International Conference on Artificial Intelligence and Machine Learning
In association withInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com) | Volume 8, Issue 5, July-August-2021

The Comprehensive Study of Facial Expression Recognition on Video

Ratnalata Gupta*[1], L. K. Vishwamitra[1]
[1]Department of Computer Science, Oriental University, Indore, Madhya Pradesh, India

ABSTRACT

In human life, facial expressions and emotions reveal external and internal responses. In human-computer interaction, the video clip plays an important part for extract the emotion of the end-user. In this type of system, it is necessary to observe rapid dynamic changes in the motion of the human face to give the essential response. A real-time application is based on identifying the face and recognizes his expressions such as happy, sad, disgruntled and tiredness, etc. for example Driver exhaustion detection to prevent road accidents on the road. We focus on the comprehensive study of different traditional and recent methodsfor preprocessing, features take out and emotion recognition is employed in facial expression and recognition systems (FER). This paper also describes various Terminologies are used in the Facial expression and recognition system (FER) system. The result compares with the Number of Expressions, algorithm accuracy, and implementation tools. Video-based facial emotion recognition is a very interesting and challenging problem hence the present study provides model complexity, implementation trends, and opportunities for researchers can consider as future research works.

Keywords: Facial Expression Recognition, Deep Learning, Classification.

I. INTRODUCTION

Today, Facial expressions are playing a prominent role in our day-to-day life for communication, human emotions, and intentions which play a vital role in decision-making perception such as physical fatigue detection [1]. Facial expression and emotions are the key features of nonverbal communication and this communication generally has eye gaze, gestures, postures, and body movements. Eyecontact is very important to interpersonal communication [2]. The face is also another important communicator. An expression is the type of emotion or feelings such as happiness, sorrow, anger, confusion, fear, disgust, and surprise. Facial expression and recognition systems (FER) have to contain feature extraction and classification (emotion detection). FER methods have two approaches first is conventional and the second one is deep learning-based approaches. We discuss methods and the Terminology used in the FER system [3].

II. RESEARCH BACKGROUND OFFER

Usually, the FER system uses a conventional method or deep learning method [3].The convention approach first detected the face and its components from the image, and then they extract the features and classify emotions. This method is less dependent on data and hardware because it is based on manual feature extraction and it required small data for analysis. The deep convolutional neural network approach removes and minimizes the dependence onpre-processing techniques by employing "end-to-end" trainable learning. The input data train with a different source for classification [4]. The convolution neural network (ConvNet) is the most popular deep learning technique [2].

A. The facial expression recognition system

The Face-detection and recognition approaches include three key stages such as image pre-processing, feature extraction, and classification. Fig.1 shows the processes of the facial recognition system [3,4].

