



HDG and HDGG: an extensible feature extraction descriptor for effective face and facial expressions recognition

Farid Ayeche¹ · Adel Alti^{2,3}

Received: 28 February 2020 / Accepted: 1 March 2021 / Published online: 17 March 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

The potential of facial and facial expression recognitions has gained increased interest in social interactions and biometric identification. Earlier facial identification methods suffer from drawbacks due to the lower identification accuracy under difficult lighting conditions. This paper presents two novel new descriptors called Histogram of Directional Gradient (HDG) and Histogram of Directional Gradient Generalized (HDGG) to extracting discriminant facial expression features for better classification accuracy with good efficiency than existing classifiers. The proposed descriptors are based on the directional local gradients combined with SVM (Support Vector Machine) linear classification. To build an efficient face and facial expression recognition, features with reduced dimension are used to boost the performance of the classification. Experiments are conducted on two public-domain datasets: JAFFE for facial expression recognition and YALE for face recognition. The experiment results show the best overall accuracy of 92.12% compared to other existing works. It demonstrates a fast execution time for face recognition ranging from 0.4 to 0.7 s in all evaluated databases.

Keywords Histogram of directional gradient · Texture feature analysis · SVM classifier · Facial expression recognition · Histogram of directional gradient generalized

1 Introduction

In recent few decades, facial recognition (FR) and facial expression recognition (FER) are becoming attractive research fields. This is due to technological advances when acquiring, preprocessing, extracting, and efficiently classifying facial features. The major purpose of these technologies is intended to provide an accurate identification rate and deep understanding of the dynamic behavior of speech and facial expressions.

In fact, recognition of facial expressions is a very important process to ensure the authenticity of people when

carrying the verbal message “face to face”. Facial expression recognition techniques can make a significant contribution to analyzing and modeling facial expressions and ensuring user safety [1]. It can be achieved using two approaches: geometric feature-based approaches as well as appearance-based approaches [2]. Facial expressions recognition with geometric feature-based approaches means localizing and extracting the face’s elements (*eyes, nose, and mouth*) directly on the face image. During appearance-based approaches, extraction of features is required to achieve the recognition of facial expressions [3] using several techniques such as color differences, texture gradient direction, Gabor features.

Note that in the classical FR and FER, the facial expressions recognition process will fit the appearance-based facial expression features in emotions. This process consists of extracting the most significant features from the face image without changing the size of the face image. However, this process was a leak of good features extraction from facial images and poor recognition rates. In this proposed work, we portray effective and efficient descriptors-based directional gradients enabled high accurate results and fast execution time for facial images. The major contributions of this proposed paper are described clearly as follows.

✉ Adel Alti
alti.adel@univ-setif.dz

¹ Mechatronics Laboratory (LMETR)-E1764200, Optics and Precision Mechanics Institute, University of SETIF-1, 19000 Setif, Algeria

² Department of Management Information Systems, College of Business & Economics, Qassim University, P.O. Box 6633, Buraidah 51452, Kingdom of Saudi Arabia

³ LRSD Laboratory Computer Science Department, Sciences Faculty, University of SETIF-1, 19000 Setif, Algeria

- Present a review of the existing techniques in addition to the emerging ones for face recognition and facial expressions recognition.
- Develop the new descriptors HDG (Histogram of Directional Gradient) and HDGG (Histogram of Directional Gradient Generalized) for classification based on SVM (Support Vector Machine). These descriptors mainly rely on a holistic approach based on magnitude and orientation maps to extract more discriminative features in the face image. They are represented as feature-to-tree defined by three hierarchical feature description levels.
- By applying these proposed descriptors, attaining the best accuracy and reduced discriminant features providing to the SVM classifier.
- Compare the proposed descriptors with the 11 standard descriptors among the most common algorithms training-tests for various performance metrics. Two benchmarks are used to evaluate descriptors, which come with a variety of images in multiple databases through an SVM classifier.

The remainder of this paper is organized to present related works of feature descriptions in Sect. 2. The different existing descriptors are presented in Sect. 3. Section 4 explains the proposed descriptors for features extraction, proposed framework in Sect. 5. Section 6 exposes details of the implementation and experimental result as well as discussions. Section 7 concludes the paper.

2 Related works

This section gives some related works in the literature and discusses their strengths and drawbacks. Many researchers focus on various classification problems including face and facial expressions recognition. Ahonen et al. [4] proposed Local Binary Pattern (LBP), which is an effective texture descriptor for facial images. It involves machine-learning models associated with feature vectors from the faces images. However, LBP relies on a high-dimensional feature vector, making it infeasible to scale for a large dataset. According to Zhang et al. [5], two descriptors, Gabor Binary Pattern (LGBPHS) and Gabor Phase Pattern (GPP) can solve the problem of changing the environmental conditions. Chen et al. [6] proposed Weber Local Descriptor (WLD) which is inspired by Weber's law providing better-resulting accuracy to face recognition issues. Other discriminative descriptors are also produced good accuracy results in facial recognition systems such as Local Gradient Pattern (LGP) [7], Gradient Direction Pattern (GDP) [8].

Rivera et al. [9] proposed Local Directional Number (LDN) pattern as a face descriptor to extract directional values. This work deals with the best directional numbers

with the most positive and negative directions of those edge responses. This work suffers from the self-imposed restriction of continuously having a leading *one* that immediately reduces the number of available combinations within the twofold vector by half. The number of *zero* and *one* has been used to process original LDP, depending on a threshold value. In [10], the authors proposed Local Phase Quantization (LPQ). They involve Gabor wavelet with blur invariance property to extract discriminative features. Therefore, both works suffer from the major drawback of a stable classification rate with a high number of features. Generally, Median Ternary Pattern (MTP) [11], Local Gradient Code (LGC) [12], and Histogram of Orientation Gradient (HOG) [13] have been widely used for facial feature extraction. More recently, the descriptor Local Gradient Neighbor (LGN) [14] have shown significant performance in the face recognition process by combining the qualities of both descriptors LBP [4] and LGC [12].

Works in [15–19] details five descriptors with machines learning are Scale Invariant Feature Transform (SIFT) with Independent Component Analysis (ICA) [15], Eigenfaces using Principle Component Analysis (PCA) [16], Gabor Wavelet (GW), and Linear Discriminant Analysis (LDA) [17], Pyramid of Histogram Oriented Gradient (PHOG) with Support Vector Machine (SVM) [18], Gabor wavelets with Principle Component Analysis (PCA) [19]. These descriptors are widely used for FR and FER systems. In general, the Eigenfaces discover the similarities between faces with minimal controlled environments. However, PCA suffers from a low recognition rate. PCA is much more dimensionality reduction than other existing recognition methods.

Other works involve a wide range of machine learning techniques on domains FR and FER from classic classifiers such as Artificial Neural Networks (ANNs) and Hidden Markov Models (HMMs) [20], to more advanced approaches such as Support Vectors Machines (SVMs) [17]. These methods aim to develop and evaluate the performance of a statistical classifier based on a new generation of neural networks using pattern code faces. However, SVMs are the most used approach for FR and FER systems. Recently deep learning with deep networks is being widely used for FER. It integrates both feature extraction and learning phases such as Cos Face and Deep Face [21]. In this context, SVM classifier has been used to improve classification performance [1], elucidating issues related to overfitting and local minima that occur with more conventional neural network algorithms. These characteristics are important in pattern recognition applications such as human face recognition.

Kasthuri et al. [22] proposed a powerful deep texture feature called DGOLOF for describing facial features. It adapted Name Semantic Network (NSN)-based face annotation to efficiently improve image classification. DGOLOF's deep texture feature is discriminant and invariant but less

performant in the partially occluded facial images. In [23], Zhu et al. proposed a collaborative deep framework for face photo–sketch synthesis. It combined collaborative loss with generative adversarial nets. This method offers good face recognition accuracy, but the processing time is high.

In [24], Kasthuri et al. compared and evaluated Name Semantic Network (NSN) with various annotation techniques. This work provides a good review of feature extraction methods as well as the recently deep feature extraction methods. Experimental results demonstrate that the deep feature methods achieved better recognition rates other than texture features using Yahoo images.

