Facial expression recognition using a combination of enhanced local binary pattern and pyramid histogram of oriented gradients features extraction

Maede Sharifnejad\textsuperscript{1} \quad Asadollah Shahbahrami\textsuperscript{1} \quad Alireza Akoushideh\textsuperscript{2} \quad Reza Zare Hassanpour\textsuperscript{3}

\textsuperscript{1} Department of Computer Engineering, University of Guilan, Rasht, Iran
\textsuperscript{2} Department of Electronic Engineering, Faculty of Shahid-Chamran, Technical and Vocational University (TVU), Guilan Branch, Rasht, Iran
\textsuperscript{3} Department of Technology and Operations Management, Erasmus University, Rotterdam, The Netherlands

Correspondence
Asadollah Shahbahrami, Department of Computer Engineering, University of Guilan, Rasht, Iran.
Email: shahbahrami@guilan.ac.ir

Abstract
Automatic facial expression recognition, which has many applications such as drivers, patients, and criminals’ emotions recognition, is a challenging task. This is due to the variety of individuals and facial expression variability in different conditions, for instance, gender, race, colour and changing illumination. In addition, there are many regions in a face image such as forehead, mouth, eyes, eyebrows, nose, cheeks and chin, and extracting features of all these regions are expensive in terms of computational time. Each of the six basic emotions of anger, disgust, fear, happiness, sadness and surprise affect some regions more than the other regions. The goal of this study is to evaluate the performance of enhanced local binary pattern, pyramid histogram of oriented gradients feature-extraction algorithms and their combination in terms of recognition accuracy, feature vector length and computational time on one, two and three combined regions of a face image. Our experimental results show that the combination of both feature-extraction algorithms yields an average recognition accuracy of 95.33% using three regions, that is, the mouth, nose and eyes on Cohn–Kanade dataset. Besides, the mouth region is the most important part in terms of accuracy in comparison to eyes, nose and combination of both eyes and nose regions.

1 | INTRODUCTION

In interpersonal communications, facial expressions have an important role and provide useful information about a person's mood condition. As an example, a smile may indicate a positive or optimistic mood, while fear, sadness, and disgust indicate a negative mood \cite{1}. Studies show that to understand an individual's emotions, 38% of information achieved through vocal cues, 7% through verbal cues, and 55% through facial expressions \cite{2}. These values indicate the importance of facial expressions to express a message adequately. Automatic classification of facial expressions has a significant influence on many areas such as human and computer interactions, intelligent transportation, medicines, tele-education. Therefore, studies of automatic facial expression recognition (AFER) and the improvement of its methods via machines have recently become a growing issue in artificial intelligence researches and machine vision \cite{3}. In 1971, regardless of various ethnic groups, six basic universal emotions were defined as anger, disgust, fear, happiness, sadness, and surprise \cite{4}.

Facial expression recognition (FER) usually has three stages, namely, face detection, features extraction and classification. The most important stage in this system is the features extraction, which extracts features from either the whole face image or some parts of the face such as the forehead, mouth, eyes, eyebrows, nose, and cheeks. The percentages of changes on human faces for each expression in different cultures and conditions such as skin colour, age, gender, race and illuminations are different. In other words, the amount of changes that occur in different parts of a face differs from presenting each expression. Consequently, it is hard to find a suitable feature-extraction algorithm to work efficiently in terms of...
recognition accuracy and computational time for each emotion. When feature extraction is performed on whole face image, it applies unwanted non-linear features to data that it may cause a problem for good-learning classification [5]. Consequently, to overcome the problem, it is concentrated on the key face areas such as the eyebrows, eyes, mouth and nose. These areas are typically affected by changing expression in different individuals.

The goal of this study is to evaluate the feature-extraction algorithms, enhanced local binary patterns (ELBP), pyramid histogram of oriented gradients (PHOG) and their combination in terms of recognition accuracy, computational time, and feature vector length for six basic facial expressions. In other words, we are trying to answer the following questions. Which regions or which combination of regions of a face image are more important regions from the feature-extraction algorithms point of view? Which feature-extraction algorithms, (ML+MR+MB=ELBP) [27], PHOG and their combination have the highest- and the lowest-average recognition accuracy? Which basic emotions can be recognised with higher accuracy compared to other emotions using a simple or single feature-extraction algorithm? Is the combination of feature-extraction algorithms needed to increase the average recognition accuracy of facial expression?

