

# Facial Emotion Recognition with Sparse Coding Descriptor

\*Aliyu S. Nuhu, Emmanuel K. Akut, Kishak Z. Cinfwat, Patrick. N. Julius, Maryam L. Imam, Solomon I. Momoh, Felix I. Egbujo and Danjuma D. Dajab

Department of Electrical and Electronics Engineering, University of Jos, Nigeria

nuhua@unijos.edu.ng | czkishak@yahoo.com |

{eakut88 | nyabvou | maryamimamlami | solomonmomoh | pheleix | danjuma.dajab}@gmail.com

Received: 19-MAY-2021; Reviewed: 18-JUN-2021; Accepted: 27-JUN-2021

<http://dx.doi.org/10.46792/fuoyejet.v6i2.642>

**Abstract-** With the Corona Virus Disease 2019 (COVID-19) global pandemic ravaging the world, all sectors of life were affected including education. This led to many schools taking distance learning through the use of computer as a safer option. Facial emotion means a lot to teacher's assessment of his performance and relation to his students. Researchers has been working on improving the face monitoring and human machine interface. In this paper we presented different types of face recognition methods which include: Principal component analysis (PCA); Speeded Up Robust Features (SURF); Local binary pattern (LBP); Gray-Level Co-occurrence Matrix (GLCM) and also the group sparse coding (GSC) and come up with the fusion of LBP, PCA, SURF GLCM with GSC. Linear Kernel Support Vector Machine (LSVM) Classifier out-performed Polynomial, RBF and Sigmoid kernels SVM in the emotion classification. Results obtained from experiments indicated that, the new fusion method is capable of differentiating different types of face emotions with higher accuracy compare with the state-of-the-art methods currently available.

**Keywords-** Face Recognition, Principal Component Analysis, Speeded Up Robust Features, Grey Level Co-Occurrence Matrix

## 1 INTRODUCTION

The classification of human facial expressions which include surprise, happiness, anger, contempt, disgust, fear is a process known as facial expression recognition (FER) (Shuaibu, et al. 2018). Facial expression is used by humans irrespective of their nationality, sex or race to express their states of emotion and what they intent (Wang, et al. 2020). Despite the FER process, recognition of the face remains a difficult task due to variations of the face that are caused by aging, varying illumination, pose etc.

Humans can easily and accurately understand facial expression, but accurate and reliable facial recognition by computer/machine still remains a difficult task (Shuaibu, et al. 2018; Alaisawi, 2017). This led into intensive research by researchers to develop a reliable system that can discriminate facial emotion with higher precision. These researches are also necessary due to the wide variety of human computer interaction applications that FER is required, these include, but not limited to pain detection in medical treatment, driver's mood in driving software, student's reaction in an online class, customer's mood in a business environment etc. (Piatkowska & Martyna, 2012). In the educational sector, many factors like the virtual learning environment, the curriculum content and the teacher's approach to students determine the success in simulating interaction and student's mood in the learning process (Yang, et al. 2017; Dass, 2015).

In a physical class, a teacher uses facial expression of students to evaluate and improve his teaching methods such as perception, understanding and expressing emotions. In human-computer interaction, FER is required by a virtual teacher in order to analyse student's emotions and adjust for a more effective teaching (Farsi & Thobaiti 2015). Darwin conducted a research in 1872 on 'The Expression of Emotions in Man and Animals' Darwin, C. (1872). In another work, Ekman and Friesen in 1978 reported on 'The Facial Action Coding System: A Technique for The Measurement of Facial Movement' with the San Francisco: Consulting Psychologists Press (Piatkowska, & Martyna, 2012; Ekman, & Friesen, 1978). These opened up a bright insight for researchers to go deeply into optimising and effective discrimination of FER using different techniques which include: Principal component analysis (PCA); Speeded Up Robust Features (SURF); Local binary pattern (LBP); Gray-Level Co-occurrence Matrix (GLCM) and also the group sparse coding (GSC) (Bengio et al., 2009). Recently, the concept of Deep learning is implemented using Convolution Neural Network (CNN), Recurrent Neural Network, Auto Encoders (AE) and Deep Belief Networks (DBN) (Hosseini, 2017).

