DEPARTMENT OF COMMUNICATIONS ENGINEERING
AHMADU BELLO UNIVERSITY, ZARIA - NIGERIA

**Proceedings of the 1st National Communication Engineering Conference 2018**

# Emotion Recognition Based on Fusion of Global and Local Grayscale Features with Sparse Coding Descriptor

Aliyu Nuhu Shuaibu
*Dept of Electrical and Electronics Engineering*
*University of Jos*
Plateau State, Nigeria
nuhua@unijios.edu.ng

Danjuma D. Dajab
*Dept of Electrical and Electronics Engineering)*
*University of Jos*
Plateau State, Nigeria

Francis E. Chinda
*Dept of Electrical and Electronics Engineering)*
*University of Jos*
Plateau State, Nigeria

***Abstract*— Facial emotion recognition study is a vital part of stress level monitoring and human machine interface. Principal component analysis (PCA) is a feature extraction technique based on statistical features which extracted the global grayscale features of an image. But the grayscale global features are sensitive to noise. Local binary pattern (LBP) extracts the local grayscale level features of the mouth region, which contribute most to facial emotion recognition, thereby minimizing the noise level in global grayscale features of facial expression recognition. Extensive experiments have shown that dictionary learning method with sparsely coded features captured vital structures of image and yielded discriminant descriptors for classifications. So fusion of PCA and LBP with sparse coding (SC) is introduced in this paper. The linear kernel multi-class support vector machine (LSVM) is used for facial emotions classification on CK+ dataset. Experimental results show that, this method can discriminate different emotions more effectively with improved recognition rate than the state-of-the-art approaches.**

***Keywords*— *facial features, emotion, recognition, grayscale, sparse coding***

## I. Introduction

Facial expression recognition is the process of classifying the human facial expressions like happiness, sadness and anger from the face images. Facial expressions recognition can be an important component of natural human-machine interfaces; it may also be used in various other fields of behavioural and medical sciences. Although humans can recognize facial expression quickly and precisely but reliable expression recognition by machine is still a challenge Emotions are feeling or response to particular situation or environment. Emotions are an integral part of our existence, as one smiles to show greeting, frowns when confused, or raises one's voice when angered. It is because we understand other emotions and reacts, based on that expression only enriches the interactions. Computers are "emotionally challenged". They neither recognize other emotions nor possess its own emotion [1].

There has been much research on recognizing emotion through facial expressions recognition via human-computer interface, non-intrusive sensors, intelligent cameras etc. Automatic detection of facial expressions has become an increasingly important research area. It involves computer vision, machine learning, health-care and behavioural sciences and can be used for many applications such as security, human-computer-interaction, driver safety, and stress level and depression analysis. Significant advances have been made in the field over the past decade with increase interest in non-posed facial behaviour in naturalistic contexts [2] and posed data recorded from multiple views with sparse appearance model [3].

Human emotional facial expressions play an important role in interpersonal relations. This is because humans demonstrate and convey a lot of evident information visually rather than verbally. Although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine remains a challenge as of today. To automate recognition of emotional state, machines must be taught to understand facial gestures vial model training and learning. In recent years, the design and implementation of ubiquitous, intelligent space and healthcare systems has become very popular in the field of human computer and human robot interaction. Such systems automatically monitor both the environment and the humans within it to provide assistance and services. Several systems provide support of the physical aspects of people at the expense of the emotional aspects. However, emotional aspects are equally important, as negative emotional health can lead to social and mental health issues.

## II. Emotion Recognition System

In general, facial expression recognition system mainly involves four parts which are face image collection, image preprocessing, features extraction and expression recognition. Facial expression recognition system chart is shown in Fig. 1. The most important parts of facial expression recognition system are feature extraction and expression classification.

In order to acquire digital face image data through a camera connected to a USB port, Adaboost classifier and Haar-Like feature [4] are adopted face detection and tracking. There are many factors of captured image, such as contrast, brightness, image size, much redundant information etc., which affect the accuracy, robustness and instantaneity of facial expression recognition system. So it is most important to preprocess the captured image before expression recognition. In this paper, the preprocessing methods including geometry normalization, brightness normalization, histogram equalization, image filtering and facial effective area segmentation based on eight eyes [5].

