International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-06, SEP 2021

A Review on Facial Expression Recognition Techniques

Methil Muley, Student, GEC Aurangabad, India, methilmuley.99@gmail.com

A. S. Bhalchandra, Professor, GEC Aurangabad, India, asbhalchandra@gmail.com

Abstract **Human facial expression is one of the most important forms of nonverbal social communication. Face expressions are a simple way for humans to express their emotions and intentions. This review paper describes the Facial Expressions Recognition techniques with its major contribution in three phases preprocessing of the input image, feature extraction from the image, and the classification of emotions on the basis of extracted features. We looked at the technologies that can be used for face expression identification in both traditional and deep learning approaches. On the basis of a static dataset and a sequence image, the various FER approaches are evaluated. This review covers JAFFE, CK+, MMI, and a number of other databases. Existing methods ignore the effects of facial characteristics such as age, but the manifestation is done to overcome difficulties. The development of facial expression recognition utilizing deep neural networks, particularly convolutional neural networks, for multiple database attributes has been noteworthy.**

***Keywords* — FER Database, Facial Expression Recognition, Feature Extraction, Deep Learning, Convolutional Neural Network, SVM**

I. INTRODUCTION

One of the most important nonverbal forms of social communication is facial expression. While communication encompasses both verbal and nonverbal expressions, nonverbal communication uses gesture, eye look, body language and posture, proxemics, haptics, and appearance to express the emotions, intentions, and behavior of a human being. By looking, staring, and blinking, the eyes are the most essential nonverbal communication body part. Various emotions are expressed by facial expressions, including happiness, sadness, fear, disgust, anger, and surprise. Happiness can convey that a person is in a pleasurable and contented state, and each emotion can describe a human's conditional state. Mehrabian et al. found that visual emotional information accounts for 55% of environmental psychology, vocal information for 38%, and verbal information for 7% [1]. In 1978, Ekman et al. established the FACS (Facial Action Coding System), which describes facial motions in Action Units (AUs). They split the human face into 46 Action Units, each of which was coded with each and every facial muscle [2]. Because nonverbal Facial emotional expression communications are part of human communication, researchers are paying more attention to calculating their emotional state using various facial expression recognition systems. There are two approaches for recognizing facial expressions: the standard method and the deep learning method. Automatic Expressions of the Face Recognition is not an easy task because each person expresses facial emotions in his or her own unique way, and there are several barriers and challenges in the FER area, such as differences in age, gender, background, head pose, and occlusions caused by skin illness, wearing glasses, scarf, and so on. From the review analysis, databases such as CK, CK+, JAFFE, MMI, YALE, MUG, RAF-DB, and AffectNet are introduced in this study. This study solely analyses the performance of various studies on the basis of the database.

Preprocessing, feature extraction, and classification are the three essential phases of the FER approach. When the system receives an image, it calculates the image's properties and outputs the feature vector using the supplied operator. When a weak feature is removed from a facial image, the emotion recognition accuracy suffers. One of the most powerful and classic face image feature extractors is the Local Binary Pattern (LBP) approach. Due to its invariance property, monotonically global grayscale level change, and changing the brightness property in real applications like Face recognition, the LBP descriptor is frequently employed in various applications. Many descriptors are defined and reviewed in the study, including Gabor Filter (GF), which yields textural features, Local Directional Pattern (LDP), curvelet transform, Active Shape Model (ASM), Local Fisher Discriminant Analysis (LFDA), and others. Deep learning is a strong technique for extracting information that can be used to recognize facial expressions. To prevent the overfitting problem, deep learning requires a large number of datasets to train the model. Existing databases are insufficient to train the model

 DOI : 10.35291/2454-9150.2021.0543

and distinguish emotion using deep learning and neural network systems. Additionally, there are variables such as occlusions, illumination, and head attitude, which are all common issues when learning the system. As a result, deep networks are introduced to overcome numerous barriers in order to make expression recognition more successful. In this paper, models such as Convolutional Neural Networks (CNN) and Deep Belief Networks (DBF) are defined and reviewed.

