Journal of King Saud University – Computer and Information Sciences xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing

L.B. Krithika, G.G. Lakshmi Priya *

SITE, VIT University, Vellore, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 25 April 2018
Revised 21 August 2018
Accepted 10 September 2018
Available online xxxx

Keywords:
Emotion recognition
Minimal-Angular Feature Oriented Neural Network based Emotion Prediction (MAFONN-EP)
Weighted Median Filtering (WMF)
Edge-Preserved Background Separation and Foreground Extraction (EPBSFE)
Minimal-Angular Deviation (MAD) and Weight Based Pointing Kernel Classification (WBPKC)

ABSTRACT

In recent days, facial emotion recognition techniques are considered important and critical due wide array of application domains use facial emotion as part of their workflow and analytics gathering. Reduced recognition rate, inefficient computation and increased time consumption are the major drawbacks of these techniques. To overcome these issues, this paper develops the new facial emotion recognition technique named as Minimal Angular Feature Oriented Neural Network based Emotion Prediction (MAFONN-EP). Initially, the input video sequence are categorized into image frames that are pre-processed by eliminating the noise by using Weighted Median Filtering (WMF) technique and to separate the background and foreground regions using Edge Preserved Background Separation and Foreground Extraction (EPBSFE) technique. Then, the set of texture patterns are extracted based on four key parts such as two eyes, nose and mouth by using the Minimal Angular Deviation (MAD) technique. Particular features are selected by employing the Cuckoo Search based Particle Swarm Optimization (CS-PSO) technique that also reduces the feature dimensionality. Finally, the Weight Based Pointing Kernel Classification (WBPKC) technique is employed for recognizing the emotion. In the experimental results, the performance of the proposed technique is analyzed and compared with different performance measures like accuracy, sensitivity, specificity, precision, recall.

1. Introduction

Facial expression recognition is a kind of computer application that is mainly used to identify the facial emotions (Redmond et al., 2017) of the persons either using an image or video (Majumder et al., 2014; Rivera et al., 2013). Video processing has been emerged as one of the major research field in the recent days and it is used in various applications that include tracking, face recognition, surveillance, action recognition and emotion recognition (Raghuvanshi and Choksi, 2016). Also, it is integrated with some electronic devices due to the advent of advanced technologies for sharing the activities of users in the social sites (Khan et al., 2016). It leads to the sharing of people's mood and perception about the events that takes place around the world. So, identifying and analysing the mood of people plays an essential role in video processing, which aims on emotion prediction. The facial features (Bobkowska et al., 2018) are expressed, when these parts contract with each other. Due to the variations of face, recognizing the emotions of a human face is a highly challenging task (Kalita and Das, 2013). The recognition ratio may be affected due to the facial features that include eyes, nose, lips, cheeks, etc. The facial expression recognition contain three stages: face detection, feature extraction and expression classification (Azeem et al., 2014; Beaudry et al., 2014; Datcu and Rothkrantz, 2014).

1.1. Problem description

Typically, a digital video is represented as a pictorial information that is spatially and temporally digitized, then it is formed as a resultant pixel intensity information (Datcu and Rothkrantz, 2014). The videos are split into frames, and its features are extracted for recognizing the emotions (Cruz et al., 2014). During this process, the frames are pre-processed and the facial expression is extracted based on the identified key parts (Li et al., 2013; Halder et al., 2013). Then, the connected component extraction method (Blazek et al., 2014) is applied to get the connected features from the frames. Finally, the classification technique is applied to classify the emotions of the persons in the frames. For emotion recog-

* Corresponding author.
E-mail address: lakshmipriya.gg@vit.ac.in (G.G. Lakshmi Priya).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2018.09.011
1319-1578/

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

nition, various techniques and algorithms are proposed in the existing works, but it requires to improve the accuracy ratio (Owusu et al., 2014).

1.2. Objectives

Based on the problem identified, this paper has the following objectives: To accurately predict the emotion from the given face image, a new Minimal-Angular Feature Oriented Neural Network based Emotion Prediction (MAFONN-EP) technique is proposed. To pre-process the video frames by performing normalization, a Weighted Median Filtering (WMF) technique (Torkhani et al., 2017) is used. To extract the foreground and background, an Edge-Preserved Background Separation and Foreground Extraction (EPBSFE) technique (Hosaka et al., 2011) is introduced. To extract the key parts, a Minimal-Angular Deviation (MAD) based feature extraction technique is utilized. Particular features are selected by employing the Cuckoo Search based Particle Swarm Optimization (CS-PSO) technique (Farag et al., 2016; Khadhraoui et al., 2016; Mlakar et al., 2017) that also reduces the feature dimensionality. To select the set of features, a Weight Based Pointing Kernel Classification (WBPKC) technique is proposed.

1.3. Organization

The remaining sections in the paper are organized as follows: Section 2 reviews the existing techniques and methodologies that are used for emotion prediction and recognition. Section 3 presents the new framework based on various image processing techniques for an accurate emotion prediction. Section 4 evaluates the results of both existing and proposed techniques based on different performance measures. Finally, the paper is concluded and the enhancement that will be implemented in future is discussed in Section 5.

2. Related works

In this section, the existing techniques and algorithms used for video processing are surveyed with its advantages and disadvantages.

2.1. Filtering

Morphological Mean (MM) filter (Lin et al., 2016) was designed for detecting the high density impulse noise based on image restoring. This work contains two modules such as Noise-free Pixel Counter (NPC) and Morphological Pixel Dilation (MPD), in which the position and number of noise free pixels were collected by examining the pixels of input image in the first module. Then, the dilation operation was performed to recover the image corrupted based on the density of the noise in the second module. These modules were iteratively performed until the target is obtained. However, the MM technique has the following drawbacks: it consumed more execution time and required to increase the quality of restored image.

A bitonic filter (Treece, 2016) was suggested for an efficient noise removal that comprises both the non-linear morphological and linear operators. This technique does not create any additional noise during the noise elimination in other areas. To evaluate the performance of this technique, it was compared with some other techniques such as gaussian linear filter, mean median filter, image guided filter, anisotropic diffusion, Non-Local Means (NLM) filter, Opening Closing – Closing Opening (OCCO) filter, and self-dual grain filter. From the analysis, it was observed that the bitonic filtering technique provided efficient results in terms of Signal to Noise Ratio (SNR) and Structural Similarity (SSIM) compared than the other techniques.

2.2. Foreground and background separation

A new segmentation algorithm (Wei et al., 2015) was suggested for separating the foreground and background regions from the video sequences. The main intention of this paper was to exactly identify the moving objects. During background extraction, a Gaussian Mixture Modeling (GMM), improved GMM and adaptive GMM techniques were employed. Then, during foreground extraction, background gradient difference, neighbourhood smooth, morphology processing, filtering in a single frame and filtering between the frames were performed. From the paper, it was inferred that the suggested segmentation technique has the ability to effectively remove the noise and to extract the foreground information of different video types in an accurate manner. However, it has an increased time complexity during segmentation, which was the major limitation of this paper.

An enhanced GMM (Nurhadiyatna et al., 2013) was integrated with the Hole Filling Algorithm (HFA) for detracting the background from the given video sequence. In this technique, the result obtained from GMM was improved by applying the principals of median filtering and HFA. The limitation of this paper was, it required to determine the parameter of radius based on the video data in non-parametric GMM-HFA.