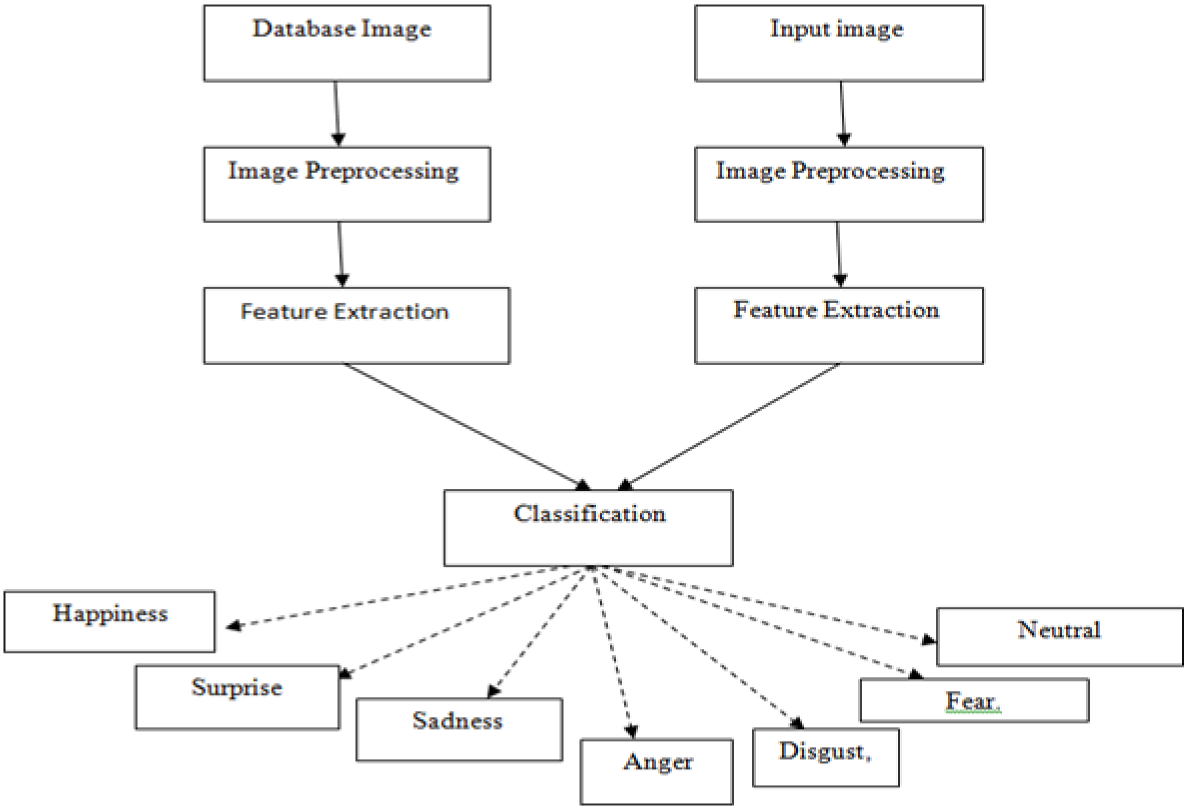


Figure1: Overview of FER system

B. Face image preprocessing

Image preprocessing is done before the feature extraction process to enhance the overall performance of the FER system. The different types of Processes included in Image preprocessing viz. image size and quality and scaling, the brightness of an image, normalization, and additional improvement processes to enhance the recognition rate. The cropping is performed to extract the detected face from the entire image, and image scaling is the process of increasing and decreasing the size of a face image, for this interpolation method is used [4, 5] and Gaussian filter (GF) is used for noisereduction and reducing the size of an image [6]. Normalization is a preprocessing method that reduces the illumination and variations of facial images, it also reduces redundant information such as background, hair, neck, shoulders, sunglasses, etc. The Viola-Jones algorithm is used for Localization and finding a face from an image. It is irrespective of its size, situation, and surroundings. The algorithm uses four key steps (1) haar-like feature (2) create an integral image (3) Adaboost training (4) cascading classifiers for face identification [7]. The Adaptive-Boosting algorithm and a Haar-like feature are used for representing the rectangular region of an image at a particular location [7.8].The Scale-Invariant Feature Transform(SIFT) algorithm is used for feature detection and face alignment. It is a technique for detection and also describes local feature points in an image. These features must not change with Scale, Illumination, and rotation [9]. The histogram equalization method is an image processing technique that is used to increase contrast and adjust images by using histograms [7]. Region of interest (ROI) is the best preprocessing algorithm. It has based on image segmentation in which different types of functions are performed such as1.The settlement of the face dimension is based on color, thresholding, 2. Forehead, eye, and Mouth segmentation. This method is suitable for the findingof a face in an image[10].

