Journal of King Saud University – Computer and Information Sciences 33 (2021) 619–628

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

A Survey on Human Face Expression Recognition Techniques

I.Michael Revina *, W.R. Sam Emmanuel

Reg No. 12417, N.M. Christian College, Marthandam Affiliated to Manonmaniam Sunadaranar University, Abishekapatti, Tirunelveli – 627012, Tamil Nadu, India
Department of Computer Science, N.M. Christian College, Marthandam Affiliated to Manonmaniam Sunadaranar University, Abishekapatti, Tirunelveli – 627012, Tamil Nadu, India

A R T I C L E I N F O

Article history:
Received 13 April 2018
Revised 24 August 2018
Accepted 3 September 2018
Available online 05 September 2018

Keywords:
Classification
Face Expression Recognition (FER)
Feature extraction
Preprocessing

A B S T R A C T

Human Face expression Recognition is one of the most powerful and challenging tasks in social communication. Generally, face expressions are natural and direct means for human beings to communicate their emotions and intentions. Face expressions are the key characteristics of non-verbal communication. This paper describes the survey of Face Expression Recognition (FER) techniques which include the three major stages such as preprocessing, feature extraction and classification. This survey explains the various types of FER techniques with its major contributions. The performance of various FER techniques is compared based on the number of expressions recognized and complexity of algorithms. Databases like JAFFE, CK, and some other variety of facial expression databases are discussed in this survey. The study on classifiers gather from recent papers reveals a more powerful and reliable understanding of the peculiar characteristics of classifiers for research fellows.

Contents

1. Introduction

Human facial expressions are extremely essential in social communication. Normally communication involves both verbal and nonverbal. Non-verbal communications are expressed through facial expressions. Face expressions are the delicate signals of the larger communication. Non-verbal communication means communication between human and animals through eye contact, gesture, facial expressions, body language, and paralanguage.

Eye contact is the important phase of communication which provides the mixture of ideas. Eye contact controls the contribution, discussions and creates a link with others. Face expressions include the smile, sad, anger, disgust, surprise, and fear. A smile on human face shows their happiness and it expresses eye with a curved shape. The sad expression is the feeling of looseness which is normally expressed as rising skewed eyebrows and frown. The anger on human face is related to unpleasant and irritating conditions. The expression of anger is expressed with squeezed eyebrows, slender and stretched eyelids. The disgust expressions are expressed with pull down eyebrows and creased nose. The surprise

* Corresponding author.
E-mail addresses: michaelrevina09@gmail.com (I. Revina), sam_emmanuel@nmcc.ac.in (W.R.S. Emmanuel).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2018.09.002
1319-1578/

or shock expression is expressed when some unpredicted happens. This is expressed with eye-widening and mouth gaping and this expression is an easily identified one. The expression of fear is related with surprise expression which is expressed as growing skewed eyebrows.

FER has the important stage is feature extraction and classification. Feature extraction includes two types and they are geometric based and appearance based. The classification is also one of the important processes in which the above-mentioned expressions such as smile, sad, anger, disgust, surprise, and fear are categorized. The geometrically based feature extraction comprises eye, mouth, nose, eyebrow, other facial components and the appearance based feature extraction comprises the exact section of the face (Zhao and Zhang, 2016).

Generally, the face offers three different types of signals such as static, slow and rapid signals. The static signals are skin color which includes the several lasting aspects of face skin pigmentation, greasy deposits, face shapes, the constitution of bones, cartilage and shape, location and size of facial features such as brows, eyes, nose, mouth. The slow signals are permanent wrinkles which include the changes in facial appearance such as muscle tone and skin texture changes that happen slowly with time.

The rapid signals are raising the eyebrows which include the face muscles movement, impermanent face appearance changes, impermanent wrinkles and changes in the location and shape of facial features. These flashes on the face remain for a few seconds. These three signals are altered with individual option while it is very hard to alter static and slow signals. Also, the face is a multi-message system and it is not only a multi-signal system. Messages are transmitted through a face which includes emotion, feel position, age, quality, intelligence, attractiveness and almost certainly other substances as well (Ekman and Friesen, 2003).

This paper mainly focuses on various FER techniques with three major steps respectively preprocessing, feature extraction and classification. Also, this paper shows the advantages of different FER techniques and the performance analysis of different FER techniques. In this paper, only the image based FER techniques are chosen for the literature review and the video based FER techniques are not chosen. Mostly FER systems meet the problems of variation in illumination, pose variation, lighting variations, skin tone variations. Also this paper gives an essential research idea for future FER research.

Rest of the paper is structured as follows. Section 2 elaborates the detailed description of face expression recognition system. Section 3 evaluates the performance of FER techniques and through different table and charts. Section 4 provides suggestions along with the conclusion of this survey.

2. Face expression recognition system

The overview of the FER system is illustrated in Fig. 1. The FER system includes the major stages such as face image preprocessing, feature extraction and classification.

2.1. Preprocessing

Preprocessing is a process which can be used to improve the performance of the FER system and it can be carried out before feature extraction process (Poursaberi et al., 2012). Image preprocessing includes different types of processes such as image clarity and scaling, contrast adjustment, and additional enhancement processes (Bashyal et al., 2008) to improve the expression frames (Taylor et al., 2014).

The cropping and scaling processes were performed on the face image in which the nose of the face is taken as midpoint and the other important facial components are included physically (Zhang et al., 2011). Bessel down sampling is used for face image size reduction but it protects the aspects and also the perceptual worth of the original image (Owusu et al., 2014). The Gaussian filter is used for resizing the input images which provides the smoothness to the images (Biswas, 2015).

Normalization is the preprocessing method which can be designed for reduction of illumination and variations of the face images (Ji and Idrissi, 2012) with the median filter and to achieve an improved face image. The normalization method also used for the extraction of eye positions which make more robust to personality differences for the FER system and it provides more clarity to the input images. Localization is a preprocessing method and it uses the Viola-Jones algorithm (Noh et al., 2007; Demir, 2014; Zhang et al., 2014; Cossetin et al., 2016; Salmam et al., 2016) to detect the facial images from the input image. Detection of size and location of the face images using Adaboost learning algorithm and haar like features (Happy et al., 2015; Mahersia and Hamrouni, 2015). The localization is mainly used for spotting the size and locations of the face from the image.

Face alignment is also the preprocessing method which can be performed by using the SIFT (Scale Invariant Feature Transform) flow algorithm. For this, first calculate reference image for each face expressions. After that all the images are aligned through related reference images (Dahmane and Meunier, 2014). ROI (Region of Interest) segmentation is one of the important type of preprocessing method which includes three important functions such as regulating the face dimensions by dividing the color components and of face image, eye or forehead and mouth regions segmentation (Hernandez-matamoros et al., 2015). In FER, ROI

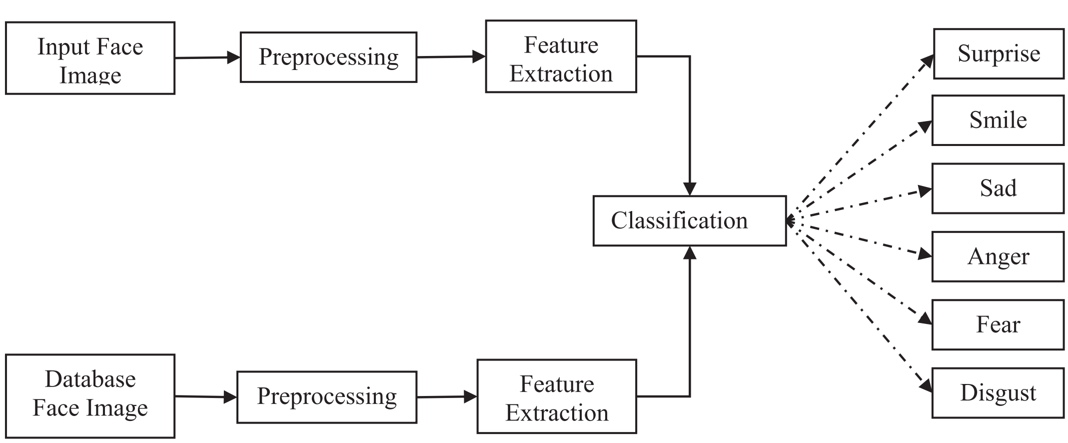

Fig. 1. Architecture of face expression recognition system.

segmentation is most popular because for convenient segmentation of face organs from the face images.

