International Journal of Mathematics, Statistics, and Computer Science
Vol. 2, 2024
ISSN: 2704-1077 eISSN 2704-1069, DOI: 10.59543/ijmscs.v2i.7815

A Framework for Recognition of Facial Expression Using HOG Features

Vaman Ashqi Saeed[1]
Technical College of Administration, Information Technology Management Department, Duhok, Kurdistan Region of Iraq;
vamanatroushi@gmail.com

Received 15.03.2023, Revised 26.04.2023, Accepted 09.05.2023, Published 27.05.2023

ABSTRACT: Recognition of facial expressions has been an important topic of study over the last several decades, and despite the advancements that have been made, it is still difficult to do because of the significant intra-class diversity. The handcrafted feature is used in traditional methods to address this issue. This feature is then preceded by a classifier that is trained using a database of pictures or videos. Most of these works do quite well on datasets of photographs that were recorded in a controlled environment. However, they do not perform as well on datasets that are more difficult to work with since they include greater image variance and partial faces. The Histogram of Oriented Gradient (HOG) descriptor is the foundation for the strategy that is suggested in this study. During the initial step of the procedure, the input picture is pre-processed to locate the datum region, which assists in the extraction of the most relevant characteristics. After that, the Random Forest (RF) algorithm was employed as a classifier for facial expressions. The Japanese Female Facial Emotions Database (JAFFE) is used to assess our technique. The experimental findings demonstrated that the suggested method is accurate and effective in identifying facial expressions.

Keywords: Facial expression, recognition, HOF features, classifications.

1. INTRODUCTION

Expressions on people's faces, along with inflections, body posture, hand gestures, and other behaviors of the people that communicate non-verbal information, are basic components of human connection. Facial expressions disclose the focus of our attention, demonstrate if we comprehend something or disagree with something, tell whether we are trying to be comical or serious, and in general, govern and deepen interpersonal communication [1]. Therefore, automatic facial expression recognition (AFER) has consumed a lot of effort over the past two decades due to the fact that it is fundamental for human-computer interactions and has applications in a range of areas and difficulties, including psychology, medicine, education, computer science, well-being assessment, and ambient intelligence [2]. This is because AFER is fundamental for human-computer interactions and has applications in a wide variety of fields and problems.

The FER issue may be tackled in a number of different ways, most of which fall into one of two broad categories. In the first category of approaches, expressions are categorized according to Action Units (AUs), which are small but distinguishable muscular motions that are associated with expressions. The FER issue is often recast as the challenge of AU identification when AU-based approaches are used [3]. However, since local changes in faces are difficult to identify, it is challenging for computers to carry out reliable AU detections. Changes in lighting or position, for example, are two examples of variations that might hinder the effectiveness of AU detection. The second kind of strategy is one that typically extracts visual characteristics by making use of hand-crafted patterns. The representation of an expressional picture and the training of a classifier for FER both make use of the extracted features. Nevertheless, it is challenging for researchers to build a hand-crafted pattern that is adaptable to a variety of situations [4]. The majority of the work done in the FER system pertains to doing research within the context of pattern recognition. Face recognition, extraction of facial features, and categorization of facial expressions are the three stages that make up this process. The study of these stages should definitely be done since it is very important. In this examination, several stages of facial expression analysis are presented along with separate classification methods for a total of six fundamental facial expressions [5]. A variety of different methods, such as the Haar classifier and the adaptive skin color algorithm, are used in the process of face identification. In order to extract features, several methods were used, such as the Gabor feature, active appearance model, principal component analysis, and others. Classifiers such as support vector machines (SVM), neural networks, closest neighbors, and others are used throughout the expression categorization process.

Expression feature extraction is the most essential component of the FER system, and efficient extraction of face characteristics will significantly increase recognition performance [6]. Emotion feature extraction is the most essential element of the FER system. In light of the drawbacks of handcrafted features and the complexity of the deep learning model during the process of training and tuning, a new model for an expression recognition system has been developed. This model combines feature extraction with the conventional learning classifier in order to circumvent the extensive amount of time required for training and to address the issue of incompleteness that is caused by handcrafted features. The design of the facial expression classifier, which is an essential component of facial expression recognition, has a significant impact on the accuracy rate of facial expression recognition. Because of this, the selection and implementation of the classifier are critical in determining the final result. The classifier for facial expression identification needs to have a high computational efficiency as well as a strong capacity to deal with a huge number of data sets [7].