C. Feature Extraction

The FER system is generallyclassified into various categories like Textureand shapedescriptor Methods, edgedetection techniques, Global and Local appearance-based approaches, analytic (geometric) approach, and small regions (patch) approach [3, 10]. Gabor filter and Local Binary Pattern(LBP)based method is used for extracting the features and it is based on the texture analysis technique [11,12]. Texture descriptors for images can be obtained by using thresholds its differences between neighborhood pixels and central pixels[13,14]. The Gaussian Laguerre (GL) [15] uses a single filter and there is no need for multiple filters for facial texture features extraction. Weber Local-Descriptor (WLD) [16] is a robust local descriptor and it depends on the actual intensity of the excitation. WLD consists of two portions 1. Differential excitation 2. Gradient orientation, by using this component a WLD histogram is made. Supervised Descent Method extracted main and related positions of the face and estimates the distance between various components [17]. Weighted Projection-based LBP (WP-LBP) extracts the instructive LBP features and that is weighted [18]. Discrete Wavelet Transform (DWT) is a method for converting image pixels to wavelets and is employed for a large image problem. It compresses the image and provides limited directional information. The feature which is not present in DWT provides a contourlet transform (CT) approach such as multi-scale and directional segregation a high degree of directionality and anisotropy which uses a double filter bank laplacian pyramid and Directional Filter Bank (DFB)[19]. The Local Curvelet Transform (LCT) is a geometric feature descriptor that uses average, entropy, and standard deviation to the local regions [20]. The edge-based methods combine with template matching and geometrical feature matching [21]. Active Shape Model(ASM) and Active Appearance Model(AAM) are used for Face alignment which is used as, statistical model. [7]. Histogram of oriented gradients (HOG) algorithm takes out features from an image which is done by the gradient and orientation of the edges. [21,22]. Features describe changes in facial texture when a particular action occurs viz crease(Jhurriyan), bulges, obverse, the space surrounding the mouth, and eyes. The feature vector is extracted by an image filter which is used to either the entire face or specific areas. Appearance-based algorithms are broad-based, it includes the principal component analysis (PCA) algorithm, which is used for the reduction of higher to lower dimensional space[2], independent component analysis(ICA)[23,24] is an extension to PCA it is the statistical and computational approach. Stepwise LinearDiscriminant Analysis (SWLDA) [25] algorithm takes out the small set of localized features by implementing (forward and backward) regression models, it is based on F-test values and is calculated by class labels. Texture feature-based descriptors are more batter than others so now day Local Directional Number Pattern (LDN), Local Ternary Pattern (LTP) [26], Karhunen-Loeve Transform Extended LBP (KELBP), and in recent years use Discrete wavelet transform (DWT) algorithm decompose input data for recognizing expression using time scale, in the FER system [7].

D. Classification

The last stage of the FER System is emotion detection and the output of the previous stage gives as an input to the classifier in which the classifier classifies expressions such as Fear, Happy, sadness, Surprise, Anger, and Disgust, etc. The Line segment Hausdorff Distance (LHD) is used for image matching and the Euclidean distance metric defines the distance between two sets of the points both methods used for classification purposes [3]. KNN (k-Nearest Neighbors) [17], SVM (Support Vector Machine) [24], Convolutional neural network (CNN) [21], Adaptive Boosting (Adaboost) [8], Bayesian, Sparse Representation based Classifier (SRC)[2]. KNN is a kind of lazy learning it takes more time for predicting. It works on instance-based learning techniques and compares test samples with training datasets, its classification decision is based on a small neighborhood of similar objects(analogy). The KNN algorithm is simple for all machine learning algorithms and easy to implement [17]. SVM (Support Vector Machine) [24] algorithm is suitable for linear and non-linear data for Classification and regression problems. It creates decision boundaries in high n-dimensional spaces and finds out the best decision boundary. This is called a hyperplane which is helpful for classification. Convolutional neural networks (CNNs) performances significantly high in classification [27]. CNNs consist of two steps 1. a convolution and pooling layer extract features and analyze image 2. the output of the first steps takes as input in a fully connected layer, this layer detects emotions from the image [28]. AdaBoost is an ensemble learning algorithm, so it builds one strong learner classifier such an algorithm may be less overfitting problem as compared to other classification algorithms [8]. Naive Bases classifiers algorithms based on Bayes' Theorem. It is highly scalable and required a set of linear parameters for the learning problems [21]. Deep Neural Network (DNN) [28] is a supervised learning method it contains multiple layers and Feed Forward Neural Networks. The Deep Belief Network (DBN) [2] is a robust learning network that contains multiple layers with values and defines as a stack of Restricted Boltzmann machines (RBM) layers that use the backpropagation layer (BP) for commutation. Which has two steps pre-training and fine. It optimizes the weights by minimizing the cross-entropy error [26].