The histogram equalization method is used to conquer the illumination variations (Demir, 2014; Happy et al., 2015; Cossetin et al., 2016). This method is mainly used for enhancing the contrast of the face images and for exact lighting also used to improve the distinction between the intensities.

In FER, more preprocessing methods are used but the ROI segmentation process is more suitable because it detects the face organs accurately which organs are is mainly used for expression recognition. Next the histogram equalization is also another one important preprocessing technique for FER because it improves the image distinction.

2.2. Feature extraction

Feature extraction process is the next stage of FER system. Feature extraction is finding and depicting of positive features of concern within an image for further processing. In image processing computer vision feature extraction is a significant stage, whereas it spots the move from graphic to implicit data depiction. Then these data depiction can be used as an input to the classification. The feature extraction methods are categorized into five types such as texture feature-based method, edge based method, global and local feature-based method, geometric feature-based method and patch-based method.

The descriptors which extract the features based on the texture feature-based methods are described as follows. Gabor filter is a texture descriptor for feature extraction and it includes the magnitude and phase information. The Gabor filter with the magnitude feature confines the information about the organization of the face image. The phase feature precincts the information about the complete description of the magnitude features (Bashyal et al., 2008; Owusu et al., 2014; Zhang et al., 2014; Hernandez-matamoros et al., 2015; Hegde et al., 2016). Local Binary Pattern (LBP) is also a texture descriptor and it can be used for feature extraction. Generally LBP features are produced with the binary code and it can be obtained by using thresholding between the center pixel and its locality pixels (Happy et al., 2015; Cossetin et al., 2016). Also LBP with Three Orthogonal Planes (TOP) features are extracted for multi resolution approaches and (Zhao and Pietikäinen, 2009). It is used for extracting non dynamic appearance based on features from the static face images (Ji and Idrissi, 2012). The facial texture features are extracted using the Gaussian Laguerre (GL) function which grants a steering pyramidal structure which extracts the texture features and the facial related occurrence information. Comparing to Gabor function GL uses the single filter instead of multiple filters (Poursaberi et al., 2012). Moreover another descriptor which is used namely Vertical Time Backward (VTB) which also extracts the texture features of face images. Moments descriptor extracts the shape related features of significant facial components. Both VTB and moments descriptors are effective on spatiotemporal planes (Ji and Idrissi, 2012). Weber Local Descriptor (WLD) is a feature extraction technique that extracts the high discriminant texture features from the segmented face images (Cossetin et al., 2016). Feature extraction is performed with three stages using Supervised Descent Method (SDM). At first, the facial main positions are extracted. Next the related positions are selected. Finally it estimates the distance between the various components of the face (Salmam et al., 2016). Weighted Projection based LBP (WPLBP) is also a feature extraction but based on the instructive regions which extracts the LBP features. After that based on the significance of the instructive regions these features are weighted (Kumar et al., 2016). Discrete Contourlet Transform (DCT) extracts the texture features which can be performed by decomposition with two key stages. The stages are Laplacian Pyramid (LP) and Directional Filter Bank (DFB) which is used in the transformed domain. In LP stage, partitions the image into low pass, band pass and confines the discontinuities position. The DFB stage processes the band pass and forms the linear composition by associating the discontinuities position (Biswas, 2015).

The descriptors which extract the features based on the edge based methods are described as follows. Line Edge Map (LEM) descriptor is a facial expression descriptor which improves the geometrical structural features by using the dynamic two strip algorithm (Dyn2S) (Gao et al., 2003). Based on the motion analysis two types of facial features are extracted such as non discriminative and discriminative facial features (Noh et al., 2007). Graphics-processing unit based Active Shape Model (GASM) is the feature extraction method which can be performed with edge detection, enhancement, tone mapping and local appearance model matching. After that the image ratio features are extracted from the expressed face images (Song et al., 2010). Histogram of Oriented Gradients (HOG) is a window supported feature descriptor which uses the gradient filter. The extracted features are based on the edge information of the registered face images. It extracts the visual features, for example a smile expression means curvature shaped eyes (Dahmane and Meunier, 2014).

The descriptors which extract the features based on the global and local feature-based methods are described as follows. Principal Component Analysis (PCA) method is used for feature extraction. It extracts the global and low dimensional features. Independent Component Analysis (ICA) is also a feature extraction method which extracts the local features using the multichannel observations (Taylor et al., 2014). Stepwise Linear Discriminant Analysis (SWLDA) is the feature extraction technique which extracts the localized features with backward and forward regression models. Depends on the class labels the F-test values are estimated for both regression models (Siddiqi et al., 2015).

The descriptors which extract the features based on the geometric feature-based methods are described as follows. Local Curvelet Transform (LCT) is a feature descriptor which extracts the geometric features which depends on wrapping mechanism. The extracted geometric features are mean, entropy and standard deviation (Demir, 2014). Addition to these geometrical features energy, kurtosis are extracted by using three stage steerable pyramid representation (Mahersia and Hamrouni, 2015).

The descriptors which extract the features based on patch-based methods are described as follows. Facial movement features are extracted as patches depending upon the distance characteristics. These are performed by using two processes such as extracting the patches and patch matching. The patch matching is performed by translating extracted patches into distance characteristics (Zhang et al., 2011).

The texture feature based descriptors are more useful feature extraction method than the others because it extracts the texture features like related to the appearance which provides the important feature vectors for FER. Also Local Directional Number (LDN) pattern (Rahul and Cherian, 2016), Local Directional Ternary Pattern (LDTP) (Ryu et al., 2017), KL-transform Extended LBP (K-ELBP) (Guo et al., 2016) and Discrete Wavelet Transform (DWT) (Nigam et al., 2018) texture feature based descriptors are used as feature descriptors in recent years FER.

Several extracted features have high dimensional vectors. Generally these feature vectors are reduced by using various dimensionality reduction algorithms such as PCA, Linear Discriminant Analysis, Whitened Principle Component Analysis and the important features are also selected with different algorithms such as Adaboost and similarity scores.

2.3. Classification

Classification is the final stage of FER system in which the classifier categorizes the expression such as smile, sad, surprise, anger, fear, disgust and neutral.