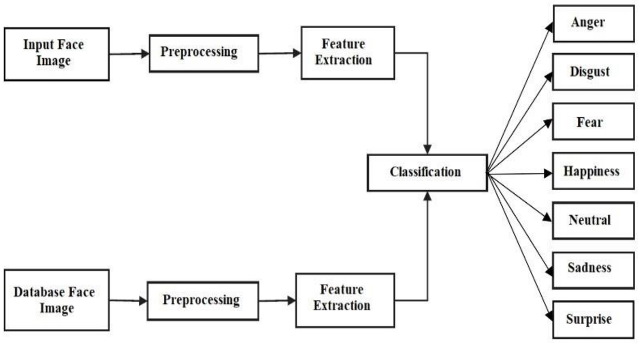

Fig.1: The basic architecture for facial expression recognition.

II. FACIAL EXPRESSION RECOGNITION

A. Preprocessing

Preprocessing is used to improve picture data that has been distorted or suppressed, as well as to enhance some of the image's properties. The initial step in feature extraction for Facial Expression is preprocessing [4]. Further face feature extraction is done using image obviousness, enhancement approaches, brightness corrections, and other preprocessing techniques [3]. Cropping and resizing of the face photos were done with the Nose as the middle significant component in mind [5]. The intensity value of the pixels is changed during image normalization [14]. To lessen the amount of fluctuation and illumination in the image of the face The median filter can be used to normalize data. Normalization can also be used to isolate the ocular element of the FER model, making it more robust. We can detect the facial region using the Viola-Jones technique [6] with the help of Localization. The noise is removed with a low pass filter, and the face image is localized. Adaboost learning for face detection using Haar-like features [12] is an alternative to the Viola-Jones technique. Histogram equalization is used for the brightness correction [7]. Haar function is used to detect lighter darker portions such as the brow, nose, and lighter darker lighter portions such as the lips [17]. To alter the image intensities the technique called as histogram equalization is applied on the image. Gaussian Filter blurs the image and reduces the noise from the input image, it also resizes and smoothen the image [8]. Bidirectional computation of a Supervised Descent Method (SDM) and Active Appearance Model (AAM) is used for extracting the facial landmarks [18].

Only the face component of the image is required when dealing with facial expression detection; the rest of the image is undesired, therefore preprocessing can be used to remove the unwanted backdrop. Preprocessing differs depending on the technique, but the other characteristics of deep neural network topologies remain the same. Data Augmentation, Data Alignment, and Normalization are three typical ways for preprocessing in the deep feature learning architecture.

Data Augmentation: To train a Deep Network, a large number of image samples or data samples are required. The artificial method of increasing the quantity of samples is known as data augmentation. The common ways to data Augmentation are translation, reflection, rotation, and scaling [20]. Offline and on-the-fly data augmentation are the two types of data augmentation available. The on-the-fly Data Augmentation approach, which is commonly integrated in deep learning tools, can overcome problems caused by overfitting. In this method, the input sample is clipped from the center and four corners and then flipped horizontally, resulting in an increase in the number of datasets [21]. Offline Data Augmentation includes things like shifting, scaling, and rotation. A robust network can be created by combining several offline Data Augmentation methods.

Face Alignment: Face Alignment enhances the features retrieved by natural and deep learning techniques. In deep learning, the Multi Task Convolutional Neural Network (MTCNN) has the greatest performance for face alignment [22].

Face Normalization Illumination and posture normalization are the two types of normalizing. It is possible to change the contrast and illumination of the same FER image. The different Illumination normalization methods used often are the Discrete Cosine Transform (DCT), Isotropic diffusion (IS), and the Difference of Gaussian (DoG) [23]. In the deep network, pose normalization produces the best results. The geometric elements of the FER approach are given by FER on pose normalization [24]. The usual method of normalization is inappropriate for the deep learning process [39].

B. Feature Extraction

Due to a lack of good facial expression recognition techniques, Facial Expression Recognition applications have yet to be realized in everyday life. Gabor Filter (GF), a texture descriptor, may provide information from face expressions such as magnitude and phase. The phase feature [3], [6] can provide all of the information regarding the magnitude feature description. The magnitude and phase features recovered by the Gabor Filter on the dataset have huge dimensions, so they were reduced. The dataset's feature vector is subjected to subspace techniques [9]. For facial emotion recognition, a combination of Gauss-Laguerre wavelet textural features and geometrical

 DOI : 10.35291/2454-9150.2021.0543

information can be used. For extracting texture features, GL wavelet has higher frequency extraction capabilities, as well as properties like rotation invariance and multiscale approach. The GL function creates a self-contained pyramidal analytical structure. Face expression detection is achieved by extracting geometric fiducial pointes and then combining these two features [4], [5]. Patch-based method for extracting and matching facial movement feature descriptors that includes both spectral and structural information and is dependent on distance characteristics [5]. By thresholding the neighborhood and treating them in binary integers, the Local Binary Pattern (LBP) may readily recover the texture operators of the pixel in an image. Only select significant facial landmarks can be retrieved for better computational complexity. Local Binary Patterns is an effective feature extraction method for low-resolution images, and the LBP histogram retrieves appearance features [7]. LBP can extract features such as spatial structure from a facial picture, but this is insufficient; therefore, LGBP can extract edge orientation and angle features, which are crucial aspects for refining facial features [13].