2.3. Feature extraction and selection

A new framework (Khan et al., 2013) was introduced for identifying the facial expressions with increased accuracy. The main objectives of this paper were to increase both the time and memory efficiency. This work includes the following stages: pre-processing, feature extraction and classification, in which the face was normalized in the pre-processing stage by cropping the background. Then, the Local Binary Pattern (LBP) and Weber Local Descriptor (WLD) pattern extraction techniques (Chao et al., 2015) were employed to extract the features of the whole face image and the most important features were obtained after zigzag scanning. Finally, the classification was performed to detect the face expressions based on the selected features. The major drawbacks of this paper were increased computational complexity and misclassification rate.

A self-organizing map (Majumder et al., 2014) was suggested for recognizing the emotions based on geometric facial features. The main intention of this work was to develop an automatic method for detecting facial expressions. Also, Paul Viola and Michael Jones face detection techniques were utilized to extract the face region. A multiple deep network learning (Yu and Zhang, 2015) was used for detecting the facial expressions from the given video sequence. Here, the pre-processing was performed to remove the irrelevant noise by applying the transformations. In this paper, the network architecture contains three stochastic pooling layers and three fully connected layers. From the paper, it was observed that the suggested model attained the accuracy of 45% for the Facial Expression Recognition (FER) dataset.

A Recurrent Neural Network (RNN) was introduced for recognizing the emotions (Ebrahimi Kahou, 2015) in a video sequence. Here, the Convolutional Neural Network (CNN) was mainly utilized to detect the facial features from the input for video recognition. A Recurrent Neural Network (RNN) was developed for recognizing the emotion based on a multi-modal dimension (Chen and Jin, 2015). Here, the prediction performance was improved based on the domain modalities, loss function, duration of features and multi-task learning and directions of Long Short Term Memory (LSTM). Moreover, the arousal dimension was predicted based on

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

the audio features, which was more suitable for valence dimension prediction.

2.4. Emotion prediction and classification

The use of machine learning algorithms (Ringeval et al., 2015) was suggested to predict the emotion based on the ratings offered by different raters. For this purpose, a fully naturalistic multimodal database, namely, a REmote COLlaborative and Affective interaction (RECOLA) was utilized in this paper. Here, multimodal fusion was performed in two levels such as, features and decision. The main aim of this paper was to increase the relevance of emotion prediction by reducing the quantity of data that were given to the machine learning algorithm.

Mid-level features (Sanchez-Mendoza et al., 2015) was utilized for recognizing the emotions in an automatic manner. Here, the histograms of action units were utilized to classify a particular set of positive and negative attitudes. Moreover, both the geometric and appearance descriptors were used to train the emotion classifier. Then, the fusion of both sources was performed during the level of emotion classification. A dynamic emotion prediction task (Dzafic et al., 2016) was designed by creating a novel Functional Magnetic Resonance Imaging (FMRI) paradigm. Here, the audio visual videos were considered for predicting the emotional information with prior information. The limitation of this paper was, this approach has the capability to pertain the results only for male population.

An ensemble deep learning model (Yin et al., 2017) was suggested for emotion recognition. In this paper, a Multiple fusion layer based Ensemble Classifier of Stacked Auto Encoder (MESAE) was developed to predict the emotion based on a physiological data driven approach. Moreover, the reliability of both structural learning and fusion methods were validated with respect to varying size of the available physiological instances. The disadvantages of this paper were, it required an increased time and memory consumption.

A new emotion recognition algorithm (Katsimerou et al., 2015) was developed for predicting the mood of a known sequence. In this paper, two analysis were performed in which, the mood estimation was facilitated at first and then the mood recognition was performed in the second analysis. Moreover, this paper classified the mood classes into the following: Positive valence high arousal that includes gladness and eagerness and negative valence high arousal which includes worry and annoyance. Positive valence low arousal includes quietness and peacefulness and negative valence low arousal includes unhappiness and seriousness. The major drawback of this paper was, it required to prove the accuracy of the suggested model, when applying the sequence of machine recognized emotions.

A dynamic regression model (Elaiwat et al., 2016) was developed for identifying the facial expression in the video sequence. The main objectives of this paper was to describe the subject specific facial feature movements of various expressions and to develop a salient longitudinal facial expression based on a sparse group wise registration method. The disadvantage of this paper was, it failed to utilize the complex image matching metrics that includes correlation coefficient and localized mutual information for proving the robustness against illumination changes.

A Deep Convolutional Neural Network (DCNN) was suggested for recognizing the emotion (Kahou et al., 2013) in the given video sequence. Here, an additional static frame training datasets were used to increase the validation set performance. From the paper, it was observed that this investigation suggested a simple averaging model without any combination, which leads to increased performance. A new image based representation (Yang and Bhanu, 2012) was introduced for the Emotion Avatar Image (EAI) for detecting the facial expressions. This work includes the following stages: Face detection, Face registration of video frames, Feature computation and Feature classification. The main intention of this work was to build a new model, namely, avatar model for recognizing the discrete facial expressions.

A rule-based decision model (Patwardhan and Knapp, 2016) was suggested for recognizing the emotion from the video sequence. In this paper, a multimodal emotion system was implemented with the help of infrared sensor. Moreover, the face recognition Application Programming Interface (API) was utilized to extract the facial features. The limitation of this work was, it has a reduced accuracy level during the recognition of emotions.

A multi-kernel learning approach (Li et al., 2015) was introduced for recognizing the emotion from the given video sequence. This paper integrated the approaches of both multimodal and hybrid features for accurately classifying the emotions. Moreover, a multi-kernel learning was utilized to fuse both the learned and engineered features from various channels. Furthermore, a Support Vector Machine (SVM) based classification was utilized to take a decision for emotion recognition.

A deep transfer learning classification technique was suggested for a video based emotion recognition. This work includes the following stages: pre-processing, image purification, feature extraction, video modelling and classification. Here, Principal Component Analysis (PCA) was utilized to remove the false detections. Then, the Convolutional Neural Network (CNN) was utilized to extract the features with flexible registration and fine tuning. Also, it integrated the audio and visual features based on the weighted score level fusion. Moreover, the fisher vector encoding model was applied to improve the diversity of learners based on the frame level features. From the paper, it was observed that the CNN (Kaya et al., 2017; Baltrusaitis et al., 2016) required large amount of data for avoiding over-fitting.

From the survey, it is observed that the existing algorithms and techniques have both advantages and disadvantages, but it has major limitations that include the following: Increased computation time and misclassification results, reduced accuracy level, not highly efficient, reduced robustness. In order to overcome these drawbacks, this paper aims to develop a new recognition system for detecting facial expressions in a given video sequence, which has the following steps:

1. Pre-processing – A Weighted Median Filtering (WMF) technique is implemented to efficiently eliminate the noise.
2. Background separation and foreground extraction – An Edge Preserved Background Separation and Foreground Extraction (EPBSFE) technique is used for suppressing the foreground region from the background.
3. Feature extraction – A new Minimal Angular Deviation (MAD) based feature extraction is employed to extract the meaningful features from the preprocessed image.
4. Feature Selection: To select the optimal features from the image, Cuckoo Search based Particle Swarm Optimization (CS-PSO) is employed.
5. Classification – A Weight Based Pointing Kernel Classification (WBPKC) technique is proposed to recognize the emotion with the classified label.

