

Research Article

A New Local Descriptor Based on Strings for Face Recognition

Hicham Zaaraoui ¹, Abderrahim Saaidi,^{1,2} Rachid El Alami ³, and Mustapha Abarkan¹

¹LSI, Department of Mathematics, Physics and Computer Science, Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P. 1223, Taza, Morocco

²LIAN, Department of Mathematics and Computer Science, Faculty of Science, Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, B.P. 1796, Atlas, Fez, Morocco

³LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, B.P. 1796, Atlas, Fez, Morocco

Correspondence should be addressed to Hicham Zaaraoui; hicham.zaaraoui@usmba.ac.ma

Received 11 April 2019; Revised 1 September 2019; Accepted 20 January 2020; Published 18 February 2020

Academic Editor: Sos S. