STM JOURNALS

International Journal of
Electronics Automation

Volume 1, Issue 1, 2023
January–June
DOI (Journal): 10.37591/IJEA

https://journals.stmjournals.com/ijea

Review IJEA

Procedure for Conventional Facial Emotion Detection Algorithms Based on Machine Learning

Sachin T.[1,*], Komala K.[2], M.Z. Kurian[3]

Abstract

Researchers in psychology, computer science, linguistics, neurology, and allied fields have become more interested in a human-computer interface system for autonomous face recognition or facial expression recognition. This study has recommended an Automatic Facial Expression Recognition System (AFERS). The proposed methodology consists of face detection, feature extraction, and facial expression identification processes. The initial phases of the face detection procedure include skin colour identification using the YCbCr colour model, illumination adjustment for face uniformity, and morphological operations for maintaining the required face region. Using the AAM (Active Appearance Model) approach, the first phase's output is utilised to extract face features such the mouth, nose, and eyes. Automatic facial expression recognition is the third stage, and it is straightforward. Method of Euclidean Distance: This method compares the Euclidean distance between the feature points on the query image and the training images. The output picture expression is chosen based on the minimal Euclidean distance. This approach has a true recognition rate of between 90 and 95%. Utilising the Artificial Neuro-Fuzzy Inference System (ANFIS), this method is further modified. In comparison to previous systems, this non-linear recognition system provides a recognition rate of close to 100%, which is satisfactory.

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

INTRODUCTION

Researchers have focused on this topic because being able to recognise one's expressions aids in human-computer interaction, helps to correct advertising campaigns, and ultimately improves human communication by modifying emotional intelligence (EQ) in people. The examination of human expression recognition can be done in a variety of methods, including by looking at facial expressions, body posture, speech tones, etc. We have emphasised the recognition of facial expressions in this research. Many technological developments, such as machine to machine communication and automatic translation systems, are taking place in the burgeoning study field of facial emotion recognition (FER). The purpose of this study, in contrast, is to survey and evaluate various facial extraction characteristics, emotional databases, classifier techniques, and so on. This article has the following format: The background material, the emotion recognition system, and the

***Author for Correspondence**
Sachin T.
E-mail: mailtosachint@gmail.com

[1]Student, Department of Electronics & Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India
[2]Assistant Professor, Department of Electronics & Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India
[3]Head of Department, Department of Electronics & Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

Received Date: May 16, 2023
Accepted Date: July 18, 2023
Published Date: August 16, 2023

Citation: Sachin T., Komala K., M.Z. Kurian. Procedure for Conventional Facial Emotion Detection Algorithms Based on Machine Learning. International Journal of Electronics Automation. 2023; 1(1): 7–13p.

applications of emotion recognition are covered in the next Section. The methods for feature selection and image optimisation are explained; various facial emotional databases are compared; numerous classifier algorithms for categorising photos in accordance with the determined expression are covered; and the last Section concludes the essay.

The next communication method for computers could be expressions. The need for automatic facial expression interpretation of emotions is growing. Most of the field's study focuses on taking human emotions out of audio or visual content. Convolutional neural networks have not been used to extract emotions from photographs; instead, most research work matches and detects faces. Emotion Recognition looks at how emotions are recognised, along with the methods and tools used to do so. Emotions can be recognised through facial expressions and speech patterns. Massive techniques, including machine learning, neural networks, artificial intelligence, and emotional intelligence, have been developed to infer emotions. Emotion Recognition is becoming increasingly significant in research, which is essential to solving many issues. When photos are used as input for the mechanisms, the primary need of psychological intelligence is the recognition of emotions from face movements.

RELATED STUDIES

An approach to real-time facial expression recognition based on adaptive Canny operator edge detection is presented in the study presented by Peng *et al*. [1]. In the procedure, a model of the skin's structure and colour was first utilised to locate faces. The AAM (Active Appearance Model) algorithm and adaptive Canny operator edge detection was then utilised to extract face emotion features, which decreased computing complexity and increased feature point location accuracy. The overall image was separated into several sub-images using Canny operator edge detection. Additionally, dynamic threshold was self-adaptively constructed in accordance with the sub-image edge gradient information and merged with the global edge gradient's characteristic information, which improved the edge detection outcomes. The characteristics were classified and identified using the least-squares approach. Finally, the attributes information was classified and identified using the least-squares approach. Studies demonstrated how well this approach recognised face expressions and could satisfy real-time system requirements.

