IET Image Processing

Review Article

Facial expression recognition techniques: a comprehensive survey

ISSN 1751-9659
Received on 17th December 2018
Revised 23rd February 2019
Accepted on 18th March 2019
E-First on 7th May 2019
doi: 10.1049/iet-ipr.2018.6647
www.ietdl.org

Saranya Rajan[1] ✉, Poongodi Chenniappan[1], Somasundaram Devaraj[2], Nirmala Madian[2]

[1]*ECE, Bannari Amman Institute of Technology, Erode, TN, India*

[2]*ECE, Sri Shakti Institute of Engineering and Technology, Coimbatore, TN, India*

✉ *E-mail: shreesaranya1987@gmail.com*

Abstract: Over the past decades, facial expression recognition (FER) has become an interesting research area and achieved substantial progress in computer vision. FER is to detect human emotional state related to biometric traits. Developing a machine based human FER system is a quite challenging task. Various FER systems are developed by analysing facial muscle motion and skin deformation based algorithms. In conventional FER system, the developed algorithms work on the constrained database. In the unconstrained environment, the efficacy of existing algorithms is limited due to certain issues during image acquisition. This study presents a detailed study on FER techniques, classifiers and datasets used for analysing the efficacy of the recognition techniques. Moreover, this survey will assist researchers in understanding the strategies and innovative methods that address the issues in a real-time application. Finally, the review presents the challenges encountered by FER system along with the future direction.

1 Introduction

Facial expression recognition (FER) has a high impact in the field of pattern recognition, and a substantial effort is made by researchers to develop an FER system for human–computer interaction applications. The facial expression provides sensitive information cues to build an FER system and considered as the best tool for recognising human emotions and intentions easily. In 1971, Ekman and Friesen [1] defined six distinct expressions (happy, sad, anger, surprise, fear, and disgust) as the basic emotions and each emotion is associated with a unique facial expression which readily recognised across different cultures. The psychologist Mehrabian [2] proposed a study on information communication between humans. The study reveals 55% of information is conveyed by facial expression, 38% by supporting language like sound, speech and so on, and only 7% by oral language.

Currently, an FER system plays a central part of artificial intelligence and serves as a potential real-world applications in different areas for psychological studies [3], driver fatigue monitoring, interactive game design, portable mobile application to automatically insert emotions in chat and assistance systems for autistic people, facial nerve grading in medical field [4], emotion detection system used by disabled to assist a caretaker, socially intelligent robot with emotional intelligence [5].

Most of the research work in FER system follows the framework of pattern recognition [6]. It consists of three phases: face detection, facial feature extraction, expression classification. It is quite substantial and noteworthy to research these phases. In this current survey, various phases of facial expression analysis are discussed with distinct algorithms to classify six basic expressions. Face detection is performed by algorithms such as Haar classifier, adaptive skin colour algorithm and so on. Gabor feature, local binary patterns (LBPs), active appearance model, principal component analysis and other algorithms exploited for feature extraction. The classifiers used for expression classification are support vector machine (SVM), neural networks, nearest neighbour and so on.

The essential step of FER is face detection. The efficiency of the classifier will be better with effective facial feature extraction, and this is achieved only by a proper face detection method. Viola and Jones [7] suggested an AdaBoost classifier that extracts and classifies the features quickly and accurately. Danti and Kadiyavar [8] represented a face detection method based on skin colour irrespective of face orientation and background. Various aspects of distinct algorithms and methods are published by researchers to identify a face from static and dynamic image frames. Followed by face detection, many researchers developed various feature extraction techniques to analyse facial expression. Several attempts were carried out to extract detailed parametric facial feature vectors in the frontal view of both still and an image sequence. Tian *et al.* [9] developed an automated face analysis system to recognise the subtle changes based on the facial action coding system (FACS). Automated action coding and facial expression detection is still a challenging problem. Chu *et al.* [10] proposed a selective transfer machine to personalise a generic classifier to overcome the challenges in video sequence frames like illumination and complex background. Despite action coding, they proposed local and global descriptors for detecting distinct facial expressions. Further, its generalisability was tested using valuable databases.

Extensive research helped to develop a better FER system in recent years, but the performance of the system is affected by various factors. As far as the FER systems concern, the existing methods handle only prototypic posed facial expressions which are captured under laboratory constraints [11]. In the case of the unconstrained environment, the use of an existing method often leads to a higher probability of misclassification due to spontaneous expressions. The recognition of spontaneous expression in real-time is a challenging issue related to changes in illumination; head pose variation, subtle facial deformations, aging, occlusion of any objects like hair, glass or scarf, skin colour variations and complex background. Most of the trained images are also misclassified by a better performing classifier owing to the challenges mentioned above. Correspondingly, the available benchmark databases are not naturally linked to an emotional state of the test image. For such reasons, Sebe *et al.* [12] created an authentic emotion database and developed a real-time automatic FER system.

This survey concentrates on the FER system in a controlled and uncontrolled environment based on their performance traits. We have discussed the current approaches to face detection and feature extraction techniques for FER and also presented the real-time applications. The survey presents state-of-the-art methods for face detection and facial feature extraction and expression classification. The paper also presents various techniques that solve

Table 1 Summary of face detection

S. no	Method/ algorithm	Accuracy	Comments
1	eigenspace method	high accuracy for face detection under variable pose conditions	head motion allowed in horizontal direction makes the system robust
2	adaptive skin colour	accuracy is good as it identifies the skin colour easily but fails due to illumination	adaptive gamma correction method is used to overcome the illumination problem
3	Haar classifier	high accuracy obtained by Haar features	computational complexity is less due to minimum features
4	Adaboost classifier	high accuracy because of the strong classifier and detects a single face	uses trained model so reduced computational cost
5	contours	accuracy is good as it uses contour points	due to minimum features, the computational cost is less.

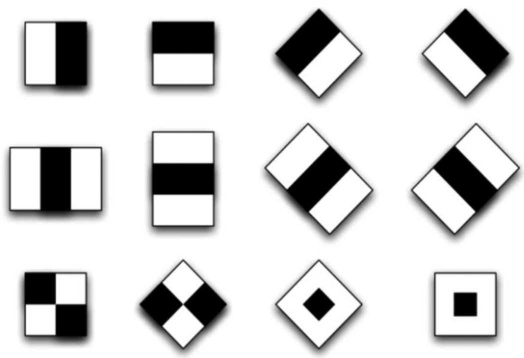

Fig. 1 *Haar features*

the issues about the FER system. The study mainly focuses on feature extraction techniques. The paper is organised as follows: Section 2 describes the face detection methods. Section 3 provides a detailed review of facial feature extraction techniques, various classifiers and frequently used datasets are discussed in Section 4. Section 5 deliberates the challenges related to FER system. Finally, in Section 6, we conclude the paper with a promising future direction.

2 Face detection methods

Face detection is a significant phase of FER. A proficient automated system can be developed for recognising the face region in static or video image spontaneously. A face region is detected in the image sequence using facial features such as edge, skin colour, texture and face muscle motion. These features easily distinguish a face region from the background. In this phase, the input image is segmented into two parts: one is the face region and other representing a non-face region. There are many face detection methods available like eigenspace method, adaptive skin colour and Viola–Jones method and their algorithms are developed based on Haar classifier, Adaboost and contour points. The survey is about the accurate identification of faces and its performance under a constrained and unconstrained environment. Table 1 gives an outline of face detection methods.

2.1 Eigenspace method

Pentland *et al.* [13] described the eigenspace technique to locate face under variable pose. Also, modular eigenspace descriptors are used to recognise the face image with salient features. Later, Essa and Pentland [14] utilised the eigenspace method to locate the face in any random image sequence; 128 image samples were obtained using principal component analysis (PCA). The eigenfaces define the subspace of sample images called 'face space' [15]. To detect the existence of the face in a single image, the distance between the observed image and face space was estimated using the projection coefficients and the signal energy. Correspondingly, the spatio-temporal filtering method was employed to detect the face in an image sequence. The thresholding concept was applied on the filtered image that causes binary motion which benefits the analyse of 'motion blobs' over a period time. Each motion blobs represent a human head for detecting the face location.