The directed Line segment Hausdorff Distance (dLHD) method is used for recognition of expressions (Gao et al., 2003). Euclidean distance metric is also used for classification purpose which uses the normalized score and similarity score matrix for estimating Euclidean distance (Hegde et al., 2016). Minimum Distance Classifier (MDC) is also one of the distance based classifier used for classification which estimates the distance between the feature vectors every sub image (Islam et al., 2018). The KNN (k – Nearest Neighbors) algorithm is a classification method in which the relationship among the assessment models and the other models are estimated during the training stage (Poursaberi et al., 2012).

Support Vector Machine (SVM) is one of the classification techniques in which two types of approaches are involved. They are one against one and one against all approaches. One against all classification means it constructs one sample for each class (Zhao and Pietikäinen, 2009; Zhang et al., 2011; Zhang et al., 2014; Biswas, 2015). One against one classification means it constructs one class for each pair of classes (Happy et al., 2015; Kumar et al., 2016; Hegde et al., 2016) and SVM is one of the strongest classification methods for advanced dimensionality troubles (Dahmane and Meunier, 2014). SVM is the supervised machine learning technique and it uses four types of kernels for its better performance (Hernandez-matamoros et al., 2015). They are linear, polynomial, Radial Basis Function (RBF) and sigmoid. The linear kernel maps the high dimensional data and it is linearly separable (Zhang et al., 2014; Kumar et al., 2016). The RBF kernel uses the function that maps the single feature into the high dimensional data (Song et al., 2010; Wang et al., 2010; Dahmane and Meunier, 2014; Happy et al., 2015; Hegde et al., 2016). The polynomial kernel learns the nonlinear models and also resolves their similarity (Zhao and Pietikäinen, 2009; Zhang et al., 2011; Ji and Idrissi, 2012; Biswas, 2015).

The Hidden Markov Model (HMM) classifier is the statistical model which categorizes the expressions into different types (Taylor et al., 2014). Hidden Conditional Random Fields (HCRF) representation is used for classification. It uses the full covariance Gaussian distribution for superior classification performance (Siddiqi et al., 2015).

Online Sequential Extreme Learning Machine (OSELM) is a method that uses RBF for classification. OSELM mainly contains two stages. They are initialization and sequential learning stages. Initialization stage includes the training samples (Demir, 2014). Pair wise classifiers are also used for expression classification. It uses the one against one classification approach so exacting separation is utilized (Cossetin et al., 2016).

ID3 Decision Tree (DT) classifier is a rule based classifier which extracts the predefined rules to produce competent rules. The predefined rules are generated from the decision tree and it was constructed by information gain metrics. The classification is performed using the least Boolean evaluation (Noh et al., 2007; Rashid, 2016). Classification and Regression Tree (CART) is a machine learning algorithm for classification. The metric likely Decision tree and Gini impurity are estimated. CART classifiers are signified by using the distance vectors (Salmam et al., 2016).

Learning Vector Quantization (LVQ) is the unsupervised clustering algorithm (Bashyal et al., 2008) which has two layers namely competitive and output layers. The competitive layer has the neurons that are known as subclasses. The neuron which is the greatest match in competitive layer then put high for the class of exacting neuron in the output layer. Multi Layer Perceptron (MLP) is also used for classification and it contains three layers such as input layer, output layer and processing layer in which neurons are present (Rashid, 2016).

The Multilayer Feed Forward Neural Network (MFFNN) classifier uses three layers such as input, hidden and output layers and back propagation algorithm for classification. In the training stage the weights are initialized and the activation units are estimated (Owusu et al., 2014). Bayesian neural network classifier is the classification method which also includes three layers such as input, hidden and output layers. The classical back propagation algorithm is used with Bayesian classifier for its better accuracy (Mahersia and Hamrouni, 2015). Convolution Neural Network (CNN) consists of two layers such as convolutional layer and subsampling layer in which the two dimensional images are taken as input. In convolutional layer the feature maps are produced by intricate the convolution kernels with the two dimensional images where as in the subsampling layer, pooling and redeployment are performed (Shan et al., 2017). The CNN also contains two important perceptions likely shared weight and sparse connectivity (Rashid, 2016). In FER, the CNN classifier used as multiple classifiers for the different face regions. If CNN is framed for entire face image then first frame the CNN for mouth area and next for eye area likely for each other area CNNs are framed (Cui et al., 2016).

Deep Neural Network (DNN) contains various hidden layers and the more difficult functions are trained efficiently comparing with other neural networks (Li and Lam, 2015). The Deep Belief Network (DBN) contains the hidden variable resides of the various number of Restricted Boltzmann Machine (RBM) which are the undirected generative pattern (Lv, 2015). DBN contains the Back Propagation (BP) layer classifies the high-level features using classification (Yang et al., 2016). DBN generally includes two phases such as pre-learning and fine-tuning (Wu and Qiu, 2017) in which RBM are developed separately in the first step whereas the BP are learning the input and output data in the last phase.

According to several classifiers SVM classifier gives better recognition accuracy and it provides better classification. The neural network based classifier CNN gives better accuracy than the other neural network based classifiers. In FER, SVM classifier is more exploitable comparing with other classifiers for recognition of expressions.

The various FER techniques with their algorithm is analyzed in Table1 which includes the algorithms that are used for three important requirements such as preprocessing, feature extraction and classification. The various preprocessing methods used in this table are, face detection, image enhancement, normalization, Gabor filter, localization, face acquisition, down sampling, histogram equalization, face region detection, face alignment, ROI segmentation and resizing. The different feature extraction methods used in this table are LEM, Action based model, Gabor filter, LBP-TOP, GASM, Patch based, GL wavelet, LBP, VTB, Moments, PCA, ICA, LCT, HOG, Steerable pyramid, DCT, SWLDA, WLD, SDM, WPLBP, haar like features, LDN, LDTP, DWT, K-ELBP, 2DPCA and eigenfaces. Classifiers used in this table are ID3 decision tree, LVQ, SVM, KNN, HMM, MFFNN, OSLEM, Bayesian neural network, HCRF, pair wise, CART, Euclidean distance, CNN, MDC, Chi square test and fisher discrimination dictionary.

In recent year papers, for preprocessing mostly the histogram equalization method is used. For feature extraction, Gabor filter, WPLBP, SDM, WLD, HOG are used. In feature extraction, the majority of the methods are based on the texture descriptor such as LBP based which gives improved results. In modern years, the classification uses the classifiers are SVM, Euclidean distance, CART, Neural network based classifiers and pair wise classifiers. The SVM classifier is highly used classifier in FER and it uses one- to –one, one- to all classification approach. Also, SVM with RBF kernel is most probably used which gives the highest classification performance comparing to other classifiers.

In 3D FER, the preprocessing of face images are performed by using the various methods such as smoothing, cropping, face align-

Table 1
Algorithm analysis of 2D FER Techniques.

