World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:4, 2019

Improved Feature Extraction Technique for Handling Occlusion in Automatic Facial Expression Recognition

Khadijat T. Bamigbade, Olufade F. W. Onifade

Abstract**—**The field of automatic facial expression analysis has been an active research area in the last two decades. Its vast applicability in various domains has drawn so much attention into developing techniques and dataset that mirror real life scenarios. Many techniques such as Local Binary Patterns and its variants (CLBP, LBP-TOP) and lately, deep learning techniques, have been used for facial expression recognition. However, the problem of occlusion has not been sufficiently handled, making their results not applicable in real life situations. This paper develops a simple, yet highly efficient method tagged Local Binary Pattern-Histogram of Gradient (LBP-HOG) with occlusion detection in face image, using a multi-class SVM for Action Unit and in turn expression recognition. Our method was evaluated on three publicly available datasets which are JAFFE, CK, SFEW. Experimental results showed that our approach performed considerably well when compared with state-of-the-art algorithms and gave insight to occlusion detection as a key step to handling expression in wild.

Keywords**—**Automatic facial expression analysis, local binary pattern, LBP-HOG, occlusion detection.

I. Introduction

THE expression on the face of a person does not only convey his/her emotional state but also provides communicative clues that tell one's level of understanding, concentration, objection and deception during social interactions with people and the environment [9]. Facial expression analysis (FEA) is a cross-field area of research as it finds its way into psychology, computer vision, pattern recognition, human computer interaction and affective computing. It becomes a very useful clue in a host of diverse areas like telecommunications, behavioral science, video games, animations, psychiatry, automobile safety, educational software, security, health care, law enforcement, etc.

The state-of-art-model to FER can be divided into three phases: the face detection and tracking, the feature extraction, and the classification/recognition of expression. Ekman and Friesen [12] designed the most widely used measuring tool for observable facial movement called the facial action coding system. They also affirm the universality of some expressions which are regarded as basic expressions. They include those of happiness, sadness, anger, fear, surprise and disgust. Till date, majority of studies in the field of facial expression recognition have been limited to the six basic emotions. Also, a major issue which is still addressed is the unique identification of expressions even when they have common facial actions as mostly found in fear - AU (1, 2, 4, 5, 20, 25, 26, 27) and surprise - AU (1, 2, 5, 26, 27). Generally speaking, expression recognition is either based on facial action units which later inform the expression displayed or expression based on taking the input image holistically to see what expression is present.

In the wild, recognizing a facial expression becomes more difficult because of conditions such as illumination, head pose orientation, and occlusion (such as eye glasses, beards, facial mask, etc.) [16]. In order to deliver a system that will function well in real life situations, such a system must have been trained with such complex data from the wild.

Several techniques have been employed in extracting features and classifying expression for the purpose of recognition. However, the feature extraction process remains the heart of any successful system in that even the best classifier will perform poorly when presented with feature vectors that do not sufficiently and accurately represent the input image.

The remaining part of this paper is organized as follows: Section II presents related feature extraction techniques that have been used in FEA, discussing the strengths and weaknesses of each of the techniques; Section III discusses the proposed methodology in a bid to handling occlusion expression recognition; Section IV presents the findings, while experimenting on three state-of-the- art datasets JAFFE [7], CK+ [13] and SFEW [3]; Section V concludes and gives further research direction.

II. Related Works

Feature extraction techniques can be broadly classified into appearance-based methods and geometry-based methods. The former applies feature descriptors to model facial texture changes but is known for their sensitivity to illumination and out of plane head pose, while the latter uses fiducial points to describe the shape of the face. Also, for adequate representation, accurate facial component detection is paramount. Reference [4] used Gabor wavelet representation and independent component analysis (ICA) in recognizing eight individual AUs and four combinatory AUs.