Pentland *et al.* [13, 16] proposed a real-time approach that was successfully tested on a database containing 7562 images of both sexes having occluded objects on the face like hair, spectacles and so on.

2.2 Adaptive skin colour method

Skin colour is an effective feature to detect face [17, 18]. Based on colour dependency, any one of the colour systems is preferred. The common colour systems are RGB, CMY, YIQ, YUV, YCbCr. Mostly, YIQ and YUV colour systems are used, where the brightness effect is removed during the procession [19], for colour intelligence. In the YIQ colour model, the component I refers to hue and Q refers to saturation and the formula to calculate I is given below:

$$I = 0.596 \times R - 0.274 \times G - 0.322 \times B$$

where I is the value of face skin colour in YIQ space that changes in a particular range between 30 and 100. Simultaneously, in YUV space, the range of face skin colour's, hue is between 105° and 150°. To establish a primary face skin-colour model, YIQ and YUV colour systems are synthesised. If an image satisfies the following conditions consecutively, $30 \le I \le 100, 105° \le \theta \le 150°$ then it is a skin colour. Most of the research adopt skin colour for face detection based on a fixed threshold scheme which causes large errors due to illumination and pose variation. An iterative thresholding algorithm [20] is proposed to acquire actual face region that satisfies the face geometric pattern. However, it is not suitable for real-time applications due to its high computational cost. An adaptive skin colour filter is introduced [21] that adaptively adjust the threshold values and use a linear discriminant function to separate the skin region from a complex background. Gamma corrective method is used to influence illumination and pose variation. Zhao-yi *et al.* [19] proposed an adaptive skin colour and structure model for multi-pose colour images in a complex background which highly improves the accuracy and effectually discards the impact of illumination level.

2.3 Haar classifier method

Haar classifier is considered as a robust face detection method in a real-time environment [6]. Haar features are considered to detect face edges, lines, motions, and skin colour. The Haar features are a black and white connected rectangular box as shown in Fig. 1, used for feature extraction. Haar features can be easily scaled, and the positions are examined by increasing or decreasing the pixel intensities at different parts of an image. The located feature value is the difference of the sum of pixels of black and white regions inside the rectangle box [22]. Haar classifier detects the features which contribute face detection problems in the training phase. Thus, reduces the computational cost and complexity in the texting phase that leads to high detection accuracy.

2.4 AdaBoost method

The AdaBoost is an ensemble approach for face detection [23, 24] and is highly used due to its improved accuracy and relatively low complexity of computation. It is a popular face detection method with a low false positive rate. The major limitation in Adaboost is sensitivity to noisy data and outliers [23]. A set of image features are trained with several classifiers in cascade using Adaboost to

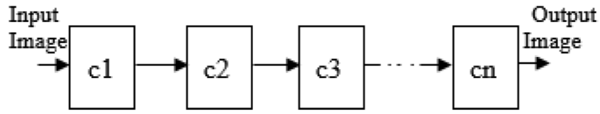

Fig. 2 *Structure of cascade classifier*

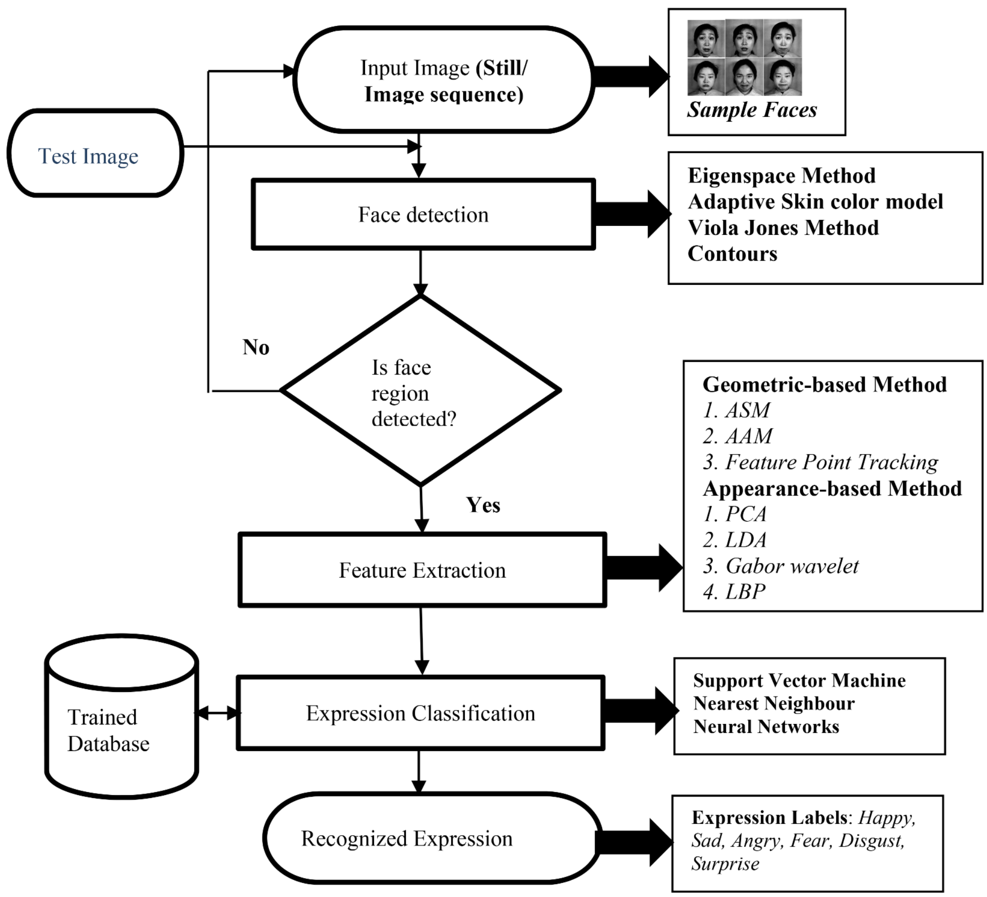

Fig. 3 *Flowchart for FER*

eliminate the negative samples. In the cascade structure as shown in Fig. 2, the output of the first classifier will be the input to the next classifier which is used to get accurate face region. Therefore, a strong classifier was built that helps in reducing the number of features and thus leads to high accuracy in detection. Kheirkhah and Tabatabaie proposed a hybrid and robust face detection system for colour and complex images [24]. The hybrid approach uses skin colour information and Adaboost-based face detection. It gives better performance in accuracy with minimum execution time.

2.5 Contours

Face detection based on contour points leads to better accuracy [25, 26]. In an image sequence, the first pixel of the first frame is scanned based on skin colour, and that point is considered as a first contour point of the head. In the same way, the remaining contour points are computed in a frame. The pixel considered is called seed point. The direction of the contour point is initialised with the identified seed point, and detection path can be clockwise or anti-clockwise. Face motion is identified, when there is a shift in two successive frames in a sequence and experience a shift in contour points beyond a threshold [25]. Aniruddha *et al.* used a contour-based procedure to detect and track human face from video frames. Logical operation and Gaussian filters are used to get proper face contour. The scalar and vector distance of a rectangular window drawn from four corner points of two consecutive frames are calculated to detect and track the face from an image sequence [26].

3 Feature extraction techniques

After face detection, the next step in FER is feature extraction. The main aim of facial feature extraction is to extract an effective and efficient representation of facial components without any loss of face information. Geometric-based and appearance-based features are the two feature extraction techniques classified based on facial motion and deformation of facial features. The input image may be either a static image or image sequence. Based on the input image, a suitable facial feature extraction algorithm is applied to extract either the local or global or hybrid features. The extracted features are considerably reduced in size, which is given as input to the classifier and that significantly helps the classifier to make the decision easier in identifying and recognising the facial expression. Fig. 3 represents the FER process.