Author, Year	Preprocessing method	Feature extraction method	Classification method
Gao et al. (2003)	Not reported	LEM	dLHD
Noh et al. (2007)	Face detection	Action based model	ID3 decision tree
Bashyal et al. (2008)	Image enhancement	Gabor filter (GF)	LVQ
Zhao and Pietikäinen (2009)	Not reported	LBP – TOP	SVM
Song et al. (2010)	Not specified	GASM	SVM
Wang et al. (2010)	Not reported	Not reported	SVM
Zhang et al. (2011)	Gabor filter	Patch based	SVM
Poursaberi et al. (2012)	Localization, Normalization	GL Wavelet	KNN
Ji and Idrissi (2012)	Face acquisition	LBP, VTB, Moments	SVM
Taylor et al. (2014)	Enhancement	PCA ICA	HMM
Owusu et al. (2014)	Down sampling	GF	MFFNN
Demir (2014)	Histogram equalization	LCT	OSLEM
Zhang et al. (2014)	Face region detection	GF	SVM
Dahmane and Meunier (2014)	Face alignment	HOG	SVM
Mahersia and Hamrouni (2015)	Normalization	Steerable pyramid	Bayesian neural network
Hernandez-matamoros et al. (2015)	ROI segmentation	Gabor function	SVM
Happy et al. (2015)	Histogram equalization	LBP	SVM
Biswas (2015)	Histogram equalization	DCT	SVM
Siddiqi et al. (2015)	Not specified	SWLDA	HCRF
Cossetin et al. (2016)	Histogram equalization	LBP, WLD	Pairwise Classifiers
Salmam et al. (2016)	Face detection	SDM	CART
Kumar et al. (2016)	Not reported	WPLBP	SVM
Hegde et al. (2016)	Resizing	GF	Euclidean distance (ED), SVM
Rashid (2016)	Balancing data	Luxand Face SDK, ED	DT, MLP, CNN
Cui et al. (2016)	Face detection, normalization	Not reported	CNN
Jain et al. (2016)	Face detection	LBP	ED, SVM, Neural Network
Rahul and Cherian (2016)	Face region cropping	LDN	Chi square test
Guo et al. (2016)	Normalization	K-ELBP	SVM
Sharma and Rameshan, 2017	Face normalization	HOG, LBP, Eigen faces	Fisher discrimination dictionary
Shan et al. (2017)	Histogram equalization	Haar like features	CNN
Nazir et al. (2017)	Face Detection	HOG, DCT	KNN
Chang (2017)	Face detection	DCT, GF	SVM
Zhang et al. (2017)	Localization	Not reported	CNN
Ryu et al. (2017)	Not reported	LDTP	SVM
Nigam et al. (2018)	Cropping, Normalization	DWT, HOG	SVM
Clawson et al. (2018)	Histogram equalization	Not reported	CNN
Islam et al. (2018)	Face detection	2DPCA	MDC

ment. The facial expressions are recognized from videos using the head gesticulation (Anisetti et al., 2005). The features such as geometric features and appearance features are extracted from 3D faces using the various descriptors likely 3D surface descriptors (Yi Sun, 2008), texture filters (Gaeta and Gerardo Iovane, 2013), and covariance region descriptors (Hariri et al., 2017). The facial landmarks eye and nose areas are localized by extracting the principal curvatures and shape index for efficient recognition of facial expressions (Vezzetti et al., 2017). The lip based features are easily extracted by using the geometric descriptors (Moos et al., 2014) and the mean, median, histogram (Vezzetti et al., 2016) features also extracted for 3D FER. The descriptors also formed from two facial components such as Basic Facial Shape Component (BFSC) and Expressional Shape Component (ESC) (Gong et al., 2009). The classification is performed in 3D FER using the different types of classifiers likely multi-SVM (Hariri et al., 2017), HMM (Yi Sun, 2008), neural networks (Hamit Soyel, 2007), Deep Fusion – CNN (Li et al., 2017), Naïve Bayes Classifier (NBC) (Arman Savran, 2017). In 3D FER experimentation, mostly the Binghamton University 3D Facial Expression (BU-3DFE) database and Bosphorus databases are used.

2.4. Database description

Experiments are performed on FER by using various databases likely Japanese Female Facial Expressions (JAFFE, 2017), Cohn – Kanade (CK, 2017), Extended Cohn – Kanade (CK+), MMI (MMI, 2017), Multimedia Understanding Group (MUG, 2017), Taiwanese Facial Expression Image Database (TFEID, 2017), Yale (Yale, 2017), AR face database (AR, 2018), Real-time database (Zhao and Pietikäinen, 2009), Own database (Siddiqi et al., 2015) and Karolinska Directed Emotional Faces (KDEF, 2018).

In most of the experiments, JAFFE database is used. JAFFE holds ten Japanese female's expressions with seven facial expressions and totally 213 images. Each image in JAFFE database contains 256×256 pixel resolution. Some of the sample images of JAFFE database are shown in Fig. 2.

CK database also has seven expressions but it contains 132 subjects that are posed with natural and smile. It contains totally 486 image sequences with 640×490 pixel resolution of gray images. Some of the sample images of the CK database are shown in Fig. 3.

Fig. 2. Sample images from JAFFE database.

Fig. 3. Sample images from CK database.

Table 2
FER Databases description.

Database Name	Origin	Acquisition	Expressions	No. of images	Resolution
Japanese Female Facial Expressions (JAFFE)	Japan	Photos are taken from Kyushu University	Smile, sad, surprise, anger, fear, disgust, neutral	213	256 × 256
Yale	California	Photos are taken from U.C. San Diego Computer vision Laboratory	Happy, normal, sad, sleepy, surprised, wink.	165	168 × 192
Cohn Kanade (CK)	United States	Photos are taken by Panasonic WV3230 cameras	Joy, surprise, anger, fear, disgust, sadness	486	640 × 490
Extended Cohn Kanade (CK+)	United States	Photos are taken by Panasonic AG-7500 cameras	Neutral, sadness, surprise, happiness, fear, anger, contempt and disgust	593	640 × 490
Multimedia Understanding Group (MUG)	Caucasian	High resolution and no occlusion photos are taken	Neutral, sadness, surprise, happiness, fear, anger, and disgust	1462	896 × 896
AR face database	Spain	Photos are taken by Sony 3CCD cameras	Neutral, smile, anger, scream	4000	768 × 576
MMI	Netherlands	Photos are taken by JVC GR-D23E Mini-DV cameras	Disgust, Happiness, surprise, neutral, surprise, sad, fear	250	720 × 576
Taiwanese Facial Expression Image Database (TFEID)	Taiwan	Photos are taken by two CCD cameras simultaneously with different angles (0°,45°)	Neutral, anger, contempt, disgust, fear, happiness, sadness, surprise	7200	600 × 480
Karolinska Directed Emotional Faces (KDEF)	Sweden	Photos are taken by Pentax LX cameras	Angry, Fearful, Disgusted, Sad, Happy, Surprised, Neutral	490	762 × 562

Table 2 shows the origin, acquisition, expression types, number of images, resolution details of the FER databases. The Real-time dataset is also used for FER which contains nearly 2250 images for six expressions and another one own dataset is used which contains 687 image pairs with 640 × 480 resolution.

3. Performance comparison

The performance comparison of this survey is based on the complexity rate, recognition accuracy on different databases, availability of preprocessing and feature extraction methods, expression count analysis, major contribution and advantages of the various FER techniques.

The complexity rates of the various FER techniques are shown in Fig. 4. The x-axis indicates the complexity value of various FER techniques and the y-axis indicates the name of the FER methods. The complexity value of each method is calculated from its own papers which is categorized into three levels are less, medium and high. In Fig. 4, the less complexity is denoted as 1, the medium complexity is denoted as 2 and the high complexity is denoted as 3. The complexity rates are less in Gabor functions, DCT, LBP and WLD comparing with other methods.

The Accuracy rates of the various FER techniques are plotted in Fig. 5 where the x-axis indicates the name of the FER methods and the y-axis indicates the percentage of accuracy acquired in FER techniques. The accuracy of each method is analyzed from its own papers and difference databases are used in every paper, so the mean of the accuracy rate is calculated. The methods such as Gabor functions and DCT with SVM classifier give better accuracy. LBP and WLD descriptors with the pair wise classifiers give better accuracy rate.

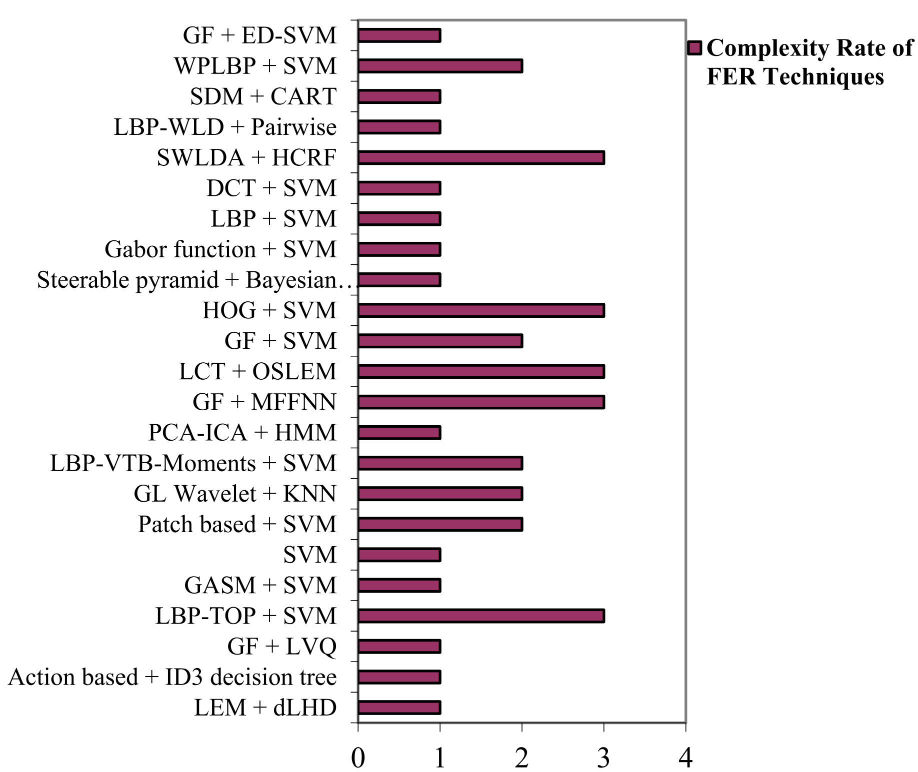

Fig. 4. Complexity rate of various FER techniques.

The availability of preprocessing and feature extraction is shown in Fig. 6. The x-axis indicates the author name of the various FER techniques. The y-axis denotes the availability of preprocessing and feature extraction methods in survey papers. The availability of preprocessing and the feature extraction calculation is based on the presence of the preprocessing and feature extraction in FER papers. If the preprocessing is a presence in the paper then it is denoted as 1 otherwise denoted as 0 and it is represented as 0.1 for visible. Likewise the same procedure for calculating the availability of feature extraction in FER papers.

The expression count analysis of FER techniques are described in Fig. 7 here the x-axis denotes the name of the FER methods and the y-axis denotes the number of recognized expressions using the FER methods. The expression count is analyzed from its own papers and the maximum of seven numbers of expressions are recognized in most papers.

The performance analysis of various FER techniques is described in Table 3. It includes the various fields such as author name, year, FER method name, database name, complexity rate, recognition accuracy, number of expressions recognized, major contributions and advantage of FER techniques. The author name and year field of the table denote the authors of various FER papers and the year denotes the publishing year of the FER papers. The FER method name field of the table describes the methods used for recognition

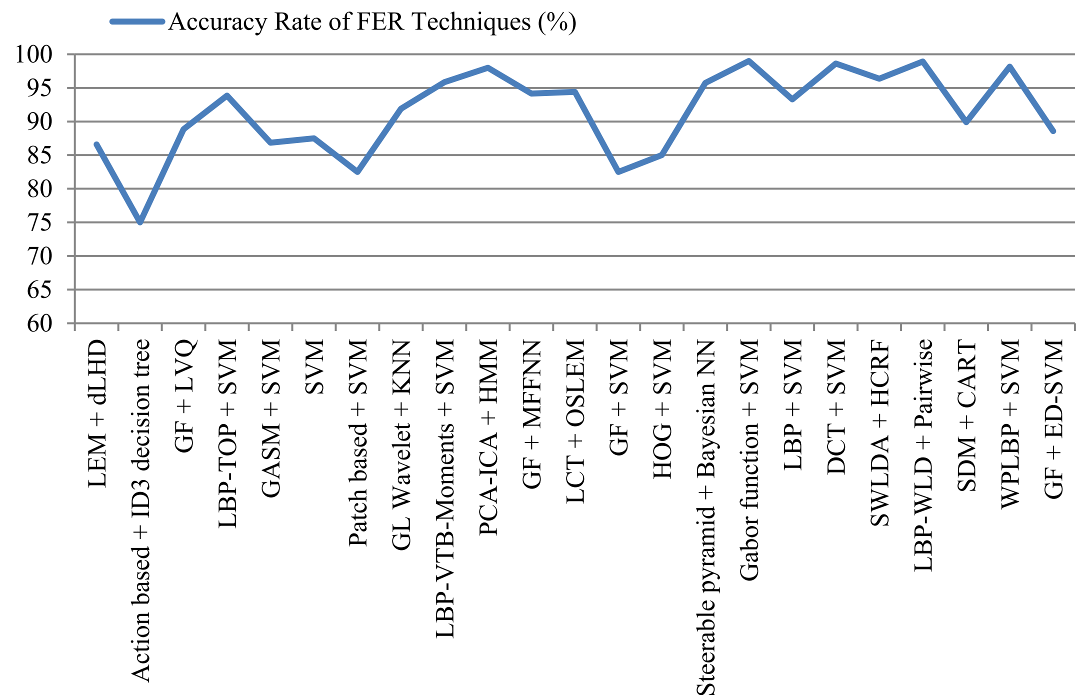

Fig. 5. Accuracy rate of various FER techniques.