On the other hand, the conventional HOG places a greater emphasis on the distribution of the features that have been extracted. A large number of feature points are extracted, much as in conventional HOG, in order to attain improved accuracy. This HOG method determines the total number of feature points that need to be extracted as well as the number of regions that need to be separated in order to compute the number of feature points that should be extracted from each area. Experiments were run using the JAFFE database, which contains information on facial expressions.

2. RELATED WORKS

The research that has been done on facial expression detection systems in the past is what we mean when we talk about related studies. The intensities of the face pictures are utilized to derive the textural pattern that is then employed to represent the various appearance aspects. Conventional methods of facial expression recognition (FER) include the detection of the face, extraction of facial features, and categorization of those features. These methods are organized into three primary groups. Not only has facial expression recognition advanced significantly over the years, but it has also been successfully used in a wide variety of applications that take place in the real world. Two of the most important aspects of FER are a feature set that is as relevant as possible and an accurate classification system. When attempting to represent pictures, it is more typical to make use of method-based features, hand-crafted features, and image expression.

The face image intensities are used to derive the appearance characteristics, which are then used to create a discriminative textural pattern. The adaptive positional thresholds for a face picture were retrieved by Mandal et al. The threshold parameters in the immediate neighborhood may be modified in an adaptive manner for various images. After that, multidistance magnitude characteristics were encoded. Facial expression recognition may be accomplished with the use of SVM as a classifier [8]. In order to recognize facial expressions, Meena et al. used graph signal processing in conjunction with the curvelet transform. Not only has the size of the feature vectors decreased, but also the identification of the facial expression has been much enhanced [9]. Compressive sensing combined with statistical analysis of the collected compressed facial signal was utilized by Ashir et al. to create much more robust visual features for each unique facial expression class [10]. Mlakar et al. [11] proposed a method that identifies emotional responses by estimating variances on a level of descriptors between such a neutral expression and a peak expression of an observed person. This method recognizes emotions by comparing an individual's neutral expression to their peak expression. However, each of the aforementioned approaches has its drawbacks. There is a wealth of information on gradients around each pixel. In conventional HOG, the gradient information surrounding pixels is solely represented by the results of calculating the gradient information for two orientations. In the meanwhile, the feature information that is retrieved by a single feature descriptor is restricted, and as a result, it is unable to satisfactorily complete the facial expression identification job. Additionally, it does not include any high-level semantic information.

An approach that could quantitatively identify reliable feature representation for many categories of facial expressions was proposed by Ashir et al. Utilizing the compressive sensing approach and performing statistical analysis on the recovered compressed facial data, a novel feature representation was developed [12]. Meena et al. presented and enhanced a FER approach in which curvelet transformations in conjunction with graph signal processing are used to construct a model that is able to distinguish facial expressions. Identifying facial expressions has also been significantly improved because of the work done in this research, which resulted in a reduction in the dimensionality of feature vectors [13]. FE was identified by Perikos et al. [14] via the use of adaptive neuro-fuzzy inference systems. FEs are responsible for identifying facial deformations in particular areas, such as the lips, eyes, and eyebrows, and then extracting parameters like position, length, breadth, and shape. Following this step, adaptive neuro-fuzzy inference methods are used to analyze the feature vectors that describe the deformations of the facial expressions in order to identify FEs. This approach demonstrated an accuracy of roughly 90% on average for analyzing the facial expressions of Japanese women (JAFFE).

3. PROPOSED METHOD

This paper presents a system for classifying different kinds of facial expressions. The fundamental objective of this line of study is to enhance the degree of precision with which various kinds of facial expressions may be categorized. To accomplish this goal, this work made use of a wide array of hand-crafted components. This goal was attained by using distinctive characteristics that were hand-picked for their appropriateness. This section provides further information on the proposed procedure.