3. Proposed method

This section presents the detailed description of the proposed Minimal-Angular Feature Oriented Neural Network based Emotion Prediction (MAFONN-EP) for video based emotion prediction system. The main aim of this work is to develop a new framework by overcoming the disadvantages of the existing works. The proposed work includes the following stages: video frames splitting,

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

pre-processing, background and foreground separation, key-points extraction, feature extraction, feature selection and classification. At first, the input video sequence is split into frames, which is pre-processed by applying the Weighted Median Filtering (WMF) technique. Then, a background and foreground region is extracted by applying an Edge-Preserved Background Separation and Foreground Extraction (EPBSFE) technique. Next, the facial key parts are extracted by using the Minimal-Angular Deviation (MAD) based feature extraction technique. Finally, the set of features are selected using Cuckoo Search based Particle Swarm Optimization (CS-PSO) and a Weight Based Pointing Kernel Classification (WBPKC) technique is implemented to classify the emotion. The overall flow of the proposed MAFONN-EP system is illustrated in Fig. 1 and its detailed description is provided in the following sections:

3.1. Pre-processing

Pre-processing is an initial stage that is mainly performed for edge preservation and noise elimination. Generally, a face occupies a small part of image and in order to extract this part, it must be pre-processed. In this work, the WMF technique is employed for smoothening the image by eliminating the noise. The main aim of this technique is to get the sequence of images with uniform size, shape and normalized intensity value. After splitting the video into frames, the laplacian model is used to sharpen the edges as shown below:

$$\mathrm{I}(\mathrm{x},\mathrm{y}) = \mathrm{J}(\mathrm{x},\mathrm{y}) + \mathrm{m}\left[\Delta^2(\mathrm{J}(\mathrm{x},\mathrm{y}))\right] \tag{1}$$

where, J (x, y) represents the original input image, I (x, y) is defined as the sharpened image and m is -1 for the filter mask values. The filter mask function is estimated as follows:

$$\Delta^2\mathrm{J} = \partial^2\mathrm{J}/\partial\mathrm{x}^2 + \partial^2\mathrm{J}/\partial\mathrm{y}^2 \tag{2}$$

The original image and the pre-processed result after applying WMF technique are shown in Fig. 2(a) and (b).

3.2. Background separation and foreground extraction

The foreground includes the region containing moving edges and the statistical information like illumination changes of foreground and background regions. After edge detection, the foreground that consists of more edge lines compared to back ground helps to classify the background scene easily. Then the edge based gradient is estimated for all detected edges, from which slow changes in gradient indicates the background region and vast changes in gradient denotes the foreground pixels. After preprocessing, background separation and foreground extraction is necessary where the new image is subtracted from the background and the foreground objects stay alone. After background extrac-

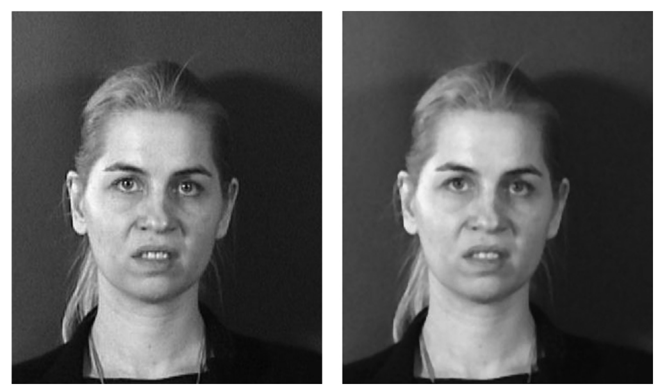

Fig. 2. (a) Original image (b) Pre-processed image.

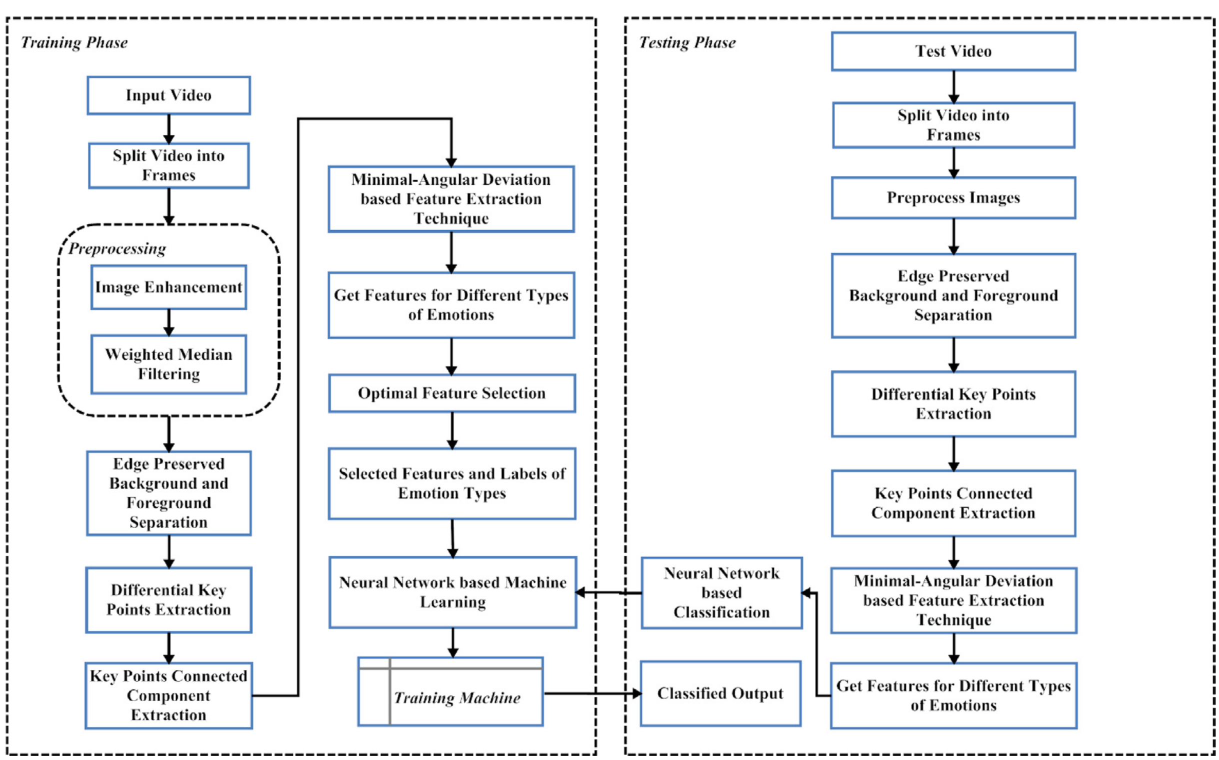

Fig. 1. Flow of the proposed system.

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

Fig. 3. Background suppressed image.

tion, pre-liminary background is formed and foreground targets can be detected. However, the detected targets are easily broken. There may be some misjudged targets as well. The background region of frames are separated which will help in extracting the facial parts from frame. To achieve this, an Edge-Preserved Background Separation and Foreground Extraction (EPBSFE) technique is used for separating the background and foreground objects by detecting the edges are shown in Fig. 3.