## 2 RELATED WORK

Right from time immemorial, both man and animal, young and old use facial expression to communicate their inner feeling about themselves or their things around them. An English man, Charles Robert Darwin in 1872 reported on 'Expression of Emotions in Man and Animals' explain on the important of facial expression in communication. In another similar report, Ekman & Friesen (1976) conducted many researches on the work of Darwin to give a better insight on FER. Recent reports on FER were computer-based research (Piatkowska & Martyna, 2012) where they proposed using a support vector machine (SVM) system that will detect and classify spontaneous emotions. They encoded the face texture with Local Binary Patterns (LBP) using face detection algorithm to create face representation. Automatic Face Analysis (AFA) system was developed by (Tian, et al.,

\*Corresponding Author

Section B- ELECTRICAL/COMPUTER ENGINEERING & COMPUTING SCIENCES  
Can be cited as:

Shuaibu A.N., Akut E.K., Cinfwat K.Z., Julius P.N., Imam M.L., Momoh S.I., Egbujo F.I. and Dajab D.D. (2021): Facial Emotion Recognition with Sparse Coding Descriptor, FUOYE Journal of Engineering and Technology (FUOYEJET), 6(2), 53-57. <http://dx.doi.org/10.46792/fuoyejet.v6i2.642>

2001) to analyse FER using both permanent and transient facial features and face image sequence. Using Neural Network classifier training, (Yang et al., 2017) proposed a learning emotion recognition model consisting of feature extraction, subset feature and emotion classifier. In the work of Wang et al (2020), the framework of FER algorithm was combined with online courses platforms using computer simulation to recognise student's emotion in an online education. Shuaibu et al. (2018) fused PCA and LBP with sparse coding to achieved FER using the linear kernel multi-class support vector machine (LSVM) classifier. Many algorithms are used to fuse multiple exposure sequence of photos to a single high quality and very low range of image contrast (Mertens et al., 2009; Mertens et al., 2007). Also, Wu (2014) used this method to deployed an algorithm that will fuse a weighted pyramid image and blend red, green, and blue colour channels of the input image into a single grayscale image (Saleem et al., 2012).

In recent time, Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN) have taken the machine learning field to a higher level and are used in many applications including deep learning is computer vision and FER (Hosseini et al. 2017; Ciresan, 2010; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014). They are scheme to instinctively learn and understand spatial hierarchies of features through back-propagation using multiple building blocks, such as pooling layers, fully connected layers, and convolution layers (Yamashita et al., 2018).

**3 FACIAL EMOTION RECOGNITION SYSTEM**

Facial expression recognition system basically involves four parts viz: face image collection, image pre-processing, features extraction and expression recognition (Shuaibu et al., 2018). Figure 1 depicts the FER system as shown below.

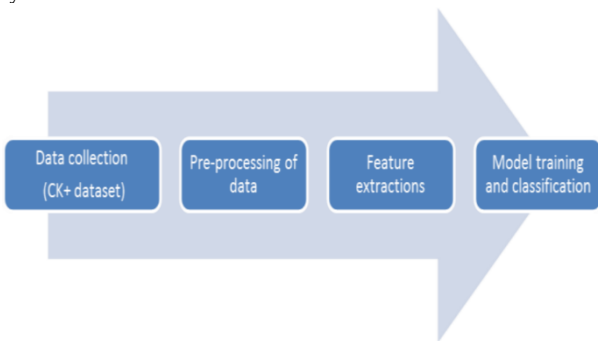


Fig. 1: Emotion recognition system flow chart (Shuaibu et al., 2018)

**3.1 DATA COLLECTION (CK+ DATASET)**

To acquired face image data digitally from a connected device, Adaboost classifier and Haar-Like feature are adopted for face detection and tracking. Instantaneity, robustness and accuracy of FER system is affected by many factors of captured image, which include contrast, brightness, image size etc. The image size  $I(x,y)$  is  $M \times N$ , while the energy normalization process is given as

$$I'(x,y) = \frac{I(x,y)}{\|I(x,y)\|} \tag{1}$$

where  $x$  and  $y$  are pixels locations.