Table 1 summarizes the strengths and drawbacks of related works. The previous related works have many limitations due to several reasons. The first reason, such proposed works suffer from a high computation time and significant memory space to build and represent facial features. The improvement is obtained by defining a set of new directional gradient-based descriptors, feature-to-tree representation, and reduced feature vectors. The second reason, the majority of solutions suffer from high computational costs. Deep learning methods for face recognition outperform feature extraction methods when giving huge faces images for training with a high rate of recognition. A study on advanced deep face deep learning techniques [21–24] indicated that these methods can take long time to train. We focus on feature extraction approach that provides a faster-training model. This performance plays a vital role when dealing with real-time applications. To this end, we extract the amplitude and orientations to train the model that best represent texture features.

In this paper, we propose new, improved face recognition descriptors based on the gradient directions that improve the classification accuracy rate for FR and FER systems. The proposed descriptors are based on the exploitation of (1) directional codes representation of facial images, (2) magnitudes and orientation maps to establish on one-hand, the facial features defined on blocks’ image, on the other hand, of those related pixels horizontal and vertical coordinates. The proposed approach consists of a reduced feature vector based on directional gradients that describes facial expressions to make a tradeoff between computational time and classification accuracy. We validate and compare it with several existing descriptors designed for face recognition and conducted several experiments on the widely used benchmarks. In the next section, we detailed the proposed descriptors.

3 Standard descriptors for features extraction

In this section, we present the most standard descriptors for extracting the features of the image. The feature extraction is the dimensionality reduction that represents the discriminative parts of an image in a reduced feature vector. There is a wide spectrum of feature vectors that can be used for many recognition tasks as shown below.

- Local Binary Pattern (LBP) [4] is a texture descriptor that matches the gray level of a pixel with the gray levels

Table 1 Related works’ comparison

Related works	Test BD	Strengths	Limitations	Results (%)
LBP [4]	CF1–LFW	Frontal robustness	Image asymmetry	13.03–90.9
WLD [6]	FERET–LFW	Difficulties Shifted	Low recognition accuracy	68.59–23.4
LGP [7]	CK+	Monotonic robustness	High space complexity	92.4
GDP [8]	FERET	Noise robustness	High computational cost	76.10
LDN [9]	CK+–JAFFE	Compact coding	Higher total execution time	92.2–89.2
LPQ [10]	FERET–LFW	Robustness to lighting	Lots of discriminating info	99.2–75.3
MTP [11]	CFAPS	Rotation invariant	High space complexity	92.4
LGC [12]	JAFFE	Lighting robustness	Only instead central pixel	88.33
HOG [13]	KIMIA	Lighting robustness	High computational cost	34.37
SIFT–ICA [15]	CMUPIE–FRGC2	Occlusion insensitive	Base Linear Method	96.6–91.92
GW–LDA [17]	AT–FACE	Luminosity invariant	High computational complexity	88.0–94.02
GW–PCA [19]	FRGC	Expression robustness		99.74
EF–PCA [15]	ORL	Rotation invariant		76.23
PHOG–SVM [18]	LFW	Pose Invariant		88.50
CosFace [21]	LFW–YTF	Good performance		99.73–97.6
DeepFace [21]	LFW–YTF	High accuracy		97.35–91.4
DGOLOF [32]	Deep Gabor	High accuracy		97.35–91.4

of its neighbors. It assigns a binary code to describe the local texture of a region.

- Weber Local Descriptor (WLD) [6] is a well-known descriptor encode the gray-level difference between the central pixel and its neighbors within a local differential orientation.
- Local Gradient Patterns (LGP) [7] descriptor is computed based on local gradient flow from one side to another side through the center pixel in a 3×3 pixels region. The center pixel of that region is represented by two bits of binary patterns.
- Gradient Direction Pattern (GDP) [8] is a more invariant feature description to noise while using edge response value instead of the intensity of pixel.
- Local Directional Number Pattern (LDN) [9] is a micro-pattern descriptor that used the top directional numbers, which is the most positive and negative directions of those edge responses.
- Local Phase Quantization (LPQ) [10] is a texture descriptor based on the blur invariance property of short-term Fourier transform (STFT) within the neighborhood.
- Local Ternary Pattern (LTP) [11] extends LBP code by using three values of encoding to provide uniform consistency regions
- Local Gradient Code (LGC) [12] descriptor describes the distribution of the gray levels in the neighborhood of the center pixel. However, it uses horizontal, vertical, and diagonal gradients instead of only the central pixel value
- Histogram of Oriented Gradient (HOG) [13] is a successful feature descriptor based on the gradient orientation that is invariant to lighting. HOG is often used with an SVM classifier to identify the face.
- Local Directional Pattern (LDP) [25] extends LBP descriptor that uses the directional responses by using Kirsch kernels. LDP represents a robust descriptor for face recognition.
- Gradient Local Ternary Pattern (GLTP) [26] descriptor is a content-based pattern that uses gradient magnitude values instead of gray levels with a three-level encoding scheme to discriminate between smooth and high textured facial regions.

4 Proposed descriptors: HDG and HDGG

We propose HDG and HDGG, two novel descriptors for extracting facial image features. The novelty of our descriptors is to provide: (1) reduced feature dimension to train the SVM classifier to distinguish between different classes of facial expressions and (2) structured tree representing discriminant features regarding the specific region of the facial image. These two descriptors improve the performance of the classification process of the different faces and the

different facial expressions while extracting discriminate information.

4.1 HDG: feature extraction with histogram of directional gradient

HDG is a new local feature descriptor for face and facial expression recognition, very simple and efficient. It just groups eight directional gradients into a vector of eight values. Each value is an information dimension of a specific direction. HDG is used to extract the distribution of directions of oriented gradients from the whole image. Then, encode the extracted features by prominent direction indices according to the comparison. This allows distinguishing among similar structural patterns that have different gradient transitions. We also include histogram reduction algorithms to enhance the execution time. HDG as HOG eliminates the problem of changing lights caused by the environment. It consists of the following steps.

Step 1 Apply Kirsch masks on each image to obtain the improved edge response value. Each pixel is represented by the eight-edge response values m_i $i = 0, 1, \dots, 7$. Figure 1a shows an example of edge response value and Fig. 1b shows an original image and filtered images resulted from Kirsh masks.

Step 2 Divide the face image into $n \times m$ blocks.

Step 3 Compute the sum of all edge response values bit to each block X in eight directions independently (one by direction) as follows:

$$b_i = \sum_{m_i \in X} |m_i| \quad i = 0, 1, \dots, 7 \quad (1)$$

where i represents a direction, X is a block of the image, and m_i is the response value of pixel for direction i .

Step 4 Represent each block by a histogram « B » of eight values. Each value is a cumulative sum of information in each direction.

$$B = \{b_i\} \quad i = 0, 1, \dots, 7 \quad (2)$$

Step 5 Concatenate all histograms to form a feature vector of size $n \times m \times 8$. This vector will use it as a face descriptor.

Step 6 Represent the facial image and HDG results after applying Kirsh by a structured tree [26]. The histogram of the facial image as a global feature vector is represented as a root of the tree. Then, the histogram of eight-values of

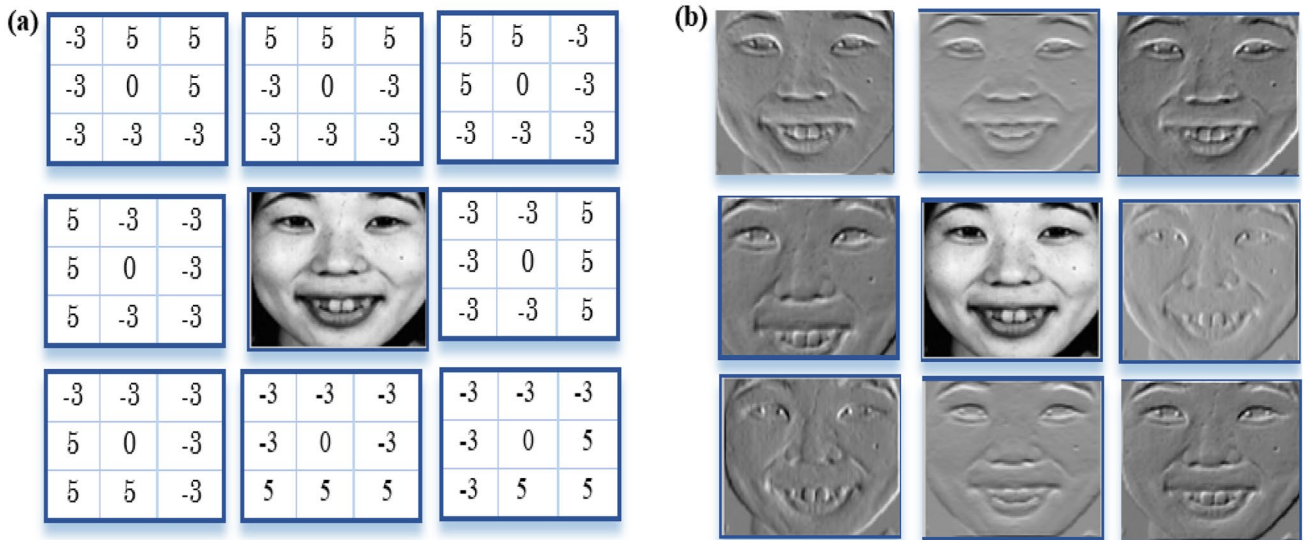


Fig. 1 a Example of edge response values. b Filtered images resulted from 8-Kirsh masks

each block of the facial image presents a specific vertex in the child’s root. Finally, the eight response values of each block are represented as eight leaves. Figure 2 presents an illustration of a global feature, $n \times m$ blocks’ features, and eight response values. The structured tree promotes the efficiency and the effectiveness of the classification process of the facial images.