LBP [10] is a non-parametric texture descriptor. The LBP operator picks each pixel of an image and compares it with its neighboring image, i.e. thresholding the eight neighbors in a 3X3 neighborhood with the central pixel value. The results are encoded into binary value 0 and 1 which is referred to as LBP

Khadijat T. Bamigbade is a PhD Student and O. F. W. Onifade is a senior lecturer of the Department of Computer Science, University of Ibadan, Nigeria (e-mail: kt.bamigbade@ui.edu.ng, ofw.onifade@ui.edu.ng).

code. The encoding process is shown in Fig. 1.

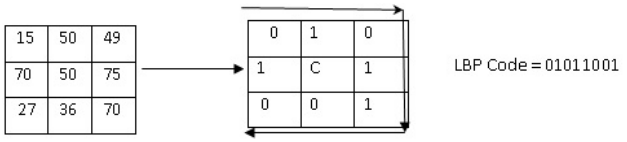

Fig. 1 LBP encoding process

An important property of LBP is that it is invariant to monotonic illumination changes and it is computationally simple. However, the selection of 3X3 neighborhood around each pixel by the original LBP cannot sufficiently represent large scale structure of dominant features resulting into the emergence of LBP variants which have evolved, either in a bid to improve on its discriminative property or enhance its robustness. Reference [5] represented selected facial parts with LBP for person independent expression recognition and claimed promising result. Reference [1] used a variant of LBP by extension called Compound Local Binary Pattern (CLBP). It differs by an additional P bit to the original LBP code which is used to preserve the descriptor that utilizes the sign and magnitude value obtained from the difference between a neighboring pixel value and its center pixel value. This is in a bid to preserve the pattern consistency of the local texture property of an input image.

Histogram of gradient on the other hand is an excellent descriptor for object detection and recognition. It computes the gradient magnitude and direction of local images (see [2] for more details).

More recently, deep learning techniques are currently being employed in facial expression analysis as found in [8] where the authors used Deep Convolution Neural Network (DCNN) been implemented on VGG-face and ExpNet network architecture for expression recognition and performed. They experimented on both CK+ and JAFFE Datasets. Deep learning techniques are yielding great recognition performance but need prior knowledge of occluded regions in an input image which is not feasible in the wild. They are also highly computationally intensive.

For classification, Support Vector Machine remains a leading and popularly used method with state-of-the-art classification rate.

III. Proposed Methodology

In this section, an LBP-HOG feature extraction technique with occlusion detection, using multiclass-SVM classifier for Action Unit recognition and a single SVM for expression recognition is presented. We employed the popularly known and widely used face detector algorithm which is based on Haar features with integral image called the Viola Jones face detection algorithm for frontal and near frontal view faces. Detailed information about Viola Jones face detection algorithm can be found in [14]. The detected face image is pre-processed, converted to gray scale, cropped and resized into a standard scale of 256*256. Reference [15] gave a detailed explanation on image resolution on facial expression analysis and reported no obvious difference in expression analysis for resolution greater than 72*96. The local feature representation of facial components is in line with the facial action coding system [12].