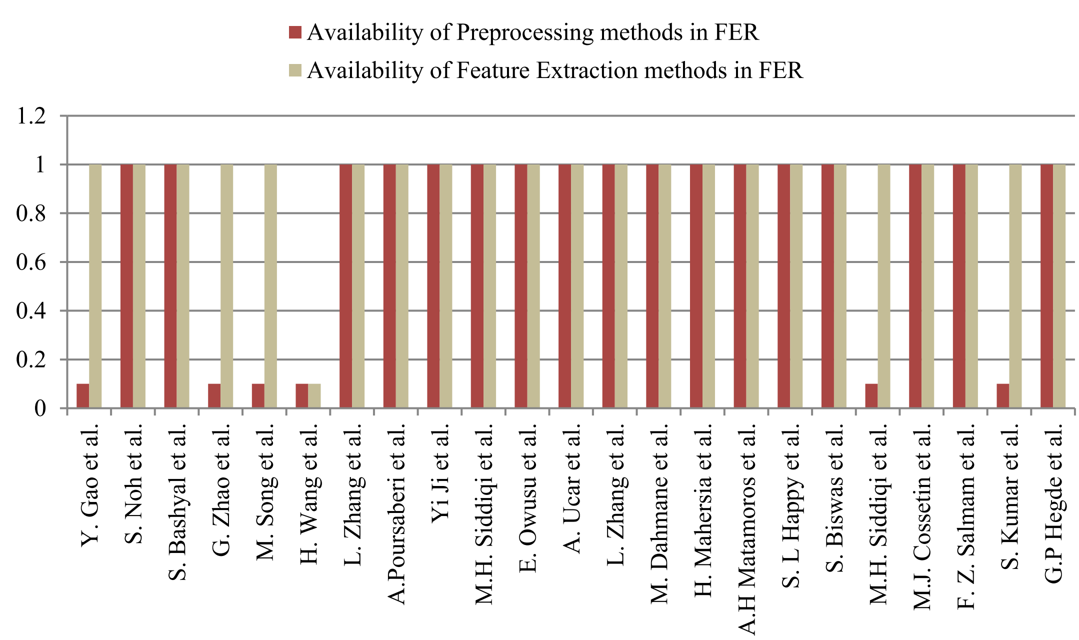

Fig. 6. Availability of preprocessing and feature extraction.

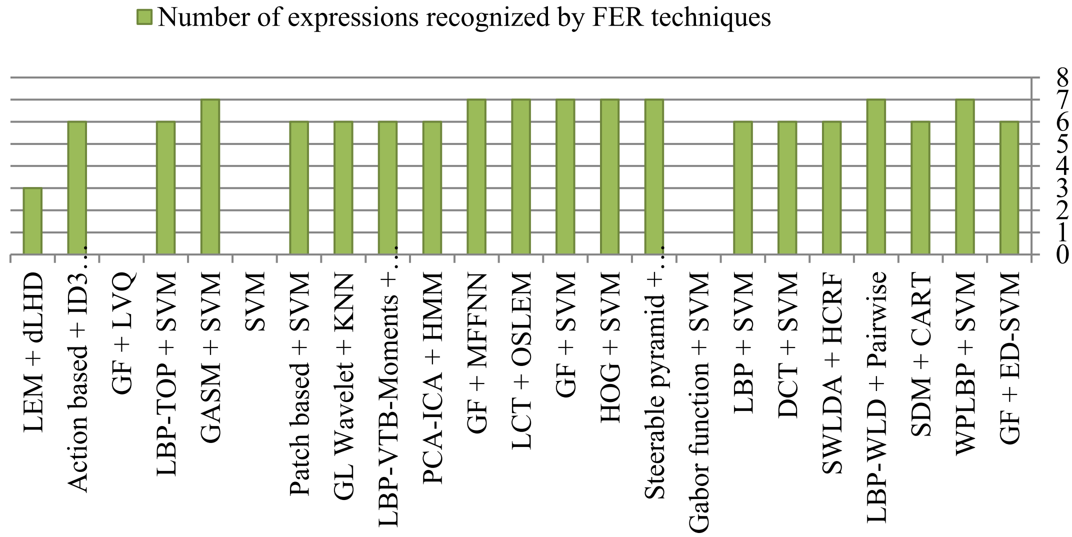

Fig. 7. Expression count analysis.

Table 3
Performance analysis of FER techniques.

Author name, year	FER method name	Database name	Complexity	Recognition accuracy (%)	No. of expressions recognized	Major contribution	Advantages
Gao et al. (2003)	LEM, dLHD	AR	Less	86.6	3	Oriented structural features are extracted	Suitable for real time applications
Noh et al. (2007)	Action based, ID3 decision tree	JAFFE	Less	75	6	Facial features are discriminative & non discriminative	Cost effective in speed and accuracy
Bashyal et al. (2008)	GF, LVQ	JAFFE	Less	88.86	Not reported	LVQ performs better recognition for fear expressions	Better accuracy for fear expressions
Zhao and Pietikäinen (2009)	GASM, SVM	CK	High	93.85	6	Adaboost learning for multi resolution features	Flexible feature selection
Song et al. (2010)	LBP-TOP, SVM	JAFFE, CK Realtime	Less	86.85	7	Detection of facial features point motion & image ratio features	More robust to lighting variations
Wang et al. (2010)	SVM	JAFFE	Less	87.5	Not reported	DKFER for emotion detection	More efficient emotion detection
Zhang et al. (2011)	Patch based, SVM	JAFFE, CK	Less	82.5	6	Capture facial movement features based on distance features	Effective recognition performance
Poursaberi et al. (2012)	GL Wavelet, KNN	JAFFE, CK, MMI	Medium	91.9	6	Extraction of texture and geometric information	Wealthy capability for texture analysis
Ji and Idrissi (2012)	LBP, VTB, Moments, SVM	CK, MMI	Medium	95.84	6	Extraction of spatial temporal Features	Effective image based recognition
Taylor et al. (2014)	PCA, ICA, HMM	Own	Less	98	6	Multilayer scheme to conquer similarity problems	High accuracy with own dataset
Owusu et al. (2014)	GF, MFFNN	JAFFE, Yale	High	94.16	7	Feature selection based on Adaboost	Lowest computational cost
Demir (2014)	LCT, OSLEM	JAFFE, CK	High	94.41	7	Extraction of statistical features mean, entropy and S.D	Reliable algorithm for recognition
Zhang et al. (2014)	GF, SVM	JAFFE, CK	Less	82.5	7	Template matching for finding similar features	High robustness & fast processing speed
Dahmane and Meunier (2014)	HOG, SVM	JAFFE	High	85	7	SIFT flow algorithm for face Alignment	Robust to rotation, occlusion & clutter
Mahersia and Hamrouni (2015)	Streerable pyramid, Bayesian NN	JAFFE, CK	Less	95.73	7	Statistical features are extracted from the steerable representation	Robust features & achieve good results
Hernandez-matamoros et al. (2015)	Gabor function, SVM	KDEF	Less	99	Not reported	Segmentation of face into two Regions	High performance with low cost

Table 3 (*continued*)

Author name, year	FER method name	Database name	Complexity	Recognition accuracy (%)	No. of expressions recognized	Major contribution	Advantages
Happy et al. (2015)	LBP, SVM	JAFFE, CK+	Less	93.3	6	Facial landmarks lip and eyebrow corners are detected	Lower computational complexity
Biswas (2015)	DCT, SVM	JAFFE, CK	Less	98.63	6	Each image is decomposed up to fourth level	Very fast & high accuracy
Siddiqi et al. (2015)	SWLDA, HCRF	JAFFE, CK+,MMI, Yale	High	96.37	6	Expressions are categorized into 3 major categories	High accuracy
Cossetin et al. (2016)	LBP, WLD, Pairwise classifier	JAFFE, CK, TFEID	Less	98.91	7	Each pair wise classifier uses a particular subset	High accuracy & less computation power
Salmam et al. (2016)	SDM, CART	JAFFE, CK	Less	89.9	6	Decision tree for training	Improved recognition accuracy
Kumar et al. (2016)	WPLBP, SVM	JAFFE, CK+, MMI	Medium	98.15	7	Extraction of discriminative features from informative face regions	Lower misclassification
Hegde et al. (2016)	GF, ED, SVM	JAFFE, Yale	Less	88.58	6	Projects feature vector space into low dimension space	Improves the recognition efficiency

of facial expressions. The databases used in the FER papers are JAFFE, CK, CK+, MMI, MUG, TFEID, AR, Yale, KDEF (Karolinska Directed Emotional Faces), Real-time and own dataset. The complexity rate of various FER techniques are denoted as less, medium, high and it is also illustrated in Fig. 4. The recognition accuracy of the different techniques is from 75% to 99% and it is also illustrated in Fig. 5. The number of expressions recognized in the FER survey papers is 7. The LEM method recognizes only 3 expressions and the majority of paper recognizes 6 or 7 expressions. The major contribution field of this table describes the major work involved in the FER papers and the advantage field indicates the benefits of the FER techniques.