4.2 HDGG: feature extraction with histogram of directional gradient generalized

HDGG is a new efficient and improved features extraction approach extended HOG that was proposed by Dala et al. [13]. HDGG extends the HDG feature local descriptor to magnitudes and orientation maps that requiring proper descriptive vector. HDGG encodes the directional information of the face’s texture in a reduced way for producing a more discriminative code than existing methods. It

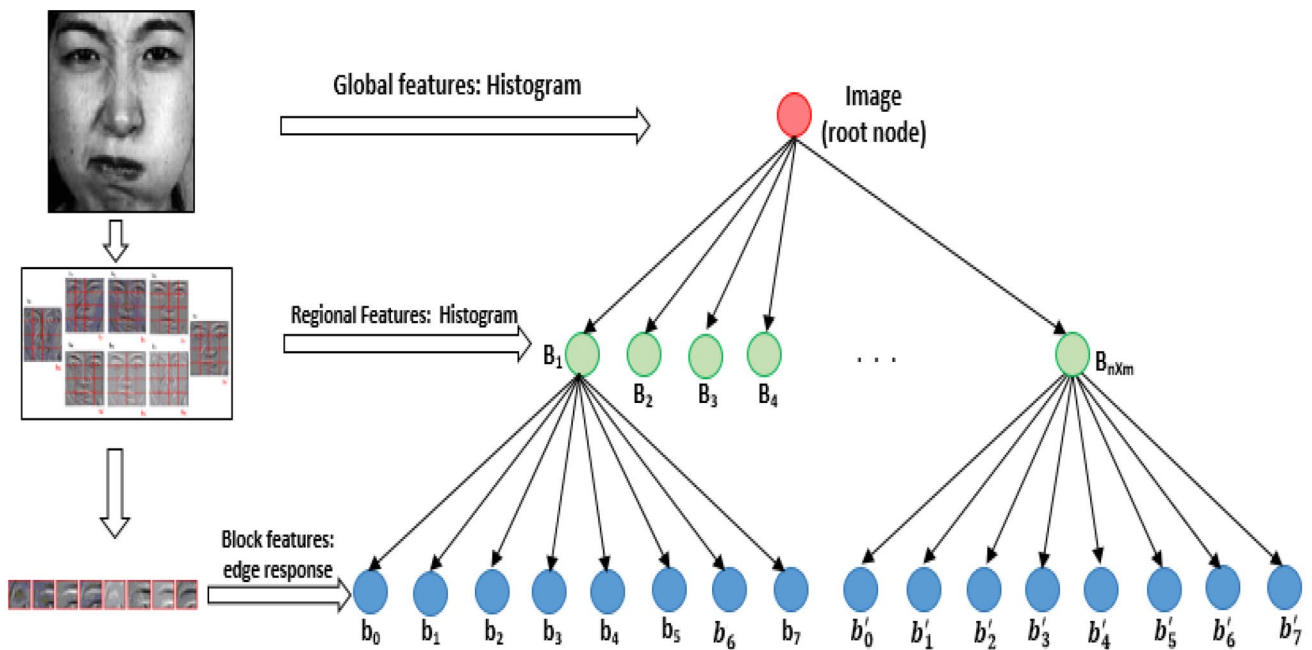


Fig. 2 Building Tree2vector after applying HDG operator

determines a new attribute that can be specified to describe long-distance relationships.

HDGG enhances the classification accuracy and execution time of FR and FER systems. They provide faster, potentially more stable computation and express more clearly object boundaries in a long-distance relationship. HDGG consists of summing all gradient values of image pixels referring to 8-pixels using Kirsch filter, which will be mapped on magnitudes and orientation maps. HDGG consists of the following steps.

Step 1 Apply Kirsch filter on each block as HDG, and then compute for each pixel, 8-gradient feature. HDGG considers these eight values as eight oriented vectors (see Fig. 3).

Step 2 Compute the sum of gradient vectors on a new pixel's vector, as shown in Fig. 3, by using the following equation:

$$x = \sum_{i=0}^7 (m_i \cos(i * \pi/4)) \quad (3)$$

$$y = \sum_{i=0}^7 (m_i \sin(i * \pi/4)) \quad (4)$$

Step 3 Perform the magnitude and orientation maps values on the horizontal and vertical coordinates of each pixel's vector according to G and Θ values:

$$G = \sqrt{x^2 + y^2} \quad (5)$$

$$\Theta = \tan^{-1}(y/x) \quad (6)$$

Step 4 Decompose the whole image into $n \times m$ non-overlapping blocks, then quantizes orientation values of each block in the histogram with 9-orientation bins, where the magnitude values are the votes.

Step 5 Normalize all of the histogram blocks to obtain the feature vector. For each face image in the training set, we have calculated and stored the associated feature vector.

To extract facial features vector from a facial image, we divide it into $n \times m$ blocks. We use 8-equally spaced intervals in the interval $[0, \pi]$. For each block, a local histogram is automatically generated and normalized. These normalized histograms are concatenated to form the image's global histogram, which may be offering a comparison between facial recognition methods. Figure 4 shows an example of HDGG magnitude and HDGG orientation results.

5 Proposed framework for face and facial expressions recognition

The proposed recognition framework (see Fig. 5) aims to find perfect facial features of the face image and provides directional gradients face recognition technique for identifying faces in images that ensures high accuracy and good effectiveness. We used two well-known datasets of JAFFE [26] and YALE [27] to do the facial features learning in the new, improved HDG and HDGG descriptors, provide suitable pre-trained model features for directional gradient training image and then use SVM to classify input tested image. We improved the recognition rates by directional oriented gradients and combined it with an SVM classifier to process the model's output. The implementation of the proposed system involves two-phases that collaborated to accomplish the system goal. The phases of the proposed system are the face training phase and the face classification phase.

The system architecture of the proposed model with the two phases is shown in Fig. 5 and described as follows:

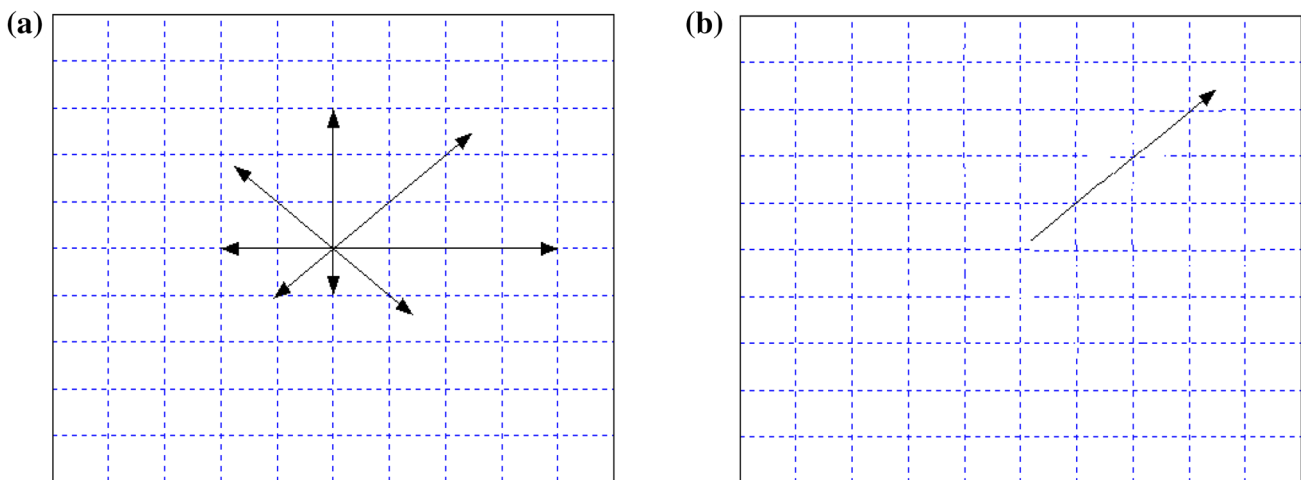


Fig. 3 a Pixel represented by eight vectors. b Pixel represented by a single vector