Our implementation results show that: Feature extraction of the mouth region has the highest-average recognition accuracy in comparison to each eyes region, nose region, and both eyes and nose regions together and whole face image. When features are extracted from the whole face, due to common features like parts of the forehead and cheeks, recognition accuracy decreases.

Feature extraction of both mouth and eye regions has the highest-average recognition accuracy in comparisons to regions of eyes and nose, and regions of mouth and nose together. In addition, features extraction of three regions–mouth, eyes, and nose, has the highest-average recognition accuracy in all algorithms in comparison with the combination of one and two regions.

The combination of ELBP+PHOG feature extraction has the highest-average recognition accuracy in comparison to each algorithm separately for all cases, one, two and three regions, and ELBP has the lowest-average recognition accuracy.

The recognition accuracy of some basic emotions such as happiness and disgust using a single-feature-extraction algorithm like PHOG is higher than the recognition accuracy of other emotions. In other words, to recognise some basic emotions, we do not need either complex feature-extraction algorithms or a combination of feature-extraction algorithms. On the other hand, a combination of feature-extraction algorithms causes the recognition accuracy of some other basic emotions such as sadness and surprise to be increased and this increases the computational time.

The remainder of this study is organised as follows. A brief review of the local binary pattern (LBP), PHOG and related study are presented in Section 2. The proposed approach is discussed in Section 3. In Section 4, evaluation and comparison results are described. Finally, the conclusions are drawn in Section 5.

2 | BACKGROUND INFORMATION AND RELATED STUDY

2.1 | Local binary pattern

LBP is a texture-feature-extraction algorithm that was introduced by Ojala et. al. in 1996 [6]. Due to its resistance to brightness changes and low-computational complexity, this descriptor is one of the most common descriptors. This algorithm is applied to a window, for instance, a window of dimension 3 x 3, which is depicted in Figure 1 and it can be extended to larger windows. For this, a neighbourhood circle \((P, R)\) is defined. The \(P\) is the number of neighbours and \(R\) is the neighbourhood radius. The centre of the window is put on each pixel and for that pixel in the window; its value is compared with the value of eight neighbours \(P_n\) \((n = 0, 1, \ldots, 7)\). If the central pixel value is less than its neighbour, value 1 is set otherwise value 0 is put into the neighbour pixel and pixels are followed as a circle clockwise or counter clockwise. The result is an eight-bit binary string and then turns an equivalent decimal value. Finally, a histogram of values is created as a features vector. Label LBP is computed for each central pixel \((x, y)\) image \(f(x, y)\) according to Equation (1) that \(s(z)\) is threshold function in this equation and is valued as Equation (2).

\[
LBP_{BR} (x, y) = \sum_{p=0}^{P-1} s \left( \left( f_{x_0, y_0} - f(x, y) \right) 2^p \right) 
\]

\[
s (z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}
\]

2.2 | Pyramid histogram of oriented gradient

The PHOG is originally used for object retrieval. Its principle is to divide a tracked and localised interest area into many cells at several pyramid levels. Gradient orientation on all pixels within each cells is accumulated to form a histogram. All the histograms are concatenated to construct the final histogram. PHOG is an image descriptor based on image direction...
gradient. This algorithm is based on division of an image, regarding to the number of different levels, into some cells in pyramid level as shown in Figure 2. Each level \( L \) has \( 2^L \times 2^L \) cells (for instance, \( L = 3 \) has \( 8 \times 8 \) cells). In each cell, gradient magnitude \( \| (x, y) \| \) and gradient direction \( \theta (x, y) \) on each pixel \( (x, y) \) are determined using Equations (3) and (4) where \( g_x (x, y) \) and \( g_y (x, y) \) are image gradients \( x \) and \( y \) directions, respectively. Histogram vector is calculated according to the size and direction values in each cell. By putting cells histogram vectors together, a bigger feature vector is made and the final feature vector is the result of PHOG method [7, 8].

\[
\begin{align*}
\| (x, y) \| &= \sqrt{g_x (x, y)^2 + g_y (x, y)^2} \\
\theta &= \arctan \frac{g_x (x, y)}{g_y (x, y)}
\end{align*}
\]