To minimize the effects of illumination in facial expression recognition system, energy normalization method is apply to the face image collected from CK+ database after the segmentation of eight eyes. The size of image I (x, y) is M × N. The process of energy normalization is shown in (1). Where x and y are pixels locations.

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


Fig. 1. Emotion recognition system flow chart

The feature extractions section is made up of implementation of principal component analysis (PCA) and local binary pattern (LBP). The essence of PCA is finding the best projection direction which represents the original data in the condition of least mean-square. The dimension of feature space, which decides the size of data, can be reduced by PCA. PCA extracts the global grayscale features of a whole image and the global features are useful and important. But the global feature of facial expression is environment sensitive. Considering the face is non-rigid and the compute capacity, we select some local features to assist the global features to resolve this problem. Therefore, local binary pattern (LBP) for local feature extraction is combined with the PCA and learns with sparse coding descriptor (SC). The SC is use for optimization of the feature descriptors for effective classifications.

LBP is an effective texture description operator. It can extract the local neighbor texture information of grayscale image. First, LBP calculates the binary relations between each pixel in the image and its local neighbor points in the grayscale. Then, the binary relationship is weighted into a LBP code according to certain rules. Finally, LBP histogram sequence, which is extracted from sub-region of facial expression image, is described as the image feature.

The aim of sparse coding is to locate an effective method of images pattern classification by fusion of multiple features selected from a dictionary [6]. Given a sparse dictionary matrix D = $d_1, d_2, \ldots d_n$ that contains K atoms as column vectors d, the sparse coding problem of extracting image descriptor T (PCA, LBP, PCA+LBP) can be stated as finding the sparsest vector y such that T is approximately as Dy. A signal T can be represented by the linear combination of atoms:

$$T = \sum_{i=1}^{n} yD_n = Dy \quad (2)$$

$l(x_c, y_c)$ is any pixel within a local area of an image, $l(x_c, y_c)$ is the center of the 3×3 window, the other eight points are $g_0,\ldots,g_7$. Define the local area texture as $T = t(g_c, g_0,\ldots,g_7)$, binary other eight pixel within the window using the threshold, here set the threshold for the gray value of center point in the window. The equation is as following:

$$T \approx t(s(g_0 - g_7), \ldots, s(g_7 - g_c)) \quad (3)$$

| | | |
|---|---|---|
| 126 | 136 | 101 |
| 145 | 130 | 128 |
| 114 | 135 | 157 |

Threshold = 130

| | | |
|---|---|---|
| 0 | 1 | 0 |
| 1 | | 0 |
| 0 | 1 | 0 |

$(01000101)_2$ = LBP = $(69)_{10}$

Fig. 2. LBP implementation and coding

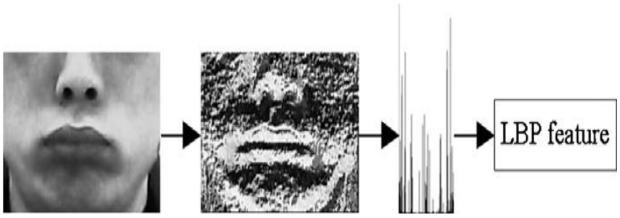


Fig. 3. LBP feature extraction

Read out an 8-bit binary number in a clockwise direction as an eigenvalue of this pixel. Convert the binary number into a decimal number by the following formula for each symbol function. Then, LBP code which is described the spatial structure of local image texture features is got as following:

$$LBP\,(x_c, y_c) = \sum_{i=0}^{7} s(g_i - g_7)2^i \qquad (4)$$

After scanning a facial expression image using LBP operator, the LBP coding image of original image is got. Then the texture feature of image can be described by counting the facial expression image histogram. The process of LBP feature extraction is shown in Fig. 3. The binary number of 3×3 patterns is 8, and the account of LBP code is 256 (28). The LBP coding image includes local micro mode information of the original image, such as edge, feature points and spot. So the local texture feature of an expression image can be described by a histogram which is formed by 256 LBP codes.