Fig. 4 HDGG magnitude and orientation results

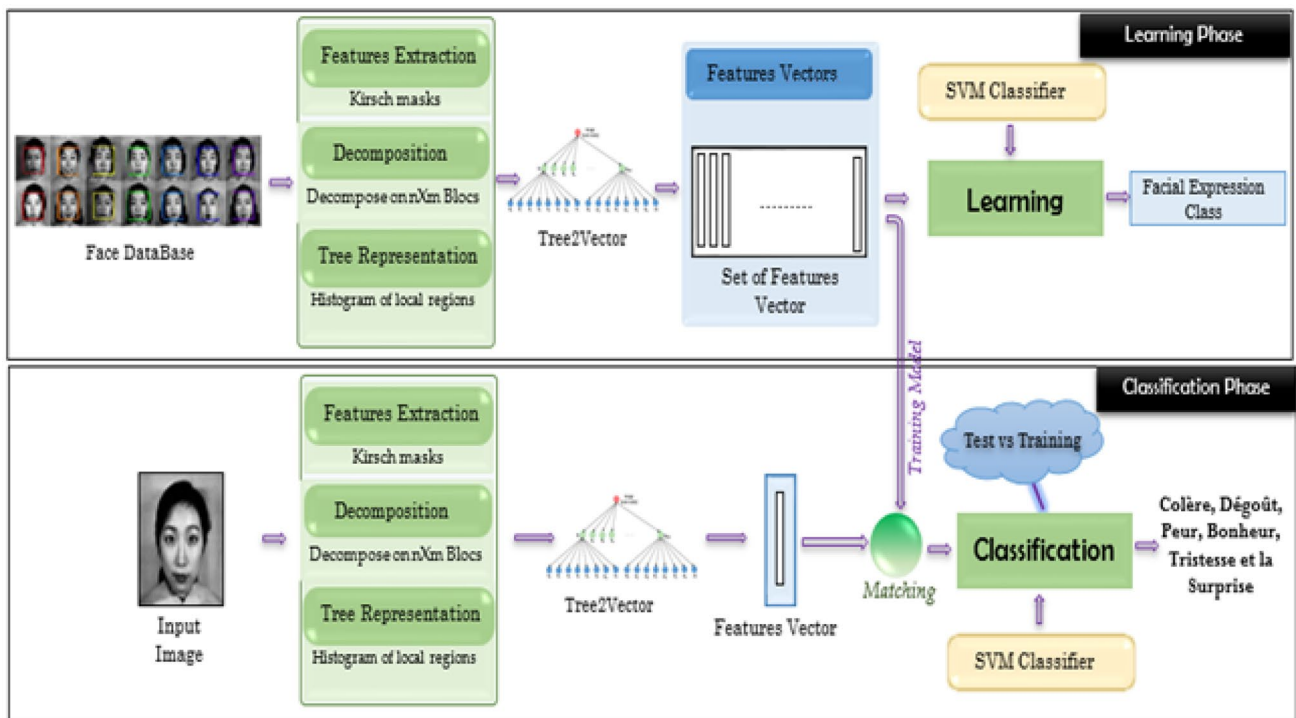
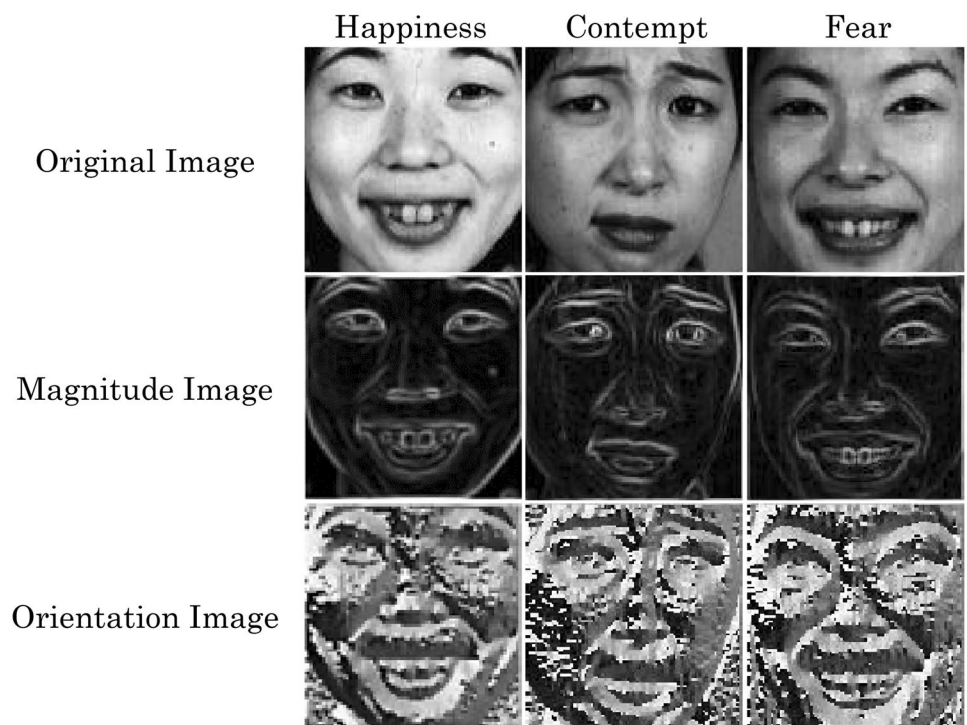


Fig. 5 Schematic representation of proposed framework

6 Proposed framework for face and facial expressions recognition

The proposed recognition framework (see Fig. 5) aims to find perfect facial features of the face image and provides directional gradients face recognition technique for identifying faces in images that ensures high accuracy and good effectiveness. We used two well-known datasets of JAFFE [26] and YALE [27] to do the facial features learning in the new, improved HDG and HDGG descriptors, provide suitable pre-trained model features for directional gradient training image and then use SVM to classify input tested image. We improved the recognition rates by directional oriented gradients and combined it with an SVM classifier to process the model's output. The implementation of the proposed system involves two-phases that collaborated to accomplish the system goal. The phases of the proposed system are the face training phase and the face classification phase.

6.1 Training process

The purpose of training is to extract the facial features of human faces from images to solve the learning problem that there is only a reduced size of labeled feature vectors. Then, the trained model can be applied directly to train the target data, so that it can be used for a large number of data training and reduce training time. Two new descriptors have been used to perform the training phase. One is HDG (Histogram of Directional Gradient) and another is HDGG (Histogram of Directional Gradient Generalized). Both are directional gradients. These are used to extract facial features, normalize, and map them as a reduced feature vector. We expanded 213 peak facial expressions in the JAFFE and 165 faces images in the YALE databases. The extracted features from these images are normalized and transferred to the SVM classifier for the training process.

The training process (see Fig. 6) consists of dividing the image into $n \times m$ blocks after applying HDG and HDGG for each of those blocks. First, we perform a processing step known as feature extraction to store the discriminate information about each face in a reduced vector. Next, we have a histogram among 9-values (i.e., 8-values for HDG) of the gradient directions and their magnitude inside each block. Finally, all 9-vectors are normalized and concatenated into a final feature vector, which is stored with a face image in a database. Then, at the same time, we use SVM with a multi-class linear kernel to fit a model of the facial appearance in the database. So that we can discriminate between different people of a database. The output of this phase is an SVM trainer, a model that will be used to recognize input images.