SVM classifier provides better classification and recognition results as compared to other classifiers and CNN gives a better result than other neural network-based classifiers Table 2 shows FER techniques and algorithm that is used in preprocessing, feature extraction, and classification. It shows the different preprocessing methods like Localization, Normalization, Face acquisition, Histogram equalization, ROI segmentation, Haar-like features, Multitask cascaded convolutional network(MTCNN), Viola-Jones, Gabor filter, and Table 2 shows different feature extraction methods like Enhanced Independent Component Analysis (EICA), PCA, PCA-FLDA, ICA, local directional ternary pattern (LDTP) GL Wavelet, Local Binary Pattern(LBP), VTB (vertical time backward), and moments, Principal Component Analysis(PCA), Supervised Descent Method(SDM), WPLBP, Contourlet transforms (CNT) and curvelet transform(CLT), CNN, Attention convolutional network, DSN, DTN, BiLSTM, DBN, HOG, DWT, local directional rank histogram pattern(LDRHP), local directional strength pattern(LDSP), Kernel principal component analysis and generalized discriminant analysis to generate(KPCA-GDA)..

Nowadays preprocessing method recently used the histogram equalization and used for feature extraction LBP method give the best result and for classification SVM and CNN give best result than others

III. TERMINOLOGIES

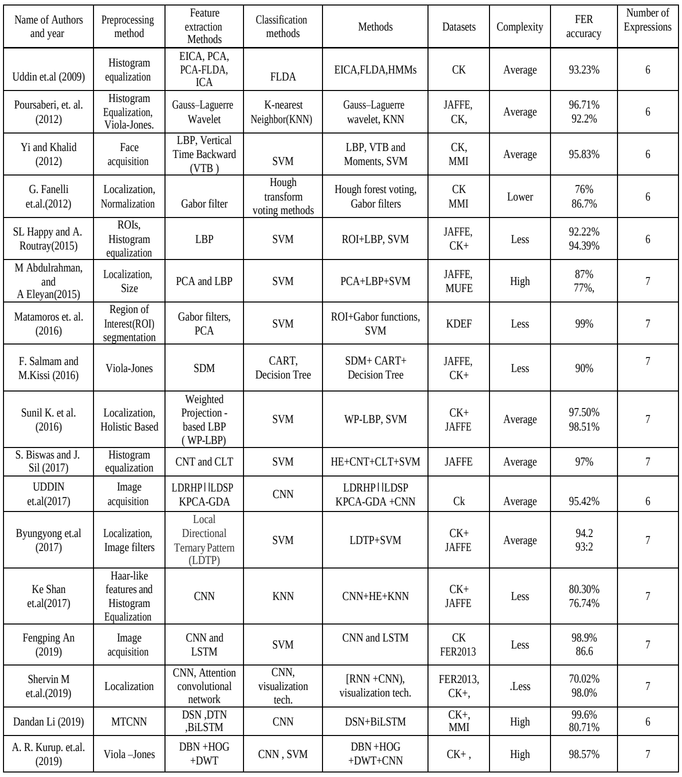
In this paper, we explain some related terminologies which are used in the theory of the FER system. Facial Landmark points (FLs)[29], Action Units (AUs), Facial Action Coding System(FACS), Basic Emotions(BE), Compound Emotions (CE), and Micro Expressions(ME) [3,29] are basic categories of expression. Facial Landmark (FLs) is eyes & eyebrows, nose, and mouth, lips, jawline, etc. are the special landmark. FLs methods are divided into Shape- and Texture- based methods. The shape-based method is split into explicit and implicit methods. The Explicit-based statistical models which can detect landmarks are the Active ShapeModel (ASM) & Active Appearance Model (AAM)[28,29]. The deep neural network identifies multiple landmarks on the entire face without state information that are implicit-based techniques.