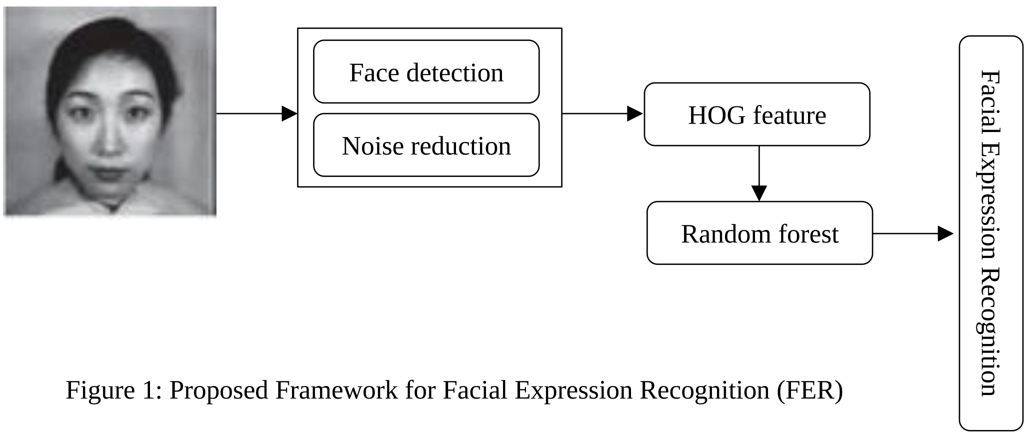

Figure 1: Proposed Framework for Facial Expression Recognition (FER)

Pre-processing

In the realm of image recognition, preprocessing is a necessary step. It is possible to increase the overall performance of the system used to identify facial emotions by removing noise from the picture and enhancing the information that is required via proper preprocessing. Methods of picture processing such as face identification and noise reduction were used in this research project. All these approaches are used in the research presented here with the goal of enhancing the performance of the algorithm. A grayscale picture with just one channel as the input to the classifier is a good example of this. Utilizing this approach has several advantages, including the reduction of differences between training samples, the acceleration of the training process, and a reduction in the amount of computational storage required. Finding and identifying people's faces in photographs using image recognition software is a challenging task in the field of computer vision. This procedure is a prerequisite for FER system implementation and must first be completed. The OpenCV Haar cascaded classifier or more recent deep learning-based algorithms making use of the OpenCV library [15] are both used in this method. Both approaches make use of traditional machine learning (feature-based). OpenCV has a cascade classifier that may be applied for the purpose of face recognition. The XML file containing the pre-trained model might be sent into the constructor in the form of a parameter. A technique might be applied to a picture to locate faces by using the detectMultiScale function, which returns boxes for each of the identified faces once the method is applied. Because the scale factor controls how the input picture is scaled before detection, the minNeighbors and scale factor parameters both need to be fine-tuned for each dataset.

During the first phase of the pre-processing of the proposed model, the Gaussian filter (GF) and the skull stripping approach are used as key phases. Reduced background noise, a simpler design process, automated filtering, and rotational symmetry are just some of the benefits that come with using the GF approach [16]. There is a possibility that the input picture contains noises of various types, including Gaussian noise, salt and pepper noise, and others. The information that is present in the input dataset is kept in a way that is comparable to that of programs that remove noise. The GF technique is used to clear the picture of any noise that may be present. This filter makes use of a 2D Gaussian distribution function, and that function may be represented in the following manner:

$$G(i,j) = \frac{1}{2\pi\sigma^2}\, e^{\frac{-i^2 + j^2}{2\sigma^2}}$$

where the σ represents how the standard deviation is distributed across the population. When developing a convolutional filter with the intention of producing a certain effect, the value of a Gaussian kernel plays an important role in the process. To accurately capture facial expressions, pictures must first have convolutional filters applied to each one of its pixels. When creating facial expression pictures, matrix multiplications are performed on the intensity and kernel values of each pixel in the image. This aids in the early detection of brain cancers. Following this step, the picture is improved with the use of Gaussian filtering, and the noise that was before visible in the facial expression has been removed as a direct result of this action.

Histogram of Oriented Gradients

HOG is a global feature descriptor that may be determined based on the orientations of the edges or the allocation of intensity gradients. The HOG algorithm was developed with the intention of quantifying directed gradients in picture segments that are spatially constrained. It is possible to characterize both the contours of the picture and its overall look by referring to the pixel value or the total quantity of light. Alteration directional in intensity is believed to be a gradient or the image's color, as well as the image's significant information, which can be retrieved by making use of the HOG standard deviation, mean, entropy, and variance characteristics that can be derived by making use of the HOG. In addition to this, HOG is used for the purpose of determining the locations of objects in digital photographs [4].

