International Journal of Technology Engineering Arts Mathematics Science
Vol. 2, No. 2, December 2022, pp. 01~04
ISSN: 2583-1224,

Facial Expression Recognition in Human using Machine Learning

Mr. P.G. Ghulaxe[1], Mrs. M.V. Shelke[2]

[12]Assistant Professor, Artificial Intelligence and Data Science Department, AISSMS IOIT, Maharashtra, India
Corresponding Author: Mrs. M.V.Shelke (mayura.shelke@aissmsioit.org)

Article Information

Article history:

Received Nov 8, 2022
Revised Dec 13, 2022
Accepted Dec 20, 2022

ABSTRACT

A human-computer interface system for autonomous face recognition or facial expression recognition is gaining interest among researchers in psychology, computer science, linguistics, neurology, and other disciplines. This paper proposes an Automatic Facial Expression Recognition System (AFERS). The proposed method has three stages: face detection, feature extraction, and facial expression identification. Face detection begins with detecting skin colour using the YCbCr colour model, followed by lighting correction to achieve face homogeneity and morphological procedures to maintain the required face area. The output of the first phase is used to extract face features such as eyes, nose, and mouth using the AAM (Active Appearance Model) approach. The third stage, automatic facial expression recognition, is simple. To compute the distance between two points, the Euclidean Distance technique is used. This method compares the Euclidean distance between the feature points of the training and query images. The minimum Euclidean distance is used to determine the output picture expression. This method has a true recognition rate of 90 to 95 percent. The Artificial Neuro-Fuzzy Inference System is used to enhance this method. When compared to previous approaches, this nonlinear recognition system achieves an approximate recognition rate of 100%.

KEYWORDS: Facial expression recognition (FER), multimodal sensor data, emotional expression recognition, spontaneous expression, real-world conditions

1. INTRODUCTION

Facial expression recognition (FER) has advanced significantly in recent years, thanks to advances in allied domains such as machine learning, image processing, and human cognition. As a result, the impact and potential use of automatic FER in a wide range of applications, such as human-computer interaction, robot control, and driver status surveillance, has grown. Strong detection of facial emotions from photos and videos, however, has remained a difficult task to date due to difficulties in reliably extracting the useful emotional elements.[1] These traits are frequently represented in a variety of ways, including static, dynamic, point-based geometric, and region-based appearances.

Facial movement aspects such as feature position and shape changes are caused by movements of facial parts and muscles during emotional expression. When participants are expressing emotions, the facial components, especially the critical elements, will constantly shift positions. As a result, the same feature appears at different times in different photos. In rare cases, the contour of the feature may be affected by small facial muscle movements. The mouths in the first two photographs, for example, differ from those in the third. As a result, in image databases and videos, the geometric-based location and appearance-based form of any componentindicating a specific emotion typically shift from one image to the next.[2]

This type of movement features a large pool of FER-relevant static and dynamic expression qualities.

The dynamics of facial expressions have been largely ignored in previous FER research. Attempts to capture and use face movement features have been made, almost all of which have been video-based. These efforts aim to incorporate geometric features of the tracked facial points (such as shape vectors, facial animation parameters, distance and angular, and trajectories), or appearance differences between holistic facial regions in subsequent frames, or texture and motion changes in local facial regions, in subsequent frames. Although these methods have produced promising results, they frequently necessitate precise placement and tracking of face locations, which remains difficult.[3]

Fig. 1 illustrates the three main processes of typical FER approaches:face and facial component detection, feature extraction, and expression categorization. Following the extraction of a face picture from an input image, facial landmarks, or components (such the eyes and nose) are identified. Second, the face components can be used to generate a variety of spatial and temporal properties. Third, pre-trained FE classifiers like SVM, AdaBoost, and random forest offer recognition results utilising the obtained features.

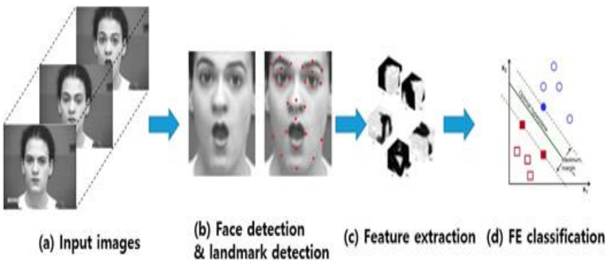

Fig. 1. Procedure used in conventional FER approaches

2. LITERATURE REVIEW

Understanding how to read emotional expressions on the face effectively is one of the characteristics that contribute to the strength of interpersonal interactions. More opportunities for such interactions arise when one's capacity to accurately read another's emotions increases. Inability to recognize facial expressions may be a symptom of certain psychopathological illnesses that cause problems in social interactions. These deficiencies have been shown in a number of clinical populations. However, there have been conflicting results from the research performed thus far regarding facial expressions.The purpose of this work is to investigate the subject of emotion and emotional facial expressions from ancient periods, to emphasize the advantages and disadvantages of related research, to contrast the results, and to bring attention to this issue.[4]