3.3. Feature extraction

The feature extraction mainly focuses to extract the most suitable features that are used for representing the characteristics of human face. It converts the image pixel into a higher level representation of colour, texture, shape, motion and spatial configurations. In this work, eight different directions are considered and minimum different angle and value of pixel blocks are subtracted from all neighbourhood. Hence the feature extraction method is been named as Minimal Angular Deviation based. The four key parts such as two eyes, nose and mouth are extracted to recognize the facial expressions. Here, the viola jones algorithm is implemented to extract the differential key parts of the face. It is the robust and fast method widely used on many face detection applications. Here, the haar-like features are extracted for detecting the face from the image. This framework has the advantage of increased detection rate, speed and high accuracy.

To extract the features of the key parts, an optimal MAD technique is developed, which takes the input as connected facial key parts that include left eye (F_{c1}), right eye (F_{c2}), nose (F_{c3}) and mouth (F_{c4}). Here, n is the number of key parts and the matrix is initialized as 5×5. The size of the window over the extracted key parts (F_{cn}) is estimated based on the row size and column size. After estimating the median value, the different angles of the images that include *0°, 45°, 90°, 135° and 180°* are considered for pattern extraction. Then, the upper and lower portion of each key part is calculated. Based on this value, the histogram features are estimated. Finally, the pattern vector is extracted by combining the histogram of all key parts. The key parts are extracted from the background separated image by using MAD is shown in Fig. 4. In this stage, 256 features are extracted with respect to each key part for both upper and lower points. So, totally there are 2048 features extracted in this stage by using the MAD algorithm. This technique extracts the most relevant features based on the key part and produced the features in the form of texture patterns. A new feature extraction algorithm is proposed (Algorithm I) which extracts features based on dissimilarity variation of visual content from the orientation of the pattern. However, the patterns are the key part of the facial features.

Algorithm I (*Minimal Angular Deviation Based Feature Extraction*).

Input: *The facial key parts such as left eye (F_{c1}), right eye (F_{c2}), nose (F_{c3}) and mouth (F_{c4});*
Output: *Texture pattern of the key parts;*
Let n be the number of key parts;
Step 1: Initialize × window matrix; //The key parts of eyes, nose, and mouth are
located within the 5 × 5 matrix, thus this window size is used in this process;
Step 2: Protect the window over the extracted key parts (F_{cn});
Step 3:[p, q] = Size of (F_{cn}),
//where, p represents the row size and q represents the column size;
For i = 3 to (p) – 2
For j = 3 to (q) – 2
Step 4: t = F_{cn} (i – 2: i + 2, j – 2: j + 2); // where, t is the temporary variable;
Step 5:t_s = median (median (t));
Step 6: Let t_k = 2; // Different angles in the image pixels like 0°, 45°, 90°, 135° and 180° are considered; Here, the value of F_{cn} index will be 0 or -1 in Abs(F_{cn}(3, 4) -($t_s + t_k$)), if the value of 2 is assigned to t_k.
Step 7: For each angle (Compute the absolute difference between the neighborhood pixels)
// For 0° cases,
If$(abs(im(i,j+2) - im(i,j+1) > abs(im(i,j+1) - ts)$
$temp\ 1(i,j+1) = 1;$
$temp\ 2(i,j+1) = 0;$
else if $(abs(im(i,j+2) - im(i,j+1))$
$< abs(im(i,j+1) - (ts+k))$
$temp\ 1(i,j+1) = 1;$
$temp\ 2(i,j+1) = 0;$
else if $(abs(im(i,j+2) - im(i,j+1)$
$< abs(im(i,j+1) - (ts+k))$
$temp\ 1(i,j+1) = 0;$
$temp\ 2(i,j+1) = 1;$
else $temp\ 1(i,j+1) = 0;$
$temp\ 2(i,j+1) = 0;$
Step 8:$L_p(i,j) = \sum l_{pt}(i,j) * 2^n$
Step 9:$U_p(i,j) = \sum u_{pt}(i,j) * 2^n$*//where, n represents the feature part number.*
Step 10: Histogram FeaturesH_1,H_2
H_1 = {Hist(L_p)}
H_2 = {Hist(U_p)}
Step 11: $H = H_1 + H_2$;
Step 12: Pattern VectorP_V;
P_V={H_{leye}, H_{reye}, H_{nose}, H_{mouth}};

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

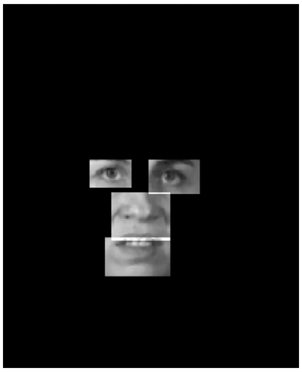

Fig. 4. Extracted key parts.

3.4. Feature selection

Feature selection is one of the special form of dimensionality reduction that is mainly used to describe a large set in an accurate manner. In this work, the Cuckoo Search (CS) based PSO technique is used for selecting the optimal features for face recognition. Typically, CS is a global search algorithm that is mainly used to find an optimal solution, in which the quality of the fitness can be determined based on the value of objective function. In this algorithm, each and every particle deposit it's flight experience and then the updates are added in the repository. It is used to solve the optimization problem by identifying the lead for providing the guide to search. It is also used for solving the non-linear problems based on the multi-objective optimizations. It works based on the following rules: Typically, each cuckoo lays one egg and dumps it into randomly selected nests. Then, the nest that have high quality eggs are selected for next generations. Based on the probability distribution, a host can discover an alien egg; in this case, the host bird can build a new nest in a new location.

The PSO is an attractive feature selection algorithm that discover the best features within the subset space. When compared to the traditional Genetic Algorithm (GA), the PSO does not require any complex operators like crossover and mutation. Moreover, it is computationally inexpensive and uses only simple and primitive mathematical operators. The cuckoo search algorithm randomly selects the features for analysing and PSO algorithm selects the features by comparing each particle to neighbourhood particle. The advantage of CS algorithm is high convergence speed. Due to this benefits of CS and PSO, both are integrated in this work for a better feature selection process and it offered effective results.

In this Algorithm II, the extracted texture patterns from the previous stage are given as the input, in which the number of cuckoos and fitness value are initialized. Then, the set of patterns C_u are estimated based on the number of training set images. Consequently, the fitness value is initialized as 0 and the radius value is initialized as 0. The fitness value is computed based on the set of patterns and the size of the particles. In this paper, the key parts are considered as patterns, which are converted into features by using the Minimal Angular Deviation (MAD) based feature extraction algorithm as illustrated in Algorithm I.

After computing the cuckoo coordinates, the best cluster point is selected and its radius is updated based on the best selected particles and radius. Furthermore, the cluster head is also selected and the mutation values are estimated with respect to the minimum and maximum values. If the mean is greater than 0, the radius value is updated with the mean value. Finally, the set of features are selected based on the size of fitness value. The total number of features selected from the 2048 features obtained from the previous stage is reduced to 60 by using the CS-PSO technique. This technique selects the most suitable features that are used for exactly recognizing the emotion. Here a new weight based Pointing Kernel Classification algorithm is proposed (Algorithm II) that classifies the emotion based on the optimized dissimilarity variation. The updated weight is used for calculating the kernel matrix.

Algorithm II (*Cuckoo Search Based Particle Swarm Optimization for Feature Selection*).