From this table clearly understand the combination of preprocessing method ROI segmentation, feature extraction method GF and classification method SVM gives better FER accuracy 99% and less complexity which are analyzed by using the KDEF database. Comparing with other FER methods SVM classification is mostly used which classifies the maximum 7 number of expressions. From this table JAFFE, CK databases are frequently used in many papers and the Real-time dataset is used with the SVM classifier which gives 86.85% accuracy.

4. Conclusion

The important future enhancements described from recent papers are FER for side view faces using the subjective information of facial sub-regions and use different parameters to represent the pose of the face for real-time applications. FER is used in real-time applications such as driver sate surveillance, medical, robotics interaction, forensic section, detecting deceptions. This survey paper is useful for software developers to develop algorithms based on their accuracy and complexity. Also, it is helpful for hardware implementation to implement with low cost depends on their need. This survey compares algorithms based on preprocessing, feature extraction, classification and major contributions. The performance analysis is done based on the database, complexity rate, recognition accuracy and major contributions. This survey discusses the properties such as availability of preprocessing and feature extraction and expression count. The power of algorithms, advantages are discussed elaborately to reach the aim of this survey. ROI segmentation method is used for preprocessing and it gives the highest accuracy 99%. According to feature extraction GF have less complexity which gives the accuracy always between 82.5% and 99%. The highest recognition accuracy of 99% is provided by the SVM classifier and it recognizes the several expressions such as disgust, sad, smile, surprise, anger, fear, neutral effectively. In 2D FER, mostly JAFFE and CK database are used for efficient performance than the other databases.

References

Arman Savran, B.S., 2017. Non-rigid registration based model-free 3D facial expression recognition. Comput. Vis. Image Underst. 162, 146–165. https://doi.org/10.1016/j.cviu.2017.07.005.

B, K.S., Rameshan, R., 2017. Dictionary Based Approach for Facial Expression Recognition from Static Images. Int. Conf. Comput. Vision, Graph. Image Process. pp. 39–49.

Bashyal, S., Venayagamoorthy, G.K.V., 2008. Recognition of facial expressions using Gabor wavelets and learning vector quantization. Eng. Appl. Artif. Intell. 21, 1056–1064. https://doi.org/10.1016/j.engappai.2007.11.010.

Biswas, S., 2015. An Efficient Expression Recognition Method using Contourlet Transform. Int. Conf. Percept. Mach. Intell. pp. 167–174.

Boqing Gong, YuemingWang, Jianzhuang Liu, X.T., 2009. Automatic Facial Expression Recognition on a Single 3D Face by Exploring Shape Deformation. Proc. 17th ACM Int. Conf. Multimed. pp. 569–572.

Chang, H.T.Y., 2017. Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Comput. 22, 4389–4405. https://doi.org/10.1007/s00500-017-2634-3.

Clawson, K., Delicato, L.S., Bowerman, C., 2018. Human Centric Facial Expression Recognition. Proc. Br. HCI pp. 1–12.

Cossetin, M.J., Nievola, J.C., Koerich, A.L., 2016. Facial expression recognition using a pairwise feature selection and classification approach. IEEE Int. Jt. Conf. Neural Networks, pp. 5149–5155.

Cui, R., Liu, M., Liu, M., 2016. Facial expression recognition based on ensemble of mulitple cNNs. Chinese Conf. Biometric Recognit. 511–518. https://doi.org/10.1007/978-3-319-46654-5.

Dahmane, M., Meunier, J., 2014. Prototype-based modeling for facial expression analysis. IEEE Trans. Multimed. 16, 1574–1584.

Demir, Y., 2014. A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. Neural Comput. Appl. 27, 131–142. https://doi.org/10.1007/s00521-014-1569-1.

Ekman, P., Friesen, W., 2003. Unmasking the Face: A Guide to Recognizing Emotions From Facial Clues.

Vezzetti, Enrico, Marcolin, Federica, Tornincasa, Stefano, Luca Ulrich, N.D., 2017. 3D geometry-based automatic landmark localization in presence of facial occlusions. Multimed. Tools Appl. 77, 14177–14205. https://doi.org/10.1007/s11042-017-5025-y.

Vezzetti, Enrico, Tornincasa, Stefano, Moos, Sandro, Marcolin, Federica, Violante, Maria Grazia, Speranza, Domenico, David Buisan, F.P., 2016. 3D human face