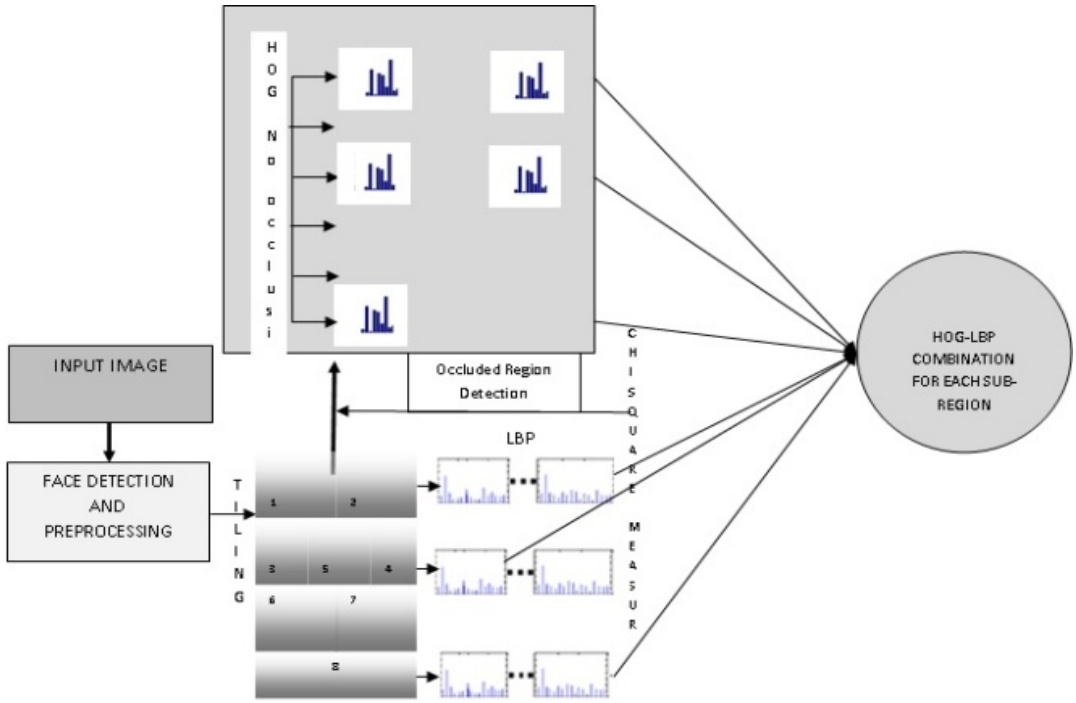

Fig. 2 Proposed HOG- LBP feature extraction techniques with occlusion detection

The face is divided into eight regions: R1- left eye; R2 - right eye; R3 - nose; R4 - left cheek; R5 - right cheek; R6 - left

part of the mouth; R7 - right part of the mouth; R8- jaw/chin, this tiling process is found in Fig. 2 below. Over 20 of the 44 action units defined by Ekman are harbored in regions R6 and R7 harbors. This approach gives the advantage of handling occluded regions and regions with participating Action Units simultaneously.

First, Uniform LBP [8] was applied on individual regions to extract the local pattern of each region. Then chi-square similarity measure was computed as shown in (1) for detecting occluded regions depicted in Fig. 2.

$$X^2(S,M) = \frac{\sum \mathrm{i}(S_i - M_i)^2}{(\mathrm{S_i} + \mathrm{M_i})} \quad (1)$$

where S and M are the LBP histograms of a region and its neighboring region in an image.

The occluded regions were removed from the set of regions where HOG features were to be extracted. The Histogram of Gradient (HOG) feature vector was then computed for every non-occluded sub-region. HOG is known for its robust edge detection capability. The feature set of both LBP and HOG from non-occluded regions is concatenated to form the singular feature vector for each region. A multi-class SVM as depicted in Fig. 3 was used for Action Unit recognition with a Radius Basis Function kernel [6]. The output of each region's classifier was combined and fed into the expression dictionary to deduce what expression the combination of AU maps to.