### 2.3 Related study

In recent years, many methods have been introduced to FER such as detecting facial actions that are related to a specific expression or making classification based on the extracted image features. The facial action coding system (FACS) [9] refers to a set of facial muscle movements that correspond to a displayed emotion. In FACS, expressions are encoded by action units (AUs), which refer to some tiny but discriminable facial muscle changes. Many algorithms based on geometry and textures have been introduced for feature extraction [10–12]. Two efficient methods for geometrical feature extraction are active appearance model (AAM) [13] and active shape model (ASM) [14]. In AAM and ASM methods, each face image needs to be manually labelled with 68 landmarks that is exhausting and time-consuming for large databases. Some of the feature-extraction methods are based on appearance, including, LBP [6], Gabor filter [15], HOG [16], local phase quantisation [17] and so forth. Deep-learning methods also are recently introduced for FER that are concentrated on how to build and learn different models of nerve net like recurrent neural network and convolutional neural network (CNN) [5, 18–21]. Some recent related studies have been compared to each other in terms of feature-extraction and classification algorithms and the datasets that has been used in their evaluations.

In [5], for better representation of facial expression and to reduce feature-vector length, the authors proposed the fusions of geometric with LBP features using auto-encoders (AE). Besides, they proposed the self organizing map (SOM)-based classifier with a soft-thresholding technique at the output nodes and improved-learning algorithm to train the parameters that reduced false prediction rate and enhanced the FER performance. Scale invariant feature transform, a meta-heuristic algorithm called grey wolf optimisation (GWO) algorithm and GWO-based neural network (NN) have been used for feature extraction, selection of the optimal feature and classification steps, respectively, in [18]. This method is carried out using two types of databases, such as the japanese female facial expression (JAFFE) and Cohn–Kanade (CK+). The GWO algorithm can select the relevant features from the feature set to be applicable for classification and reduce the error of weights applied for training during the classification process. In order to determine the accurate position of facial in the image to choose 51 landmarks, the AAM algorithm was used [19].

Then facial features are extracted and reduced by HOG. However, the usage of the high-dimensional feature will bring great challenges on training, calculation, and storage. Therefore, a principal component analysis algorithm for dimension reduction was used. Finally, for expression classification, deep sparse auto-encoders is utilised. A feature redundancy-reduced (FRR) CNN has been used for feature extraction and expression classification in [20]. Convolutional kernels of FRR-CNN are improved, which results in generating less redundant features and yields a more compact representation of an image.

Furthermore, the transformation-invariant-pooling strategy is used to feature extractions. In [21], local regions of a facial image are detected by a Canny edge detection algorithm. Then multi-layer perperptron (MLP) NN with backpropagation-learning algorithm was applied to recognise six basic facial expressions on the JAFFE dataset. However, using the MLP-NN to classify and determine the number of hidden nodes by experience required the high-calculating cost for the learning process. The deep-based framework method has suggested two branches of CNN [22]. One of them extracts local features from image patches, and other branch extracts holistic features from the whole expressional image. These two types of hierarchical features represent expressions in different scales and being complementary to each other.

Additionally, the typical CNN structure is modified with the proposed expression transformation-invariant (ETI)-pooling strategy, which reduces the impact of nuisance variations such as rotations, noises in classification. Also, the authors [22] proposed a method to learn salient expressional image patches based on the L2-norm of local feature vectors and visualise the active regions relevant to expression changes. In [23], the authors proposed a robust vectorised CNN model that introduces an attention mechanism for extracting features in the region of interests (ROIs) of the face. In particular, the attention concept is adopted to perform ROIs-related convolution calculation, and ROIs-related convolution calculation...
results of the specific fields in the ROIs are increased by extracting more robust features. In general, in all NN method for facial expressions recognition, determining which features help identify different expressions are essential, and there is a massive information transmission loss between layers of network with multiple layers. In [24], the authors proposed a framework for facial feature recognition that consists of MobileNet-like CNN and k-means clustering that is grouping faces based on their characteristics on the CelebA [25] dataset. This model performed on 37 facial attributes such as wearing hat, blonde-hair, eyeglasses, bald, smiling and so forth, but the k-means clustering is slow and requires the high-calculating costs because of the large representation space of images. Table 1 shows a comparison between recent studies in term of feature extraction, classification and how images are selected, while the CK+ database has been used in most of the studies.

## 3 | PROPOSED METHOD

The general block diagram of the proposed approach to recognise the six basic emotions is depicted in Figure 3. There are four main steps in the figure, image pre-processing operations, feature-extraction algorithms ELBP and PHOG, combination unit for the extracted features and expression classification unit using SVM.