analysis: automatic expression recognition. Proc. Biomed. Eng. https://doi.org/10.2316/P.2016.832.
Gao, Y., Leung, M.K.H., Hui, S.C., Tananda, M.W., 2003. Facial expression recognition from line-based caricatures. IEEE Trans. Syst. Man Cybern. A Syst. Hum. 33, 407–412.
Guo, M., Hou, X., Ma, Y., 2016. Facial expression recognition using ELBP based on covariance matrix transform in KLT. Multimed. Tools Appl. 76, 2995–3010. https://doi.org/10.1007/s11042-016-3282-9.
Hamit Soyel, H.D., 2007. Facial expression recognition using 3D facial feature distances. Int. Conf. Image Anal. Recognit., 831–838
Happy, S.L., Member, S., Routray, A., 2015. Automatic facial expression recognition using features of salient facial patches. IEEE Trans. Affect. Comput. 6, 1–12.
Hegde, G.P., Seetha, M., Hegde, N., 2016. Kernel locality preserving symmetrical weighted fisher discriminant analysis based subspace approach for expression recognition. Eng. Sci. Technol. Int. J. 19, 1321–1333. https://doi.org/10.1016/j.jestch.2016.03.005.
Hernandez-matamoros, A., Bonarini, A., Escamilla-hernandez, E., Nakano-miyatake, M., 2015., A Facial Expression Recognition with Automatic Segmentation of Face Regions. Int. Conf. Intell. Softw. Methodol. Tools, Tech. 529–540. doi: 10.1007/978-3-319-22689-7.
Li, Huibin, Sun, Jian, Zongben, Xu., Liming Chen, M., 2017. Multimodal 2D + 3D facial expression recognition with deep fusion convolutional neural network. IEEE Trans. Multimed. 19, 2816–2831. https://doi.org/10.1109/TMM.2017.2713408.
Islam, D.I., Anal, S.R.N., Datta, A., 2018. Facial expression recognition using 2DPCA on segmented images. Adv. Comput. Commun. Paradig. 289–297. https://doi.org/10.1007/978-981-10-8237-5.
Jain, S., Durgesh, M., Ramesh, T., 2016. Facial expression recognition using variants of LBP and classifier fusion. Proc. Int. Conf. ICT Sustain. Dev. 725–732. https://doi.org/10.1007/978-981-10-0129-1.
Ji, Y., Idrissi, K., 2012. Automatic facial expression recognition based on spatiotemporal descriptors. Pattern Recognit. Lett. 33, 1373–1380. https://doi.org/10.1016/j.patrec.2012.03.006.
Kumar, S., Bhuyan, M.K., Chakraborty, B.K., 2016. Extraction of informative regions of a face for facial expression recognition. IET Comput. Vis. 10, 567–576. https://doi.org/10.1049/iet-cvi.2015.0273.
Li, J., Lam, E.Y., 2015. Facial expression n recognition using deep neural networks. IEEE Int. Conf. Imaging Syst. Tech.
Lv, Y., 2015. Facial expression recognition via deep learning. Int. Conf. Smart Comput. doi: 10.1109/SMARTCOMP.2014.7043872
M. Anisetti, V. Bellandi, F.B., 2005. Accurate 3D Model based Face Tracking for Facial Expression Recognition. Proc. Int. Conf. Vis. Imaging, Image Process. 93–98.
Mahersia, H., Hamrouni, K., 2015. Using multiple steerable fi lters and Bayesian regularization for facial expression recognition. Eng. Appl. Artif. Intell. 38, 190–202. https://doi.org/10.1016/j.engappai.2014.11.002.
Gaeta, Matteo, Gerardo Iovane, E.S., 2013. A 3D geometric approach to face detection and facial expression recognition. J. Discret. Math. Sci. Cryptogr. 9, 39–53. https://doi.org/10.1080/09720529.2006.10698059.
Nazir, M., Jan, Z., Sajjad, M., 2017. Facial expression recognition using histogram of oriented gradients based transformed features. Cluster Comput. https://doi.org/10.1007/s10586-017-0921-5.
Nigam, S., Singh, R., Misra, A.K., 2018. Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. Multimed. Tools Appl., 1–23
Noh, S., Park, H., Jin, Y., Park, J., 2007. Feature-adaptive motion energy analysis for facial expression recognition. Int. Symp. Vis. Comput., 452–463
Owusu, E., Zhan, Y., Mao, Q.R., 2014. A neural-ada boost based facial expression recognition system. Expert Syst. Appl. 41, 3383–3390. https://doi.org/10.1016/j.eswa.2013.11.041.
Poursaberi, A., Noubari, H.A., Gavrilova, M., Yanushkevich, S.N., 2012. Gauss – Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. EURASIP J. Image Video Process., 1–13
Rahul, R.C., Cherian, M., 2016. Facial expression recognition using PCA and texture-based LDN descriptor. Proc. Int. Conf. Soft Comput. 113–122. https://doi.org/10.1007/978-81-322-2674-1.
Rashid, T.A., 2016. Convolutional neural networks based method for improving facial expression recognition. Intell. Syst. Technol. Appl. 73–84. https://doi.org/10.1007/978-3-319-47952-1.
Ryu, B., Member, S., Kim, J., 2017. Local directional ternary pattern for facial expression recognition. IEEE Trans. Image Process. 26, 6006–6018. https://doi.org/10.1109/TIP.2017.2726010.
Salmam, F.Z., Madani, A., Kissi, M., 2016. Facial Expression Recognition using Decision Trees. IEEE 13th Int. Conf. Comput. Graph. Imaging Vis., 125–130. doi: 10.1109/CGiV.2016.33.
Marcolin, Federica, Moos, Sandro, Tornincasa, Stefano, Vezzetti, Enrico, Violante, Maria Grazia, Fracastoro, Giulia, Domenico Speranza, F.P., 2014. Cleft lip pathology diagnosis and foetal landmark extraction via 3D geometrical analysis. Int. J. Interact. Des. Manuf. 11, 1–18. https://doi.org/10.1007/s12008-014-0244-1.
Shan, K., Guo, J., You, W., Lu, D., Bie, R., 2017. Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural-Network Structure. IEEE 15th Int. Conf. Softw. Eng. Res. Manag. Appl. 123–128.
Siddiqi, M.H., Ali, R., Khan, A.M., Park, Y., Lee, S., 2015. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans. Image Process. 24, 1386–1398.
Song, M., Tao, D., Liu, Z., Li, X., Zhou, M., 2010. Image ratio features for facial expression recognition application. IEEE Trans. Syst. Man Cybern. Part B Cybern. 40, 779–788.
Taylor, P., Siddiqi, M.H., Ali, R., Sattar, A., Khan, A.M., Siddiqi, M.H., Ali, R., Sattar, A., Khan, A.M., Lee, S., 2014. Depth camera-based facial expression recognition system using multilayer scheme. IETE Tech. Rev. 31, 277–286. https://doi.org/10.1080/02564602.2014.944588.
Hariri, Walid, Tabia, Hedi, Farah, Nadir, Abdallah Benouareth, D.D., 2017. 3D facial expression recognition using kernel methods on Riemannian manifold. Eng. Appl. Artif. Intell. 64, 25–32. https://doi.org/10.1016/j.engappai.2017.05.009.
Wang, H., Hu, Y., Anderson, M., Rollins, P., 2010. Emotion Detection via Discriminative Kernel Method. Proc. 3rd Int. Conf. PErvasive Technol. Relat. to Assist. Environ.
Wu, Y., Qiu, W., 2017. Facial Expression Recognition based on Improved Deep Belief Networks. AIP Conf. Proc. 1864. doi: 10.1063/1.4992947.
Yang, Y., Office, T.A., Fang, D., Zhu, D., Science, I., 2016. Facial expression recognition using deep belief network. Rev. Tec. Ing. Univ. Zulia 39, 384–392.
Yi Sun, L.Y., 2008. Facial expression recognition based on 3D dynamic range model sequences. Eur. Conf. Comput. Vis., 58–71
Zhang, C., Wang, P., Chen, K., 2017. Identity-aware convolutional neural networks for facial expression recognition. J. Syst. Eng. Electron. 28, 784–792 https://doi.org/10.21629/JSEE.2017.04.18.
Zhang, L., Member, S., Tjondronegoro, D., 2011. Facial expression recognition using facial movement features. IEEE Trans. Affect. Comput. 2, 219–229.
Zhang, L., Tjondronegoro, D., Chandran, V., 2014. Random Gabor based templates for facial expression recognition in images with facial occlusion. Neurocomputing 145, 451–464. https://doi.org/10.1016/j.neucom.2014.05.008.
Zhao, G., Pietikäinen, M., 2009. Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recognit. Lett. 30, 1117–1127. https://doi.org/10.1016/j.patrec.2009.03.018.
Zhao, X., Zhang, S., 2016. A review on facial expression recognition : feature extraction and classification. IETE Tech. Rev. 33, 505–517. https://doi.org/10.1080/02564602.2015.1117403.
The Japanese Female Facial Expression (JAFFE) Database, "http://www.kasrl.org/jaffe.html", accessed on 2017.
Cohn-Kanade AU-Coded Expression Database, "http://www.pitt.edu/~emotion/ckspread.htm", accessed on 2017.
MMI facial expression database, "https://mmifacedb.eu/", accessed on 2017.
Multimedia Understanding Group (MUG) Database, https://mug.ee.auth.gr/fed, accessed on 2017.
Taiwanese Facial Expression Image Database (TFEID), http://bml.ym.edu.tw/tfeid, accessed on 2017.
Yale face database, http://vision.ucsd.edu/content/yale-face-database, accessed on 2017.
AR face database, http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html, accessed on 2018.
The Karolinska Directed Emotional Faces (KDEF) database, http://www.emotionlab.se/kdef/, accessed on 2018.