In this section, we discuss on generalised view of facial feature extraction methods and have an extensive review on recent feature extraction techniques in FER.

3.1 Geometric-based method

Geometric-based algorithms focus on permanent features (eyes, eyebrows, forehead, nose and mouth) which describe the shape and location of facial components using predefined geometric landmark position. These facial components are extracted to form a feature vector that represents the face geometry. However, the expressions affect the relative shapes and positions of various face features. Consequently, underlying facial expressions can be identified by measuring the displacement of significant facial components. In the

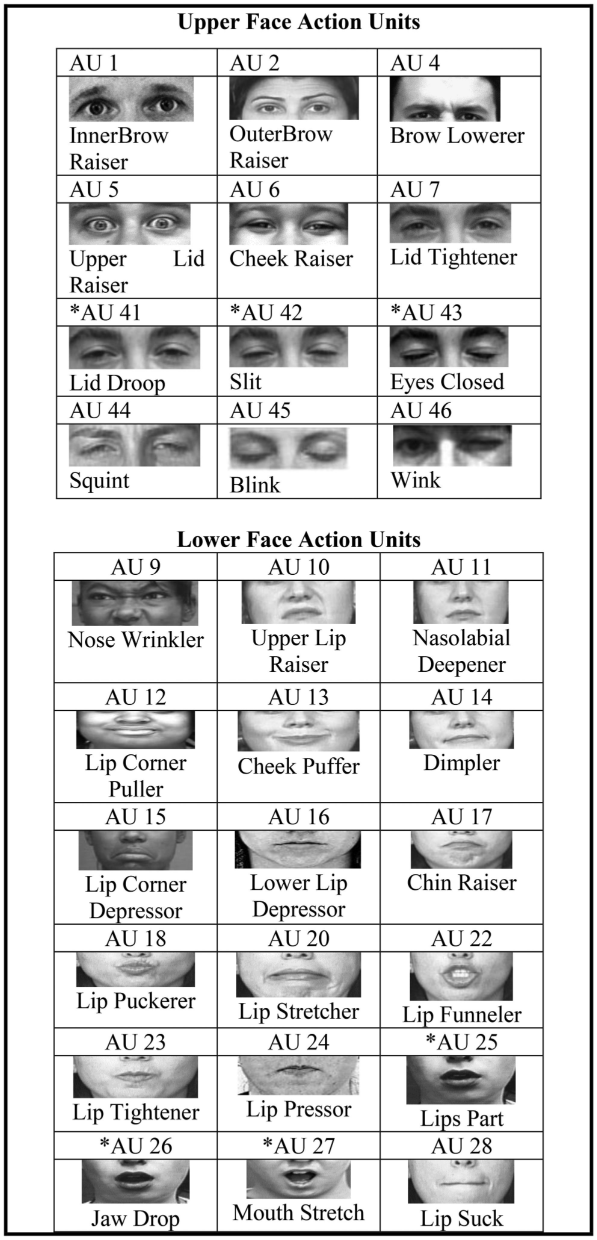

Fig. 4 *FACS AUs by Ekman and Friesen. 'Asterisk' indicates AU 25, 26 and 27 are now coded according to criteria of intensity (25 A–E) and also AU 41, 42 and 43 are now coded according to criteria of intensity [7]*

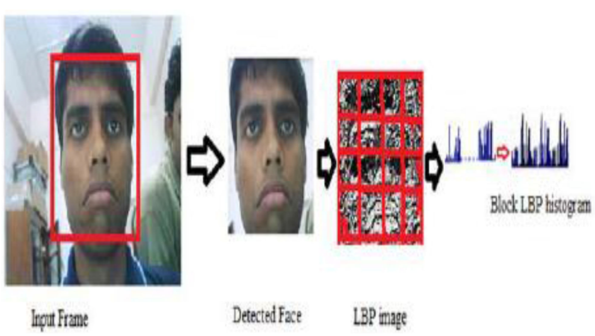

Fig. 5 *Calculation of block LBP histogram [22]*

case of an image sequence as input, FACS [27] is used that helps in differentiating facial movements based on the analysis of facial actions. FACS contains various action units (AUs) related to specific muscle contractions. Tain *et al.* developed an automated face analysis system to analyse the subtle changes in facial expressions which are further converted to AUs [7]. The expressions may contain single AU or combination of AU. Some basic upper and lower face AUs are shown in Fig. 4. On the other hand, static image input use model-based approaches, such as active shape model (ASM) [28], active appearance model (AAM) [29] and scale invariant feature transform (SIFT) [30, 31] algorithm to extract facial features. The geometric-based method is more suitable for real-time face images where features can be identified and tracked easily, but it requires an accurate face detection technique.

3.2 Appearance-based method

Appearance-based algorithms focus on transient features (wrinkles, bulges, forefront) which describe the changes in face texture, intensity, histograms and pixel values. In this method, PCA, linear discriminant analysis (LDA), independent component analysis (ICA), Gabor wavelet, LBP are the algorithms considered to extract feature descriptors. In recent years, Gabor wavelet and LBP are extensively used to extract feature descriptors. Gabor wavelets are a well-known representative feature for extracting texture information effectively. Zhang *et al.* [32] investigated and compared geometry-based and Gabor-based method and the result shows that Gabor wavelet outperforms in performance and considered as a more powerful tool for feature extraction. Many research outcomes are favoured using a Gabor filter bank to detect lines and edges over multiple scales and orientations [33], it has a good time–frequency localisation and multi-resolution characteristics [34]. The limitation of the filter is high computational time due to the large size of filtered vectors [35]. LBP [36, 37] is a non-parametric descriptor whose aim is to efficiently detect the local structure of images. Due to low computational cost and high invariance property, the LBP feature extraction is widely used for feature extraction. In LBP [16, 22], the image is divided into sub-blocks, and the histograms are calculated for each block. Later, the histograms of each block are concatenated to obtain global features. Fig. 5 explains the LBP histogram technique.

Further, a brief survey is done on recent feature extraction techniques for FER system as shown in Table 2.

Duong *et al.* [38] constructed a dimensionality reduction method by an unsupervised learning framework called projective complex matrix factorisation (proCMF). The method is similar to proNMF and cosine dissimilarity metric where it transforms real data into a complex domain. A projective matrix was found by solving an unconstrained complex problem, and cost function was reduced using a gradient decent optimisation technique. The proposed method performs well compared to proNMF and is more robust to extract discriminant facial features and is potentially superior to FER, even under noise and outliers. It shows better performance than other baseline methods and achieves the recognition rate of about 97.51%.

Fatima *et al.* [39] suggested an approach to enhance the accuracy of emotion recognition from facial expression based on fiducial points. Viola–Jones algorithm was used and 49 fiducial points were tracked using SDM [40], the obtained point represents face parts like eyebrows, eyes, nose and mouth. The proposed method calculates Euclidean distance between each pair of points. By the calculated distance ratio of first and last frames, the dynamic features were obtained. Most relevant features were selected using CfsSubsetEval evaluator and used neural network classifier for expression recognition. Hence achieved an accuracy of 99% on Cohn–Kanade (CK+), 84.7% on Oulu-CASIA VIS and 93.8% on Japanese female facial expression (JAFFE) databases, respectively.