### 3.1 | Pre-processing operations

Pre-processing operations are an important and essential stage to increase the accuracy and efficiency of the proposed method. In this stage, first, faces are detected and some filtering operations such as average or Gaussian filtering and resizing...
operations are applied to remove noises. Primarily to separate the ROI from the background and to reduce input images noise, image corruption is applied to get an exact face in the input image. For this purpose, the AAM and Viola–Jones algorithm [18] have been used. The AAM algorithm, as shown in Figure 4 [26] has 68 landmarks that each consist of image width and length coordinates. Based on these landmarks, face and its parts are cropped. Viola–Jones algorithm has also been tested. But it was observed that this algorithm takes relatively much time to extract the face areas and also extracts multiple images of the face areas in different dimensions for each face image. Applying face landmarks method to crop images is simple and fast and has no error compared to Viola–Jones algorithm. The feature should be extracted from a whole face image and also selected regions. Figure 5 depicts a sample of three key areas of face image–eyes, nose, and mouth.

3.2 | Feature extraction and classification

In this study, Enhanced LBP (ELBP = ML+MR+MB) [27] and PHOG feature extraction algorithms are used. The ELBP consists of multilevel (ML), multi-band (MB), and multi-resolutions (MR) that extract texture features from different frequency bands of different pyramid levels of wavelet coefficients as is depicted in Fig. 6. To evaluate ELBP operator efficiency, multiple sampling filters of the low level were used such as Gaussian filter and different kinds of wavelet transform functions, Haar, Coiflets, Daubechies with different decomposition levels of 2 and 3. Table 2 depicts the recognition accuracy of ELBP feature extraction for the whole face using different Gaussian pyramid transform and wavelet transform functions in different pyramid levels. As this table shows, the highest average recognition accuracy is obtained for three decomposition levels, Gaussian filter and eight-number of resolutions that extract image texture feature only from two-dimensional low band–low pass. Figure 7 depicts this operator, and each level makes a feature vector with dimensions \( f_1, \ldots, f_{59} \) that all feature vectors are combined at the end.

Table 3 represents the evaluation of recognition accuracy and length of feature vector of PHOG operator with eight neighbours and different parameters of levels number from 0 to 3 and directions of 180 and 360 degrees for whole face. Considering the table, the highest accuracy is related to decomposition level 3 and direction of 180 degrees. In conclusion, these parameters values are used in the proposed method and for each image feature vector PHOG is evaluated with the length \( f_1, \ldots, f_{680} \).

After creating feature vectors ELBP and PHOG to combine these features, all resulted vectors are normalised in zero to one range and to improve the proposed system efficiency, they are combined as Figure 8.

To evaluate the performance of the proposed method, classification of multiple classes of linear support-vector machine (SVM) with 10 different validations was used. SVM is a strong classifier and one of supervised learning methods that have attracted many attentions in pattern recognition and FER [28, 29]. SVM algorithm separates two classes by using a linear border. In this condition, it is supposed that there are a set of separable learning samples that they can be labelled by \( y \). Samples are expressed in ordered pairs \((x_i, y)\) that \( i = 1, \ldots, n \) and \( y \) have the values \(-1, 1\). The data can be separated and classified by multiple separators. Some of these learning samples with least distance to decision making border are considered as support vectors.

4 | PERFORMANCE EVALUATION

4.1 | Experimental setups

The proposed algorithm has been implemented using MATLAB R2015b programming language on a system with the Intel Core i7 processor, 6 GB main memory and Windows 7 operating system. The CK+ dataset was used for evaluation
![Figure 6](image) Performance of enhanced local binary pattern (ELBP) operator [27]

### Table 2: Recognition accuracy of ELBP operator using different Gaussian pyramid transform and wavelet transform functions

| Pyramid transform filter | Feature-vector length | # of levels | Recognition accuracy (%) | Pyramid transform filter | Feature-vector length | # of levels | Recognition accuracy (%) |
|--------------------------|-----------------------|-------------|---------------------------|--------------------------|-----------------------|-------------|---------------------------|
| DB5 wavelet              | 295                   | 2           | 56.66                     | Gaussian                 | 118                   | 2           | 57.77                     |
| DB5 wavelet              | 531                   | 3           | 54.04                     | DB1 wavelet              | 295                   | 2           | 60.37                     |
| DB6 wavelet              | 295                   | 2           | 61.48                     | DB2 wavelet              | 295                   | 2           | 57.77                     |
| DB6 wavelet              | 531                   | 3           | 57.77                     | DB3 wavelet              | 295                   | 2           | 55.18                     |
| Sym wavelet              | 295                   | 2           | 57.77                     | DB4 wavelet              | 295                   | 2           | 54.07                     |
| Sym wavelet              | 531                   | 3           | 55.18                     | DB4 wavelet              | 531                   | 3           | 54.07                     |
| Coif wavelet             | 295                   | 2           | 56.66                     | DB2 wavelet              | 531                   | 3           | 55.18                     |
| Coif wavelet             | 531                   | 3           | 57.77                     | DB3 wavelet              | 531                   | 3           | 58.88                     |
| Haar wavelet             | 295                   | 2           | 61.79                     | DB4 wavelet              | 295                   | 2           | 54.07                     |
| Haar wavelet             | 531                   | 3           | 54.07                     | DB4 wavelet              | 531                   | 3           | 54.07                     |