For facial expression recognition, the experiments show that mouth contributes most to facial expression. Based on the experimental observations and the consideration of compute capacity, the region of mouth is selected for local feature extraction in this paper. Through LBP extracts the local feature of the region of mouth and PCA extracts the global feature of the whole image, the recognition rate and the ability of robust can be improved to some extent. The extracted features of LBP and PCA descriptors are learn with SC dictionary for effective features representations then follow by the LSVM classification model.

## III. Emotion Classification with Linear SVM

SVM is a kind of data learning method based on the theory of statistical learning. The LSVM can deal with spurious regression problem and pattern recognition successfully. The mechanism of LSVM is looking for a hyper-plane meet for the requirement of classification, which is a best support vector to distinguish two different classes, under the condition of limited information based on small sample. The support vector maximizes the gap between classes and ensures the accuracy of classification at the same time. LSVM can resolve the problems of insufficient sample of facial expression and big difference of quantity between different expressions.

Facial expression recognition is a multi-class classification model. In this model, LSVM perform input data into a high dimensional feature space, where linear algebra and geometry may be used to separate data that is only separable with nonlinear rules in input space, through a selected nonlinear mapping function. The nonlinear mapping function is called kernel function. The learning algorithm is formulated to make use of kernel functions, allowing efficient computation of inner products directly in feature space. Typically used kernel functions include linear, polynomial, RBF and sigmoid. In this paper, we compared linear kernel with polynomial, RBF and sigmoid functions.

Fig. 4. Examples of facial emotions obtained from CK+ database [7]. (a) Disgust, (b) Happy, (c) Surprise, (d) Fear, (e) Angry, (f) Contempt, (g) Sadness, and (h) Neutral.

If we are happy, corners of the lips are drawn back and up, mouth may or may not be parted, teeth exposed, a wrinkle runs from outer nose to outer lip, cheeks are raised, lower lid may show wrinkles or be tense, crow's feet near the outside of the eyes. Disgust has been identified as one of the basic emotions. Its basic definition is "bad taste" secondarily to anything which causes a similar feeling, through the sense of smell, touch and even of eyesight. The well-defined facial expression of disgust is characterized by furrowing of the eyebrows, closure of the eyes and pupil constriction, wrinkling of the nose, upper lip retraction and upward movement of the lower lip and chin, drawing the corners of the mouth down and back.

## IV. 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 [7]. Total sample of 800 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 36 × 36. 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 I. AVERAGE RECOGNITION RATE (ARR) FOR EXPERIMENTS (EXPT. 1 AND EXPT. 2)

| Methods | Expt. 1: ARR Normal | Expt. 2: ARR Worse |
|---|---|---|
| PCA + LSVM [8] | 91.25 | 88.05 |
| LBP + LSVM | 92.52 | 90.31 |
| PCA + LBP +LSVM [8] | 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 reported in [8] 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 II. 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.

## V. RESULTS ANALYSIS AND DISCUSSIONS

In this article, we proposed an approach for facial emotion recognition using 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.

## REFERENCES

[1] A. Saeed, A. Al-Hamadi, R. Niese, M. Elzobi, Frame-based facial expression recognition using geometrical features, Adv. Hum. Comput. Interact. 14 (2014) 1–14.

[2] S. Zafeiriou, M. Petrou, Sparse Representation for Facial Expression Recognition via L1 Optimization, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., 2010.

[3] B. Jiang, F.Valstar, M. Pantic, Action Unit Detection Using Sparse Appearance Descriptors in Space-Time Video Volumes, IEEE Int'l Conf. on Automatic Face and Gesture Recognition (AFGR), 2011.

[4] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137 154.

[5] Y. Luo, C.M. Wu, Y. Zhang, Facial expression recognition based on principal component analysis and support vector machine applied in intelligent wheelchair, Appl. Res. Comput. 29 (8) (2012) 3166–3168.

[6] Aliyu Nuhu Shuaibu, Nidal Kamel, Mohd Naufal Saad, Aamir Saeed, Ibrahima Faye and Yasir Salih Ali, Sparse Representation for Crowd Attributes Recognition, IEEE Access Journal, 2017.

[7] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, The Extended Cohn-Kanade Dataset(CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression, IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR), 2010.

[8] Yuan Luo, Cai-ming Wu, Yi Zhang, Facial expression recognition based on fusion feature of PCA and LBP with SVM, Optik 124 (2013) 2767– 2770.