Figure 2 depicts an example of facial emotions obtained from CK+ dataset.



Fig. 2: Example of facial emotions obtained from CK+ dataset (Lucey, et al., 2010) (a) Disgust, (b) Happy, (c) Surprise, (d) Fear, (e) Angry, (f) Contempt, (g) Sadness, and (h) Neutral.

**3.2 PRE-PROCESSING OF DATA**

To pre-process the image and extract features at different orientations where the output is computed as in Equation (2) (Almohamedh & Almotairi, 2019).

$$Z(i) = D(i) \times IG \times I \tag{2}$$

where  $I$  is a two dimensional input face array,  $Z(i)$  the response if  $i^{th}$  is filtered,  $D$  and  $G$  are the filter coefficients which is done by using two dimensional convolution pixelwise.

In this work, pre-processing methods which include brightness normalization, image filtering, geometry normalization, histogram equalization and facial effective area segmentation based on eight eyes (Luo, 2012).

**3.3 FEATURE EXTRACTION**

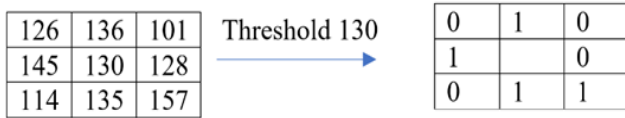
Implementing the PCA and LBP constitute the feature extractions section. The core function of PCA is to obtain the choice projection direction which represents the foremost data in the condition of least mean-square. PCA also extracts the global grayscale features of an image and reduced the size of feature space, which decides the size of data. Since the global feature of FER is sensitive to environment, LBP is used for local feature extraction is combined with PCA and learns with sparse coding descriptor (SC). SC is mostly used for the optimization of feature descriptors in image classifications. SC aim at locating an effective method of images pattern classification by fusion of multiple features selected from a dictionary (Shuaibu et al., 2017).

Given a sparse dictionary matrix  $\mathcal{D} = d_1, d_2, \dots, d_n$  containing  $K$  atoms as column vectors  $d$ , the sparse coding problem of extracting image descriptor  $T$  (PCA, LBP, PCA+LBP ) can be described as finding the most sparse vector  $y$  by concatenating the extracted features of LBP and PCA descriptor such that  $T \approx Dy$ . The signal  $T$  can be represented using linear combination of atoms as described by the following:

$$T = \sum_{i=1}^n y_i D_n = Dy \tag{3}$$

If  $I(x,y)$  is any pixel inside a local area of an image,  $I(x_c, y_c)$  is the centre of the  $3 \times 3$  window, the other eight points are  $g_0, \dots, g_7$ . The local area texture is define as  $T = t(g_c, g_0, \dots, g_7)$ . The equation describing eight pixels within the window using a set threshold is as following:

$$T \approx t(s(g_0 - g_7), \dots, s(g_7 - g_c)) \tag{4}$$



(01001101)<sub>2</sub> = LBP = (77)<sub>10</sub>  
 Fig. 3: LBP implementation and coding

The eigenvalue of this pixel is read from the threshold table of the LBP code in a clockwise direction to obtain the 8-bit binary code. The binary bit is then converted into decimal number for each symbol function as shown in figure 3. LBP code which is described as the spatial structure of local image texture features is described by the following:

$$LBP(x_c, y_c) = \sum_{i=0}^7 s(g_i - g_7)2^i \tag{5}$$

Figure 4 depicts the process of LBP feature extraction where LBP coding image of the original image is gotten after scanning a facial expression image using LBP operator.

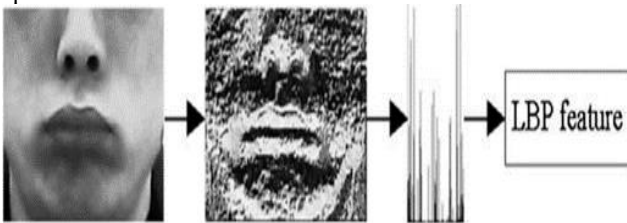


Fig. 4: LBP feature extraction

Local micro mode information of the original image which include edge, feature points and spot are included in the LBP coding image. Therefore, the local texture feature of an expression image is described with a histogram formed by 256 LBP codes.