[30]. This dataset has various races of African, Asian and Caucasian people. An example of each facial emotion of this dataset is depicted in Figure 9. This dataset includes different men and women from 18- to 30-year-old and an emotion has been labelled in the dataset. Every image sequence in dataset starts with a natural expression and finishes with one of the six

### Table 3: Recognition accuracy of PHOG operator with different parameters using eight neighbours

| Degrees | # of levels | Feature-vector length | Recognition accuracy (%) |
|---------|-------------|-----------------------|---------------------------|
| 180     | 0           | 8                     | 50.45                     |
| 180     | 1           | 40                    | 60.2                      |
| 180     | 2           | 168                   | 73.12                     |
| 180     | 3           | 680                   | 85.43                     |
| 360     | 0           | 8                     | 42                        |
| 360     | 1           | 40                    | 57.3                      |
| 360     | 2           | 168                   | 65.1                      |
| 360     | 3           | 680                   | 78.8                      |

### Table 4: Sitting parameters to implement the proposed algorithms for features extraction and classification

| ELBP | 1. Gaussian pyramid transform tool  
|      | 2. Feature vector length 177  
|      | 3. Number of levels is 3  
|      | 4. Number of resolution is 8  
|      | 5. Number of band is 1  
| PHOG | 1. Degrees is 180  
|      | 2. Feature vector length 680  
|      | 3. Number of levels is 5  
|      | 4. Number of neighbours is 8  
| SVM  | 1. Linear kernel  
|      | 2. Ten-fold cross validation  

---

SHARFNEJAD ET AL.
### TABLE 5  Comparison of ELBP and PHOG algorithms and their combination in terms of average accuracy, computational time and feature vector length using CK+ dataset

| Selected parts of a facial image | Whole face | Eyes | Nose | Mouth | Eyes, mouth | Nose, mouth | Eyes, nose, mouth |
|----------------------------------|------------|-----|------|-------|-------------|-------------|------------------|
| Number of regions                | 1          | 1   | 1    | 1     | 2           | 2           | 2                |
| ELBP Average accuracy (%)        | 64.48      | 64.98 | 60.4 | 74.85 | 82          | 79.1        | 72.04            |
| Computational time (s)           | 21.09      | 18.37 | 18.68 | 18.3  | 24.23       | 24.27       | 24.09            |
| Feature vector length            | 354        | 354  | 354  | 354   | 708         | 708         | 708              |
| PHOG Average accuracy (%)        | 85.43      | 73   | 67.27 | 88    | 90.8        | 88.12       | 77.67            |
| Computational time (s)           | 24.66      | 22.02 | 22.21 | 21.44  | 33.57       | 33.9        | 33.1             |
| Feature vector length            | 1360       | 1360 | 1360 | 1360   | 2720        | 2720        | 2720             |
| ELBP + PHOG Average accuracy (%) | 89.52      | 74.15 | 68.21 | 90     | 91.2        | 90.4        | 81.1             |
| Computational time (s)           | 32.27      | 25.35 | 25.37 | 25.3   | 39.39       | 39.53       | 45.37            |
| Feature vector length            | 1714       | 1714 | 1714 | 1714   | 3428        | 3428        | 3428             |