The Discrete Contourlet Transform (DCT) is mostly employed as a frame operator, which means that it can quickly rebuild the original image. Decomposing the facial picture using discrete contourlet transform yields components such as various size and angles with high and low frequency. Directional sub bands are determined by analyzing the entropy of feature vectors to reduce the aspects of the feature vectors [8]. To classify as a feature vector, Euclidean Distance (ED) estimates between the similarity score and the normalized vector [9]. Local Fisher Discriminant Analysis (LFDA) is a supervised dimension reduction feature extractor that automatically learns the metric data to capture the characteristics from the input image [11]. The Active Shape Model (ASM) is a statistical shape-based model that disfigures an item repeatedly to fit it into a new image, focusing on the boundary curve edges and estimating pose and shape features [12]. Based on the graphical representation features retrieved from the other image, the Elastic Bunch Graph Matching (EBGM) algorithm recognizes the object. Temporal features are given by triangular geometric features [14]. Curvelet Transforms are intended to extract features by performing localization in both the time and frequency domains, hence reducing the number of dimensions. It's a multi-scale, multi-direction tool with the property of optimal sparse curve representation [15]. The Local Directional Pattern (LDP) assigns a value to each and every pixel in an image, which is then encoded to build a histogram, which is then concatenated to form the feature vector, which is made up of edge responses from various pixel directions [16].

Learning the expression state and learning using LSTM are two methods for representing spatial temporal characteristics. The LSTM [30] learns dynamic temporal facial features frame by frame. The state-of-the-art method that computes the explanation on Ekman's Facial Action Coding System (FACS) using visualization techniques of CNN reveals the ability of network both cross task and cross data connected to Facial Expression Recognition [31].

The overfitting and underfitting difficulties are caused by combining the two separate features, which increases the dimension of the feature vectors. To reduce the dimensions, two feature vectors are fine-tuned. CNN uses a Deep Temporal Geometric Network (DTGN) to extract Facial Landmark features [32]. By splitting up into separate local patches, which are cited by facial muscle action, facial image area decomposition provides the location of occlusions. For the learning framework, attention mechanisms identify the occluded and un-occluded region of interest [33]. In the field of face and object identification, deep learning has had a lot of success. Convolutional Neural Networks (CNN), Deep Belief Networks (DBF), Deep Autoencoders (DAE), Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN) are well-known deep learning networks with excellent performance in a variety of applications. In comparison to other multilayer perceptron models, a CNN has multiple layers that are good enough for location change and invariance behavior. The network is divided into three layers: a convolutional layer with learnable filters that converts results into feature maps, a pooling layer that reduces the spatial dimensions of the feature maps, and fully connected layers at the network's end that verify that all neurons are fully connected to activation and, finally, convert into single dimension feature maps for classification.

For these restricted Boltzmann Machines (RBM), a DBN is a way of stacking that uses each hidden layer as the input layer for the next layer. Hinton et al. [25] suggested a statistical generative approach with graphical model consisting of multiple hidden layers.

DAE learns how to reduce the dimensions by using well-organized coding. It reconstructs the inputs by minimizing the errors. It regards the two layers next to it as RBM, resulting in very good solutions as a result of the pretraining. It is primarily used in information retrieval applications [26]. RNN primarily uses temporal information, which is more suitable for forecasting sequential data with random length. Each layer has a single directed connection. The Long Short-Term Memory (LSTM) is used by RNN to avoid the vanishing gradient property. LSTM extracts the temporal properties of continuous frames in general. The accuracy of detecting audio-visual structures is frequently improved by using

 DOI : 10.35291/2454-9150.2021.0543

LSTM. For the sequence of video-based expression recognition, RNN LSTM are used [27]. GAN is an unsupervised learning model that detects and learns patterns in input data on its own and creates new output that is based on the original data. The random noise is generated periodically by the GAN network. The missing facial traits were predicted using a trained GAN network [32].