The research by Abdat *et al*. described a method for facial expression-based emotion recognition [2]. Face detection, facial characteristic extraction, and facial expression classification form the basis of a fully automatic facial expression recognition system. In conjunction with the various approach, Abdat *et al*. have created an anthropometric model to identify face feature points [2]. To code the facial expression, the changes of 21 distances that define the deformations of facial features from the neutral face were used. The SVM (Support Vector Machine) approach was used in the classification step. Results from experiments show that the suggested strategy is a useful way to identify emotions from facial expressions, with an emotion detection rate of greater than 90% in real time. Utilising variations in the computer user's emotional state, this method is utilised to control the music player.

Most of the prior research on the automatic analysis of facial expressions concentrated on identifying prototypical expressions of fundamental emotions like happiness and rage. By identifying the motions of the face muscles that combine expressions, the method shown by Valstar and Pantic makes it possible to detect a far wider spectrum of facial behaviour [3]. Because AUs are agnostic, higher level decision-making processes, such emotion recognition, should be used to infer intentions that were transmitted. In addition to allowing the recognition of 22 AUs, the proposed fully automatic method clearly describes their temporal properties (i.e., sequences of temporal segments: neutral, onset, apex, and offset). It accomplishes this by autonomously localising 20 face fiducial points using a facial point detector based on Gabor-feature-based boosted classifiers. The technique used to follow these spots is known as particle filtering with factorised likelihoods. It uses a mix of GentleBoost, SVMs, and hidden Markov models to encode AUs and their temporal activation models depending on the tracking data. When evaluated on a benchmark set of consciously produced facial expressions, we achieve an average AU identification rate of 95.3%, and when tested on spontaneous expressions, we achieve a rate of 72%.

More than 750,000 photos of 337 people, each taken in up to four sessions over a 5-month period, are available in the CMU Multi-PIE face database. The subjects were captured while exhibiting a variety of facial emotions under 15 different viewpoints and 19 different lighting situations. High-definition frontal photos were also taken in addition. More than 305 GB of face data is available in the database. The database is detailed on the Content page [4].

PROPOSED SYSTEM

The two datasets we employed in our research are the Karolinska Directed Emotional Faces (KDEF) dataset and Kaggle's Facial Expression Recognition Challenge. This study extensively investigates this dataset because it is rarely used. The 15 actors in the sessions that make up Corpus Data have markers placed on their faces, heads, and hands that record every facial emotion and hand motion.

Data Flow Diagram

Figure 1 shows the present diagram for showing capturing the emotion algorithm.

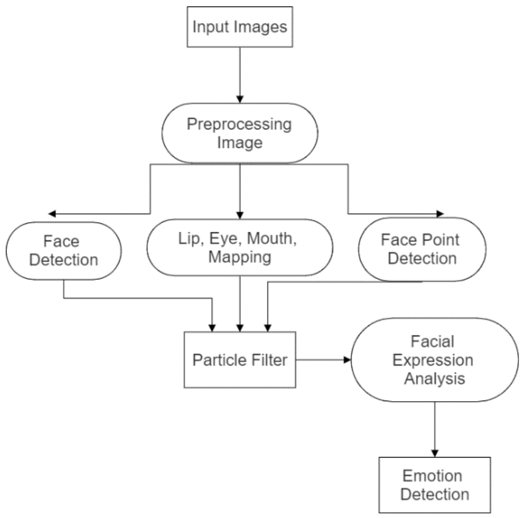

Figure 1. Data flow diagram.

Use Case

Figure 2 shows the Used case for showing capturing the emotion of the individuals.

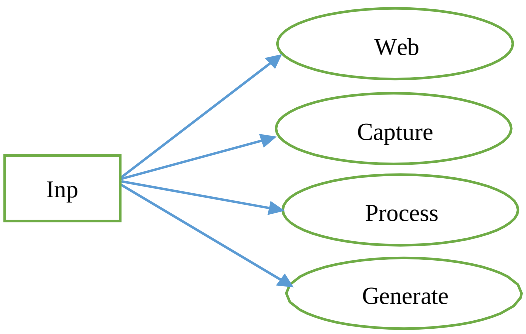

Figure 2. Use case for system.