Experiments show that mouth contributes most to facial expression in FER (Shuaibu et al., 2018); therefore, in this paper, the region of mouth is used for local feature extraction. LBP and PCA which are used for local and global extracts of mouth and whole image are learn with SC dictionary to obtain high effective features representations and then follow by the LSVM classification model.

**3.4 MODEL TRAINING AND CLASSIFICATION**

According to the theory of statistical learning, SVM is defined as a data learning method which deal with bogus regression problem and pattern recognition successfully. LSVM mechanism is looking for hyper-plane to meet for the requirement of classification, which is a best support vector to distinguish two different classes. LSVM performed input data into a high dimensional feature space, in this model. Linear algebra and geometry will be used in separating data that is only separable with nonlinear rules through a choice non-linear mapping function. This non-linear mapping function is called kernel function. In extracting the fundamental facial features in distinct directions, kernel is expressed as M<sup>th</sup> order Gaussian derivative (Sánchez, A. 2011). The coefficient is defined by the following:

$$Gs(x, y) = G^s(x) \times G^s(y) \tag{6}$$

Where G<sub>s</sub>(x) and G<sub>s</sub>(y) are Gaussian functions in a variable such that

$$G^s(x) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{x^2}{2s^2}\right) \tag{7}$$

Fractional derivative with respect to length *x* or *y* is figured as the product of Hermite polynomial and Gaussian function as [31][32]:

$$\frac{\partial G}{\partial x}(x, t) = \sum_{n=1}^{\infty} He'_n(x) \frac{t^n}{n!} = \sum_{n=1}^{\infty} He'_n(x) \frac{t^{n+1}}{n!} = tG(x, t) \tag{8}$$

where H is the Hermite polynomial.

The learning algorithm make use of kernel functions, it allows computational efficiency of inner products in feature space. Characteristic used of kernel functions include polynomial, RBF, linear and sigmoid. In this paper, we compared linear kernel with polynomial, RBF and sigmoid functions.

**4 RESULTS ANALYSIS AND DISCUSSIONS**

In this section, we describe the evaluations of PCA with LBP descriptors, and present the results for combined sparse coding (SC) descriptors for emotion classifications. In particular, we show the discrimination between: (a) Disgust, (b) Happy, (c) Surprise, (d) Fear, (e) Angry, (f) Contempt, (g) Sadness, and (h) Neutral. Finally, we use multi-class support vector machine which is implemented with linear complexity in training and testing phases.

The experiments are designed based on the set-up in CK+ dataset (Lucey, et al., 2010). Total sample of 1000 images of different facial expressions were used for the evaluations. The training sample set includes eight different facial expressions and the sum of each is 100. Half of them is in better illumination, the others is in worse illumination. The training set images were normalized into small sizes of dimensions 56 × 56. Two experiments were carried out under different illumination conditions. One is normal and the other is worse. Leave-one- out evaluation method was used with the multi-class LSVM classifier. Average recognition results for normal and worse illumination conditions are presented in Table 1.

Table 1. Average Recognition Rate (ARR) for Experiments (Expt. 1 and Expt. 2)

Methods	Expt. 1: ARR Normal	Expt. 2: ARR Worse
PCA + LSVM	91.25	88.05
LBP + LSVM	92.52	90.31
PCA + LBP +LSVM	93.75	90.35
SC + PCA + LBP + LSVM	96.25	91.75

Table 1 shows the average recognition rate for facial emotions. Based on the evaluation, results of the combined descriptor with sparse coding outperform the baseline methods PCA + LBP + SVM and PCA + SVM on both normal and worse illumination conditions



computed as 96.25% and 91.75% respectively. In addition, comparison of different kernel functions: linear SVM (LSVM), Polynomial SVM (PSVM), RBF SVM (RSVM) and Sigmoid SVM (SSVM) were conducted as reported in Table 2 under normal and worse illumination conditions.