Michael Revina and Sam Emmanuel [41] proposed EMDBUTMF for reducing the noisy pixel in an image. The method was robust to eliminate salt and pepper noise. After noise removal, the feature vectors were obtained LDN (local directional number) pattern and DGLTP (directional gradient local ternary pattern). The DGLTP calculates the directional pattern of its neighbour and quantises into the three-bias level to encode the

Table 2 Review of recent techniques under the FER system

Author/year	Methodology	Facial features	Classifier	Dataset	Accuracy	Advantage/ disadvantage
Duong/2018	projective complex matrix factorisation under unsupervised learning	local features	nearest neighbour	CK+ and JAFFE	97.51 and 82.10%	this method applied on positive and negative data
Fatima/2018	supervised decent method based on Euclidean distance of fiducial points	eyes, eyebrows, nose and mouth	neural network	CK+ Oulu-CASIA and JAFFE	99, 84.7 and 93.8%	achieves higher recognition rate in real-time
Revina/2018	enhanced modified decision based unsymmetric trimmed median filter, local directional number pattern, dominant gradient local ternary pattern	dots, edges and local features	SVM	JAFFE and CK	88.63%	robust against noisy faces than illumination
Ding/2017	logarithm Laplace-double local binary pattern and Taylor feature pattern	global and local features	nearest neighbour	JAFFE and CK	93.0 and 91.4%	satisfied recognition result under an uncontrolled environment
Arshid/2017	MSBP	eyebrows, eyes, mouth, bulges and wrinkles	simple logistic classifier	wild dataset	96 and 60%	resolves issues of illumination
Munir/2017	fast Fourier transform and contrast limited adaptive histogram equalisation and merged binary pattern code	eyebrows, eyes, mouth, bulges and wrinkles	SMO, KNN, simple logistic MLP	SFEW	96.2% holistic 65.7% division based	suits for poor illumination
Hasani/2017	modified inception-ResNet layers and conditional fields for sequence labelling	spatial relation and temporal relation of labels	deep neural network	CK+, MMI and FERA	93.04, 78.68 and 66.66%	improves the recognition rate
Khadija/2017	IntraFace (IF) facial decomposition method	global feature with seven ROIs	multiclass SVM	CK and FEED	94.1 and 87.5%	highest recognition rate
Holder/2017	improved gradient local ternary pattern	eyes, nose and mouth	SVM	CK+ and JAFFE	97.6 and 86.8%	robust against varying illumination and random noise
Qayyum/2017	stationary wavelet transform	face image is decomposed into subbands	feedforward neural network	JAFFE and CK+	98.83 and 96.61%	HCI and Kinect based applications
Du/2017	LBPs and supervised descent method	global and local features	M-CRT	JAFFE and CK+	89.45 and 90.72%	improves to classify expression
Liu/2017	LBP and HoG features with gamma correction	salient features	linear SVM	CK+ and JAFFE	96.6 and 63.4%	avoids overfitting and low noise impacts
Kumar/2016	weighted-projection based LBP	discriminative features	SVM	MUG, JAFFE and CK+	98.44, 98.51 and 97.50%	discriminative information improves recognition
Majumder/2016	1. Geometric features extraction 2. Regional LBP 3. Fusion of 1 and 2	local features	SOM-based classifier	MMI and CK+	97.55 and 98.95%	more efficient and accurate
Kamarol/2016	STTM	eyes and mouth	SVM	CK+, CASME II and AFEW	95.37, 98.56 and 84.52%	captures subtle motions
Tang/2016	DGFN	local and global features	ANN	Oulu-CASIA and CK+	DGFN-78 and 93.81%	achieves a higher recognition rate
	DFSN				DFSN-86.88 and 98.10%	
	fusion of DGFN and DFSN (DFSN-I)				DFSN-I 87.50 and 98.73%	
Hsieh/2015	directional gradient operators like Gabor filters and Laplacian of Gaussian	frown, nose wrinkle, two nasolabial folds, two eyebrows and mouth	SVM	CK+ and online members	94.7 and 93%	effectively represents the change in expression
Zhang/2015	PHRNN and MSCNN	eyes, nose and mouth	neural network	CK+, Oulu-CASIA and MMI	98.50, 86.25 and 81.18%	robust against dynamic image
Kumbhar/2012	Gabor filter and PCA	local and global features	feedforward neural networks	JAFFE	70%	proves to achieves FER in practical application

local texture. The subsequent patterns are exploited as facial feature descriptors. Further, SVM supervised machine learning classifier was used to map the labelled trained data into a higher-dimensional feature space with an optimal hyperplane for expression classification. The proposed approach achieved 88% of accuracy with CK and JAFEE databases.

Ding *et al.* [42] proposed a method to detect and classify facial expressions from the video. 24-dimensional DLBP was proposed to detect the peak frame from the image sequence which extracts efficient facial features. Further, Taylor's theorem was used to expand the peak frame feature pixels to extract the discriminative features. To overcome the illumination variation in real-time application, the logarithm-Laplacian was proposed. The proposed method TFP outperformed the existing LBP-based feature extraction methods and suitable for real-time applications and were tested on JAFFE and CK datasets.

Arshid *et al.* [43] proposed a multi-stage binary pattern (MSBP) feature extraction technique to handle the illumination in the real-world scenario by using sign and gradient difference. Holistic and division based methods were applied on existing methods and also compared with the proposed technique in a wild dataset by using different classifiers such as BF tree, Bagging, Naïve Bayes, Simple logistic and KNN. The result shows that the MSBP method achieved 96% of accuracy rate in the holistic approach and 60% of accuracy using division based approach.

Munir *et al.* [44] proposed a merged binary pattern coding to extract local facial features using sign and gradient difference. The method was robust against illumination and pose variation. Before extraction, certain pre-processing steps were done using FFT plus CLAHE and histogram equalisation. The performance of classifier was improved using PCA, a feature extraction method. The real-world image dataset was used and tested with the existing and proposed method. The result shows that the proposed method outperforms all with an accuracy of 96.5%. From the analysis, holistic approach performed better than division based approach.

Hasani and Mahoor [45] introduced a new framework of deep neural network cascaded with a conditional random field model to increase the recognition accuracy rate in an image sequence. Modified Inception-ResNet modules were proposed to extract the spatial relationships within an image, whereas the temporal relations were extracted between the successive frames and labelled by linear-chain conditional random field. The proposed method was evaluated using CK+, MMI and FERA databases and mainly dealt with subject-independent and cross-database validation cases.

Mahmud and Al Mamun [46] suggested FER based on an extreme learning machine. Face regions were detected using Viola–Jones method. Feature vectors were found using morphological operations and edge detection techniques. Finally, the feature vectors were given as an input to a feedforward neural network classifier to perform expression classification. The proposed methodology was tested on a publically available JAFFE database and acquired satisfied accuracy.

Khadija *et al.* [47] proposed a novel facial decomposition for expression recognition. IntraFace algorithm was used to detect seven regions of the face called ROI with the aid of facial landmarks. The feature vectors are extracted using different local descriptors like LBP, CLBP, LTP and Dynamic LTP. Then extracted feature vectors are fed as input to SVM classifier and tested on CK and FEED databases.

Holder and Tapamo [48] suggested gradient local ternary pattern [49] to overcome illumination and noise in a real-world scenario. In this method, they have used enhanced pre-processing techniques, Scharr gradient operator, PAC for feature selection to improve the performance and tested on CK+ and JAFFE databases. The result of the proposed method provides a promising efficiency for expression recognition.

Qayyum *et al.* [50] suggested stationary wavelet transform (SWT) extract the facial features in spectral and spatial domains. Each subband of SWT contains different image information and most of the information was retained in LL subband. The extracted features retain the original image size and the DCT feature reduction was employed. Further, the reduced feature vectors were allowed as input to feedforward neural network classifier and trained with the backpropagation algorithm. CK+ and JAFFE datasets were engaged in testing the accuracy and efficiency of this technique and the results were compared with existing methods.

Du and Hu [51] proposed a modified classification and regression tree (M-CRT) to identify the unpredictable changes in expression. The proposed method was modelled with a regressive segmental threshold which minimises both intra-class impurity and inter-class difference to attain better feature representation and classification. JAFFE and CK+ datasets are used to test the performance of the proposed method and also compared with classical recognition methods. The result shows the significant improvement in the recognition rate of the proposed method.

Liu *et al.* [52] suggested a simple framework algorithm to extract facial features from the salient face area. The proposed algorithm normalises the salient features to the same size and extracts similar face features of different subjects. Later, the algorithm compares the different subject features with neutral face features. LBP and gamma correction method was also employed to obtain the finest recognition rate, respectively. The extracted features from LBP and histogram of orientation gradients (HOG) were fused and applied PCA for dimensionality reduction. The reduced feature vectors were trained using different classifiers and tested with CK+ and JAFFE databases. The results of the proposed framework attained the performance as the state-of-the-art methods.