Sequence Diagram

Due to its tremendous academic and business potential, facial emotion recognition (FER) is a crucial topic in the domains of computer vision and artificial intelligence. Although FER can be carried out with a variety of sensors, this review concentrates on research that only employs facial images because facial expressions are one of the primary information routes in interpersonal communication. This study provides a brief review of research in the field of FER conducted over the past decades. First, a synopsis of the representative categories of FER systems and their primary algorithms is given together with a description of typical FER approaches as shown in Figure 3.

Then, FER approaches based on deep learning that use deep networks to enable "end-to-end" learning are discussed. This study also focuses on a modern hybrid deep learning strategy that combines long short-term memory (LSTM) for the temporal aspects of successive frames and a convolutional neural network (CNN) for the spatial features of a single frame. A brief overview of publicly accessible evaluation metrics is provided in the later section of this study, and a comparison with benchmark results, a standard for a quantitative comparison of FER research, is explained. This review can act as a quick reference for both experienced researchers seeking fruitful avenues for future study and newbies to the field of FER, giving fundamental knowledge and a general overview of the most recent state-of-the-art investigations.

Human communication relies heavily on facial expressions to grasp one another's intentions. People use their speech tones and facial expressions to infer other people's emotional states, such as happiness, sadness, and rage. Various studies have found that nonverbal components make up two-thirds of human communication while verbal components only account for one third [1, 2]. Facial expressions are one of the primary information carriers in interpersonal communication among several nonverbal cues because they convey emotional significance. So, it makes sense that investigation on facial expression has become more popular over the past few decades, with implications not just in the perceptual and cognitive sciences but also in affective computation and computer animations [2].

There is interest in automatic facial emotion recognition (FER), also known as facial emotion recognition and facial expression recognition (expanded form of the acronym FER). Although a variety

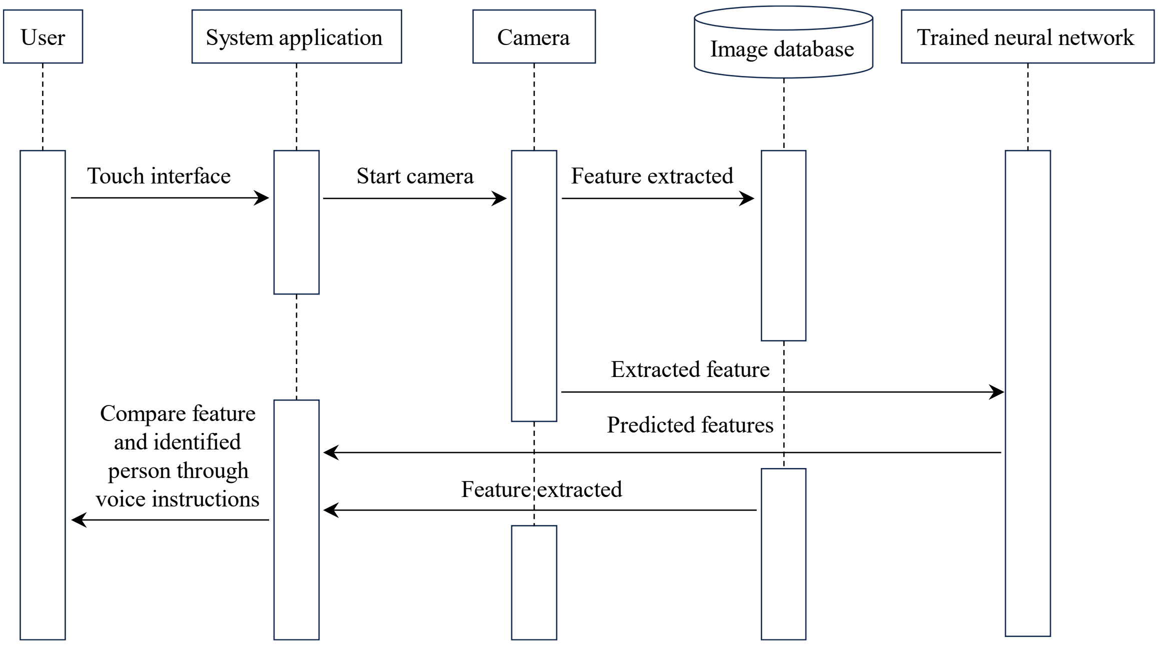

Figure 3. Sequence diagram for application.

of sensors, including an electromyography (EMG), electrocardiogram (ECG), electroencephalograph (EEG), and camera, can be utilised for FER inputs, the camera is perhaps the most promising kind of sensor because it offers the most useful hints for FER and does not require wearing anything. This study first categorises research on automatic FER into two groups based on whether the features are created manually or because of a deep neural network's output.