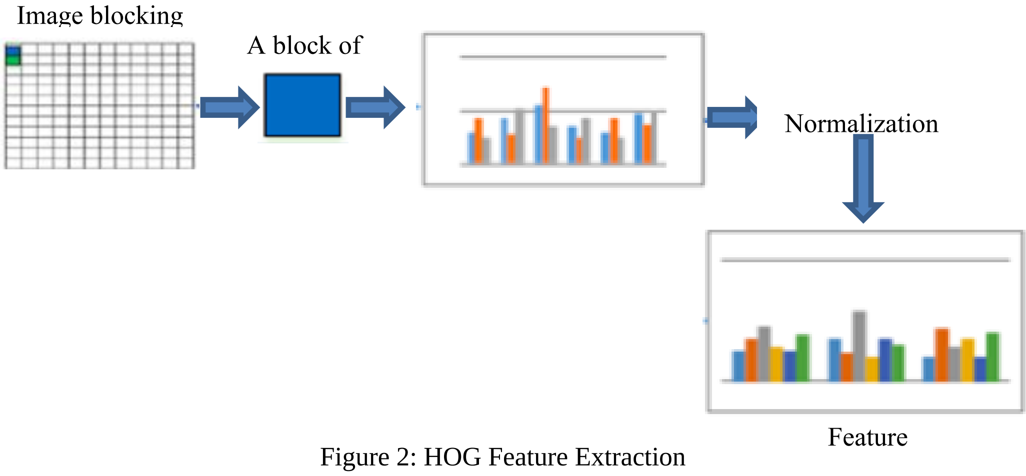

Figure 2: HOG Feature Extraction

The mask coordinates (-1, 0, 1) are used in the computation of the image gradients. This research properly employed the gradient at each mammography site for the appropriate directions based on this extraction technique as it is displayed in Figure 4, for the purpose of identifying the direction. Figure 4 shows this. To achieve higher invariance of lighting, shadowing, and other factors, the contrast-normalization technique of the domestic histogram is used. Figure 4 illustrates the steps involved in extracting HOG features. The following is a rundown of the most significant components of the HOG feature descriptors:

• The HOG descriptor is concerned with an object's structure or form rather than its color. You may be wondering how this is different from the edge characteristics that are extracted from photos at this point. When it comes to edge features, all that must be determined is whether the pixel in question is an edge. HOG is also able to offer information on the edge's direction. To do this, the gradient and orientation of the edges, sometimes referred to as their magnitude and direction, are extracted.

• Furthermore, these orientations are computed in "localized" chunks of the whole. This implies that the whole picture is segmented into smaller sections, and then the gradients and orientations of each segment are determined individually. In the next parts, we will go into much more detail about this topic.

• At the end of the process, the HOG would produce a different histogram for each of these sections. Because the histograms are constructed utilizing gradients and orientations of the pixel values, they are referred to as "Histograms of Oriented Gradients."

Random Forest

The RFC is a powerful decision tree ensemble that is utilized for large-scale and multivariate pattern detection. The idea of random subspace recognition serves as the foundation for this kind of ensemble learning that has been created. This ensemble learning is based on the notions of the random subspace technique and the stochastic discrimination method of classification. Both methods are methods of classification. When used to generate an ensemble of classification trees, a random forest model emerges as a strong instrument [39]. Several studies have revealed that RFC has been successfully used in a variety of contexts. RFC is comprised of dozens upon dozens of trees of decision. Each node of the decision tree poses a question to be answered by the data, and each branch of the tree contains possible solutions to the issue. A random forest is comprised of a hundred different decision trees. RFC is a sort of learning that is supervised and can cope with classification and regression issues [17].

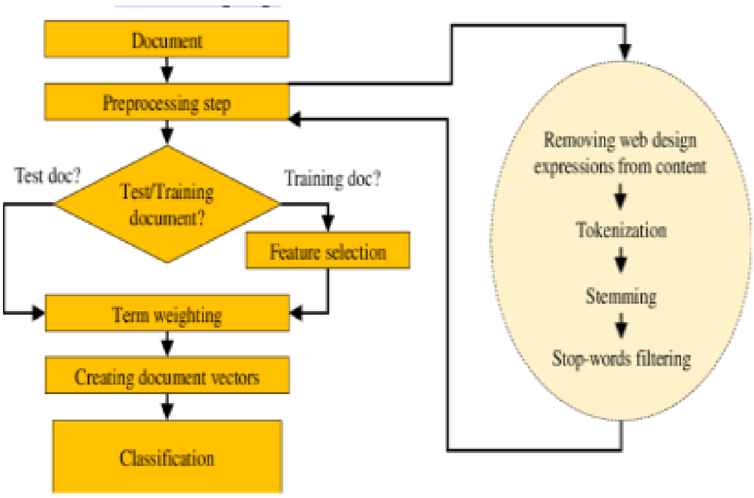

Figure 3: Architecture of RF Classification [18].