PCA & ICA [23]. Facial Action Units (AUs)[28] are related to the contraction of one or more specific facial muscles. Facial expressions are broken down into individual components of muscle movement [29]. The facial Action coding System (FACS) [3] was developed by Paul Ekman and Wallace Friesen, which describes each and every observable change in facial movement based on Action Units (AUs)[3,29].Compound emotions are those emotions that are formed by the combination of basic emotions such as Happily surprised and Happily disgusted there are two different emotional sense categories. In which 22 Compound Emotions are expressed by humans in 7 basic emotions and 12 compound emotions 3 additional emotions are also included appall, hate, and awe[3].

IV. DISCUSSION AND COMPARISON OF RESULTS

We have studied and compared FER techniques to findscope for improvement in video-based FER system.

Table 2 Study of FER Techniques.



Name of Authors and year	Preprocessing method	Feature extraction Methods	Classification methods	Methods	Datasets	Complexity	FER accuracy	Number of Expressions
Uddin et.al (2009)	Histogram equalization	EICA, PCA, PCA-FLDA, ICA	FLDA	EICA,FLDA,HMMs	CK	Average	93.23%	6
Poursaberi, et. al. (2012)	Histogram Equalization, Viola-Jones.	Gauss–Laguerre Wavelet	K-nearest Neighbor(KNN)	Gauss–Laguerre wavelet, KNN	JAFFE, CK,	Average	96.71% 92.2%	6
Yi and Khalid (2012)	Face acquisition	LBP, Vertical Time Backward (VTB)	SVM	LBP, VTB and Moments, SVM	CK, MMI	Average	95.83%	6
G. Fanelli et.al.(2012)	Localization, Normalization	Gabor filter	Hough transform voting methods	Hough forest voting, Gabor filters	CK MMI	Lower	76% 86.7%	6
SL Happy and A. Routray(2015)	ROIs, Histogram equalization	LBP	SVM	ROI+LBP, SVM	JAFFE, CK+	Less	92.22% 94.39%	6
M Abdulrahman, and A Eleyan(2015)	Localization, Size	PCA and LBP	SVM	PCA+LBP+SVM	JAFFE, MUFE	High	87% 77%,	7
Matamoros et. al. (2016)	Region of Interest(ROI) segmentation	Gabor filters, PCA	SVM	ROI+Gabor functions, SVM	KDEF	Less	99%	7
F. Salmam and M.Kissi (2016)	Viola-Jones	SDM	CART, Decision Tree	SDM+ CART+ Decision Tree	JAFFE, CK+	Less	90%	7
Sunil K. et al. (2016)	Localization, Holistic Based	Weighted Projection - based LBP (WP-LBP)	SVM	WP-LBP, SVM	CK+ JAFFE	Average	97.50% 98.51%	7
S. Biswas and J. Sil (2017)	Histogram equalization	CNT and CLT	SVM	HE+CNT+CLT+SVM	JAFFE	Average	97%	7
UDDIN et.al(2017)	Image acquisition	LDRHP||LDSP KPCA-GDA	CNN	LDRHP||LDSP KPCA-GDA +CNN	Ck	Average	95.42%	6
Byungyong et.al (2017)	Localization, Image filters	Local Directional Ternary Pattern (LDTP)	SVM	LDTP+SVM	CK+ JAFFE	Average	94.2 93:2	7
Ke Shan et.al(2017)	Haar-like features and Histogram Equalization	CNN	KNN	CNN+HE+KNN	CK+ JAFFE	Less	80.30% 76.74%	7
Fengping An (2019)	Image acquisition	CNN and LSTM	SVM	CNN and LSTM	CK FER2013	Less	98.9% 86.6	7
Shervin M et.al.(2019)	Localization	CNN, Attention convolutional network	CNN, visualization tech.	[RNN +CNN), visualization tech.	FER2013, CK+,	.Less	70.02% 98.0%	7
Dandan Li (2019)	MTCNN	DSN ,DTN ,BiLSTM	CNN	DSN+BiLSTM	CK+, MMI	High	99.6% 80.71%	6
A. R. Kurup. et.al. (2019)	Viola –Jones	DBN +HOG +DWT	CNN , SVM	DBN +HOG +DWT+CNN	CK+ ,	High	98.57%	7

We observed the precision and execution ofvarious strategies. Figure2 shows the Accuracyrate of different FER methods.