Table 2. ARR Comparison of Kernel Functions for Normal and Worse Illumination Conditions

Types of kernel function	Expt. 1: ARR Normal	Expt. 2: ARR Worse
Propose LSVM +	96.25	91.75
Propose PSVM +	94.50	88.35
Propose RSVM +	95.82	90.25
Propose SSVM +	93.50	86.25

Table 2 shows the comparison average recognition rate for normal and worse illuminations conditions with different SVM kernels. It is worth mentioning that the LSVM with ARR 96.25% and 91.75% for normal and worse conditions respectively, outperforms PSVM, RSVM and SSVM kernel function. In a similar study in Adeyanju, Omidiora & Oyedokun (2015) carried out a performance comparison of facial emotions on locally acquired dataset and experimented on four different SVM kernels (RBF, Linear, Quadratic and Polynomial). PCA was used for feature extraction and dimensionality reduction. However, in this study, comparison of fusion of features was based on PCA, PCA + LBP and PCA + LBP + SC in each case with different kernels of SVM (Linear, Polynomial, RBF, and Sigmoid) as presented in Tables 1 and 2.

## 5 CONCLUSION AND FUTURE WORK

In this article, facial emotion recognition was proposed based on fusion of features from PCA and LBP descriptors. A sparse coding dictionary learned and optimized the robust facial extracted features of image and provides discriminate descriptors for emotions recognition with multi-class linear support vector machine. The proposed method with SC performed better than the state-of-the-art methods on CK+ dataset. Moreover, different kernel functions of the linear multi-class SVM were compared. The propose method with LSVM outperforms PSVM, RSVM and SSVM kernel functions on both normal and worse illumination conditions experiments.

The generic convolutional neural network (CNN) architecture performed well on large images. Researchers can explore transfer learning of FER and hybrid fusion of handcrafted engineering features with the CNN architectures for better recognition performance.

## ACKNOWLEDGMENT

The authors wish to thank University of Jos. This work was supported by the Nigerian Communication Commissions (NCC) under the grant ref. NCC/R & D/UNIJOS/010.

## REFERENCES

- Adeyanju, I. A., Omidiora, E. O., & Oyedokun, F. O. (2015) "Performance evaluation of different support vector machine kernels for face emotion recognition" IEEE Intelligent Signal Conference, London, UK.
- Alaisawi, S. K. M. (2017) "Comparison of Face Recognition Methods" <https://www.researchgate.net/publication/320133783>
- Almohamedh H. & Almotairi S. (2019) "Facial Emotion Recognition Using Eigenface and Feature Optimization" International Journal of Advanced Trends in Computer Science and Engineering. <https://doi.org/10.30534/ijatcse/2019/28842019>. Vol. 8 No. 4.
- Aneesh, K. U., Hasan., A. Z., & Smitha, N. P. (2020) "Facial Emotion Recognition – A gift for the visionless" International Journal of Scientific & Technology Research Volume 9, Issue 06, June 2020.
- Bengio, S., Pereira, F., Singer, Y., & Strelow, D. (2009) "Group Sparse Coding" Advances in Neural Information Processing Systems
- Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010) "Deep, big, simple neural nets for handwritten digit recognition," Neural computation, vol. 22, no. 12, pp. 3207–3220.
- Darwin, C. (1872) "The Expression of Emotions in Man and Animals", reprinted by Univ. of Chicago Press, 1965
- Das. (2015) "Virtual Classroom for Effective Learning in IT Industry", IEEE International Conference on Information Technology. 221-226.
- Dores, A. R., Barbosa, F., Queirós, C., Carvalho, I. P., & Gri, M. D. (2020) "Recognizing Emotions through Facial Expressions: A Largescale Experimental Study" International Journal of Environmental Research and Public Health 2020, 17, 7420; doi:10.3390/ijerph17207420
- Ekman, P., & Friesen, W. V. (1976) "Pictures of Facial Affect" Palo Alto, Calif.: Consulting Psychologist Press
- Ekman, P., & Friesen, W. V. (1978) "The Facial Action Coding System: A Technique for The Measurement of Facial Movement". San Francisco: Consulting Psychologists Press
- Farsi, M., & Thobaiti (2015) "The Effects of Teaching Primary School Children the Islamic Prayer in a Virtual Environment", Science and Information Conference, 765-769.
- Hosseini, H., Xiao, B., Jaiswal, M. & Poovendran, R. (2017) "On the Limitation of Convolutional Neural Networks in Recognizing Negative Images", arXiv:1703.06857v2
- Hosseini, H., Xiao, B., Jaiswal, M. & Poovendran, R. (2017) "On the Limitation of Convolutional Neural Networks in Recognizing Negative Images", arXiv:1703.06857v2 [cs.CV]
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012) "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105.
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., & Ambadar, Z. (2010) "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression", IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).
- Luo, Y., Wu, C. M., & Zhang, Y. (2012) "Facial Expression Recognition Based on Principal Component Analysis and Support Vector Machine Applied in Intelligent Wheelchair" Application Resource Computation 29 (8), 3166–3168.
- Mertens, T., Kautz, J., & Reeth, F. V. (2007) "Exposure fusion" in Proceeding of the 15th Pacific Conference on Computer Graphics and Applications PG '07, pp. 382–390, IEEE Computer Society, Washington, DC
- Mertens, T., Kautz, J., & Reeth, F. V. (2009) "Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography," Computer. Graphics Forum, 28(1), 161–171.
- Piatkowska, E., & Martyna, J. (2012) "Computer Recognition of Facial Expressions of Emotion" P. Perner (Ed.): MLDM 2012, LNAI 7376, pp. 405–414, Springer-Verlag Berlin Heidelberg 2012
- Saleem, A., Beghdadi, A., & Boashash, B. (2012) "Image fusion-based