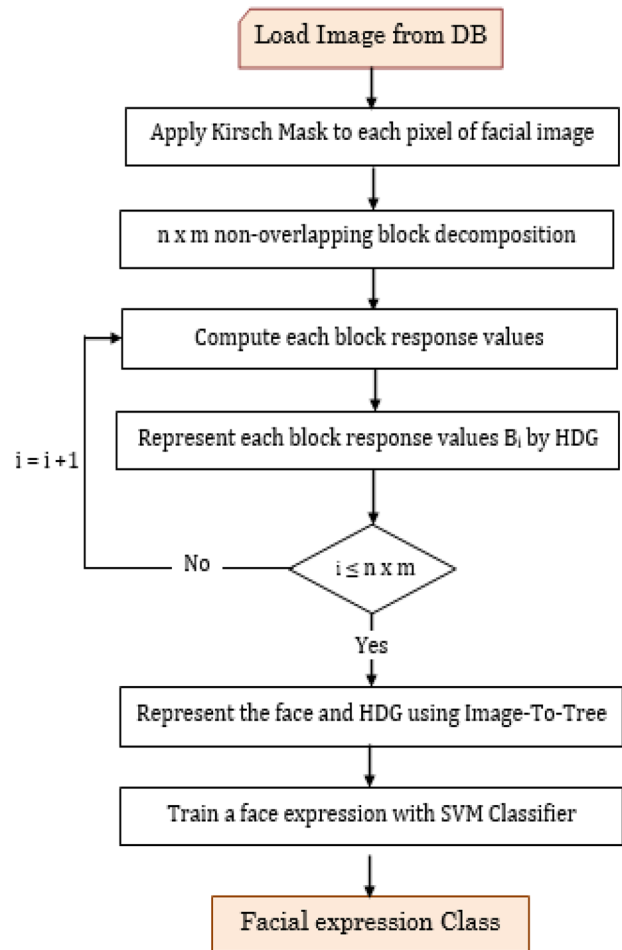


Fig. 6 Functional model of training phase

6.2 Classification process

The classification process recognizing the face and classifying the facial expressions of the input image, by computing the HDG (or HDGG) vector, applying the SVM classifier [1] to find the matched class as illustrated in Algorithm 1. The first step consists to compute the response value of each pixel of the input image using HDG (i.e., HDGG). Then, we divide the input image is into $n \times m$ blocks. A histogram is constructed for each block of the input image. Next, we concatenate the histograms of each block to get the feature vector for the input image. We normalize and reduce the facial vector to keep those related to the training phase.

The classification process consists of transforming the features into a structured tree. The structured tree consists of three layers: a global feature layer, a regional feature layer, and a response values layer for all blocks. The classification process creates a new root branch for the global feature histogram of the input image. Next, it assigns regional features to its parent global feature histogram. Then, it aggregates the

response values of each block, which are created in the tree’s leaves and assigned to its regional feature. This process is repeated until all the blocks have been treated. Finally, it compares the previously structure tree and those of the train

faces, to determine the distance between them, which will be thereafter scored and returned within a label or an indicator to signify which person from the database.

Algorithm 1 Facial expression classification using HDG-SVM	
Input :	
<i>Query Image</i>	<i>% query image to be classified</i>
<i>Training Set</i>	<i>% Set of training images</i>
<i>Test Set</i>	<i>% Set of tested images</i>
Output :	
<i>Person Label, matched Class</i>	<i>% founded expression class</i>
Begin	
1: <i>person</i> ← 1;	
2: <i>query Image</i> ← Read (test(<i>person</i>), 1);	
3: <i>query Features</i> ← extract HDG Features (<i>query Image</i>);	<i>% case of HDG descriptor</i>
4: <i>person Label</i> ← predict (SVM, <i>query Features</i>);	
5: <i>index</i> ← find (<i>person Label</i>);	
6: <i>matched Class</i> ← image (<i>index</i>);	
End	



Fig. 7 Sample Images from the JAFFE database and their facial expressions

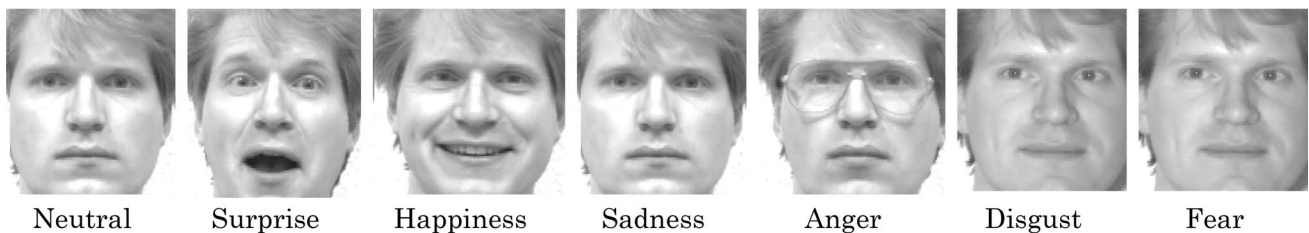


Fig. 8 Sample Images from the YALE database and their facial expressions

7 Experimental results and analysis

In this section, we present the experiments on two well-known datasets. The main purpose of these experiments is to evaluate the performance of the proposed recognition system based on the extended HOG descriptors. We will validate the effectiveness and efficiency of our recognition system and compare it with several descriptors in terms of several metrics such as accuracy, precision, recall, F1-score, execution time.

7.1 Datasets

To build and evaluate our proposed recognition system, two well-known benchmarks of face images have been used. One is JAFEE (Japanese Female Facial Expression) [28] and another is the YALE face database [29]. The JAFEE database contains 213 peak facial expressions from ten subjects, seven emotion categories are considered. They are happiness, sadness, surprise, anger, disgust, and natural. The gray level images are of size 256×256 . We used the `fdlibmex` library, free code available for MATLAB for face detection. We normalized all the evaluated face images before the experimentation in size of 128×128 pixels. Figure 7 presents samples of the JAFEE dataset. The YALE Face dataset [28] contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, 1 per different facial expression or configuration: center light, no glasses, happy, left light, and normal, right light, sad, sleepy, surprised, and wink. All images are of size 64×64 divided into blocks with 8×8 . In our work, we have only considered the frontal images. All images are 128×128 size and divided into 8×8 equally blocks. Figure 8 presents samples of the YALE dataset.

7.2 Experimental setup

All experiments are carried out with an Intel i5, 3 GHz processor, 12 GB of RAM Intel Core i3-5005U and 4 GB of RAM Windows 10 64 bits, C++ Builder and Matlab. We used Library for Multiclass Support Vector Machine (LIB-SVM) [30] which is used for extracting various features. These features are trained and classified through pairwise approach (one vs one) applied with Linear Kernel function. The LIB-SVM library is a modern toolkit that contains several machine-learning algorithms that help in writing sophisticated C++ based face recognition applications.

7.3 Evaluation metrics

We evaluated and compared the performance of the proposed HDG and HDGG descriptors with some standards descriptors in terms of effectiveness and efficiency.

7.3.1 Effectiveness metrics

We evaluated the relevance of the proposed descriptors throughout the following metrics.

- **Accuracy (A)** is the measure of how a classifier can predict the correct predictions. It is also the ratio of the correct predictions to the total predictions

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

where True Positives (TP) is the number of relevant faces, True Negatives (TN) is the number of non-relevant faces, False Positives (FP) is the number of relevant faces that are not classified by the given approach, False Negatives (FN) is the number of non-relevant faces that are not classified by the given approach.

- **Precision (A)** is the measure to identify the number of relevant faces among the classified ones. It is also the ratio of the correct predictions to the total predictions

$$P = \frac{TP}{TP + FP} \quad (8)$$

- **Recall (R)** calculates the correctly classified faces images over all the faces images in the dataset

$$R = \frac{TP}{TP + FN} \quad (9)$$

- **F1-score** evaluates a weighted average of P and R . It is an important factor based on weighted recall. The F-score is computed as follows:

$$F_score = 2 \times \frac{P \times R}{P + R} \quad (10)$$

- **Peak signal to noise ratio (PSNR)** [31, 32] is a logarithmic function of Mean Square Error (MSE) interpreted as a corrected version of the Signal-to-Noise Ratio. A high PSNR means that the two images are identical. MSE and PSNR can be calculated using Eqs. (13) and (14), respectively

$$MSE(x, y) = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (x_{ij} - y_{ij})^2 \quad (11)$$

Table 2 Accuracy of HDG and HDGG using different block sizes on the JAFFE database

Block dimension	Accuracy HDG (%)	Accuracy HDGG (%)
1 × 1	38.57	39.52
2 × 2	53.33	66.67
4 × 4	73.80	82.38
8 × 8	90.00	91.43
16 × 16	89.04	88.57

Bold indicates the highest results or scores

$$PSNR(X, Y) = 10 \log_{10} \frac{255^2}{MSE} \text{dB} \tag{12}$$

where x_{ij}, y_{ij} are the pixel (i, j) in the original image X and the distorted image Y , respectively. $M \times N$ is the size of the image.