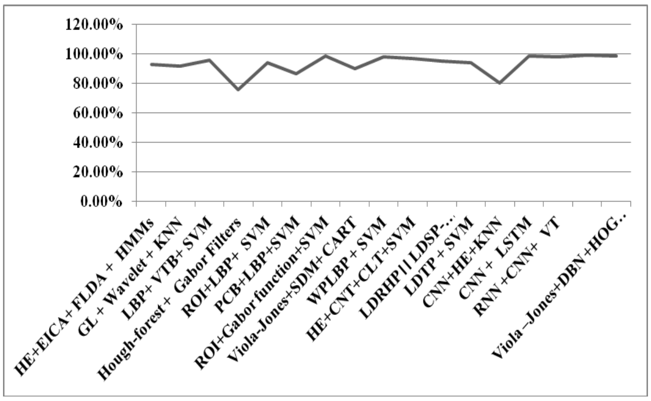


Figure 2 Accuracy rate of various techniques

V. CONCLUSION AND FUTURE WORK

Nowadays Facial Expression Recognition System (FER) is the centre of attraction and for this many algorithms have been developed using different types of parameters to show emotion in real-life applications viz driver monitoring, medicine, robotics Conversation, forensic section, and fraud detection. This paper is useful for analyzation and evolution of any algorithm so researchers can use it for further development of conventional methods and deep learning-based methods of the FER system and introduce some related FER terminology. We study and analyze all various possible methods for pre-processing, feature extraction, and classification. The end of the paper presents a performance analysis of FER techniques that require future research.

The goal of this survey paper is to find the power of algorithms for preprocessing, feature extraction, and classification. ROI segmentation, WPLBP, Histogram equalization with LBP method, CNT and CLT gives high accuracy of 99%, 98.51%, and 97% respectively for preprocessing and feature extraction and GF provides lower complexity which gives accuracy between 76 % to 86%, and the SVM and MTCNN classifier provided 98.9% and 99.62% accuracy respectively with some basic 7 universal emotion such as disgust, sadness, happiness, surprise, anger, fear, and neutral in which CK+ database gives the best result as compared to another database.

VI. REFERENCES

[1]. Chris Frith,” Role of facial expressions in social interactions”, Phil. Trans. R. Soc, pp 3453–3458, (2009).

[2]. S Zhang, X Pan, Y Cui, X Zhao, L Liu., “Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning”, IEEE Access vol 7, pp.32297-32304, (2019).

[3]. F. Khan, “Facial Expression Recognition using FacialLandmark Detection and Feature Extraction via Neural Networks”, Xiv:1812.04510v3 [cs.CV], pp.1-7,(2020).

[4]. M. Yeasin, B. Bullot, R. Sharma, “Recognition of Facial Expressions and Measurement of Levels of Interest from Video”, IEEE transactions on multimedia, Vol. 8, pp.500- 508(2006).

[5]. S Bashyal, GK Venayagamoorthy, “Recognition of facial expressions using Gabor wavelets and learning vector quantization”, Engineering Applications of Artificial Intelligence Elsevier, pp 1056–1064, (2007).

[6]. S. Biswas, J. Sil, “An efficient face recognition method using contourlet and curvelet transform” Journal of King Saud University Elsevier, pp. 1-12, (2017).

[7]. Ebenezer Owusu, Yongzhao Zhan, Qi Rong Mao, “A neural- AdaBoost based facial expression recognition system” Expert Systems with Applications Elsevier, pp. 3383-3390(2014).

[8]. P Viola, M Jones, Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, pp. 511-518, (2001).

[9]. J. Yang, S.Liao, Stan Z. Li “Automatic Partial Face Alignment in NIR Video Sequences”, International Conference on Biometrics Springer, pp. 249–258, (2009).

[10]. Hernandez- Matamoros, A. Bonarini, E. Hernandez, M Nakano- Miyatake, H.Meana “ Facial Expression Recognition with Automatic Segmentation of Face Regions using a Fuzzy based Classification Approach” Knowledge-Based Systems, Vol- 110, Pages 1-14,(2016)

[11]. Ke Shan, Junqi Guo, Wenwan You, Di Lu, RongfangBie “Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural-Network Structure”¸ IEEE 15thInternational Conference on Software Engineering Research, Management and Applications (SERA) ¸ pp: 123-128(2017).