The algorithm is used to illustrate the stages that are listed below, as well as Figure 3:

- Select a sample at random from among a predetermined collection of the findings.
- Establish decision trees for each of the samples, then anticipate the results based on the outputs of each decision tree.
- Choose the result that is most likely to occur based on the most recent projection.

4. EXPERIMENTAL SETUP

This section covers the particulars of the experiments that were conducted and how they were conducted. To evaluate how well the suggested FER approach works, a test that makes use of the JAFFE databases has been carried out. The JAFFE dataset samples that were used for training purposes throughout the experiment were also used for testing purposes and included in the testing sample sets. In the previous studies, the most basic kind of classifier that was used was an RF classifier; this research likewise made use of this type of classifier. During the training phase, measurement techniques are quite important, and the selection of such methods is extremely important for discriminating between them and achieving the most effective classifier feasible for the job at hand. FER is fundamentally a multi-class classification problem, with Acc (accuracy), which is the percentage of correctly classified datasets, acting as the key performance metric for the system. This is because Acc measures the percentage of properly classified datasets. The final accuracy might also be obtained by taking the average of the recognition rate for each category of expression within each category of expression. This would allow for the recognition effect of each category of expression to be taken into consideration in a way that is both complete and comprehensive in nature. The two approaches to calculating accuracy that have been discussed so far are referred to as overall accuracy and average accuracy, respectively. In most cases, improved classification performance is directly proportional to increased accuracy. The following equation may be used to determine the level of accuracy of a measurement:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where *TP*, *TN*, *FP*, and *FN indicate* true positive, true negative, false positive, and false negative, respectively.

Dataset

The JAFFE Database aims to collect information on the facial expressions of Japanese women. These individuals are part of the JAFFE database, which has a total of 213 photographs taken of 10 distinct Japanese female subjects. Each of these subjects exhibits three or four examples of one of seven distinct facial expressions (six basic facial emotions and one neutral emotion). The Tisa database stores grayscale pictures, and the resolution of each image is 256 pixels by 256 pixels [4]. In our experiment, there are a total of 213 images are evaluated using the newly developed algorithm. These images depict a range of emotions, such as anger (30 images), disgust (29 images), fear (32) images, happiness (31) images, neutrality (30 images), sadness (31) images, and surprise (30 images). The JAFFE database is the source of the pictures seen in Figure 4.

Figure 4: Samples of Used Dataset

Classification result

According to the findings of this research project, it is abundantly obvious that the strategy that was suggested can deal with the identification of facial emotions in a timely way. The methods of feature extraction that have been implemented so far are broken out in great length in the section that came before this one. One of the strategies that have been put into action is referred to as the HOG feature. After that is complete, a method of classification called an RF classifier is employed on the FER to divide it into the appropriate categories. This work builds the network for assessing the detection accuracy of the presented approach for FER based on all the picture observations that have been gathered up to this point. FER stands for fuzzy error rate estimation. Because it has a greater number of facial expressions than the other datasets used for facial expression identification, the JAFFE dataset is a more challenging one on which to train. Furthermore, this dataset also has a significant issue with the unequal nature of the many emotional classes. This is one of the key problems with the dataset. Three of the classes, "neutrality," "anger," and "happiness," have a disproportionately high number of examples in comparison to the other classes. The findings of the experiment are shown in Table 1.