- contrast enhancement" *EURASIP Journal on Image and Video Processing*, Springer.
- Sánchez, A., Ruiz, J. V. Moreno, A. B., Montemayor, A. S., Hernández, J., & Pantrigo, J. J. (2011) "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 74, no. 8, pp. 1272-1282. <https://doi.org/10.1016/j.neucom.2010.07.017>
- Shuaibu, A. N. & Kamel, N., Saad, M. N., Saeed, A., Faye, I., & Ali, Y. S. (2017) "Sparse Representation for Crowd Attributes Recognition", *IEEE Access Journal*.
- Shuaibu, A. N., Dajab, D. D., & Chinda, F. E. (2018) "Emotion Recognition Based on Fusion of Global and Local Grayscale Features with Sparse Coding Descriptor". *Proceedings of the 1st National Communication Engineering Conference*, Ahmadu Bello University, Zaria, Nigeria.
- Simonyan, K., & Zisserman, A. (2014) "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*.
- Tian, Y., Kanade, T., & Cohn, J. F. (2001) "Recognizing Action Units for Facial Expression Analysis" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2
- Viola, P., & Jones, M. J. (2004) "Robust Real-Time Face Detection," *International Journal of Computation* Vol. 57 (2), 137-154.
- Wang, W., Xu, K., Niu, H., & Miao, X. (2020) "Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation", *Hindawi Complexity*, 9 pages. <https://doi.org/10.1155/2020/4065207>
- Wu, T., & Toet, A. (2014) "Color-To-Grayscale Conversion Through Weighted Multiresolution Channel Fusion" *Journal of Electronic Imaging* 23(4), 043004. SPIE and IS&T [DOI: 10.1117/1.JEI.23.4.043004]
- Yamashita R., Nishio M., Do R. K. G. & Togashi K. (2018) "Convolutional Neural Networks: An Overview and Application in Radiology" *Insights into Imaging* <https://doi.org/10.1007/s13244-018-0639-9> ©Springer2018
- Yang, D., Alsadoon, A., Prasad, P. W. C., Singh, A. K. & Elchouemi, A. (2017) "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment". *6th International Conference on Smart Computing and Communications, ICSCC 2017*, 7-8 December 2017, Kurukshetra, India.