The earliest physiological explanation of emotion was put forth by William James in 1884. According to James, emotion is rooted in physical experience. He claims that after noticing the object, we experience a physical reaction before being emotionally interested. For instance, when we see a bear, our hearts start to race, we start to run, and then we become afraid. Our fear drives us to run, not because we run because we are frightened. The hypothesis is known as the James-Lange theory of emotions since it was first put forth in 1885 by his Danish colleague Carl Lange, who held a similar perspective.[5]

An alternative interpretation was offered by Cannon, who asserted that rather than being a physiological state of attention, emotions are a cognitive one. His perceptions of the sequence of events were external stimuli, cerebral processing, and physiological responses. This novel hypothesis states that emotional inputs cause both an emotional experience, such as dread, and a physiological response, such as shivering.

The study looked at how well children with autism (N = 20) and kids without autism (N = 20) recognizedstandardized facial expressions of emotion (angry, fear, disgust, happy, sorrow, and surprise) at a perceptual level (experiment 1) and at a semantic level (experiments 2 and 3). The findings showed that children with autism were equally capable of recognizing all six emotions at various intensities as controls and that they committed the same kinds of mistakes.

These unfavorable results are examined in light of the following: (1) earlier evidence showing a specific deficit in autism in identifying the belief-based expression of surprise; (2) prior research showing that individuals with autism, like patients with brain injury, pass a basic emotion recognition test but fail to recognize more complicated stimuli including the perception of faces or parts of faces. 3) The data that are consistent with the idea that people with autism, like people who have had brain damage, may recognize fundamental emotions but struggle to identify more complicated inputs, such as the perception of faces or parts of faces.[6]

Since Kanner's original clinical account of autistic children noticed their extraordinary lack of affective contact with other people, psychologists have been evaluating the social and emotional disorders in autism. It is not surprising that the findings are so divergent because the empirical research on cognitive impairment in children and individuals with autism is so extensive and diverse.

The ideas of a general psychological weakness and a specific emotion perception impairment were studied by Howard et al. (2019). Additionally, the theory of mind (ToM) deficit hypothesis of autism permits study on selective emotion processing failure by contrasting recognized tasks that do and do not involve the ability to define mental states. The goal of the ongoing study is to confirm and build on these results in children with autism.[7]

A paradigm of visual object identification that draws inspiration from biology is used to address the problem of multiclass object categorization. The models of Serre, Wolf, and Poggio are modified by ours. To increase feature complexity and position/scale invariance, researchers employ Gabor filters at all positions and scales, much like in that paper, and then alternate template matching and max pooling techniques. Researchers improve the method utilising straightforward forms of scarification and lateral inhibition in a number of biologically reasonable ways. They demonstrate how crucial it is to save some location and scale data above the intermediate feature level.[8]

Using feature selection, researchers create a model that performs better with fewer attributes. Their final model is tested on the UIUC automotive localization challenge and the Caltech item categories, and it performs admirably in both cases. The outcomes support the application of this class of models in computer vision.

Multiple item classes in natural images have shown to be challenging for computer vision to identify. Given that human vision does this task substantially better than other visual systems, it makes sense to look to physiology for guidance. On some of the most popular recognition datasets, Serre, Wolf, and Poggio recently showed that a computational model based on our understanding of the visual cortex can compete with the finest extant computer vision systems. By including biologically inspired characteristics like feature sparsification, lateral inhibition, and feature localization, our research builds on their methodology. We demonstrate that these adjustments improve recognition even more, expanding our understanding of the computational constraints both biological and artificial vision systems must contend with.[9]

3. ANALYSIS AND DESIGN OF THE APPLICATION

A. EXISTING WORK

The three key steps of AFERS

1. To identify a face in an input image or video,
2. To identify facial features such as the eyes, nose, and mouth from the identified face.
3. Classify different facial expressions into groups such as surprise, anger, sadness, fear, and happiness. A subset of object detection is the detection of faces. Additionally, it makes use of linguistic structures and light adjustment techniques to preserve the face of the provided image.

B. DRAWBACKS

The system serves as a means of communication in interpersonal relationships since it can provide information about a person's affective state, cumulative activity, personality, intention, and psychological state. The suggested system consists of three modules. To convert a given image into a binary image, which is subsequently used to detect faces, the face detection module uses an image segmentation algorithm.

C. PROPOSED WORK

In order to further raise the system's recognition rate, the third stage makes use of the Artificial Neuro-Fuzzy Inference System (ANFIS). The expressions in this method can be tested using both static photographs and video input. To identify human facial expressions including happiness, fear, sadness, rage, contempt, and surprise, a neuro-fuzzy based automatic facial expression identification system has been developed. A video of numerous expressions is initially framed into several images.

After that, the chosen image sequence is stored in a database folder. Using the AAM method, the features of every photograph are located and saved as ASF files. All the photographs in the data folder are then given a mean shape. The shift in the AAM shape model in response to variations in facial expressions serves as a proxy for the distance or difference between Neutral and other facial emotions. Each expression is given a specific value during ANFIS training, and these values are kept in a.MAT file. These differential values are subsequently provided as input to the ANFIS (Artificial Neuro-Fuzzy Inference System).

