities retrieve data regarding changes such location, distance, and shape, whereas facial muscular movements create expressions. Geometric feature-based methods frequently begin by identifying face-relevant points or regions, then retrieve features. Face photos are difficult to recognise expressions in because they are high dimensional, involve a lot of data that are redundant, and have a lot of unnecessary noise. The extraction of features based on subspace learning is a more abstract approach. This method makes accurate expression recognition simpler, transforms the original data into a more condensed and useful representation, and employs an improved model to train the mapping function. Instances of such methods include PCA, LDA, and manifold learning and improvements. In order to derive facial expression components for use in practical applications, conventional facial expression detection systems mostly require an artificial design. Artificially created features frequently fall short of adapting to complex and evolving expressions, leading to a subpar categorization of expressions as a whole.

VII. Recurrent Neural Network (RNN) BASED EMOTION RECOGNITION

Recurrent neural network (RNN) is a type of RNN with a memory function and the ability to learn knowledge about the dynamic evolution of time series data. The sequence of input data is processed iteratively in the time evolution direction through a chain of connections between the nodes of the RNN. The NLP and video comprehension are two examples of temporal logic applications that frequently use RNNs. For recognising facial emotions, many research use RNN or its variants, such as long short-term memory (LSTM) and bidirectional LSTM. In these methods, the feature vector of the image is often extracted manually or using another deep learning network as input data.

VIII. Deep Belief Network BASED EMOTION RECOGNITION

A constrained Boltzmann machine is adjacent to a generative model called a deep belief network (DBN). Data satisfying a certain distribution may be produced by layering greedy training. When solving classification problems, DBNs are typically coupled to classifiers or serve as better initialization parameters for DNN that are transformed into discriminative models. DBNs have been employed to find solutions in a number of domains, including as face expression recognition.

IX. Broad LEARNING BASED EMOTION RECOGNITION

The general DL network is distinct from the broad learning system (BLS) network [22]. It creates a horizontal network framework, and during training, this network structure is fluid. The expanding network has achieved the best categorization outcome. The BLS contains fewer variables with horizontal expansion and a simpler topology than the typical deep learning network. The random vector functional-link neural network (RVFLNN) serves as the BLS network's foundation [23]. A typical neural network will add a bias to the next hidden layer and multiply the input by a weight for a given input. The RVFLNN network, however, is more than that. By using a nonlinear mapping of the activation function, it also multiplies the input by an array of random weights and adds a bias to create an enhancement layer. The data from the enhancement layer and the input are then linked to the output layer. The BLS network has been enhanced using the RVFLNN network as a base.

Alternatively, BLS employs the altered feature information as the network's feature layer rather than using the initial input data explicitly. Instead, it performs a linear transformation, which is comparable to feature extraction. The enhancement layer is created by nonlinearly mapping the feature layer, and then the feature layer and the enhancement layer merge to form the input layer for the network. The main component of BLS is incremental learning, and the network topology is dynamic. By adding more feature nodes or enhancement nodes to the network model in training, incremental learning can be accomplished.

X. CONCLUSION

This work presents a thorough analysis and comparison of FER techniques. We divided these methods into two main categories: (1) traditional ML-based methods, and (2) DL-based methods. Face detection, feature collection from observed faces, and classification of emotions using features retrieved make up the traditional ML technique. Standard ML for FER employs a number of classification techniques, including random forest, AdaBoost, KNN, and SVM. The dependence on face physics-based models is greatly diminished in comparison to DL-based FER approaches. In order to enable "end-to-end" learning in the input photos, they also shorten the preprocessing time. Yet, these techniques take longer throughout both the testing and training stages. By executing end-to-end feature training explicitly deep learning effectively prevents human influence. The precision of facial expression detection is improved by the use of DL technology, which makes gathering expression characteristics simpler and more efficient. Because of its deep layers and numerous neural nodes, the generic neural network has a large number of network parameters, making training exceedingly time-consuming. The BLS network model is appropriate for data with small feature dimensions since it has a straightforward structure and few configurable parameters. The number of nodes in the feature layer or enhancement layer can be suitably raised to improve the recognition impact when the network training fit is inadequate, and the training procedure is straightforward and quick.

References

1. Ankit Jain, Kirti Bhatia, Rohini Sharma, Shalini Bhadola, An Overview on Facial Expression Perception Mechanisms, SSRG International Journal of Computer Science and Engineering (SSRG - IJCSE) - Volume 6 Issue 4 - April 2019, pp. 19-24.
2. Ankit Jain, Kirti Bhatia, Rohini Sharma, Shalini Bhadola, An emotion recognition framework through local binary patterns, Journal of Emerging Technologies and Innovative Research, Vol - 6, Issue-5, May 2019.