Table 1: Obtained Performance Accuracy for JAFFE based on Different Features

Expression	Accuracy
Neutral	97.53%
Angry	97.07%
Disgust	91.44%

Happy	98.05%
Sad	84.92%
Fear	89.21%
Surprise	92.63%
Average	92.97%

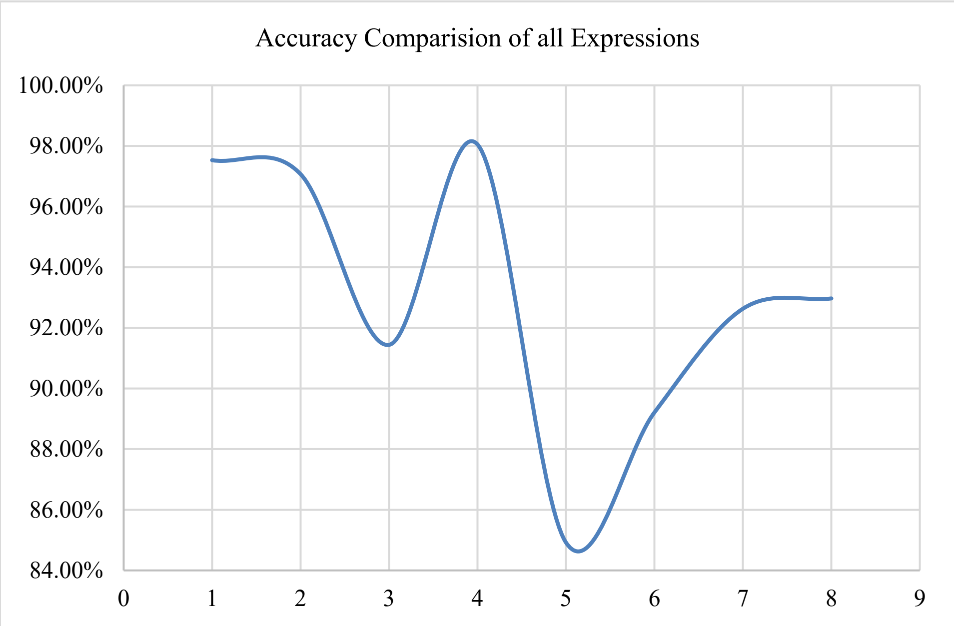

Figure 6: Accuracy Comparison between all Expressions

In [19], there is a proposal for a dependable feature descriptor for the recognition of facial expressions. This descriptor is called regional adaptive affinitive patterns (RADAP). This approach computes positionally adaptive thresholds in the immediate neighborhood and encodes multi-distance magnitude characteristics in a picture in a way that is resistant to changes within the same class as well as variations in irregular lighting. In addition to this, they made use of logical operators in RADAP to construct cross-distance co-occurrence connections, which they then put to the test. It was suggested that three distinct kinds of radar be used. Some examples of these kinds of radar are the xor, adder, and decoder. Using pairwise co-occurrence, XRADAP incorporates into the RADAP feature the quality of resilience to intra-class changes in the RADAP features. On the other hand, the ARADAP and DRADAP descriptors can extract characteristics that are more consistent and unaffected by changes in illumination while simultaneously being capable of collecting minute expression cues that traditional descriptors often neglect. The suggested methods are tested using nine different benchmark datasets, such as the Cohn–Kanade+ (CK+), Japanese female facial expression (JAFFE), and Multimedia Understanding Group (MUG), to assess the overall performance of these methods. An efficient technique for conducting privacy-preserving (PP) facial expression categorization (FEC) in the client-server architecture, which is based on a distributed computing model, is presented in a paper that is cited as [20]. This work presents a lightweight approach that relies on the randomization method, which is meant to be quick and efficient. This method is provided because of this effort. Utilizing the JAFFE and MUG FaE databases, which are readily accessible to the public, it is shown that the proposed method is effective.

Table 2: Comparison Between Some of the Recent Previous Studies and the Proposed Study

Algorithm	**Accuracy**
[19]	90.5%
[20]	94.37%
Proposed method	92.97%

5. CONCLUSION

The results of this study were used to construct a model for recognizing facial expressions by drawing on a variety of appearance-based elements. This model was then tested in an experiment. To extract features, the HOG descriptor is used, and RF classification was implemented to identify the many ways in which the face might express itself. The results of the experiments showed that the proposed framework performed better on the JAFFE database than several other methods that are frequently utilized. The bulk of this piece was dedicated to an examination of different facial emotions that may be seen in still images. In the work that we are going to do in the future, we are going to take into consideration a greater number of facial expressions from a variety of databases, as well as computation times for embedded systems.

Funding: This research received no external funding.
Conflict of interest: The authors declare no conflicts of interest.