- **Average difference (AD)** [33] expresses the similarity between the original image and the distorted one. Good classification means having the lowest difference value. Usually, it can be considered good if AD close to or equal to 0. This metric is computed as follows:

$$AD(x, y) = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M |x_{ij} - y_{ij}| \tag{13}$$

- **Maximum difference (MD)** [33] provides the maximum difference between the original image and the distorted one. It used to measure the maximum accuracy of the proposed descriptors. MD is defined as follows:

$$MD(x, y) = \max |x_{ij} - y_{ij}| \tag{14}$$

- **Structural content (SC)** [33] is expressed as follows:

$$SC(x, y) = \frac{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^2}{\sum_{i=1}^N \sum_{j=1}^M y_{ij}^2} \tag{15}$$

- **Normalized absolute error (NAE)** [33] is used to evaluate the linear relation between the host image and the distorted image. It could be used also to evaluate the closeness of prediction to outcomes. A lower NAE value shows that image is of good classification. It is defined as follows:

$$NAE(x, y) = \frac{\sum_{i=1}^N \sum_{j=1}^M |x_{ij} - y_{ij}|}{\sum_{i=1}^N \sum_{j=1}^M |x_{ij}|} \tag{16}$$

Table 3 Recognition accuracy of different methods on (a) JAFFE database and (a) YALE database

(a)		(b)	
Method	JAFFE	Method	YALEE
LTP	40.95	LTP	48.48
GLTP	38.09	GLTP	36.36
LBP	42.00	LBP	87.88
LDP	52.38	LDP	56.36
LDN	68.57	LDN	79.39
LGC	78.57	LGC	70.00
LPQ	69.52	LPQ	73.93
WLD	81.90	WLD	92.12
GDP	87.14	GDP	48.48
LGP	83.80	LGP	36.36
HOG	88.57	HOG	87.88
HDG	90.00	HDG	56.36
HDGG	91.43	HDGG	79.39

Bold indicates the highest results or scores

7.3.2 Efficiency metrics

We evaluated the efficiency of the proposed descriptors throughout the execution time (T). It is the time needed to accomplish the classification process of a specific face of faces dataset.

7.4 Results and analysis

7.4.1 Efficiency evaluation

To evaluate the impact of block-size on the accuracy, HDG and HDGG operators are applied on faces images at different sizes: $1 \times 1; 2 \times 2; 4 \times 4; 8 \times 8$, and 16×16 in Table 2. The results show that 8×8 block combination is the best for the 128×128 original image size for both HDG and HDGG descriptors. Moreover, our proposed descriptors also ignore the object boundaries with small block dimensions.

To evaluate the accuracy of the proposed HDG and HDGG operators in classifying faces, their results are compared with ten standard descriptors namely HOG [13], LBP [4], LDN [9], LPQ [10], WLD [6], LGP [7], GDP [8], LGC [12], LTP [11], and GLTP [26] in Table 3a using JAFFE dataset. The results show that our proposed descriptors yield much better accuracy as compared to other methods.

Table 3b presents the accuracy rates of different operators on the YALE database. Based on the mentioned results in Table 3b, we can note that the proposed HDG and HDGG successfully recognize the facial expression features from the extracted ones with 92.12, due to the effective block segmentation and detection method. These results prove the

Table 4 Confusion matrix of -class facial expressions recognition using SVM with HDG in the JAFFE database

(%)	AN (%)	CO (%)	DI (%)	FE (%)	HA (%)	SA (%)	SU (%)
AN	100	0.00	0.00	0.00	0.00	0.00	0.00
CO	0.00	89.65	0.00	3.22	0.00	6.45	0.00
DI	0.00	0.00	84.38	0.00	3.33	9.68	3.70
FE	0.00	0.00%	0.00	80.65	0.00	16.12	3.70
HA	0.00	0.00	3.13	6.45	90.00	0.00	0.00
SA	0.00	0.00	6.25	3.20	0.00	87.09	0.00
SU	0.00	0.00	0.00	0.00	0.00	0.00	100

Bold indicates the highest results or scores

Table 5 Confusion matrix of -class facial expressions recognition using SVM with HDG in the YALE database

(%)	AN (%)	CO (%)	DI (%)	FE (%)	HA (%)	SA (%)	SU (%)
AN	96.66	3.45	0.00	0.00	0.00	0.00	0.00
CO	0.00	93.10	3.25	0.00	0.00	3.23	0.00
DI	0.00	3.45	81.25	0.00	3.33	9.68	3.70
FE	0.00	0.00	0.00	87.09	0.00	9.68	3.70
HA	0.00	0.00	6.25	3.22	86.67	3.23	0.00
SA	0.00	0.00	0.00	3.22	0.00	96.77	0.00
SU	0.00	0.00	0.00	0.00	0.00	0.00	100

Bold indicates the highest results or scores

Table 6 Comparison of precision, recall, and F1-score on the JAFFE database for several methods

Method	JAFFE database			YALE database		
	P (%)	R (%)	F1-score (%)	P (%)	R (%)	F1-score (%)
HDG	90.83	90.2532	90.5427	90.83	90.25	90.54
HDGG	92.03	91.65	91.84	92.03	91.65	91.84
HOG	88.93	88.80	88.86	88.93	88.80	88.86
LGP	86.49	84.08	85.26	86.49	84.08	85.26
GDP	86.49	84.08	85.26	86.49	84.08	85.26
WLD	84.29	82.26	83.26	84.29	82.26	83.26
LPQ	81.34	70.15	75.33	81.34	70.15	75.33
LGC	80.71	68.93	74.35	80.71	68.93	74.35
LDN	79.95	42.00	55.07	79.95	42.00	55.07
LDP	72.18	52.86	61.03	72.18	52.86	61.03
LBP	76.15	60.65	67.52	76.15	60.65	67.52
LTP	41.45	62.59	49.87	41.45	62.59	49.87
LGTP	80.11	39.47	52.88	80.11	39.47	52.89

Bold indicates the highest results or scores

efficiency of the proposed face descriptors in achieving a higher accuracy rate compared to other methods on YALE.

Next, the proposed descriptors are evaluated through the confusion matrix (Tables 4 and 5). The results show that the proposed operators' HDG or HDGG yields the best accuracy on all images involved in the experiments. These results validate and confirm the effectiveness of HDG and HDGG descriptors to identify facial expressions on the JAFFE dataset.

Moreover, our model also dominates over those descriptors in terms of other evaluation metrics namely precision, recall, and F1-score. As shown in Table 6, we can notice

the highest precision, recall, and F1-score for the proposed descriptors HDG and HDGG over other traditional descriptors on the JAFFEE database. The improvement in such performances of the HDG and HDGG classifiers with the inclusion of magnitude and orientations map can be observed with 91.65% precision, 92.03% recall, and 91.84% F-score for HDGG, 90.83% precision, 92.03% recall 90.25% and 91.84%, respectively for HDG. These results confirm that the proposed SVM classifier with HDG effectively helps to recognize the face successfully.

In addition, the results of the experiments on the YALE database also show the highest performance for HDG and

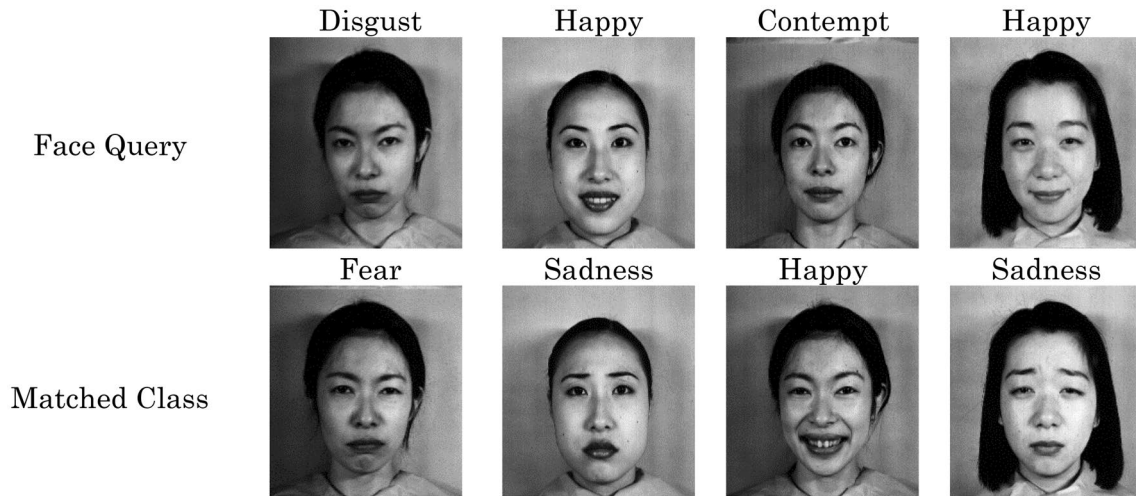


Fig. 9 False classified facial expressions using HDG code in the JAFEE database

Fig. 10 False recognized faces using HDG code in the YALE database



Table 7 Comparison of the similarity measures on the JAFEE database for several methods