### TABLE 6  Confusion matrix for ELBP algorithm to recognise the six basic emotions using three regions, eyes, nose and mouth

|                  | Anger | Disgust | Fear | Happiness | Sadness | Surprise | Accuracy (%) |
|------------------|-------|---------|------|-----------|---------|----------|--------------|
| Anger            | 39    | 1       | 0    | 0         | 0       | 5        | 86.66        |
| Disgust          | 0     | 57      | 0    | 1         | 1       | 0        | 96.61        |
| Fear             | 2     | 2       | 13   | 3         | 2       | 3        | 52           |
| Happiness        | 0     | 0       | 0    | 68        | 0       | 1        | 98.55        |
| Sadness          | 7     | 0       | 0    | 0         | 0       | 20       | 71.43        |
| Surprise         | 0     | 0       | 1    | 1         | 0       | 81       | 97.59        |
| Accuracy (%)     | 83.81 |

mentioned expressions. In our evaluation and comparison, 309 image sequences have been used with 45, 59, 25, 69, 28, and 83 expressing anger, disgust, fear, happiness, sadness and surprise, respectively. Dataset was divided into 10 parts with almost equal numbers that nine sets were used for training and the other part was applied as test dataset. This process has been repeated for k-fold cross-validation. In order to evaluate the proposed technique, we have used recognition accuracy, computational time and feature vector length criteria [31, 32]. In order to evaluate the effectiveness of each region on a facial image and also their combination on the measurements parameters, we have divided a facial image into one, two and three regions. The whole face has been considered as one region. Three most important regions, that is, the eye, nose and

### TABLE 7  Confusion matrix for PHOG algorithm to recognise the six basic emotions using three regions, eyes, nose and mouth

|                  | Anger | Disgust | Fear | Happiness | Sadness | Surprise | Accuracy (%) |
|------------------|-------|---------|------|-----------|---------|----------|--------------|
| Anger            | 40    | 2       | 0    | 0         | 3       | 0        | 88.88        |
| Disgust          | 0     | 59      | 0    | 0         | 0       | 0        | 100          |
| Fear             | 1     | 0       | 21   | 1         | 0       | 2        | 84           |
| Happiness        | 0     | 0       | 0    | 69        | 0       | 0        | 100          |
| Sadness          | 3     | 2       | 0    | 0         | 23      | 0        | 82.14        |
| Surprise         | 0     | 0       | 0    | 1         | 0       | 82       | 98.79        |
| Accuracy (%)     | 92.3  |
TABLE 8  Confusion matrix for ELBP + PHOG algorithm to recognise the six basic emotions using three regions, eyes, nose and mouth

|       | Anger | Disgust | Fear | Happiness | Sadness | Surprise | Accuracy (%) |
|-------|-------|---------|------|-----------|---------|----------|--------------|
| Anger | 41    | 2       | 0    | 0         | 2       | 0        | 91.11        |
| Disgust | 0     | 59      | 0    | 0         | 0       | 0        | 100          |
| Fear  | 1     | 0       | 22   | 0         | 0       | 2        | 88           |
| Happiness | 0    | 0       | 0    | 69        | 0       | 0        | 100          |
| Sadness | 3     | 0       | 0    | 0         | 25      | 0        | 92.86        |
| Surprise | 0    | 0       | 0    | 0         | 83      | 0        | 100          |

FIGURE 7  Performance of ELBP operator with three-level, Gaussian filter and eight-number of resolutions

FIGURE 8  Combination of extracted features of ELBP and PHOG

FIGURE 9  One sample of each facial expression—anger, disgust, fear, happiness, sadness and surprise, in Cohn–Kanade (CK+) database

FIGURE 10  A comparison of vector-feature length for one, two, and three regions of a face image for each feature-extraction algorithm and their combination

4.2 Experimental results

The obtained results, recognition accuracy, feature-vector length and computational time for both proposed feature-extraction algorithms and their combination for different parts of a facial image for six basic expressions are depicted in Table 5. As can be seen in the table, from one region point of view among the

mouth have been considered separately and their combination too. The computational time has been measured in seconds. The selected parameters for the proposed algorithms are depicted in Table 4. As this table shows, SVM algorithm was used for classification step.
FIGURE 11 | Recognition accuracy of each ELBP and PHOG feature-extraction algorithms and their combination for six basic emotions—anger, disgust, fear, happiness, sadness and surprise, using three regions—eyes, nose and mouth.

whole face, eye, nose, and mouth, the mouth has the highest-average recognition accuracy for each feature-extraction algorithm and also their combinations. There are three combinations of two regions: Eyes and mouth, nose and mouth, eyes and nose. As the table shows, eyes and mouth combination has the highest-average recognition accuracy for each feature-extraction algorithm and also their combinations in comparison to the other two combinations. It should be noted that the average accuracy of one region, the mouth, is higher than the combination of two regions of eyes and nose for all algorithms. As can be seen in the rightmost column, the combination of the three, the eyes, mouth and nose regions, have the highest-average recognition accuracy for each feature-extraction algorithm and also their combinations. The results are 83.81, 92.3 and 95.33% for ELBP, PHOG and their combination, respectively.