Face and facial component identification, feature extraction, and categorization of expressions are the first three steps. To identify facial landmarks or facial components (such as the eyes and nose), a face image must first be extracted from an input image. The facial components are then used to extract a variety of spatial and temporal data. Third, using the extracted features, the pre-trained FE classifiers, including a support vector machine (SVM), AdaBoost, and random forest, provide the recognition results. With the availability of massive data, deep learning has become a generic technique to machine learning in contrast to older approaches employing handcrafted features, producing cutting-edge findings in several computer vision studies.

By allowing “end-to-end” learning to take place in the pipeline directly from the input photos, deep-learning-based FER approaches significantly minimise the dependence on face-physics-based models and other pre-processing techniques. The convolutional neural network (CNN), a specific kind of deep learning, is the most well-known network model among the various deep-learning models that are now accessible. In CNN-based methods, a feature map is created by convolving the input picture through a set of filters in the convolution layers. The output of the Softmax technique is used to identify the face expression as being a member of a specific class once each feature map has been joined to fully connected networks. The process followed by CNN-based FER techniques is depicted in Figure 4.

FER can also be split into two categories based on whether it employs frame-based or video-based imagery. First, static (frame-based) FER only uses static face traits that were manually extracted from peak expression frames in selected image sequences. To capture the dynamics of expression in facial expression sequences, dynamic (video-based) FER makes use of spatiotemporal characteristics. Because it offers more temporal information than static FER, dynamic FER is reported to have a greater recognition rate than static FER, but it has certain downsides as well. For instance, depending on the specific faces, the extracted dynamic features have variable transition times and feature properties of the facial expression. Additionally, the process of temporal normalisation, which is used to produce expression sequences with a predetermined number of frames, may cause the loss of temporal scale information.

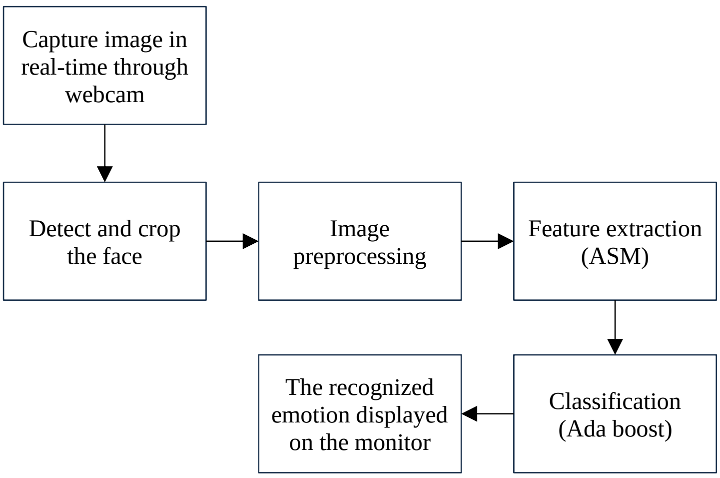

Figure 4. CNN-based FER techniques.

IMPLEMENTATION

The goal of the suggested method is to create a real-time emotion detection system that can identify fundamental emotions like anger, disgust, happiness, surprise, and neutrality from facial photos. We used the CMU MultiPIE database, which consists of 337 participants' photos exhibiting a range of distinct facial emotions, including neutral, happy, surprised, disgusted, and angry. There are 102 girls and 235 male individuals in various stances and lighting conditions. Software for recognising emotions has been created using the Viola-Jones face recognition method, Active Shape Model (ASM) for collecting facial points, and AdaBoost classifier [5–8]. The Raspberry Pi II is a credit card-sized computer with the Broadcom BCM2835 system on a chip. It has a video core 4 GPU and an ARM1176JZFS with floating point running at 900 MHz.