Kumar *et al.* [53] proposed an Informative extraction region model to extract expedient regions of a face based on a neutral image. In the absence of a neutral image, a common reference image was proposed based on the Procrustes analysis. Later, the discriminative features of LBP [54] were enhanced to minimise the wrong classification. The proposed method was based on projection analysis and LBP features, and hence it was called as weighted-projection based LBP. The experimental result of the proposed method was analysed using various datasets that significantly provides better recognition rate than other methods.

Majumder *et al.* [55] proposed deep network-based automatic FER system which was an ensemble of Geometric and LBP based feature extraction, data fusion using autoencoders and SOM-based classifier. The proposed method was validated on two popular datasets namely CK+ and MMI. The trained datasets were allowed to SVM classifier and proposed SOM-based classifier. The result of the proposed fused features method was compared with individual and concatenated features which outperform with a high recognition rate.

Kamarol *et al.* [56] presented a novel appearance-based feature extraction method for video input called spatiotemporal texture map (STTM) which generates a 3D texture map. The competency to capture the subtle variations of facial expressions in both spatial and temporal domains was better with the proposed method. The extracted spatiotemporal features were classified with SVM classifier. The proposed method was evaluated on real-world datasets containing posed and spontaneous micro-expressions with low computational cost. The observed result shows a better performance than the existing methods, respectively.

Tang *et al.* [57] proposed three video-based models for FER. A differential geometric fusion network (DGFN) model was hand-crafted in which various combinations of geometric features with the outline of action units characteristics, both local and global features are taken into account. With the advent of DGFN, deep facial-sequential network (DFSN) model was designed as an auto-extraction method based on multi-dimensional convolutional neural network. Finally, the combination of DGFN and DFSN models called DFSN-1 model was proposed to acquire better performance. The result shows that DFSN and DFSN-1 models performed better on Oulu-CASIA dataset and achieved an average performance on CK+ and MMI datasets.

Hsieh *et al.* [58] focused on semantic facial features which were acquired using directional gradient operators like Gabor filter and Laplacian of Gaussian. The feature extraction was based on geometric features in which facial action units and ROI were adopted. The changes in facial components were measured using Euclidean distance. A multi-class support vector machine classifier

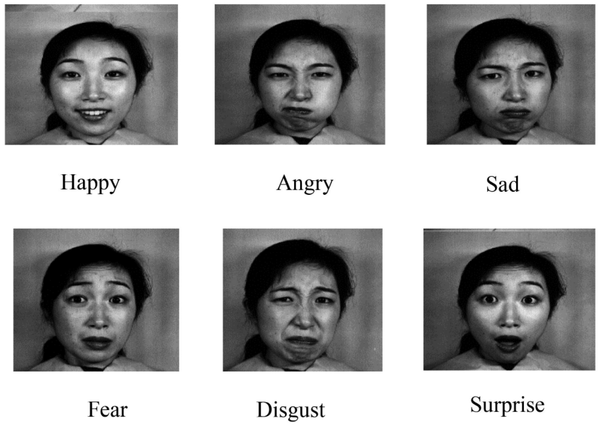

Fig. 6 *Sample facial expression from JAFFE dataset*

was used to classify the facial expressions. The method was tested on the CK+ database, and few were tested on an online dataset. The experimental results demonstrate that the proposed facial feature could represent subtle changes in expressions and reduced time consumption.

Zhang *et al.* [59] presented PHRNN to handle the subtle variations in facial expressions from video frames. A Part-based Hierarchical Bidirectional Recurrent Neural Network (PHRNN) was used to extract temporal features based on facial landmarks from subsequent frames. Similarly, the multi-signal convolutional neural network (MSCNN) was proposed to extract spatial features from static images. Both spatial and temporal features increase the difference between the subtle and identical expressions. The deep evolutional spatial-temporal network significantly enhances the performance of FER. The proposed method was tested on widely used databases (CK+, Oulu-CASIA and MMI) and the observed result proved to minimise the recognition error.

Kumbhar *et al.* [60] discussed the application of 2D Gabor function to extract the salient features, where the high-dimensional feature vectors are reduced to lower-dimensional vectors using PCA. The extracted feature vectors were given as input to a feedforward neural network to classify facial expression. JAFFE database was considered, and recognition rate of 60–70% was achieved.

4 Frequently used classifiers and datasets

4.1 Classifiers

Classification is a supervised learning model to predict the output based on the observed values. In pattern recognition, the classifier plays a vital part either to draw a conclusion or to predict the class labels of unseen data based on trained images. Many classification algorithms were developed to achieve the highest recognition rate. Indeed, all the approaches are not suitable for analysing different facial feature extraction techniques. A dataset is required with different spontaneous expressions to develop an efficient real-time FER system. Different expressions considered are labelled as happy, angry, sad, fear, disgust and surprise that are shown in Fig. 6. The labelled images are trained and given to the classifier. In classification, the test and trained images are compared to predict the expected output. The classifiers or classification approaches used in FER systems are SVM, neural network, Adaboost, eigenfaces, logistic regression, decision tree, nearest neighbour, hidden Markov models and so on.

4.1.1 Support vector machine: SVM [41, 47, 48, 52, 53, 56, 58] is a well-known supervised classification algorithm and also called a maximum margin classifier [61]. The hyperplane is selected with higher margin to discriminate two classes in n-dimensional space. Facial expression classification [47] employs a linear SVM algorithm based on one-against-one approach [41, 56] uses radial bias function (RBF) kernel function based on one-against-rest approach. Both the approaches distinguish the samples from one class to rest of available class labels with maximum margin. Moreover, there is no significant difference from one approach to another. Kernels are the one who decides the performance of classifiers and are of four types namely, linear, polynomial, RBF and sigmoid. RBF kernel is mostly used due to its superiority in classification performance.

4.1.2 Nearest neighbour: Nearest neighbour is the simplest classifier with non-parametric algorithm [62]. The decision of classifier was based on the entire training set. If the training set data is small, there exists a trade-off in accuracy, and maximum classification error occurs. Ding *et al.* [42] utilise nearest neighbour classifier with different distance measures to evaluate the similarity between the training and testing samples. In the same way, Doung *et al.* [38] use nearest neighbour classifier to decide the projected expressions. Here, a Euclidean distance was computed for each training images. Thus, the projected samples were assigned to the class that appears closest to the training image.

4.1.3 Neural network: The neural network has different types based on its characteristics and performance. In FER, feedforward neural network and convolutional neural networks are most frequently used. Fatima *et al.* [39] proposed an emotion recognition system that uses a single network with one hidden layer neural network classifier to classify facial expressions with selected dynamic features and achieved 99% accuracy in the CK+ facial expression dataset. The network was trained on multi-class recognition task with backpropagation algorithm and sigmoid activation function. Correspondingly, Qayyum *et al.* utilised feedforward neural network that was trained using backpropagation algorithm and achieved recognition rate up to 98.3% in JAFFE dataset. Zhang *et al.* [59] proposed the expression recognition system based on spatial–temporal networks. An MSCNN was used to extract facial features from static images, and loss function was calculated.

According to the facial expression, classifiers SVM gives a better classification performance. In most cases, SVM exploits prove better than the other classifiers, if execution time is not a constraint. The convolutional neural network also gives better accuracy using backpropagation algorithms and activation function. The classifiers were chosen based on the feature extraction methods and the nature of datasets. In real-time applications, there is a decrease in the performance of classifiers.

4.2 Datasets

The database contains either static images or a sequence of images. The CK database [41], extended CK (CK+) database [38], JAFFE database [42] and so on, contains potential information to recognise different facial expressions under constrained environment. However, posed expressions may occur in the real-time environment due to different illumination conditions, pose variations, and occlusions. Databases like MMI database [55], Oulu-CASIA database, AFEW & SFEW database [3, 44] and so on, were considered to handle not only frontal-view but also dual-view with varying lighting conditions. Here we discuss a few facial expression databases elaborately.