C. Classification

Learning Vector Quantization (LVQ) has two layers, the first of which is a competitive layer and the second of which is a linear layer. Because the self-organizing feature maps learn to classify input vectors in the same way as the competitive layer does, the input vectors are classified in a similar way. The linear layer transforms the first layer into the user-defined target classification [3], [4]. The oldest classification method [11] is 1-Nearest-Neighbor, which matches image features to the test picture. The K-Nearest Neighbors (KNN) algorithm stores all available examples and categorizes them using similarity measurements [4].

Support Vector Machine [17] is a paradigm for supervised learning. It is used for data classification and regression analysis, and it can quickly execute a non-linear classification by learning the input model completely and finding the solution based on similarities [5]. One-against-one (OAO) and one-against-all (OAA) are two different types of SVM classification algorithms [8]. For each emotion, a one-to-all strategy is used, which reduces calculation time [6]. Radial basis function (RBF) kernels were chosen for better classification performance [14]. For multi-class classification, a one-against-one technique is used [7]. By thorough mapping of feature vector, SVM discovers the liner separating hyper-plane [16].

Linear Discriminant Analysis (LDA) is a multivariate statistical model that seeks to differentiate the variable boundary between two or more groups by reducing the number of variables [12]. When compared to a single model, an ensemble classifier enhances classification predictive accuracy by first training the model and then making the prediction, allowing for overfitting of the training data [13]. The Online Sequential Extreme Learning Machine (OSLEM) is based on the approximation and interpolation theorem, which converts a single hidden layer into a linear equation using the recursive least square technique to calculate the weights. Spherical Clustering (SC) is used to find the optimal node, and the OSLEM-SC classifier is tested on a benchmark database [15].

Predicting one of the classes from the current categories' multiple classes. The neurons in a deep neural network take the input as a real number, multiply it by the weights, and then pass it through a non-linear activation function, allowing the deep neural network to learn very complex functions and pass them on to the next layers. Finally, the classified output is given by preventing under and overfitting. Back propagation faults are controlled using loss layers. The SoftMax layer is utilized as a loss layer in CNN. Instead of cross entropy, SVM can be utilized for an end-to-end training process that reduces the loss margin [28]. Alternatively, CNN can be used alone to extract features, after which an independent classifier such as SVM can be used to classify the data. By replacing the SoftMax layer with symmetric positive definite (SPD) with gaussian kernel and covariance features generated using DCNN [29], one can achieve efficient classification. Under cross-dataset estimation, ACNN performs exceptionally well in the FER lab agreement [33]. The regularization or dropout layer is particularly effective in overcoming the overfitting problem [41].

III. Database

Recognition of Facial Expressions The gathering of photos for training and testing, as well as the development of the Expression recognition model, are referred to as datasets. There are six primary emotions, according to Paul Ekman (anger, surprise, sadness, fear, joy, disgust).

Fig.2: Seven expression with each sample from the database.

Japanese Female Facial Expression (JAFFE): The JAFFE image dataset contains 213 static Grayscale posed photographs with seven Facial Emotions, six of which are basic emotions, and one neutral emotion, captured by 10 different Japanese females utilizing 60 participants. The photos have a resolution of 256 x 526 pixels [43].

The Extended Cohn-Kanade Dataset (CK+) is a complement to the CK dataset, which is largely grayscale with 8-bit accuracy and contains 123 subjects with a resolution of 640 x 490 pixels per [44].

MMI Dataset consists of 43 topics with high resolution 1280 videos and 250 still photos in the MMI Dataset. The MMI dataset has a resolution of 720 x 576 pixels. In the sequence of photos, action units are fully annotated with the facial expression [45].

FER2013 Dataset: FER2013 contains 35,887 still photos of

 DOI : 10.35291/2454-9150.2021.0543

faces, with 28,709 for training, 3589 for validation, and 589 for testing. All of the image sets are grayscale images with a resolution of 48 x 48 pixels. It comprises of seven emotions: angry, afraid, neutral, amazed, disgusted, pleased, and sad, each with a particular age group [46].

Affective Database in the Real-World RAF Database: There are 29,672 facial images in the RAF database. There are 40 annotations in this database, as well as seven fundamental emotions, twelve compound emotions, 37 preprogrammed emotions, and five ideal landmark locations. It has a wide range of illuminations, age, gender, occlusions, and head positions [47].