These colour spaces are already used to encode the majority of video files, which is one benefit of using them. Any of these spaces can be converted from RGB with a straightforward linear process.

Face detection, feature extraction, and facial expression recognition are the first three steps. Face detection begins with the YCbCr colour model for skin colour recognition, lighting adjustments for uniformity on the face, and morphological techniques to maintain the appropriate face portion.[10]

4. SYSTEM IMPLEMENTATION

Segmenting skin tone involves first contrasting the image. The segmentation of skin tone comes next.

Face Recognition: In order to recognize faces, we first need to convert an RGB image to a binary image. In order to convert a binary image, we compute the average RGB value for each pixel and, if the average value is less than 110, then replace it with a black pixel; otherwise, we replace it with a white pixel. This technique allows us to transform an RGB image into a binary image.

Eyes Detection: To detect eyes, we convert the RGB face to the binary face. W is now multiplied by face width. We scan from W/4 to (W-W/4), determining the position of the two eyes in the midpoint. The highest white continuous pixel along the height between the ranges serves as a representation of where the two eyes are located in the middle.

Bezier Curves should be used in the lip box, the lip, and possibly a small portion of the nose. As a result, the box has skin or skin of a certain colour. As a result, we convert the other pixel to black and the skin pixel to white. Additionally, we recognize and change skin-like pixels to white pixels. Similar pixels are those in which the difference in RGB values between two pixels is 10 or less. The histogram is used to calculate the distance between the lower average RGB value and the higher average RGB value.

Database and Training: Our database contains two tables. People's names are kept in one table called "Person," together with their indexes for four various sorts of emotions, which are kept in another table called "Position." There are 6 control points for the left eye Bezier curve and 6 control points for the lip Bezier curve for each index in the "Position" table. For the right eye, a Bezier curve with six control points The Bezier curve includes the length and width of the lips, the left and right eyes, and the width and length of the right eye. This tactic allows the programme to learn people's feelings.

Emotion Recognition: To recognize emotion in an image, we need to locate the Bezier curves of the mouth, left eye, and right eye. Then, we set the width and height of each Bezier curve to 100 and 100, respectively. The computer will compare which emotion's height is closest to the current height and output that one if the database includes information on the person's feelings.

CONCLUSION & FUTURE SCOPE

Several researchers' efforts were assessed in this essay, which placed a strong emphasis on citing as many sources from the last few years as was practical. Reviews indicate that the study solved certain issues with facial expression identification by using a variety of face detection, feature extraction, analysis, and classification methodologies. The paper gives in-depth details on the methods that are currently used for Facial Expression Recognition (FERs) at all stages. The study is very helpful to both seasoned and beginning researchers in the field of FER since it offers thorough information about approaches that are currently used at different stages of the field, enhancing their awareness of current trends, and assisting them in setting their future research objectives.In order to enhance the functionality of Facial Expression Recognition in image processing, the study also covered a variety of their technological solutions, along with advantages and disadvantages.

REFERENCES

[1] Eldar, C. Yonina, "Compressed sensing: theory and applications," Cambridge University, 2012.

[2] Solomon, Chris, Fundamentals of Digital Image Processing: A practical approach with examples in Matlab. John Wiley & Sons, 2011.

[3] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." In bmvc, Vol. 1, Iss. 3, p. 6. 2015.

[4] Grafsgaard, Joseph, Joseph B. Wiggins. "Automatically recognizing facial expression: Predicting engagement and frustration." In Educational Data Mining 2013.

[5] C. N. Moridis and A. A. Economides, "Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions," in IEEE Transactions on Affective Computing, Vol. 3, Iss. 3, pp. 260-272, July-September 2012, doi: 10.1109/T-AFFC.2012.6.

[6] G. Brodny, A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch and M. R. Wróbel, "Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions," 2016 9th International Conference on Human System Interactions (HSI), Portsmouth, UK, 2016, pp. 397-404, doi: 10.1109/HSI.2016.7529664.

[7] Nitin Shekapure*, Sandeep Wankhade, Vipin Gawai, Swati Shekapure&SachinKallurkar, "Survey Paper on Extraction of 3D image Data for Detecting Chest Diseases", Journal of Optoelectronics Laser, Vol. 41, Iss. 8, pp. 249–253, 2022, http://gdzjg.org/index.php/JOL/article/view/905.

[8] H. Ding, S. K. Zhou and R. Chellappa, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 118-126, doi: 10.1109/FG.2017.23.

[9] Y. Wu, T. Hassner, K. Kim, G. Medioni and P. Natarajan, "Facial Landmark Detection with Tweaked Convolutional Neural Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, Iss. 12, pp. 3067-3074, 1 Dec. 2018, doi: 10.1109/TPAMI.2017.2787130.

[10] Swati Shekapure and Dipti D. Patil, "Enhanced e-Learning Experience using Case based Reasoning Methodology", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 10, Iss. 4, 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100428.