Computational time increases as the feature-vector length increases, while increasing the feature-vector length does not always increase recognition accuracy. The way of choosing regions and feature-extraction algorithms have a direct impact on recognition accuracy. For example, extracting several features, ELBP and PHOG from the mouth region, which has less feature-vector length and computational time compared to extracting a feature such as PHOG from both the nose and mouth regions, has led to increased recognition accuracy.

The ELBP feature extraction is faster than PHOG for each evaluation. One main reason for this is that the ELBP feature-vector length is smaller than PHOG feature-vector length. As was expected, in each algorithm, the computational time and feature-vector length in one region is smaller than the combination of two and three regions. The combination of both algorithms for three regions has a vector length of 5142 and its computational time is 53.17 s, while these measurements for ELBP and PHOG are 1062, 35.03 and 4080, 45.73, respectively. In other words, the highest-average accuracy of the combination of both feature-extraction algorithms is because of its larger-vector length, which has been computed using three regions, the mouth, eyes, and nose.

Figure 10 depicts the feature-vector length for one, two and three regions for each algorithm. As can be seen, the feature-vector length of the ELBP is the minimum in all cases and the combination algorithm, ELBP+PHOG has the maximum-feature vector length.

As was mentioned, the best-average recognition accuracy for each algorithm, ELBP, PHOG and their combination, which is 83.81, 92.3 and 95.33%, was obtained using three regions. It is observed that classification from the combination of data outperforms from the individual features in terms of average recognition accuracy. In order to understand the effectiveness of feature-extraction algorithms to recognise each facial expression, we have presented the confusion matrix. In the confusion matrix, the recognition accuracy of each algorithm for each basic emotion is presented. For example, Table 6 depicts the confusion matrix for ELBP to recognise each of the six basic emotions. As this table shows, this algorithm has the highest-recognition accuracy for happiness that is 98.55%, while it has the lowest accuracy for fear, which it is 52%.

Confusion matrix of the PHOG algorithm is depicted in Table 7. The highest-recognition accuracy is 100% and it is related to disgust and happiness emotions, while the lowest-recognition accuracy is 82.14% and it is related to sadness emotion.

The confusion matrix of the combination of ELBP and PHOG is depicted in Table 8. This table shows that all six expressions are distinguished with high accuracy. Among six expressions, happiness, disgust and surprise achieve excellent performances with 100% accuracy, owing to their distinctive features in the regions of the eyes and mouth. Meanwhile, the expressions of anger and sadness provide satisfactory results, while they are easily confused with the disgust. In addition, the anger expression is easily misclassified as sadness. More precisely, the percentages of the anger expression falsely classified as the disgust and sadness are 4.44%. The percentage of the sadness expression falsely classified as the disgust is 7.14%. Meanwhile, the lowest accuracy is 88% and it is related to fear emotion. Some samples of fear are misclassified as anger and
CONCLUSION

AFER has many applications in machine-vision systems, while it is a challenging task. This is because face images differ, for instance, in terms of gender, race, age, and colour. Besides, there are six basic emotions and different regions on the face. Each emotion has its impact on distinct regions of the face. The recognition accuracy and computational time of the whole system should be considered for real-time systems. In this study, recognition accuracy and computational time of ELBP and PHOG feature-extraction algorithms and their combination on some regions of the face, mouth, eyes, nose and their different combinations have been evaluated. Our implementation results, which have been obtained using the CK+ dataset, show that feature extraction of the mouth region has the highest average recognition accuracy in comparison to others such as the eyes, nose, and both eyes and nose regions and whole face image. In addition, features extraction of three regions, the mouth, eyes, and nose, have the highest average recognition accuracy in all algorithms in comparison with the combination of one and two regions. The combination of ELBP+PHOG feature-extraction algorithms has the highest average recognition accuracy in comparison to each algorithm separately for all cases, and it is 95.33%. Furthermore, this combination achieves 100% recognition accuracy for three basic emotions, that is, disgust, surprise, and happiness. In other words, texture features and their combination are more useful for these three facial expressions compared to others.