AffectNet database: The AffectNet database contains facial expression annotations. These datasets contain 1 million photographs, half of which have been humanly labelled (440 images) for seven face expressions, while the remaining dataset has been automatically annotated. It includes the words neutral, distaste, glad, fear, sadness, surprise, rage, and disgust [48].

Yale Database: The Yale dataset contains 165 grayscale photos in GIF format of 15 distinct people. It has eleven photos per topic, with happy, with and without glasses, normal, sad, drowsy, center and right light, shocked, and wink expressions [42]. The MUG database (Multimedia Understanding Group database) contains 86 subjects of facial expressions at a resolution of 896 x 896 pixels. The frame rate is 19 frames per second, and each sequence contains thousands of images. The six facial emotions anger, fear, sadness, disgust, happiness, and surprise are all represented in the MUG database [49].

Table I: Analysis of the performance on facial expression recognition technique.

Database	Author, Year	Method	Classes	Accuracy (%)
CK	(Jabid, Kabir, & Chae, 2010)	LDP, SVM	7	86.9
	(Zhang & Tjondronegoro, 2011)	Patch Based, SVM	6	87.4
	(Poursaberi, Noubari, Gavrilova, & Yanushkevich , 2012)	GL Wavelet, KNN	6	92.2
	(Neoh , et al., 2015)	LGBP, Ensemble Classifier	7	96.8
	(Jain, Mishra , & Thakur, 2016)	LBP, ED, SVM, NN	6	97.20
CK+	(Liu, Han, Meng, & Tong, 2014)	DBN(BDBN)	7	96.7
	(Jung, Lee, Yim, Park, & Kim, 2015)	CNN(DTAJN)	7	97.3
	(Happy & Routray, 2015)	LBP, SVM	6	89.6
	(Mollahosseini, Chan, & Mahoor, 2016)	CNN	7	93.2
	(Breuer & Kimmel, 2017)	CNN	7	98.6
	(Gupta, 2018)	Haar-Features, SVM	8	94.1
	(Liliana, 2018)	CNN	8	92.8
	(Minaee & Abdolrashidi, 2019)	CNN	7	98.1
	(Miao, Dong, Jaam, & Saddik, 2019)	CNN(MobileNet)	6	96.9
	(Bursic S. , Boccignone, Ferrara, D'Amelio, & Lanzarotti, 2020)	CNN	7	94.6
	(Wang, Xu, Niu, & Miao, 2020)	CNN	8	-
JAFFE	(Bashyal & Venayagamoorthy, 2008)	LVQ, GF	-	90.2
	(Jabid, Kabir, & Chae, 2010)	LDP, SVM	7	82.6
	(Zhang & Tjondronegoro, 2011)	Patch Based, SVM	6	93.9
	(Poursaberi, Noubari, Gavrilova, & Yanushkevich , 2012)	GL Wavelet, KNN	6	96.7
	(Rahulamathavan, Phan, Chambers, & Parish, 2013)	LFDA, 1-NN	7	94.4
	(Liu, Han, Meng, & Tong, 2014)	DBN(BDBN)	7	91.8
	(Happy & Routray, 2015)	LBP, SVM	6	85.1
	(Biswas & Sil, 2015)	DCT, SVM	6	98.7
	(Hegde , Hegde, & Seetha , 2016)	GF, ED, SVM	6	97.1
	(Uçar, Demir, & Güzeliş , 2016)	Curvelet Transform, OSELM-SC	7	94.6
	(Refat & Azlan, 2019)	CNN	7	97.1
	(Minaee & Abdolrashidi, 2019)	CNN	7	92.8
	(Miao, Dong, Jaam, & Saddik, 2019)	CNN(MobileNet)	6	95.2
FER-2013	(Mollahosseini, Chan, & Mahoor, 2016)	CNN	7	66.4
	(Breuer & Kimmel, 2017)	CNN	7	72.2
	(Minaee & Abdolrashidi, 2019)	CNN	7	70.1