Input: *Extracted texture patterns (P_V);*
Output: *Selected feature index (S_{idx});*
Consider x = 1;
Step 1: Initialize the number of cuckoos;
Step 2: Initialize the fitness value;
Step 3: Initialize the cuckoo particles as,
$Cu = \{P_{V1}, P_{V2}, P_{V3} \cdots P_{Vn}\}$
// Where, Cu is the set of patterns, and n is the number of training set images for extracting the features; In this step, the cuckoo particles are initialized for all the key part features, but in general the single particles are only considered.
Step 4: Initialize Fit = 0 and rad = 1 // Where, Fit represents the fitness values and rad represents the radius values;
Step 5: Compute the fitness values based on the following:
$Fit = [Cu(x), (m \times n)]$
Where, $m = rad/\sum_{y=1}^{D} Cu_y * (D-1) + 1$*, and D represents the size of particles;*
Step 6: Let, $n = \sqrt{Cu_y/m} - 1$ *//Where, y represents the initial point;*
Step 7: The computed cuckoo coordinates are as follows:
$(a, b) = Cu(Fit);$
Step 8: Best cluster point is,
$Bs_{cp} = \begin{cases} if\,(a \approx b < rad) Cu_y \\ else 0 \end{cases}$
Step 9: The radius is updated as follows:
$rad = (en(1) \approx en(2) * Bs_{cp})/rad + en(1);$
Step 10: Update the cuckoo particles:
For i = 1: G //where, G represents the iterations;
If $(Fit_{i-1} > Fit_i)$
Step 11: Cluster head selection formulation;
$C_{head} = Cu_{idx};$
Where, $idx = \begin{cases} if 0 > (Cu * e^{-\varphi D}) 1 \\ else 0 \end{cases}$
$Cu_i = i - 1/(G-1);$
// where, $e^{-\varphi D}$ *defines the parameter that is based on the size of particle for estimating the cluster head location;*
Step 12: Swarming phase:
If $Cu(i) > Cu(Fit)$
Step 13: Fit_{swarm} *= [Cu(1),(m*n)]*
End if;
Step 14: If mean value (Fit_{swarm}*) greater than 0*
Update radius, as rad = rad + mean value (Fit_{swarm}*)*
End if;
End for;
Step 15: $Lev = \sum Fit/Size(Fit)$ *//Where, Lev represents the value that is calculated based on the mean of the fitness obtained for G iterations.*
Step 16: $S_{idx} = idx(Fit > Lev)$

3.5. Classification

After selecting the features, the new Weight Based Pointing Kernel Classification (WBPKC) technique, is proposed in this work for accurate face emotion recognition. The proposed classification

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

technique is developed based on the functionalities of traditional Neural Network (NN) classification technique. So, it is termed as Minimal-Angular Feature Oriented Neural Network based Emotion Prediction (MAFONN-EP). In this Algorithm III, a set of weights are randomly selected during forward propagation. This classifier calculates the weight value for identifying the emotion. Also, the margin error of the output is measured, which is reduced by adjusting the weights for the entire training features.

This classification technique contains three layers such as input layer, hidden layer and output layer. The input layer takes the input features and the output layer can produce the result as a classified label. In the hidden layer, the classification process is performed based on the input features obtained from the input layer. Here, the selected optimal training features, training labels and testing features are given as an input for classification. Then, the minimum and maximum values of the training features are estimated from the selected index. Consequently, the obtained features are divided based on the number of labels and the rule is generated for the training feature by considering the minimum and maximum values. After that, the weight value is computed with respect to the size of training and testing features. Moreover, the kernel function is generated and for each training feature the value of kernel function and weights are added. Finally, the recognition is estimated by using the classified label. Here, the emotions such as angry, disgust, fear, sad, happy are recognized by analysing the variations of the mouth and eyebrows of the face based on the features extracted. The proposed WBPKC accurately classifies the emotion by efficiently calculating the kernel function.

Algorithm III (*Weight Based Pointing Kernel Classification*).

Input: *The selected optimal training Features $P_V(S_{idx})$, training labels Tr_{label}, Testing feature*
Output: *classification label as output*
Anger – 1;
Disgust – 2;
Fear – 3;
Sad– 4;
Happy – 5;
Step 1: Initialize number of labels as n.
Step 2: $Tr_{feat}=P_V(S_{idx})$;
Step 3: $Ts_{feat}=Test_{feat}(S_{idx})$;
Step 4: $Mx\_Val= Max\ (Tr_{feat})$;
Step 5: $Mn\_Val= Mean\ (Tr_{feat})$;
Step 6: Subdivide features based on the number of labels
$L = 1 > 1/n > N$;
// For different images, the emotions can vary that has certain features. Then, the extracted features are trained and the labels are assigned based on the emotions. Also, the rules are estimated and the weight value is updated that predicts the label with respect to the testing features.
Step 7: Generate the Rules for the training features
$Rl = Tr_{feat}(Mx_val - Mn_val) * L$;
Step 8: Let $Cl = \text{size of} Tr_{feat}$;
For v = 1 to Cl
Step 9: $\rho_v = Tr_{feat}{}^{-1}(Ts_{feat}(v))$
Compute weight
Step 10: $wt = 1/P^* \left(\sum Tr_{feat}(Cl)/size(Tr_{feat}(Cl))\right)$
Step11: Generate Kernel function
$$\hat{a} \pm \cdot(v) = \sum_{i=1}^{N_i} e^{\left[-\frac{\left(CT_{feat}(v)-Tr_{feat}(v)\right)^{-1}\left(CT_{feat}(v)-Rl\right)}{2\sigma^2}\right]}$$
Step 12: $Kr=Rl^{-1}\ (\varphi(v))\ ^*\ wt$;
Step 13: For each training feature e = 1: size (L)
$CT_{feat}(e)=Kr^e + \rho_e + wt$;
End for;
Step14: if $(Tr_{feat}(v) > CT_{feat})$
Label = $L(\hat{a} \pm \cdot(v))$; *// If the trained features of the particular part is greater than* CT_{feat}*, the value of* $L(\hat{a} \pm \cdot(v))$ *is considered as the predicted label;*
End if
End for

4. Performance analysis

This section presents the performance evaluation results of the proposed system with respect to various measures such as sensitivity, specificity, accuracy, precision, recall and average recognition rate. The proposed method is evaluated over MMI dataset (MMI Facial Expression Database, 2016) which contains 2900 videos with high resolution images of 75 subjects with the facial expressions of different persons. This dataset is mainly developed for solving the issues of automatic human behaviour analysis. It is a building block of evaluating facial expression recognition algorithms. From the MMI database, 175 facial expression sequences containing emotions of 33 subjects were chosen. The selected conditions are such that each sequence can be labelled as one of the five emotions: anger, disgust, fear, sad, happy.

Moreover, the betterment of the proposed technique is proved by comparing it with the traditional face emotion recognition techniques. The existing techniques (Guo et al., 2012; Guo et al., 2016) considered in this paper are Dynamic Facial Expression Recognition (DFER), Longitude Facial Expression Atlases (LFEA), Dynamic Texture Recognition using Local Binary Pattern (DTR–LBP), Interval Temporal Bayesian Network (ITBN), Hidden Markov Model (HMM) and Feature Based Detection of Facial Landmarks (FBDFL).