REFERENCES

1. Tang, Y., et al.: Geometric-convolutional feature fusion based on learning propagation for facial expression recognition. IEEE Access 6, 42532–42540 (2018)
2. a) Mehrabian, A.: Communication without words. In: Mortensen D.C. (ed.) Communication Theory, pp. 193–200. Taylor and Francis, Milton, UK (2008) b) A. Mehrabian: Communication without words. Psychol. Today 2(4), 33–55 (1968)
3. Qi, C., et al.: Facial expressions recognition based on cognition and mapped binary patterns. IEEE Access 6, 18795–18803 (2018)
4. Fasel, B., Luettin, J.: Automatic facial expression analysis: A survey. Pattern Recognit. 36(1), 259–275 (2003).
5. Majumder, A., et al.: Automatic facial expression recognition system using deep network-based data fusion. IEEE Trans. Cybern. 48(1), 103–114 (2018).
6. Ojala, T., et al.: A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In: International Conference on Advances in Pattern Recognition, Rio de Janeiro, Brazil, pp. 399–408 (2001).
7. Yang, B., et al.: A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition. In: 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt (2009).
8. Wang, J., et al.: Human action recognition based on pyramid histogram of oriented gradients. In: IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, Alaska (2011).
9. Cohn, J.F., et al.: Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. Psychophysiology 36(1), 35–43 (1999).
10. Mehta, D., et al.: Facial emotion recognition: A survey and real-world user experiences in mixed reality. Sensors 18(2), 416 (2018).
11. Corneanu, C.A., et al.: Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE Trans. Pattern Anal. Mach. Intell. 38(8), 1548–1568 (2016).
12. Kumari, J., et al.: Facial expression recognition: A survey. Procedia Comput. Sci. 58, 486–491 (2015).
13. Cootes, T.F., et al.: Active appearance models. In: European Conference on Computer Vision, Freiburg, Germany, pp. 484-498 (1998).
14. Cootes, T.F., et al.: Active shape models-their training and application. Comput. Vision Image Understanding 61(1), 38-59 (1995).
15. Senechal, T., et al.: Facial action recognition combining heterogeneous features via multikernel learning. IEEE Trans. Syst. Man Cybern. 42(4), 993–1005 (2012).
16. Yanwei, P., et al.: Robust CoHOG feature extraction in human-centered image/video management system. IEEE Trans. Syst. Man Cybern. 42(2), 458–468 (2012).
17. Zhou, S.-R., et al.: Local binary pattern (LBP) and local phase quantization (LBQ) based on gabor filter for face representation. Neurocomputing 116, 260–264 (2013).
18. Nirmala Sreedharan, N.P., et al.: Grey wolf optimisation-based feature selection and classification for facial emotion recognition. IET Biom. 7(5), 490–499 (2018).
19. Zeng, N., et al.: Facial expression recognition via learning deep sparse autoencoders. Neurocomputing 273, 643–649 (2018).
20. Xie, S., Hu, H.: Facial expression recognition with FRR-CNN. Electron. Lett. 53(4), 235–237 (2017).
21. Thai, L.H., et al.: A facial expression classification system integrating canny, principal component analysis and artificial neural network. Int. J. Mach. Learn. Comput. 388–393 (2011).
22. Xie, S., Hu, H.: Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans. Multimedia 21(1), 211–220 (2019).
23. Sun, X., et al.: ROI-attention vectorized CNN model for static facial expression recognition. IEEE Access 8, 7183-7194 (2020).
24. Anzalone, L., et al.: Transfer learning for facial attributes prediction and clustering, Smart City and Informatization, Guangzhou, China, pp. 105–117 (2019).
25. Liu, Z., et al.: Large-scale celebfaces attributes (CelebA) dataset. Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong (2015).
26. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio (2014).
27. Akoushideh, A., Mazloom-Nezhad Maybodi, B.: Efficient levels of spatial pyramid representation for local binary patterns. IET Comput. Vision 9(6), 871–883 (2015).
28. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Trans. Affective Comput. 2(4), 219–229 (2011).
29. Zhao, G., Pietikainen, M.: Dynamic texture recognition Using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007).
30. Lucey, P., et al.: The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, San Francisco, California (2010).
31. Ko, B.: A brief review of facial emotion recognition based on visual information. Sensors 18(2), 401 (2018).
32. Benitez-Quiroz, C.F., et al.: EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada (2016).

How to cite this article: Sharifnejad M, Shahbahrami A, Akoushideh A, Hassanpour RZ. Facial expression recognition using a combination of enhanced lbp and PHOG features extraction. IET Image Process. 2021;15:468–478. https://doi.org/10.1049/1pr2.12037