4.2.1 JAFFE database: The JAFFE database [42, 63] is extensively used dataset for facial expression analysis. It includes 213 static images with six basic expressions (happy, sad, fear, angry, disgust and surprise) along the neutral face that were posed under a controlled environment. Each subject has posed three or four samples for every single expression. The image resolution is 256×256 pixels, and the semantic ratings of the expressions were calculated on seven emotion categories by 60 female subjects as ground truth. The expression images are annotated based on the predominant expression of a particular image.

4.2.2 CK database: The CK database [41] is a popular benchmark dataset. The main purpose of this database is to evaluate face recognition and FER. It contains 486 video sequences from 97 subjects with neutral to apex displays. The frames have a

Fig. 7 *Sample images from the CK+ database*

resolution of 640 × 480 or 640 × 490 pixels and are fully FACS coded. Annotation of six basic expressions is also provided. The extended version (CK+) includes 593 posed expression sequence from 122 spontaneous smile sequences of 66 subjects. CK and CK + do not include occluded faces. Some sample images of CK+ database are shown in Fig. 7.

4.2.3 MMI database: The MMI database [55] contains both static and video images of 43 different subjects. It includes 1280 image sequence and 250 images. Various face expressions are recorded for each image sequence from neutral to apex and vice versa. Each subject posed in frontal-view or dual-view displaying 79 AU, single AU and combination of AU of different facial expressions. The performance of posed basic expression and also to laboratory emotions is considered for analysis. Equally, SFEW dataset was static one which handles tough real-time situations like illumination variation, resolutions and occlusions.

4.2.4 Oulu-CASIA database: The Oulu-CASIA database [64] consists of 80 subjects labelled with six basic expressions (80 subjects × 6 expressions) totally 480 video sequences. All expressions of every subject are captured with three different illumination condition using NIR and VIS imaging systems. Now the database includes 2880 image sequence. The first expression of each image sequence is neutral, and the last frame has the apex expression. The image resolution is 320 × 240 pixels. The database can be used in studying facial expression under varying illumination effects and face recognition.

4.2.5 Binghamton University 3D facial expression (BU-3DFE) database: The BU-3DFE database [65] includes 606 facial expression sequences captured from 100 people. For each subject, six prototypic facial expressions are elicited by various manners with multiple intensities. Typically, the dataset is used for multi-view 3D facial expression analysis.

4.2.6 Acted Facial Expression in the Wild (AFEW) database: AFEW database [66, 67] is a video sequence. The dataset is collected from different movie scenes with a large age range of subjects from 1 to 70 years of age. It is the first database with realistic real-world scenarios containing spontaneous facial expressions; natural head pose movements, illumination and occlusions. Using this database, various face analysis research like aging, the study of facial expression in children and so on, can be done extensively. The subjects in AFEW are labelled with six basic expressions: happy, sad, fear, angry, disgust, surprise and neutral expression. Typically, the dataset contains 957 videos containing 28,287 frames, with the duration of 10–150 frames per video. Unlike the other databases, the first frame in AFEW does not start from neutral expression. In the current version database, the annotation of expressions has been updated simultaneously, and it is the only database with multiple labelled subjects in the same frame.

4.2.7 Static facial expressions in the wild (SFEW) database: The SFEW database [67, 68] has been derived by extracting static image frames from AFEW which are similar to real-world images. Mostly, it covers various head pose, close to real-world illumination, larger age range, occlusions, different resolution image of the same face. SFEW consists of 700 images labelled with seven different expressions: happy, sad, surprise, disgust, angry, fear and neutral. The expression labels are publicly available for training and validation, but for testing, it depends on challenge organiser.

5 Challenges

In this section, we elaborate on the distinct problems of FER system in terms of illumination, occlusion, pose variation and so on. These problems are discussed with the solutions provided in recent research.

5.1 Subjectivity

To create subjective experience of the human face, one should be aware of subtle changes beyond the prototypical emotions such as contempt, attentiveness, and curiosity. The expression prediction is mostly on person-dependent where the trained human face is used to recognise the facial expression, the so-called as a subject-dependent approach. Person-independent or subject-independent expression recognition is a highly challenging task due to a significant distinction in the real-time environment of facial data, and thus requires a robust classifier [22]. Algorithms such as LBP, PCA and so on are person dependent, whereas Haar classifier, active appearance model, PPBTF, LK-flow and so on are person-independent algorithms, which builds a typical facial feature model for expression recognition.

5.2 Illumination

Illumination is one of the common problems in FER system. Variations in illumination typically influence the feature extraction task which leads to misclassification and inefficiency in feature analysis. Some pre-processing techniques are employed in handling illumination, but still, robust methods are required to overcome these issues. In this regard, several feature extraction techniques are addressed to achieve robustness against illumination. Ouyang [69] suggests LBP maps an effective method under different illumination conditions. PPBTF is an appearance-based algorithm which uses PCA training model to map 5-pixel pattern of facial expression. Its computational speed is high and robust to illumination.

5.3 Occlusion

The presence of an occlusion in FER system leads to an imprecise emotion labelling in humans. A change in facial appearance due to the overlapping of different objects that may lead to loss of informative facial features. Once the informative region is occluded, it is hard to identify the predominant visual features from the non-occluded region. This causes a serious issue on tracking the face region and that significantly degrades the FER system performance. Zhuang *et al.* [70] suggested the prior knowledge about occlusion in a specific region that helps to predict the type, location, shape, appearance and time of occlusion which occurs on a face. However, the feature vectors are obtained from different states of facial sequences. Also, the trained feature vector is matched with FER feature, and the performance is not affected over the occlusion. HSMM algorithm gives better results for occluded face [69].

5.4 Pose variation

It's a self-occlusion due to changes in the head pose which frequently occurs in the image sequence. Most of the automated FER systems are regulated to recognise the frontal views of facial expressions. However, the absence of frontal view leads to posing variation. Numerous researches are carried out for the frontal view, and only a few types of research were on multi-pose expression recognition [22]. The overall system performs poor due to loss of informative face region across distinct poses. Kangkan [71] suggests that the problem arising from pose variations are compensated with 3D features by reconstructing the face to frontal view.

6 Discussion and future direction

In recent years, FER is an active research area due to its numerous real-time applications. Many authors adopted different image acquisition, extraction and classification approach for building a robust real-time FER system. In this survey, we studied the state-of-the-art techniques from this perspective and also focused to some extent on the controlled and uncontrolled environment. In this section, we discuss the approaches that suit best for real-time applications without performance trade-off. Also, we discuss the methods to deal with issues such as subjectivity, illumination, occlusion, pose variation and complex background. Finally, we comment on the scope of future research.

Face detection is not an easy task [19] since it is influenced by various factors such as non-uniform illumination, pose variation, occlusion, complex background and so on. Adaptive skin colour model [17, 19] is suitable to detect the face in the case of complex background and also shows high accuracy. In most of the research, algorithms such as Haar classifier and AdaBoost are used to detect faces in an image or video sequence due to its high accuracy and low-computational cost. Under the unconstrained environment, Haar classifier detects face efficiently and quickly than skin colour model [72]. For real-time face tracking and motion detection, LK optical flow algorithm and MRASM [73] methods are used in real time as it provides better efficiency and has increased stability in the extraction of facial landmarks. Geometric-based feature method describes the shape and structure of face components such as eyebrows, eyes, nose, and mouth. The feature extraction algorithms of these methods such as ASM and AAM are well suited in real-time applications due to efficient tracking. Appearance-based feature method defines the texture that appears on the face during expression and it has high discriminative power compared to geometric-based method but not suitable for real-time due to high computational cost and memory storage. The LBP, HOG, Gabor filter-based texture information, local directional ternary pattern provide the appearance features were employed to detect the facial expression of a static image and video.