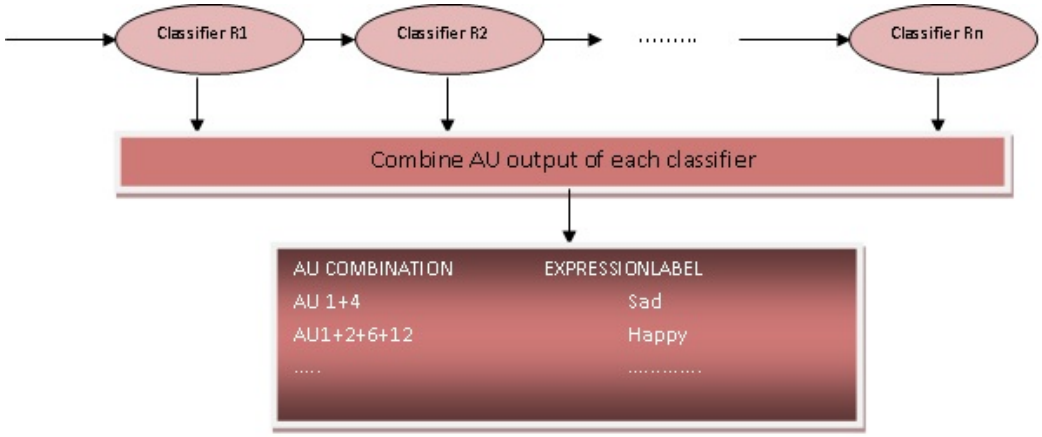

Fig. 3 AU/Expression recognition using a multi-classifier

IV. Result Discussion

The model presented in Fig. 2 was set-out principally to deal with the problem of occlusion in facial expression recognition. To this end, we set out to correct this anomaly, as seen in other techniques, by facial tiling into regions, followed by the application of LBP operators on the various regions on the face-image. Chi-square similarity check was performed to determine the occluded region which was so labeled and removed from the input region before HOG was applied. It is stressed that the inclusion of the occluded regions as found in related work has contributed, in no small measure, to the low level of accuracies as earlier witnessed.

Following the above, three groups of experiments were carried out to validate the proposed method. First, the performance of the method on SFEW dataset was reported because the target mainly was to be able to detect occluded regions from a face image and then extract features from the rest of the face region with facial actions. Only SFEW dataset has this property, so the confusion matrix derived from the proposed LBP-HOG was presented as depicted in Table I.

The Static Facial Expression in the Wild (SFEW) dataset contains 700 images representing one of the six known basic emotions. Images of this dataset possess variations such as presence of eye glasses, facial masks, beards, illumination, head orientation, image resolution and subject's age, as found in real life situations. For four consecutive years starting from 2013, SFEW has been employed as the testbed dataset in the EmotiW challenge [16].

Other publicly available datasets used for evaluating this method include the JAFFE Dataset and CK Dataset (details are in the next subsection). The Japanese Female Facial Expression Dataset consists of 10 Japanese students displaying each of the basic expressions in a controlled environment (laboratory setting), with a total of 213 facial expression images, while the CK dataset consists of 97 participants (also in a laboratory setting). The Leave-One-Subject-Out (LOSO) was adopted for a person independent validation method.

TABLE I
Confusion Matrix Derived from LBP-HOG w occ* on SFEW Dataset

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	63.60	1.80	0.00	0.00	14.60	20.00
Disgust	24.50	61.40	14.10	0.00	0.00	0.00
Fear	0.00	0.00	67.20	29.10	0.00	3.70
Happy	0.00	0.00	29.10	70.90	0.00	0.00
Sad	15.50	10.60	0.80	0.00	65.20	7.90
Surprise	0.00	0.00	4.10	28.80	0.00	67.10

Comparative Analysis

The performance of method on the three publicly available datasets mentioned earlier compared. Reference [5] extracted

LBP of important regions with considering whether the said regions are occluded or not, while the methods employed by [8], [11] require prior knowledge of which part of the face is occluded. The results shown in Tables II A-C demonstrate that the chosen method performs considerable well with or without the presence of occluded region. Also, it can do much better with larger training set.

TABLE II.A
CLASSIFICATION ACCURACY OF LBP-HOG WITH OCCLUSION DETECTION (LBP+HOG *W* OCC*), DEEP CONVOLUTION NEURAL NETWORK (DCNN) AND DEEP COVARIANCE DESCRIPTOR (DCD) ON JAFFE DATASET

Approaches	Recognition Accuracy (%)
LBP [5]	69.25
DCNN [8]	98.12
DCD [11]	98.25
LBP+HOG *w* occ.*	97.86