4.1. Overall performance evaluation

Sensitivity and specificity are the statistical measures that are mainly used to evaluate the classification performance in which, sensitivity is defined as the ability of a test that is correctly recognized by the classifier. The sensitivity and specificity are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

where, TP represents true positive, TN indicates true negative, FP represents false positive and FN indicates false negative. Based on these measures, the accuracy of the recognition system can be calculated. It states that how accurately the proposed technique can recognize the emotions. It is calculated as follows:

$$Accuracy = \frac{TN + TP}{(TN + TP + FN + FP)} \tag{3}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

The existing emotion recognition methods (Guo et al., 2012; Guo et al., 2016) namely Dynamic Facial Expression Recognition (DFER) and Longitude Facial Expression Atlases (LFEA) has been considered for comparing with the proposed method. In both the

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

Table 1
Confusion Matrix obtained by proposed method on MMI database.

	Anger	Disgust	Fear	Sad	Happy
Anger	96.3	0	0.6	0	1.3
Disgust	0.5	98.3	0	0.4	0.9
Fear	0.5	0	94.4	0.4	0.5
Sad	0	1.4	0.6	98.8	0.5
Happy	1.6	0.3	3.6	0.2	96.7

Table 2
Overall performance of existing & proposed recognition system.

Measures	LFEA	DFER	Proposed
Sensitivity/Recall	94.07	96.50	97.33
Specificity	98.45	99.10	99.32
Precision	94.06	96.48	97.35
Accuracy	94.06	96.48	97.33

existing methods, each image sequence was labelled with six emotions: anger, disgust, fear, sad, happy and surprise. To investigate a better average emotion recognition rate only five emotions namely anger, disgust, fear, sad, happy are used for comparison purpose. The confusion matrix of the proposed method over 10 fold cross validation method is adopted for the five emotions are listed in Table 1.

Table 2 shows the overall performance of the existing DFER and LFEA and proposed emotion recognition system. From the evaluation, it is observed that the proposed technique provides better performance results. In this analysis, the sensitivity is increased to 97.33%, the specificity is increased to 99.32% and the accuracy is increased to 97.33%.The proposed work is implemented using Matlab R2015a version with system configuration of Intel core i5-4200U of RAM 8 GB with speed of 2.5 GHz processor.

4.2. Average recognition rate

Fig. 5 shows the average recognition rate of the existing and proposed facial emotion recognition techniques. When compared to the existing techniques, the proposed MAFONN-EP increased the recognition rate above 97%.In the proposed system, the features of the key parts are separately extracted, which efficiently detects the emotion of the face. Thus, the recognition rate is improved above 97%. The main reason for the improvement is, the proposed technique preserves the salient information by efficiently discriminating the expressions. Also, it suppresses the facial shape variations for an accurate recognition.

Also, the recognition rate is calculated with respect to varying emotions for both existing and proposed techniques is shown in Fig. 6. From the results, it is observed that the proposed MAFONN-EP provides an increased recognition rate other than 'fear' and 'happy' emotions when compared to the other existing techniques.

In addition, the recognition rate of existing DFER, LFEA and proposed MAFONN-EP techniques are separately calculated with respect to the emotions anger, disgust, fear, sad, happy over 10 fold cross validation is shown in Fig. 7. The number of folds represent the training set and the proposed classification method attains above 90% recognition rate, when using 5 folds as the testing set and 5 folds as the training set. From the evaluation, it is observed that the proposed MAFONN-EP technique provides an improved recognition rate, when compared to the other existing techniques.

4.3. Comparison analysis of CS, PSO and proposed MAFONN-EP

This section depicts the comparison analysis of CS, PSO and proposed MAFONN-EP. Table 3 shows the accuracy results of existing CS, PSO, RNN and proposed MAFONN-EP. As per the results obtained using the proposed method, the 'Fear' emotion offered the highest value and the 'Disgust' emotion offered the lowest value compared to other algorithms.

Table 4 shows the sensitivity results of existing CS, PSO, RNN and proposed MAFONN-EP. As per the results obtained using the proposed method, the 'fear' emotion offered the highest value and the 'anger' offered the lowest value compared to other algorithms.

Table 5 shows the specificity results of existing CS, PSO, RNN and proposed MAFONN-EP. As per the results obtained using the proposed method, the 'Fear' emotion result offered the highest in

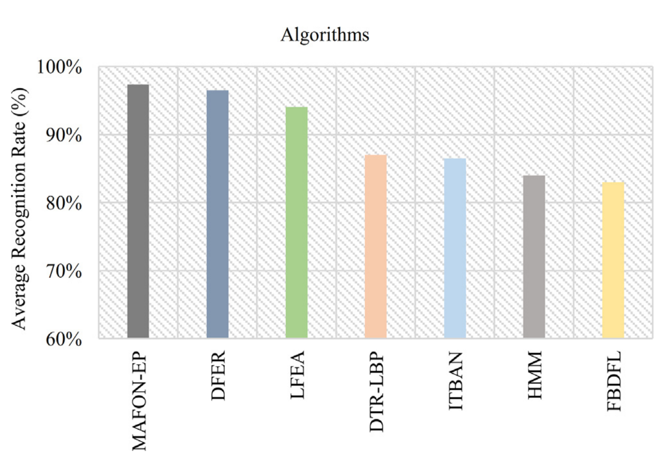

Fig. 5. Recognition rate of existing and proposed techniques.

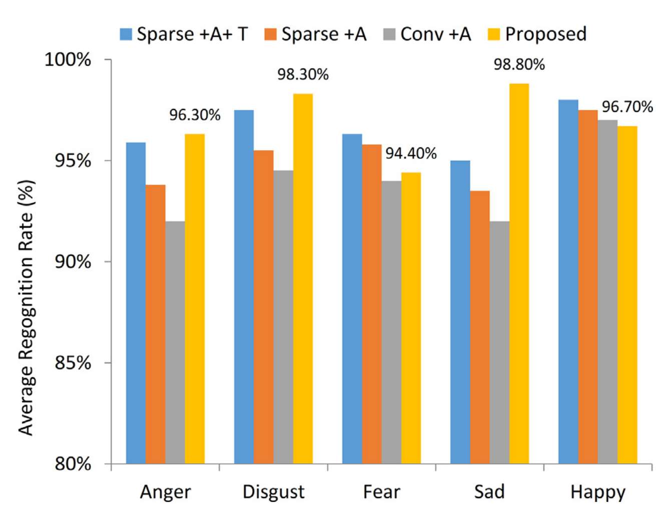

Fig. 6. Recognition rate with respect to different emotions.