Furthermore, classification is an important scheme to improve the reliability of the FER system and also reduces classification error. Some of the classifiers such as SVM, neural networks, nearest neighbour, Haar classifiers and so on, were chosen as per the feature extraction techniques and other constraints by the researchers. From the above survey, we suggest neural network classifier achieves a higher recognition rate in real time compared to other classifiers and deals mainly with large datasets. However, SVM also outperforms in many applications, but it evolves based on the discriminative information from feature extraction and also the user. Haar classifier with AdaBoost gives better results in a real-time environment as it deals with illumination and complex background. Also, neural network deals with large datasets and requires minimum execution time. Finally, the facial expression datasets used in FER system are CK, CK+, JAFFE, Oulu-CASIA VIS, MUG and so on, all these are restricted to few constraints like a frontal view, simple background and pre-processed image. However, the method and classifiers are trained with the restricted labelled dataset which is difficult to handle real-time image sequence. There are some database with real-time video sequence and frames which are extensively used to develop an algorithm for an unconstrained environment. LFW is a face recognition dataset created from internet images with the definite protocol for training and testing. AFEW and SFEW datasets are the dynamic temporal real-time dataset used for both static and image sequences.

In future, we can extend the work in developing an FER system to recognise different other facial or sign expressions rather than the basic expressions. For instance, exhaustion, frustration, anticipation, aggressiveness and so on helps in real-time applications. The present system uses available datasets that limit the efficiency when applied the same in real time. It is a known fact that the available facial expression datasets are not only frontal-view and also pre-processed under the controlled environment. Moreover, they have still image and image sequences are of fixed sizes, which enable the feature extraction process simple whereas in the real-time environment it's not possible. To cope with existing datasets and in a real-time environment, compatible new techniques should extensively develop from the baseline.

7 References

[1] Ekman, P., Friesen, W.V.: 'Constants across cultures in the face and emotion', *J. Pers. Soc. Psychol.*, 1971, **17**, (2), pp. 124–129

[2] Mehrabian, A.: 'Communication without words', *Psychol. Today.*, 1968, **2**, (4), pp. 53–56

[3] Lau, B.T.: 'Portable real-time emotion detection system for disabled', *Expert Syst.*, 2010, **37**, (9), pp. 6561–6566

[4] Dulguerov, P., Marchal, F., Wang, D., *et al.*: 'Review of objective topographic facial nerve evaluation methods', *AM. J. Otol.*, 1999, **20**, (5), pp. 672–678

[5] Alazrai, R., George Lee, C.: 'Real-time emotion identification for socially intelligent robot'. IEEE Int. Conf. Robotics and Automation, USA, May 2012, pp. 14–18

[6] Tian, Y.L., Kanade, T., Cohn, J.F.: 'Facial expression analysis', in '*Handbook of face recognition*' (Springer, New York, NY, 2005), pp. 247–275

[7] Viola, P., Jones, M.: 'Robust real-time face detection', *Int. J. Comput. Vis.*, 2004, **57**, (2), pp. 137–154

[8] Danti, V.N., Kadiyavar, P.T.: 'An adaptive skin color model for a human face detection'. IEEE Int. Conf. Contemporary Computing and Informatics (IC3I), Mysore, November 2014, pp. 526–532

[9] Tian, Y., Kanade, T., Cohn, J.F.: 'Recognizing action units for facial expression analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (2), pp. 97–115

[10] Chu, W.S., Torre, F.D., Cohn, J.F.: 'Selective transfer machine for personalized facial expression analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (3), pp. 529–545

[11] Zeng, Z., Pantic, M., Roisman, G., *et al.*: 'A survey of affect recognition methods: audiovisual and spontaneous expressions', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (1), pp. 39–58

[12] Sebe, N., Lew, M.S., Sun, Y., *et al.*: 'Authentic facial expression analysis', *Image Vis. Comput.*, 2007, **25**, (12), pp. 1856–1863

[13] Pentland, A., Moghaddam, B., Starner, T.: 'View-based and modular eigenspaces for face recognition'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, USA, August 2002, pp. 84–91

[14] Essa, A., Pentland, A.P.: 'Coding, analysis, interpretation, and recognition of facial expressions', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 757–763

[15] Turk, M., Pentland, A.: 'Eigenfaces for recognition', *J. Cogn. Neurosci.*, 1991, **3**, (1), pp. 71–86

[16] Pantic, M., Rothkrantz, L.J.M.: 'Automatic analysis of facial expressions: the state of the art', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (12), pp. 1424–1445

[17] Zhao-yi, P., Yan-hui, Z., Yu, Z.: 'Real-time facial expression recognition based on adaptive canny operator edge detection'. Int. Conf. Multimedia and Information Technology (MMIT), Kaifeng, April 2010, pp. 154–157

[18] Kheirkhah, E., Tabatabaie, Z.S.: 'A hybrid face approach in color images with complex background', *Indian J. Sci. Technol.*, 2015, **8**, (1), pp. 49–60

[19] Zhao-yi, P., Zhou, Y., Wang, P.: 'Multi-pose face detection based on adaptive skin color and structure model'. Proc. IEEE Int. Conf. Computational Intelligence and Security, Beijing, China, 2009, pp. 325–329

[20] Gargesha, M., Panchanathan, S.: 'Face detection from color images by iterative thresholding on skin probability maps'. Proc. IEEE Int. Symp. Circuits and Systems, Arizona, USA, August 2002

[21] Cho, K.M.J., Jang, H., Hong, K.S.: 'Adaptive skin-color filter', *Pattern Recognit.*, 2001, **34**, (5), pp. 1067–1073

[22] Happy, S.L., George, A., Routary, A.: 'A real time facial expression classification system using local binary patterns'. Int. Conf. Intelligent Human Computer Interaction, Kharagpur, December 2012, pp. 1–5

[23] Hassan, I., Maqbool, O., Ashan, Q., *et al.*: 'Cascading neural network with AdaBoost for face detection'. Proc. Int. Conf. Applied Sciences & Technology, Islamabad, Pakistan, January 2010, pp. 223–229

[24] Dabhi, M.K., Pancholi, B.K.: 'Face detection system based on Viola-Jones algorithm', *Int. J. Sci. Res.*, 2016, **5**, (4), pp. 62–64

[25] Geetha, A., Ramalingam, V., Palanivel, S., *et al.*: 'Facial expression recognition- a real-time approach', *Expert Syst. Appl.*, 2009, **36**, (1), pp. 303–308

[26] Dey, A.: 'Contour based procedure for face detection and tracking from video'. Proc. Int. Conf. Recent Advances in Information Technology, Dhanbad, July 2016, pp. 483–488

[27] Ekman, P., Friesen, W.V.: 'Measuring facial movement', *Environ. Psychol. Nonverbal Behav.*, 1976, **1**, (1), pp. 56–75

[28] Cootes, T., Taylor, C., Cooper, D., *et al.*: 'Active shape models-their training and applications', *Comput. Vis. Image Underst.*, 1995, **61**, (2), pp. 38–59

[29] Cootes, T., Edwards, G., Taylor, C.: 'Active appearance models', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (6), pp. 681–685

[30] Lowe, D.G.: 'Distinctive image features from scale invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

[31] Berretti, S., Del Bimbo, A., Pala, P., *et al.*: 'A set of selected SIFT features for 3D facial expression recognition'. Proc. Int. Conf. Pattern Recognition, Istanbul, Turkey, 2010, pp. 4125–4128

[32] Zhang, Z., Lyons, M., Schuster, M., *et al.*: 'Comparison between geometry-based and Gabor-based expression recognition using multi-layer perceptron'. Proc. Int. Conf. Automatic Face and Gesture Recognition, Japan, April 1998, pp. 26–39

[33] Rosdiyana, S., Hideyuki, S.: 'Edge-based facial feature extraction using Gabor wavelet and convolution filters'. Proc. Int. Conf. Machine Vision Applications, Japan, June 2011, pp. 430–433