 DOI : 10.35291/2454-9150.2021.0543

	(Bursic S. , Boccignone, Ferrara, D'Amelio, & Lanzarotti, 2020)	CNN	7	73.1
	(Ahmed & Ponmaniraj, 2020)	CNN	7	70
	(Wang, Xu, Niu, & Miao, 2020)	CNN	8	-
MMI	(Poursaberi, Noubari, Gavrilova, & Yanushkevich , 2012)	GL Wavelet, KNN	6	87.6
	(Jung, Lee, Yim, Park, & Kim, 2015)	CNN(DTAJN)	7	70.2
	(Mollahosseini, Chan, & Mahoor, 2016)	CNN	7	77.9
	(Kim, Baddar, Jang, & Ro, 2019)	CNN	7	78.6
YALE	(Hegde , Hegde, & Seetha , 2016)	GF, ED, SVM	6	83.8
MUG	(Rahulamathavan, Phan, Chambers, & Parish, 2013)	LFDA, 1-NN	7	95.24
	(Caroline , Sobral , & Vieira, 2014)	ASM, LDA	7	99.7
	(Ghimire, et al., 2015)	Geometric Features, SVM	6	95.5
RAF-DB	(Li, Zeng, Shan, & Chen, 2019)	CNN(ACNN)	7	80.5
AffectNet	(Li, Zeng, Shan, & Chen, 2019)	CNN(ACNN)	7	54.8

IV. Discussion

FER for side view faces leveraging subjective information from facial sub-regions and using alternative parameters to indicate the position of the face for real-time applications are two major future developments outlined in recent works. FER is utilized in real-time applications like driver monitoring, medical, robotics interaction, forensics, and deception detection. Software engineers can use this review paper to build algorithms based on their accuracy and complexity. Also, depending on their needs, it is beneficial for hardware implementation to implement at a cheap cost. Preprocessing, feature extraction, classification, and key contributions are all compared in this review. The database, complexity rate, recognition accuracy, and important contributions are used to evaluate performance. The availability of preprocessing, feature extraction, and expression count are all discussed in this review. The conventional and nonconventional methods are review here. Conventional methods like LDP, patch based, GL wavelet, Haar features are reviewed for feature extraction and for classification SVM, KNN are also reviewed. In this paper the non-conventional method reviewed are deep convolutional neural network and deep belief network which has the better performance in the facial expression recognition.

V. Conclusion

The research on Facial Expression Recognition has piqued the interest and piqued the interest of researchers. Due to the dataset's attribution issues, the FER algorithms are given a boost. Convolutional neural networks are the ideal deep learning option for training and testing facial expression datasets since they boost speed and reduce feature computation time. We begin by describing the preprocessing procedures required for classic facial expression detection systems, as well as the requirements for deep learning approaches such as data augmentation, data alignment, and face picture normalization. Then there are feature extraction processes such as LBP, LDA, Haar like feature extraction, geometrical features, Gabor filter, and so on, and in deep learning, the features are extracted autonomously by the deep learning algorithm. Deep learning categorization can be both an independent and dependent process. We have analyzed the recent few years' research publications on FER approaches based on the range of datasets that have been explained. This paper gives a detailed review of past work on FER in order to organize and gather it for future development. As a result of this assessment, future work on face expression recognition will focus on overcoming the constraints and barriers posed by the photographs in the collection, as well as developing more accurate algorithms.

Appendix

FER: Facial Expression Recognition
CNN: Convolutional Neural Network
SVM: Support Vector Machine
LDP: Local Directional Pattern
KNN: K-Nearest Neighbours
LBP: Local Binary Pattern
DBN: Deep Belief Network
LVQ: Learning Vector Quantization
GF: Gabor Filter
LFDA: Local Fisher Discriminant Analysis
1-NN: First Nearest Neighbour
NN: Neural Network

Acknowledgment

I sincerely thank my Guide, Prof. Dr. A. S. Bhalchandra, for her consistent and incredible ideas in carrying out my research work.

References

[1] Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

 DOI : 10.35291/2454-9150.2021.0543

[2] Ekman, P., & Friesen, W. V. (1978). *Facial action coding system : investigator's guide.* Palo Alto, Calif. : Consulting Psychologists Press.

[3] Bashyal , S., & Venayagamoorthy, G. K. (2008). Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence, 21*, 1056 - 1064. doi:https://doi.org/10.1016/j.engappai.2007.11.010

[4] 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 Journal on Image and Video Processing, 17*. doi:https://doi.org/10.1186/1687-5281-2012-17

[5] Zhang, L., & Tjondronegoro, D. (2011). Facial Expression Recognition Using Facial Movement Features. *IEEE Transactions on Affective Computing, 2*, 219-229. doi:10.1109/T-AFFC.2011.13

[6] Zhang, L., Tjondronegoro , D., & Chandran, V. (2014). Random Gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing, 145*, 451 - 464. doi:https://doi.org/10.1016/j.neucom.2014.05.008