REFERENCES

[1] Bursic, S., Boccignone, G., Ferrara, A., D'Amelio, A., & Lanzarotti, R. (2020). Improving the accuracy of automatic facial expression recognition in speaking subjects with deep learning. Applied Sciences, 10(11), 4002.
[2] Boccignone, G.; de'Sperati, C.; Granato, M.; Grossi, G.; Lanzarotti, R.; Noceti, N.; Odone, F. Stairway to Elders: Bridging Space, Time and Emotions in Their Social Environment for Wellbeing. In International Conference on Pattern Recognition Applications and Methods; SciTePress: Setubal, Portugal, 2020; pp. 548–554.
[3] Xie, S., Hu, H., & Wu, Y. (2019). Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. Pattern recognition, 92, 177-191.
[4] Liu Y, Zeng J, Shan S, Zheng Z (2018) Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018), pp 458–465.
[5] Rajan, S., Chenniappan, P., Devaraj, S., & Madian, N. (2019). Facial expression recognition techniques: a comprehensive survey. IET Image Processing, 13(7), 1031-1040.
[6] Wang, Y., Li, Y., Song, Y., & Rong, X. (2019). Facial expression recognition based on random forest and convolutional neural network. Information, 10(12), 375.
[7] Yuan X, Xie L, Abouelenien M (2018) A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recognit 77:160–172
[8] M. Mandal, M. Verma, S. Mathur, S. K. Vipparthi, S. Murala, and D. Kranthi Kumar, "Regional adaptive affinitive patterns (RADAP) with logical operators for facial expression recognition," IET Image Processing, vol. 13, no. 5, pp. 850–861, 2019.
[9] H. K. Meena, K. K. Sharma, and S. D. Joshi, "Effective cur- velet-based facial expression recognition using graph signal processing," Signal, Image and Video Processing, vol. 14, no. 2, pp. 241–247, 2020.
[10] A. M. Ashir, A. Eleyan, and B. Akdemir, "Facial expression recognition with dynamic cascaded classifier," Neural Computing and Applications, vol. 32, no. 10, pp. 6295–6309, 2020.
[11] Mlakar, U., Potocˇnik, B., (2015). "Automatedfacialexpressionrecognitionbasedonhistogramsoforiented gradient feature vector differences". Signal Image Video Process 9:245–253.
[12] H. K. Meena, K. K. Sharma, and S. D. Joshi, "Effective curvelet-based facial expression recognition using graph signal processing," Signal, Image and Video Processing, vol. 14, no. 2, pp. 241–247, 2020.
[13] A. M. Ashir, A. Eleyan, and B. Akdemir, "Facial expression recognition with dynamic cascaded classifier," Neural Computing and Applications, vol. 32, no. 10, pp. 6295–6309, 2020.
[14] Perikos, I.; Paraskevas, M.; Hatzilygeroudis, I. Facial expression recognition using adaptive neuro-fuzzy inference systems. In Proceedings of the 17th International Conference on Computer and Information Science (ICIS), Singapore, 6–8 June 2018; pp. 1–6.
[15] Xie, S., Hu, H., & Wu, Y. (2019). Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. Pattern recognition, 92, 177-191.
[16] K. Subramanian, V.B. Radhakrishnan, S. Ramasamy, Database independent human emotion recognition with meta-cognitive neuro-fuzzy inference system, in Proceedings of the IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), IEEE, 2014, pp. 1–6.
[17] A. Dapogny and K. Bailly, "Investigating deep neural forests for facial expression recognition," in Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on. IEEE, 2018, pp. 629–633.
[18] Wang, H., Wei, S., & Fang, B. (2020). Facial expression recognition using iterative fusion of MO-HOG and deep features. The Journal of Supercomputing, 76, 3211-3221.
[19] M. Mandal, M. Verma, S. Mathur, S. K. Vipparthi, S. Murala, and D. Kranthi Kumar, "Regional adaptive affinitive patterns (RADAP) with logical operators for facial expression recognition," IET Image Processing, vol. 13, no. 5, pp. 850–861, 2019.
[20] Y. Rahulamathavan, M. Rajarajan, Efficient privacy-preserving facial expression classification, IEEE Trans. Dependable Secure Comput. 14 (3) (2017) 326–338, doi: 10.1109/TDSC.2015.2453963.