%	MSE	PSNR	MNCC	AD	SC	MD	NAE
HDG	3.0612	43.2719	0.8976	0.7347	1.2212	6	0.2000
HDGG	2.1633	44.7797	0.9121	0.9796	1.1890	6	0.1714
HOG	3.4286	42.7797	0.8835	0.9796	1.2602	6	0.2286
LGP	6.3673	40.0912	0.8355	1.3878	1.3881	9	0.3238
GDP	3.9184	42.1998	0.8691	1.5510	1.3012	8	0.2571
WLD	9.3061	38.4431	0.8154	1.5510	1.4225	15	0.3619
LPQ	29.9184	33.3714	0.6890	2.6122	1.6388	23	0.6095
LGC	44.5306	31.6442	0.5937	3.4286	1.8764	24	0.8000
LDN	26.4490	33.9067	0.6822	2.6939	1.7554	18	0.6286
LDP	60.6531	30.3023	0.5192	4.0816	1.9652	28	0.9524
LBP	111.7959	27.6465	0.3988	5.1020	1.5038	32	1.1905
LTP	108.2857	27.7851	0.4048	5.0612	1.5390	32	1.1810
LGTP	121.0612	27.3008	0.3765	5.3061	1.4446	32	1.2381

Bold indicates the highest results or scores

Table 8 Comparison of the similarity measures on the YALE database for several methods

%	MSE	PSNR	MNCC	AD	SC	MD	NAE
HDG	0.0800	59.0999	0.9576	0.0622	1.0810	2	0.0848
HDGG	0.1244	57.1810	0.9273	0.1067	1.1495	2	0.1455
HOG	0.1511	56.3378	0.9333	0.1067	1.1294	3	0.1333
LGP	0.3822	52.3076	0.9091	0.1333	1.1553	6	0.1818
GDP	0.2844	53.5908	0.9091	0.1333	1.1717	4	0.1818
WLD	0.2044	55.0251	0.9212	0.1156	1.1524	4	0.1576
LPQ	1.5289	46.2870	0.7394	0.3822	1.4963	6	0.5212
LGC	2.1422	44.8222	0.6970	0.4444	1.5163	7	0.6061
LDN	1.0756	47.8145	0.7939	0.3022	1.3866	5	0.4121
LDP	3.3422	42.8905	0.5636	0.6400	1.8464	9	0.8727
LBP	5.6622	40.6009	0.4788	0.7644	1.5163	11	1.0424
LTP	5.1111	41.0457	0.4848	0.7556	1.6575	10	1.0303
LGTP	6.7022	39.8686	0.3636	0.9333	1.7917	11	1.2727

Bold indicates the highest results or scores

Table 9 The feature vector sizes and length using different methods

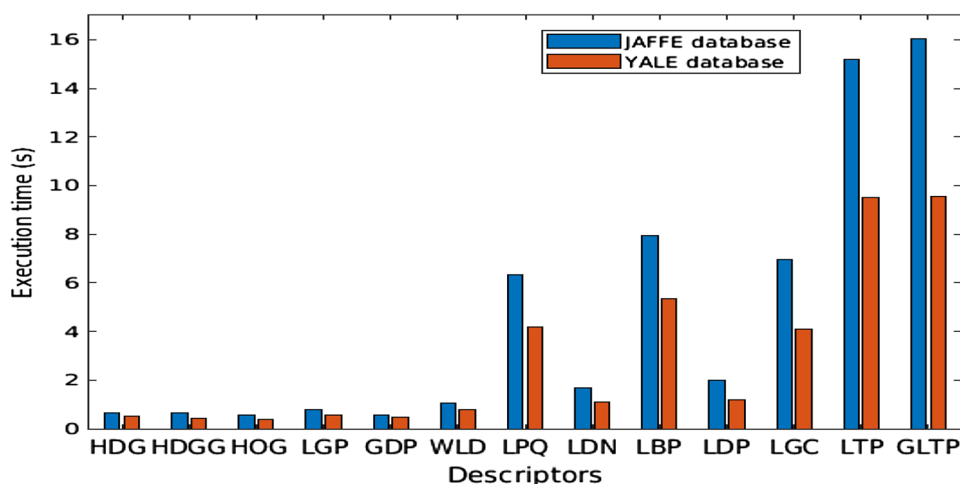
Method	Vector size	Features vector length
LTP	512	$8 \times 8 \times 512 = 32,768$
GLTP	512	$8 \times 8 \times 512 = 32,768$
LBP	256	$8 \times 8 \times 256 = 16,384$
LDP	56	$8 \times 8 \times 56 = 3584$
LDN	56	$8 \times 8 \times 56 = 3584$
LGC	256	$8 \times 8 \times 256 = 16,384$
LPQ	256	$8 \times 8 \times 256 = 16,384$
WLD	32	$8 \times 8 \times 32 = 2048$
GDP	8	$8 \times 8 \times 8 = 512$
LGP	7	$8 \times 8 \times 8 = 448$
HOG	9	$8 \times 8 \times 9 = 576$
HDG	8	$8 \times 8 \times 8 = 512$
HDGG	9	$8 \times 8 \times 9 = 576$

Bold indicates the highest results or scores

HDGG in terms of precision, recall and F1-score. HDG achieves 90.83% precision and 90.54% F1-score when directional gradients are included. Similarly, HDGG achieves 92.03% recall and 91.84% F1-score. However, HDG descriptor shows some failure cases in the JAFFE dataset as illustrated in Fig. 9 due to smoother faces. Similarly, HDG descriptor also shows some failure cases in YALE dataset as illustrated in Fig. 10. This requires the need to match stronger features in the face.

Experimental results of the similarity metrics on the database JAFFE with standard for several methods are presented in Table 7. The obtained results demonstrate high values when applying HDG and HDGG in terms of MSE, PSNR, NCC, AD, SC, MD, and NAE. Besides, we noted that the consideration of directional gradient in facial recognition system is the most influence on similarity measures. It is very interesting that SVM classifier with HDG is used to determine the identities of both faces and facial expressions,

Fig. 11 Execution time comparison using different descriptors on the JAFFE and YALE database



and the corresponding images are matched to achieve a high degree of classification against invariant illumination conditions. Results given in Table 7 indicate that as compared with other twelve descriptors in terms of similarity measures on the JAFFE database, HDGG has a lower MSE value while PSNR value is higher.

The results of the effect of including the directional gradient, magnitude and orientations map on HDG descriptor built on the database YALE are recorded in Table 8. It can be observed that when a new directional gradient using edges response values are considered, HDG descriptor has the highest rate of 59.09 in PSNR dB and the lowest rate of 0.08% in MSE of over the same model when implemented other eleven descriptors. In the case of the HDG descriptor implemented with a new magnitude and orientations map, HDGG model under the same case has an enhancement of 57.18 dB in PSNR and 0.124 in MSE.

7.4.2 Effectiveness evaluation

Now, to evaluate the proposed HDG-SVM and HDGG-SVM models, the impact of vector size is evaluated which shows the changes in learning time. Table 9 represents the feature vector size and the feature dimension for 64 blocks for the proposed HDG operator compared to other existing descriptors. In cases of LBP or LGC descriptor, the feature vector size is $8 \times 8 \times 256 = 16,384$ for LBP, LGC and LPQ, $8 \times 8 \times 56 = 3584$ for LDP, $8 \times 8 \times 8 = 512$ for HDG and $8 \times 8 \times 9 = 432$ for HDGG. The execution time depends on the vector size, the fast time its lower vector size.