Please cite this article in press as: Krithika, L.B., Lakshmi Priya, G.G. MAFONN-EP: A Minimal Angular Feature Oriented Neural Network based Emotion Prediction system in image processing. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.09.011

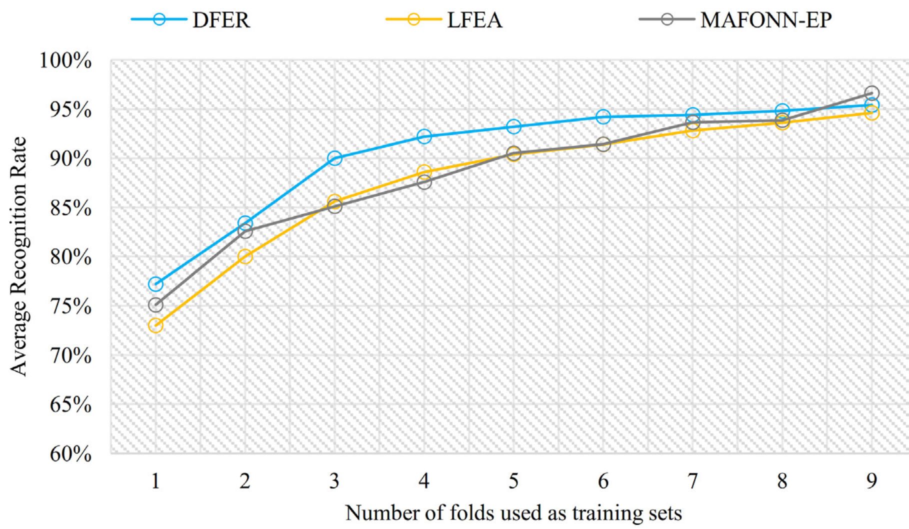

Fig. 7. Overall Recognition rate.

Table 3
Accuracy.

Accuracy	CS	PSO	RNN	MAFONN-EP
Anger	85.81	85.81	87.83	90.5
Disgust	75.37	75.67	84.45	85.13
Fear	85.81	90.54	70.94	93.24
Sad	68.24	70.94	64.18	90.18
Happy	68.91	70.94	75	84.81

Table 4
Sensitivity.

Sensitivity	CS	PSO	RNN	MAFONN-EP
Anger	29.41	18.18	25	62.8
Disgust	24	18.51	33.33	65
Fear	61.53	69.23	19.37	76.92
Sad	27.02	30	27.58	75.33
Happy	40	43.75	56.25	70.12

Table 5
Specificity.

Specificity	CS	PSO	RNN	MAFONN-EP
Anger	90.98	90.83	91.3	99.35
Disgust	89.43	87.4	88.37	99.43
Fear	88.15	87.1	81.9	99.58
Sad	80.83	81.25	81.51	99.29
Happy	76.53	72.8	76.12	98.64

proposed method and the minimum result in 'Happy' emotion compared to other algorithms.

Table 6 shows the precision results of existing CS, PSO, RNN and proposed MAFONN-EP. As per the results obtained using the proposed method, the 'Fear' emotion offered the highest value and the minimum result in 'Disgust' compared to other algorithms.

Table 7 shows the recall results of existing CS, PSO, RNN and proposed MAFONN-EP. As per the results obtained, the 'Happy' emotion offered the highest in proposed method and the minimum result in 'Disgust' emotion compared to other algorithms.

Table 6
Precision.

Precision	CS	PSO	RNN	MAFONN-EP
Anger	35.71	21.42	14.28	70.14
Disgust	31.57	26.05	21.05	67.36
Fear	41.36	75	25	83.33
Sad	33.33	30	36.66	82
Happy	41.02	35.89	25.07	75.10

Table 7
Recall.

Recall	CS	PSO	RNN	MAFONN-EP
Anger	29.41	18.18	25	70
Disgust	25	20	33.3	66.8
Fear	61.53	69.23	19.375	80
Sad	27.02	30	27.58	78
Happy	40	43.78	56.25	82.13

Table 8
Overall results for existing and proposed technique.

Overall	CS	PSO	RNN	MAFONN-EP
Accuracy	77.70	80.18	75.90	97.33
Sensitivity	34.8022	39.30	31	99.32
Specificity	86.31	87.89	85.37	97.35
Precision	32.48	39.34	26.83	97.33
Recall	34.8022	39.30	30.56	97.33

Table 8 shows the overall results for existing and proposed technique. The proposed MAFONN-EP offers better results when compared to other existing CS, PSO and RNN techniques.

5. Conclusion and future work

This paper proposed a new MAFONN-EP technique for recognizing the emotion from the given video sequence. Here, the WMF technique is used for pre-processing the image frame and the EPBSFE technique is used for separating the background and foreground regions. Then, the texture patterns of the face key parts are extracted with the help of MAD and also the CS-PSO based optimization algorithm is used to select the optimal features from the extracted patterns. Finally, the WBPKC classifier produces the classified label with the recognized emotion. The major advantages of this paper are, it utilizes different and efficient image processing techniques for facial emotion recognition. During experimentation, various performance measures such as sensitivity, specificity, accuracy, precision, recall and recognition rate are used for evaluating the proposed system. Moreover, the betterment of the proposed technique is proved by comparing it with other facial emotion recognition techniques. From the results, it is evident that the MAFONN-EP technique provides better results compared to the existing techniques.

In future, this work can be enhanced by implementing highly efficient and accurate feature extraction and optimization techniques for improving the level of accuracy. Also, it can be implemented for a real time application oriented scenarios.

References

Azeem, A., Sharif, M., Raza, M., Murtaza, M., 2014. A survey: face recognition techniques under partial occlusion. Int. Arab J. Inf. Technol. 11 (1), 1–10.

Baltrusaitis, T., Robinson, P., Morency, L.P., 2016. Openface: an open source facial behavior analysis toolkit. In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on (pp. 1–10). IEEE. DOI: 10.1109/WACV.2016.7477553.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., Tapp, R., 2014. Featural processing in recognition of emotional facial expressions. Cognition Emotion 28 (3), 416–432. https://doi.org/10.1080/17521742.2010.571963.

Bobkowska, K., Przyborski, M., Skorupka, D., 2018. Emotion Recognition-the need for a complete analysis of the phenomenon of expression formation. In: E3S Web of Conferences (Vol. 26, p. 00013). EDP Sciences. https://doi.org/10.1051/e3sconf/20182600013.

Blazek, M., Kazmierczak, M., Janowski, A., Mokwa, K., Przyborski, M., Szulwic, J., 2014. An unorthodox view on the problem of tracking facial expressions. Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on (pp. 85-91). IEEE.

Chen, S., Jin, Q., 2015. Multi-modal dimensional emotion recognition using recurrent neural networks. Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (pp. 49-56). ACM.

Chao, W.L., Ding, J.J., Liu, J.Z., 2015. Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection. Signal Process. 117, 1–10. https://doi.org/10.1016/j.sigpro.2015.04.007.

Cruz, A.C., Bhanu, B., Thakoor, N.S., 2014. Vision and attention theory based sampling for continuous facial emotion recognition. IEEE Trans. Affective Comput. 5 (4), 418–431. https://doi.org/10.1109/TAFFC.2014.2316151.

Semantic audio-visual data fusion for automatic emotion recognition. Emotion Recognition: A Pattern Analysis Approach, pp.411–435.

Dzafic, I., Martin, A.K., Hocking, J., Mowry, B., Burianová, H., 2016. Dynamic emotion perception and prior expectancy. Neuropsychologia 86, 131–140. https://doi.org/10.1016/j.neuropsychologia.2016.04.025.

Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R., Pal, C., 2015. Recurrent neural networks for emotion recognition in video. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 467-474). ACM.

Elaiwat, S., Bennamoun, M., Boussaid, F., 2016. A spatio-temporal RBM-based model for facial expression recognition. Pattern Recogn. 49, 152–161. https://doi.org/10.1016/j.patcog.2015.07.006.

Farag, M.M.M., Elghazaly, T., Hefny, H.A., 2016. Face recognition system using HMM-PSO for feature selection. In: Computer Engineering Conference (ICENCO), 2016 12th International. IEEE, pp. 105–110. https://doi.org/10.1109/ICENCO.2016.7856453.