[12]. Caifeng Shan, Shaogang Gong, Peter W. McOwan, “Facial expression recognition based on Local Binary Patterns: A comprehensive study”, Image and Vision Computing, pp. 803– 816, (2009).

[13]. Aliaa A. A. Youssif, Wesam A. A. Asker “, Automatic Facial Expression Recognition System Based on Geometric and Appearance Features”, Canadian Center of Science and Education, Vol. 4, pp 115-124 (2011).

[14]. SL Happy, A. Routray,” Automatic Facial Expression Recognition Using Features of Salient Facial Patches”, IEEE Transactions on Affective Computing, pp. 1-12, (2013).

[15]. A. Poursaberi, A Noubari, M Gavrilova, and N Yanushkevich, ”Gauss–Laguerre wavelet textural feature fusion with geometrical information for facial expression Identification”, EURASIP Journal on Image and Video Processing Springer, pp: 1-13, (2012).

[16]. J Chen, Shan, Chu, Zhao, Matti, Xilin, Wen, “WLD: A Robust Local Image Descriptor”, ieee transactions on pattern analysis and machine intelligence,pp. 1-15, (2009).

[17]. Fatima Salmam, Mohamed KISSI, “Facial Expression Recognition using Decision Trees”,13th International Conference on Computer Graphics, Imaging, and Visualization IEEE, pp. 125–130, (2016).

[18]. Sunil Kumar, M.K. Bhuyan, B. K. Chakraborty, ” Extraction of informative regions of a face for facial expression recognition”, IET Computer Vision Journals vol 10, pp. 567 – 576(2016).

[19]. S. Katsigiannis, G. Keramidas, D. Maroulis, “Contourlet Transform for Texture Representation of Ultrasound Thyroid Images” 6th IFIP WG 12.5 International Conf., AIAI 2010 Springer, pp. 138–145.,(2010).

[20]. Aysegu, Y. Demir, C. Guzelis, ”A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering”, Neural Comput&Applic Springer, pp. 131–142,(2016).

[21]. Md. Uddin, W Khaksar, J Torresen, “Facial Expression Recognition Using Salient Features and Convolutional Neural Network”, a special section on visual surveillance and biometrics: practices, challenges, and possibilities IEEE Access Vol 5, pp: 26146 – 26161, (2017).

[22]. M. Nazir, Z. Jan, M. Sajjad,” Facial expression recognition using histogram of oriented gradients based transformed features”, Cluster Comput DOI 10.1007/s10586-017-0921-5 Springer, pp: 1-10, (2017).

[23]. Yi Ji, Khalid Idrissi," Automatic facial expression recognition based on spatiotemporal descriptors", Pattern Recognition Letters Elsevier, pp 1373-1380, (2012).

[24]. M. Abdulrahman, Alaa Eleyan,"Facial Expression Recognition Using Support Vector Machines", 23nd Signal Processing and Communications Appl. Conf. IEEE, pp.1-4(2015).

[25]. Md. Zia Uddin, J. J. Lee, T.-S. Kim," An enhanced independent component-based human facial expression recognition from the video", IEEE Transactions on Consumer Electronics, Vol. 55 IEEE Transactions, pp. 2216-2224,(2009).

[26]. Byungyong Ryu, A Rivera, J. Kim, O. Chae" Local Directional Ternary Pattern for Facial Expression Recognition" IEEE Transactions on Image Processing, pp: 6006–6018, (2017).

[27]. Shervin Minaee, AmiraliAbdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network" arXiv:1902.01019v1 [cs.CV] ,pp. 1-8,4 Feb (2019).

[28]. Fengping An, Zhiwen Liu," Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM", International Journal of Computer Graphics Springer, pp.1-16, (2019).

[29]. S. Du, Y. Tao, A. M. Martinez, "Compound facial expressions of emotion". PNAS, pp.1454–1462, (2014).