Based on the above descriptors, the extracted faces features are realized into the SVM method, which successfully classifies the faces features with minimum time. The comparison of the execution time in seconds (see Fig. 11) demonstrates that HDG and HDGG effectively classify the extracted facial features with minimum time when compared to other traditional descriptors. We observe that the execution time of HDG and HDGG is 0.526 s and 0.4067 ms, respectively. HDG and HDGG have a less vector size compared to other descriptors, which may decrease the classification-processing time and the learning time. We evaluate also the execution time of our approach using the YALE dataset. Figure 11 presents the results of variation of the execution time according to applied feature extraction descriptor. In this figure, we can see that the response time significantly decreases according to varied tested facial images. As the results of execution time in cases of YALE face images, Fig. 11 ensures that the proposed features extraction operators' model provides fast execution in the case of YALE face images. While comparing two face recognition operators HDG and HDGG will serve the fastest execution time when compared to the descriptors LBP and HOG.

7.5 Lessons learned and discussion

The experiment of the SVM classifier combined with HDG (or HDGG) using two benchmarks JAFFE and YALE attain promising results compared to other existing methods in terms of accuracy precision, F1-score and similarity measures, and execution time. As a result, the proposed descriptors are much better than existing methods and demonstrate the effectiveness to recognize faces and facial expressions in two-well known datasets YALE and JAFEE, and provide the user with correct prediction in fast execution time. HDG and HDGG are more efficient and increases the precision ratio, the recall ratio, and the considerable execution time, but also make it practicable in spite of using a very large-size faces images database. However, the execution time must be improved in future works with large-size datasets using advanced Face Deep Learning [34]. Additionally, the proposed descriptors are simple and effective. These descriptors are generic and universal, and can applied in other research fields including object detection, object tracking, object recognition regardless of the nature of the object.

We notice that the space complexity of the proposed descriptors is $O(n)$, this one is $n \times m \times 8$ for the HDG descriptor and $n \times m \times 9$ for the HDGG descriptor, where n, m are the height and the width of the image, respectively, 8 and 9 are the vector lengths for HDG and HDGG, respectively, although this is considered better than some other related methods such as LBP and HOG.

8 Conclusion

In this paper, we have proposed novel feature extraction descriptors HDG and HDGG based on the gradient directions. These descriptors are based on discriminable edge response value features to the facial images and facial expressions. The proposed descriptors are concretely validated and tested using two well-known benchmarks. Experimental results have shown that the proposed approach provides better efficiency while ensuring fast execution time. The feature vector size not exceeding 512, the recognition rate reaching 92.12%, and the error rate ranged from 0.08 to 0.1. A comparison with the sequential approach shows that the proposed SVM classifier with HDGG is more efficient and enhances the similarity measures. However, the execution time must be improved in future works with large-size dataset using advanced face Deep Learning. Future work can focus on the validation of the proposed descriptors by implementing a global framework using real-time multimedia applications.

References

- Cruz AC, Bhanu B, Thakoor NS (2014) Vision and attention theory-based sampling for continuous facial emotion recognition. *IEEE Trans Affect Comput* 5(4):418–431
- Zhang X, Guan Y, Wang S, Liang J, Quan T (2006) Face recognition in color images using principal component analysis and fuzzy support vector machines. In: *First IEEE international symposium on systems and control in aerospace and astronautics*, pp 4
- Valstar MF, Pantic M (2011) Fully automatic recognition of the temporal phases of facial actions. *IEEE Trans Syst Man Cybern Part B (Cybern)* 42(1):28–43
- Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
- Zhang X, Shan S, Gao W, Chen X, Zhang H (2005) Local Gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. *Computer Vision. ICCV 2005*. In: *Tenth IEEE international conference on, IEEE, 2005*, pp 786–791
- Chen J, Shan S, He C, Zhao G, Pietikainen M, Chen X, Gao W (2009) WLD: a robust local image descriptor. *IEEE Trans Pattern Anal Mach Intell* 32(9):1705–1720
- Islam MS (2014) Local gradient pattern—a novel feature representation for facial expression recognition. *J AI Data Mining* 2:33–38
- Islam MS (2013) Gender classification using gradient direction pattern. *Sci Int Lahore (Lahore)* 25(4):797–799. <https://doi.org/10.1109/ICPR.2010.373>
- Rivera AR, Castillo JR, Chae OO (2012) Local directional number pattern for face analysis: face and expression recognition. *IEEE Trans Image Process* 22(5):1740–1752
- Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. In: *International conference on image and signal processing*, Springer, pp 236–243
- Bashar F, Khan A, Ahmed F, Kabir MH (2014) Robust facial expression recognition based on median ternary pattern (MTP). In: *International conference on electrical information and communication technology (EICT)*. IEEE, pp 1–5
- Tong Y, Chen R, Cheng Y (2014) Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle. *Optik* 125(16):4186–4189
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *2005 IEEE Comput Soc Conf Comput Vision Pattern Recognit (CVPR'05)* 1:886–893
- Farid A, Adel A (2020) Improved face and facial expression recognition based on a novel local gradient neighborhood. *J Digital Inf Manag* 18:33–34
- Siddiqi MH, Lee S, Lee YK, Khan AM, Truc PTH (2013) Hierarchical recognition scheme for human facial expression recognition systems. *Sensors* 13(12):16682–16713
- Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Philos Trans R Soc A Math Phys Eng Sci* 374(2065):20150202
- Fathima AA, Ajitha S, Vaidehi V, Hemalatha M, Karthigaiveni R, Kumar R (2015) Hybrid approach for face recognition combining Gabor Wavelet and Linear Discriminant Analysis. In: *2015 IEEE international conference on computer graphics, vision and information security (CGVIS)*, pp 220–225
- Huang HM, Liu HS, Liu GP (2012) Face recognition using pyramid histogram of oriented gradients and SVM. *Int J Adv Inf Sci Serv Sci* 4(18):1–8. <https://doi.org/10.4156/AISS.VOL4.ISSUE18.1>
- Cho H, Roberts R, Jung B, Choi O, Moon S (2014) An efficient hybrid face recognition algorithm using PCA and Gabor wavelets. *Int J Adv Rob Syst* 11(4):59
- Kar A, Bhattacharjee D, Nasipuri M, Basu DK, Kundu M (2013) High-performance human face recognition using Gabor based pseudo hidden Markov model. *Int J Appl Evolut Comput (IJAE)* 4(1):81–102
- Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Liu W (2018) Cosface: large margin cosine loss for deep face recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 5265–5274
- Kasthuri A, Suruliandi A, Raja SP (2019) Gabor-oriented local order feature-based deep learning for face annotation. *Int J Wavelets Multiresolution Inf Process* 17(05):1950032
- Zhu M, Li J, Wang N, Gao X (2019) A deep collaborative framework for face photo-sketch synthesis. *IEEE Trans Neural Netw Learn Syst* 30(10):3096–3108
- Anburajan K, Andavar S, Elango P (2020) An empirical evaluation of name semantic network for face annotation. *Recent Adv Comput Sci Commun (Formerly Recent Patents Comput Sci)* 13(4):557–571
- Jabid T, Kabir MH, Chae O (2010) Local directional pattern (LDP) for face recognition. In: *2010 digest of technical papers international conference on consumer electronics (ICCE)*. IEEE, pp 329–330
- Ahmed F, Hossain E (2013) Automated facial expression recognition using gradient-based ternary texture patterns. *Chin J Eng* 2013:831747. <https://doi.org/10.1155/2013/831747>
- Zhang H, Wang S, Xu X, Chow TW, Wu QJ (2018) Tree2Vector: learning a vectorial representation for tree-structured data. *IEEE Trans Neural Netw Learn Syst* 99:1–15
- The Jaffe Face Database (2020). <http://www.kasrl.org/jaffe.html>. Accessed 4 Jan 2020
- The Yale Face Database (2020). <http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html>. Accessed 4 Jan 2020
- LIB-SVM-A Library for Support Vector Machines (2019). <https://www.csie.ntu.edu.tw/~cjlin/libsvm>. Accessed 6 July 2019
- Huynh-Thu Q, Ghanbari M (2008) Scope of validity of PSNR in image/video quality assessment. *Electron Lett* 44(13):800–801
- Hu J, Lu J, Tan YP (2014) Discriminative deep metric learning for face verification in the wild. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1875–1882
- Eskicioglu AM, Fisher PS (1995) Image quality measures and their performance. *IEEE Trans Commun* 43(12):2959–2965
- Farid A, Adel A (2020) Performance evaluation of machine learning for recognizing human facial emotions. *Rev Intell Artificielle* 34(3):267–278. <https://doi.org/10.18280/ria.340304>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.