TABLE II.B
CLASSIFICATION ACCURACY OF LBP-HOG WITH OCCLUSION DETECTION (LBP+HOG *W* OCC*), DEEP CONVOLUTION NEURAL NETWORK (DCNN) AND DEEP COVARIANCE DESCRIPTOR (DCD) ON CK DATASET

Approaches	Recognition Accuracy (%)
LBP [5]	69.02
DCNN [8]	97.08
DCD [11]	98.40
LBP+HOG *w* occ.*	97.86

TABLE II.C
CLASSIFICATION ACCURACY OF LBP-HOG WITH OCCLUSION DETECTION (LBP+HOG *W* OCC*), DEEP CONVOLUTION NEURAL NETWORK (DCNN) AND DEEP COVARIANCE DESCRIPTOR (DCD) ON SFEW DATASET

Approaches	Recognition Accuracy (%)
LBP [5]	42.90
DCNN [8]	44.30
DCD [11]	49.18
LBP+HOG *w* occ.*	65.90

V. CONCLUSION

In this paper, an augmented LBP-HOG feature extraction technique with the capability of handling occluded face region (which is a common occurrence for facial expression analysis in the wild) has been presented. The developed model showed remarkable improvement in recognition accuracy when compared with similar existing techniques. Therefore, it is asserted that facial expression recognition, without proper consideration for occlusion, cannot give the desired result.

REFERENCES

[1] Ahmed, F., Bari, H. & Hossain, E. (2014) Person-independent facial expression recognition based on compound local binary pattern (CLBP). Int. Arab J. Inf. Technol., 11(2): 195-203.
[2] Dalal, N. & Triggs, B. (2005, June) Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference, 1: 886-893. IEEE.
[3] Dhall, A., Goecke, R., Lucey, S. & Gedeon, T. (2011, November) Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, pp. 2106-2112. IEEE.
[4] Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P. and Sejnowski, T. J. (1999) Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10): 974–989.
[5] Hablani, R., Chaudhari, N. & Tanwani, S. (2013) Recognition of facial expressions using local binary patterns of important facial parts. *International Journal of Image Processing (IJIP)*, 7(2): 163-170.
[6] Jayasumana, S., Hartley, R., Salzmann, M., Li, H. & Harandi, M. (2015) Kernel methods on Riemannian manifolds with Gaussian RBF kernels. IEEE transactions on pattern analysis and machine intelligence, 37(12): 2464-2477.
[7] Lyons, M., Akamatsu, S., Kamachi, M. & Gyoba, J. (1998, April). Coding facial expressions with gabor wavelets. In Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference, pp. 200-205. IEEE.
[8] Mayya, V., Pai, R.M., & Pai, M.M. (2016) Automatic facial expression recognition using DCNN. Procedia Computer Science, 93: 453-461.
[9] Mistry, V.J. & Goyani, M.M. (2013) A literature survey on facial expression recognition using global features. Int. J. Eng. Adv. Technol, 2(4): 653-657.
[10] Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transaction on Pattern Analysis Analysis and Machine Intelligence, 24(7): 971-987.
[11] Otberdout, N., Kacem, A., Daoudi, M., Ballihi, L. & Berretti, S. (2018) Deep Covariance Descriptors for Facial Expression Recognition. arXiv preprint arXiv:1805.03869.
[12] P. Ekman and W. V. Friesen (1978) "Facial Action Coding System: A Technique for the Measurement of Facial Movement," Consulting Psychologists Press.
[13] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews (2010) The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In CVPR4HB10, p 94–101.
[14] P. Viola and M. Jones (2001) Robust real-time object detection In International Workshop on Statistical and Computational Theories of Vision - Modelling, Learning, Computing and Sampling.
[15] Tian, Y. & Chen, S. (2012) Understanding effects of image resolution for facial expression analysis. *Journal of Computer Vision and Image Processing*.
[16] Zhang, L., Verma, B., Tjondronegoro, D. & Chandran, V. (2018) Facial Expression Analysis under Partial Occlusion: A Survey. ACM Computing Surveys (CSUR), 51(2): 25.