[7] Happy, S. L., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing, 6*, 1-12. doi:10.1109/TAFFC.2014.2386334

[8] Biswas , S., & Sil, J. (2015). *An Efficient Expression Recognition Method using Contourlet Transform.* doi:10.1145/2708463.2709036

[9] Hegde , G. P., Hegde, N., & Seetha , M. (2016). Kernel Locality Preserving Symmetrical Weighted Fisher Discriminant Analysis based subspace approach for expression recognition. *Engineering Science and Technology, an International Journal, 19*, 1321-1333. doi:https://doi.org/10.1016/j.jestch.2016.03.005

[10] Jain, S., Mishra , D., & Thakur, R. (2016). Facial Expression Recognition Using Variants of LBP and Classifier Fusion. In S. Satapathy , A. Joshi , N. Modi , & N. Pathak (Ed.), *Proceedings of International Conference on ICT for Sustainable Development. Advances in Intelligent Systems and Computing. 408*, pp. 725-732. Singapore: Springer. doi:https://doi.org/10.1007/978-981-10-0129-1_75

[11] Rahulamathavan, Y., Phan, R. C., Chambers, J. A., & Parish, D. J. (2013). Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis. *IEEE Transactions on Affective Computing, 4*, 83-92. doi:10.1109/T-AFFC.2012.33.

[12] Caroline , S., Sobral , A., & Vieira, R. (2014). *An automatic facial expression recognition system evaluated by different classifiers.*

[13] Neoh , S. C., Zhang , L., Mistry , K., Hossain , M. A., Lim , C. P., Aslam , N., & Kinghorn, P. (2015). Intelligent facial emotion recognition using a layered encoding cascade optimization model. *Applied Soft Computing, 34*, 72-93. doi:https://doi.org/10.1016/j.asoc.2015.05.006.

[14] Ghimire, D., Lee, J., Li, Z.-N., Jeong, S., Park, S. H., & Choi, H. S. (2015). Recognition of Facial Expressions Based on Tracking and Selection of Discriminative Geometric Features. *International Journal of Multimedia and Ubiquitous Engineering, 10*, 35-44. doi:10.14257/ijmue.2015.10.3.04

[15] Uçar, A., Demir, Y., & Güzeliş , C. (2016). A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. *Neural Computing and Applications, 27*, 131-142. doi:https://doi.org/10.1007/s00521-014-1569-1

[16] Jabid, T., Kabir, M., & Chae, O. (2010). Robust Facial Expression Recognition Based on Local Directional Pattern. *ETRI Journal, 32*(5), 784-794. Retrieved from https://doi.org/10.4218/etrij.10.1510.0132

[17] Gupta, S. (2018). Facial emotion recognition in real-time and static images. *2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 553-560). IEEE. doi:10.1109/ICISC.2018.8398861.

[18] Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE. doi:10.1109/WACV.2016.7477450

[19] Lopes, A. T., Aguiar, E. d., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with Convolutional Neural Networks. *Pattern Recognition, 61*, 610-628. Retrieved from https://doi.org/10.1016/j.patcog.2016.07.026

[20] Paulin, M., Revaud, J., Harchaoui, Z., Perronnin, F., & Schmid, C. (2014). Transformation Pursuit for Image Classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 3646-3653. doi:10.1109/CVPR.2014.466

[21] Kim, B.-K., Roh, J., Dong, S.-Y., & Lee, S.-Y. (2016). Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *Journal on Multimodal User Interfaces*(2), 427-434. doi:https://doi.org/10.1007/s12193-015-0209-0

[22] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters, 23*, 1499-1503. doi:10.1109/LSP.2016.2603342

[23] Chen, W., Meng Joo Er, & Wu, S. (2006). Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36*, 458-466. doi:10.1109/TSMCB.2005.857353

[24] Hassner, T., Harel, S., Paz, E., & Enbar, R. (2015). Effective face frontalization in unconstrained images. (pp. 4295-4304). IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2015.7299058.

[25] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation, 18*(7), 1527-1554.

[26] Goodfellow , I., Bengio , Y., & Courville, A. (2016). *Deep Learning.* MIT Press. Retrieved from http://www.deeplearningbook.org

[27] Hochreiter , S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780.

 DOI : 10.35291/2454-9150.2021.0543

ISSN : 2454-9150 Vol-07, Issue-06, SEP 2021

Retrieved from http://dx.doi.org/10.1162/neco.1997.9.8.1735

[28] Tang, Y. (2013). Deep Learning using Linear Support Vector Machine. *arXiv: Learning*. doi:arXiv:1306.0239.