3. Ubaid, M.T.; Khalil, M.; Khan, M.U.G.; Saba, T.; Rehman, A. Beard and Hair Detection, Segmentation and Changing Color Using Mask R-CNN. In Proceedings of the International Conference on Information Technology and Applications, Dubai, United Arab Emirates, 13–14 November 2021; Springer: Singapore, 2022; pp. 63–73.
4. Meethongjan, K.; Dzulkifli MRehman, A.; Altameem, A.; Saba, T. An intelligent fused approach for face recognition. *J. Intell. Syst.* 2013, *22*, 197–212.
5. Elarbi-Boudihir, M.; Rehman, A.; Saba, T. Video motion perception using optimized Gabor filter. Int. J. Phys. Sci. 2011, 6, 2799–2806.
6. Joudaki, S.; Rehman, A. Dynamic hand gesture recognition of sign language using geometric features learning. Int. J. Comput. Vis. Robot. 2022, 12, 1–16.
7. Abunadi, I.; Albraikan, A.A.; Alzahrani, J.S.; Eltahir, M.M.; Hilal, A.M.; Eldesouki, M.I.; Motwakel, A.; Yaseen, I. An Automated Glowworm Swarm Optimization with an Inception-Based Deep Convolutional Neural Network for COVID-19 Diagnosis and Classification. Healthcare 2022, 10, 697.
8. Afza, F.; Khan, M.A.; Sharif, M.; Kadry, S.; Manogaran, G.; Saba, T.; Ashraf, I.; Damaševičius, R. A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. Image Vis. Comput. 2021, 106, 104090.
9. Haji, M.S.; Alkawaz, M.H.; Rehman, A.; Saba, T. Content-based image retrieval: A deep look at features prospectus. Int. J. Comput. Vis. Robot. 2019, 9, 14–38.
10. Saleem, S.; Khan, M.; Ghani, U.; Saba, T.; Abunadi, I.; Rehman, A.; Bahaj, S.A. Efficient facial recognition authentication using edge and density variant sketch generator. CMC-Comput. Mater. Contin. 2022, 70, 505–521.
11. Kołakowska, A. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In Proceedings of the 6th IEEE International Conference on Human System Interactions (HSI), Sopot, Poland, 6–8 June 2013; pp. 548–555.
12. Ko, B.C. A brief review of facial emotion recognition based on visual information. Sensors 2018, 18, 401.
13. Ghayoumi, M. A quick review of deep learning in facial expression. *J. Commun. Comput.* 2017, *14*, 34–38.
14. Cornejo, J.Y.R.; Pedrini, H.; Flórez-Revuelta, F. Facial expression recognition with occlusions based on geometric representation. In Iberoamerican Congress on Pattern Recognition; Springer: Cham, Switzerland, 2015; pp. 263–270.
15. Cornejo, J.Y.R.; Pedrini, H.; Flórez-Revuelta, F. Facial expression recognition with occlusions based on geometric representation. In Iberoamerican Congress on Pattern Recognition; Springer: Cham, Switzerland, 2015; pp. 263–270.
16. Siddiqi, M.H.; Ali, R.; Khan, A.M.; Kim, E.S.; Kim, G.J.; Lee, S. Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. Multimed. Syst. 2015, 21, 541–555.
17. Abbas, N.; Saba, T.; Mohamad, D.; Rehman, A.; Almazyad, A.S.; Al-Ghamdi, J.S. Machine aided malaria parasitemia detection in Giemsa-stained thin blood smears. Neural Comput. Appl. 2018, 29, 803–818.
18. Rehman, A.; Abbas, N.; Saba, T.; Mehmood, Z.; Mahmood, T.; Ahmed, K.T. Microscopic malaria parasitemia diagnosis and grading on benchmark datasets. Microsc. Res. Tech. 2018, 81, 1042–1058.
19. Saba, T.; Haseeb, K.; Ahmed, I.; Rehman, A. Secure and energy-efficient framework using Internet of Medical Things for e-healthcare. J. Infect. Public Health 2020, 13, 1567–1575.

20. Li, J.; Wang, Y.; See, J.; Liu, W. Micro-expression recognition based on 3D flow convolutional neural network. Pattern Anal. Appl. 2019, 22, 1331–1339.
21. Li, B.Y.; Mian, A.S.; Liu, W.; Krishna, A. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In Proceedings of the IEEE workshop on applications of computer vision (WACV), Clearwater Beach, FL, USA, 15–17 January 2013; pp. 186–192.
22. X. Gong, T. Zhang, C. P. Chen, and Z. Liu, "Research review for broad learning system: algorithms, theory, and applications," IEEE Transactions onCybernetics, pp. 1–29, 2021.
23. M. A. Boucher, J. Quilty, and J. Adamowski, "Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons," Water Resources Research, vol. 56, no. 6, article e2019WR026226, 2020.