Guo, Y., Zhao, G., Pietikäinen, M., 2012. Dynamic facial expression recognition using longitudinal facial expression atlases. In: Computer Vision–ECCV 2012. Springer, Berlin, Heidelberg, pp. 631–644.

Guo, Y., Zhao, G., Pietikäinen, M., 2016. Dynamic facial expression recognition with atlas construction and sparse representation. IEEE Trans. Image Process. 25 (5), 1977–1992. https://doi.org/10.1109/TIP.2016.2537215.

Halder, A., Konar, A., Mandal, R., Chakraborty, A., Bhowmik, P., Pal, N.R., Nagar, A.K., 2013. General and interval type-2 fuzzy face-space approach to emotion recognition. IEEE Trans. Syst., Man, Cybern.: Systems 43 (3), 587–605.

Hosaka, T., Kobayashi, T., Otsu, N., 2011. Object detection using background subtraction and foreground motion estimation. IPSJ Trans. Comput. Vision Appl. 3, 9–20.

Kahou, S.E., Pal, C., Bouthillier, X., Froumenty, P., Gülçehre, C., Memisevic, R., Vincent, P., Courville, A., Bengio, Y., Ferrari, R.C., Mirza, M., 2013. Combining modality specific deep neural networks for emotion recognition in video. In: Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, pp. 543–550. https://doi.org/10.1145/2522848.2531745.

Kalita, J., Das, K., 2013. Recognition of facial expression using eigenvector based distributed features and euclidean distance based decision making technique. arXiv preprint arXiv:1303.0635. https://doi.org/10.14569/issn.2156-5570.

Katsimerou, C., Heynderickx, I., Redi, J.A., 2015. Predicting mood from punctual emotion annotations on videos. IEEE Trans. Affective Comput. 6 (2), 179–192.

Kaya, H., Gürpınar, F., Salah, A.A., 2017. Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image Vis. Comput. 65, 66–75.

Khadhraoui, T., Ktata, S., Benzarti, F., Amiri, H., 2016, March. Features selection based on modified PSO algorithm for 2D face recognition. In: Computer Graphics, Imaging and Visualization (CGiV), 2016 13th International Conference on (pp. 99-104). IEEE. DOI: 10.1109/CGiV.2016.28.

Khan, R.A., Meyer, A., Konik, H., Bouakaz, S., 2013. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recogn. Lett. 34 (10), 1159–1168. https://doi.org/10.1016/j.physletb.2010.09.059.

Khan, S.A., Hussain, A., Usman, M., 2016. Facial expression recognition on real world face images using intelligent techniques: a survey. Optik-Int. J. Light Electron Optics 127 (15), 6195–6203. https://doi.org/10.1016/j.ijleo.2016.04.015.

Li, W., Abtahi, F., Zhu, Z., 2015, November. A deep feature based multi-kernel learning approach for video emotion recognition. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 483-490). ACM, DOI: http://dx.doi.org/10.1145/2818346.2830583.

Li, Y., Wang, S., Zhao, Y., Ji, Q., 2013. Simultaneous facial feature tracking and facial expression recognition Digital Object Identifier IEEE Trans. Image Proc. 22 (7), 2559–2573. https://doi.org/10.1109/TIP.2013.2253477.

Lin, P.H., Chen, B.H., Cheng, F.C., Huang, S.C., 2016. A morphological mean filter for impulse noise removal. J. Disp. Technol. 12 (4), 344–350. DigitalObjectIdentifier10.1109/JDT.2015.2487559.

Majumder, A., Behera, L., Subramanian, V.K., 2014. Emotion recognition from geometric facial features using self-organizing map. Pattern Recogn. 47 (3), 1282–1293. doi.org/10.1016/j.patcog.2013.10.010.

Mlakar, U., Fister, I., Brest, J., Potočnik, B., 2017. Multi-objective differential evolution for feature selection in facial expression recognition systems. Expert Syst. Appl. 89, 129–137. https://doi.org/10.1016/j.eswa.2017.07.037.

MMI Facial Expression Database.2016.Available: https://mmifacedb.eu/.

Nurhadiyatna, A., Jatmiko, W., Hardjono, B., Wibisono, A., Sina, I., Mursanto, P., 2013, October. Background subtraction using gaussian mixture model enhanced by hole filling algorithm (gmmhf). In: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on (pp. 4006-4011). IEEE, DOI: 10.1109/SMC.2013.684.

Owusu, E., Zhan, Y., Mao, Q.R., 2014. A neural-AdaBoost based facial expression recognition system. Expert Syst. Appl. 41 (7), 3383–3390. https://doi.org/10.1016/j.eswa.2013.11.041.

Patwardhan, A., Knapp, G., 2016. Augmenting Supervised Emotion Recognition with Rule-Based Decision Model. arXiv preprint arXiv:1607.02660.

Raghuvanshi, A., Choksi, V., 2016. Facial Expression Recognition with. Convolutional Neural Networks.

Redmond, M., Salesi, S., Cosma, G., 2017. A novel approach based on an extended cuckoo search algorithm for the classification of tweets which contain Emoticon and Emoji. In: Knowledge Engineering and Applications (ICKEA), 2017 2nd International Conference on (pp. 13–19). D10.1109/ICKEA.2017.8169894.

Ringeval, F., Eyben, F., Kroupi, E., Yuce, A., Thiran, J.P., Ebrahimi, T., Lalanne, D., Schuller, B., 2015. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recogn. Lett. 66, 22–30.

Rivera, A.R., Castillo, J.R., Chae, O.O., 2013. Local directional number pattern for face analysis: Face and expression recognition. IEEE Trans. Image Process. 22 (5), 1740–1752. https://doi.org/10.1109/TIP.2012.2235848.

Sanchez-Mendoza, D., Masip, D., Lapedriza, A., 2015. Emotion recognition from mid-level features. Pattern Recognit. Lett. 67, 66–74. https://doi.org/10.1016/j.patrec.2015.06.007.

Treece, G., 2016. The bitonic filter: linear filtering in an edge-preserving morphological framework. IEEE Trans. Image Process. 25 (11), 5199–5211.

Torkhani, G., Ladgham, A., Sakly, A., 2017, December. 3D Gabor-Edge filters applied to face depth images. In: Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2017 18th International Conference on (pp. 578–582). IEEE. DOI: 10.1109/STA.2017.8314888.

Wei, Z., Li, P., Yue, H., 2015, August. A Foreground-Background Segmentation Algorithm for Video Sequences. In: Distributed Computing and Applications for Business Engineering and Science (DCABES), 2015 14th International Symposium on (pp. 340-343). IEEE.

Yang, S., Bhanu, B., 2012. Understanding discrete facial expressions in video using an emotion avatar image Digital Object Identifier IEEE Trans. Syst., Man, Cybern., Part B (Cybernetics) 42 (4), 980–992. https://doi.org/10.1109/TSMCB.2012.2192269.

Yin, Z., Zhao, M., Wang, Y., Yang, J., Zhang, J., 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Comput. Methods Programs Biomed. 140, 93–110. https://doi.org/10.1016/j.cmpb.2016.12.005.

Yu, Z., Zhang, C., 2015. Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 435-442). ACM, http://dx.doi.org/10.1145/2818346.2830595.