It is described as follows: A webcam is used to capture the input image, which is then provided as input to emotion detection software in real time. The Raspberry Pi II uses emotion recognition software, which outputs categorised emotions, and displays the actions taken by the software installed on the Raspberry Pi II. The following steps provides an explanation of the algorithm behind the Raspberry Pi II's real-time implementation of detecting emotions:

1. *Step 1:* First, a webcam is used to capture the input image.
2. *Step 2:* The facial image is detected using the machine learning approach. Researchers developed an integral image to recognise faces using the Haar wavelet idea. As certain areas of the face have varying levels of intensity from other regions, Haar features consider the different intensities of adjacent rectangular regions. After detection, non-face region is deleted, and the facial image is retained for further processing.
3. *Step 3:* During image pre-processing, the image is cropped to the necessary size and made into a grey image. This clipped image serves as the starting point for the Sobel filter's smoothing and noise-removal process.
4. *Step 4:* Active Shape Model (ASM), a geometric technique, is employed for feature extraction. A picture of a facial expression is first subjected to the ASM automatic fiducial point location algorithm, and then Euclidean distances between the centre gravity coordinate and the annotated fiducial points coordinates of the face image are computed. The system extracts the geometric deformation difference characteristics between a person's neutral expression and the other fundamental expressions in order to extract the discriminate deformable geometric information. To obtain the shape model in ASM, the input face shape is iteratively deformed. The input facial image's feature point is extracted after comparison with the form model. Frontal photographs of five emotions, 60 participants, and the CMU MultiPIE database were used to train the model. Then a feature vector is created by normalising these points. The feature points form a single feature vector that is provided to the classifier for training after this step is done for all the subjects and moods.

CONCLUSION

In this study, we provide a method for Raspberry Pi II-based real-time emotion identification based on geometric features. Using the Raspberry Pi II (ARM1176JZF, 900 MHz), we were able to analyse data with an average processing time of 120 ms and an overall accuracy of 94% on the Linux platform. The Raspberry Pi II is a lightweight, extremely compact hardware kit that may be placed on a moving robot. If a mobile robot with a small, portable display screen is attached, it can dynamically display a person's feelings in social and surveillance settings like nursing homes and hospitals. Our suggested solution is very beneficial to society for a variety of applications where emotion recognition is crucial. Future research could use a different algorithm to increase recognition precision. Neurological inspiration can also be used to teach robots to recognise emotion. For identifying emotions, additional modalities like voice can be coupled with images.

REFERENCES

1. Peng Zhao-Yi, Zhu Yan-hui, Zhou Yu, Real-time Facial Expression Recognition Based on Adaptive Canny Operator Edge Detection. IEEE, Multimedia, and Information Technology (MMIT), 2nd International Conference on. 2010; 2: 154–157.
2. Abdat F, Maaoui C, Pruski A. Human-computer interaction using emotion recognition from facial expression. IEEE, Computer Modeling and Simulation (EMS), 5th UK Sim European Symposium. 2011; 196–201.
3. Valstar Michel F, Maja Pantic. Fully Automatic Recognition of The Temporal Phases of Facial Actions. IEEE Trans Syst Man Cybern B: Cybernetics. 2012; 42(1): 28–43.
4. Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, Simon Baker. Multi-PIE. Proc Int Conf Autom Face Gesture Recognit. 2010 May 1;28(5):807-813.
5. Kotsia I, Zafeiriou S, Pitas I. Texture and shape information fusion for facial exp ression and facial action unit recognition. Pattern Recognit. 2008; 41(3): 833–851.
6. Peng Zhao-Yi, Zhou Yu, Wang Ping. Multi-pose face detection based on adaptive skin color and structure model. Proceedings of the 5th International Conference on Computational Intelligence and Security (CIS 2009). 2009; 325–329.
7. Wang Zhi, He Sai-Xian. An adaptive edge-detection method based on canny algorithm. Chinese Journal of Image and Graphics. 2004; 9(8): 957–962.
8. Lyons M, Akamatsu S, Kamachi M, *et al*. Coding facial expressions with Gabor wavelets. Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. 1998; 200–205.