[34] Dongcheng, S., Fang, C., Guangyi, D.: 'Facial expression recognition based on Gabor wavelet phase features'. Proc. Int. Conf. Image and Graphics (ICIG), Qingdao, China, July 2013, pp. 520–523

[35] Praseed, L.V., Sasikumar, M.: 'Facial expression recognition from global and a combination of local features', *IETE Tech. Rev.*, 2009, **26**, (1), pp. 41–46

[36] Lu, H.C., Huang, Y.J., Chen, Y.W.: 'Real-time FER based on pixel-pattern-based texture feature', *Electron. Lett.*, 2007, **43**, (17), pp. 916–918

[37] Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 971–987

[38] Duong, V.H., Lee, Y.S., Ding, J.J.: 'Projective complex matrix factorization for facial expression recognition', *EURASIP J. Adv. Signal Process.*, 2018, pp. 1–11

[39] Fatima, Z.S., Abdellah, M., Mohamed, K.: 'Emotion recognition from facial expression based on fiducial points detection and using neural network', *Int. J. Electr. Comput. Eng.*, 2018, **8**, (1), pp. 52–59

[40] Xiong, X., Torre, F.: 'Supervised decent method and its applications to face alignment'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Portland, Oregon, 2013, pp. 532–539

[41] Michael Revina, I., Sam Emmanuel, W.R.: 'Face expression recognition using LDN and dominant gradient local ternary pattern descriptors', *J. King Saud Univ., Comput. Inf. Sci.*, 2018

[42] Ding, Y., Zhao, Q., Li, B., *et al.*: 'Facial expression recognition from image sequence based on LBP and Taylor expansion', *IEEE Access*, 2017, **5**, pp. 19409–19419

[43] Arshid, S., Hussain, A., Munir, A., *et al.*: 'Multi-stage binary patterns for facial expression recognition in real world', *Cluster Comput.*, 2017, **21**, (1), pp. 323–331

[44] Munir, A., Hussain, A., Khan, S.A., *et al.*: 'Illumination invariant facial expression recognition using selected merged binary patterns for real world images', *Optik*, 2018, **158**, pp. 1016–1025

[45] Hasani, B., Mahoor, M.H.: 'Spatial-temporal facial expression recognition using convolutional neural networks and conditional random fields', arXiv: 1703.06995, 2017

[46] Mahmud, F., Al Mamun, M.: 'Facial expression recognition system using extreme learning machine', *Int. J. Sci. Eng. Res.*, 2017, **8**, (3), pp. 26–30

[47] Khadija, L., Yassine, R., Rochidi, M., *et al.*: 'Facial expression recognition using face regions'. Proc. Int. Conf. Advanced Technologies for Signal and Image Processing, Fez, Morocco, May 2017, pp. 1–6

[48] Holder, R.P., Tapamo, J.R.: 'Improved gradient local ternary patterns for facial expression recognition', *EURASIP J. Image Video Process.*, 2017, **42**, p. 42

[49] Ahmed, F., Hossain, E.: 'Automated facial expression recognition using gradient-based ternary texture patterns', *Chin. J. Eng.*, 2013, **2013**, pp. 1–8

[50] Qayyum, H., Majid, M., Anwar, S.M., *et al.*: 'Facial expression recognition using stationary wavelet transform features', *Math. Prob. Eng.*, 2017, **2017**, pp. 1–9

[51] Du, L., Hu, H.: 'Modified classification and regression tree for facial expression recognition using difference expression images', *Electron. Lett.*, 2017, **53**, (9), pp. 590–592

[52] Liu, Y., Li, Y., Ma, M., *et al.*: 'Facial expression recognition with fusion features extracted from salient facial areas', *Sensors*, 2017, **17**, (4), 712, pp. 1–18

[53] Kumar, S., Bhuyan, M.K., Chakraborty, B.K.: 'Extraction of informative regions of a face for facial expression recognition', *IET Comput. Vis.*, 2016, **10**, (6), pp. 567–576

[54] Ojala, T., Pietikainen, M., Harwood, D.: 'A comparative study of texture measure with classification based on featured distribution', *Pattern Recognit.*, 1996, **1**, (29), pp. 51–59

[55] Majumder, A., Behera, L., Subramanian, V.: 'Automatic facial expression recognition system using deep network-based data fusion', *IEEE Trans. Cybern.*, 2016, **99**, pp. 1–12

[56] Kamarol, S.K.A., Jaward, M.H., Parkkinen, J., *et al.*: 'Spatiotemporal feature extraction for facial expression recognition', *IET Image Process.*, 2016, **10**, (7), pp. 534–541

[57] Tang, Y., Zhang, X.M., Wang, H.: 'Geometric-convolutional feature fusion based on learning propagation for facial expression recognition', *IEEE Access*, 2018, **6**, pp. 42532–42540

[58] Hsieh, C.C., Hsih, M.H., Jiang, M.K., *et al.*: 'Effective semantic features for facial expression recognition using SVM', *Multimedia Tools Appl.*, 2015, **75**, (11), pp. 6663–6682

[59] Zhang, K., Huang, Y., Du, Y.: 'Facial expression recognition based on deep evolutional spatial-temporal networks', *IEEE Trans. Image Process.*, 2017, **26**, pp. 4193–4203

[60] Kumbhar, M., Jadhav, A., Patil, M.: 'Facial expression recognition based on image feature', *Int. J. Comput. Commun. Eng.*, 2012, **1**, (2), pp. 117–119

[61] Suma, K., Lakshminarayana, M.: 'Survey on face expression recognition from human emotions by using SVM classifier', *Int. J. Cur. Eng. Sci. Res.*, 2018, **5**, (5), pp. 105–110

[62] Duda, R.O., Hart, P.E., Strok, D.G.: '*Pattern classification*' (Wiley Publications, New York, 2001, 2nd edn.)

[63] Lyons, M.: 'Coding facial expressions with Gabor wavelets'. Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 20–205

[64] Zhao, G., Huang, X., Taini, M., *et al.*: 'Facial expression recognition from near-infrared videos', *Image Vis. Comput.*, 2011, **29**, (9), pp. 607–619

[65] Yin, L., Wei, X., Sun, Y., *et al.*: 'A 3D facial expression database for facial behaviour research'. IEEE Int. Conf. Automatic Face and Gesture Recognition (FGR 2006), Southampton, UK, 2006, pp. 211–216

[66] Dhall, A., Goecke, R., Lucey, S., *et al.*: 'Acted facial expressions in the wild database'. Technical Report, 2011

[67] Dhall, A., Goecke, R., Lucey, S., *et al.*: 'Collecting large, richly annotated facial-expression databases from movies', *IEEE Multimedia*, 2012, **19**, (3), pp. 33–41

[68] Dhall, A., Goecke, R., Lucey, S., *et al.*: 'Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark'. Proc. IEEE Int. Conf. ICCV Workshops, Barcelona, Spain, 2011, pp. 2106–2112

[69] Ouyang, Y.: 'Robust automatic facial expression detection method based on sparse representation plus LBP map', *Opt.-Int. J. Light Electron Opt.*, 2013, **124**, pp. 6827–6833

[70] Zhuang, L., Chan, T., Yang, A.: 'Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment', *Int. J. Comput. Vis.*, 2015, **114**, (2), pp. 272–287

[71] Kangkan, W.: 'A two-stage framework for 3D face reconstruction from RGBD images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, **36**, (8), pp. 1493–1504

[72] Anisha, A., Swati, M.: 'Comparative analysis of Haar and skin color method for face detection'. IEEE Int. Conf. Recent Advances and Innovations in Engineering, Jaipur, India, May 2014

[73] Byungtae, A., Han, Y., Kweon, I.S.: 'Real-time facial landmarks tracking using active shape model and LK optical flow'. Int. Conf. Ubiquitous Robots and Ambient Intelligence (URAI-2012), Daejeon, Korea, November 2012, pp. 541–543