[29] Otberdout, N., Kacem, A., Daoudi, M., Ballihi, L., & Berretti, S. (2018). Deep Covariance Descriptors for Facial Expression Recognition. *BMVC*.

[30] Kim, D. H., Baddar, W. J., Jang, J., & Ro, Y. M. (2019). Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition. *IEEE Transactions on Affective Computing, 10*(2), 223-236. doi:10.1109/TAFFC.2017.2695999

[31] Breuer, R., & Kimmel, R. (2017). A Deep Learning Perspective on the Origin of Facial Expressions. *arXiv: Learning*. doi:arXiv: 1705.01842

[32] Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. (2015). Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition. *IEEE International Conference on Computer Vision (ICCV)* (pp. 2983-2991). Santiago, Chile: IEEE. doi:10.1109/ICCV.2015.341

[33] Li, Y., Zeng, J., Shan, S., & Chen, X. (2019). Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism. *IEEE Transactions on Image Processing, 28*(5), 2439-2450. doi:10.1109/TIP.2018.2886767.

[34] Liu, P., Han, S., Meng, Z., & Tong, Y. (2014). Facial Expression Recognition via a Boosted Deep Belief Network. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1805-1812). IEEE. doi:10.1109/CVPR.2014.233

[35] Refat, C. M., & Azlan, N. Z. (2019). Deep Learning Methods for Facial Expression Recognition. *7th International Conference on Mechatronics Engineering.*

[36] Minaee, S., & Abdolrashidi, A. (2019, February). Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *Computer Vision and Pattern Recognition*. Retrieved from arXiv:1902.01019v1 [cs.CV]

[37] Miao, Y., Dong, H., Laam, l., & Saddik, A. (2019, june). A Deep Learning System for Recognizing Facial Expression in Real-Time. *ACM Transactions on Multimedia Computing, Communications, and Applications, 15*. Retrieved from https://doi.org/10.1145/3311747

[38] Bursic, S., Boccignone, G., Ferrara, A., D’Amelio, A., & Lanzarotti, R. (2020). *Improving the accuracy of automatic facial expression recognition in speaking subjects with deep learning* (Vol. 10). Applied Sciences. doi:10.3390/app10114002

[39] Ahmed, S., & Ponmaniraj, S. (2020, April 21). Facial Expression Recognition using Deep Learning. *International Journal of Engineering and Advanced Technology (IJEAT), 9*(4). doi:10.35940/ijeat.D8901.049420.

[40] Wang, W., Xu, K., Niu, H., & Miao, X. (2020, September 11). Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation. *2020*, 9. (Z. Lv, Ed.) Retrieved from https://doi.org/10.1155/2020/4065207.

[41] Liliana, D. Y. (2018). Emotion recognition from facial expression using deep convolutional neural network. *Journal of Physics: Conference Series International Conference of Computer and Informatics Engineering (IC2IE)*, 12-13. doi:10.1088/1742-6596/1193/1/012004

[42] Gross, R. (2005). Face Databases. In R. Gross, S. Li, & A. Jain (Eds.), *Handbook of Face Recognition*. New York: Springer.

[43] Lyons, M., Kamachi, M., & Gyoba, J. (1998). *The Japanese Female Facial Expression (JAFFE) Dataset*. Zenodo. doi:10.5281/zenodo.3451524

[44] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar , Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. *3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis, 2010.* IEEE. doi:10.1109/CVPRW.2010.5543262

[45] Valstar , M., & Pantic, M. (2010). Induced disgust, happiness and surprise: an addition to the MMI facial expression database. *Proc. Int. Conf. Language Resources and Evaluation, 2010.*

[46] Carrier, P. L., Courville, A., Goodfellow, I. J., Mirza, M., & Bengio, Y. (2017, November 16). FER-2013 Face Database. *Universit de Montral: Montreal, QC, Canada, 2013.*

[47] Li, S. (n.d.). "RAF-DB" Real-world Affective Faces Database. Retrieved from http://www.whdeng.cn/raf/model1.html

[48] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 18-31. doi:10.1109/TAFFC.2017.2740923

[49] Aifanti, N., Papachristou , C., & Delopoulos, A. (2010). The MUG Facial Expression Database. *Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Desenzano, Italy.*

 DOI : 10.35291/2454-9150.2021.0543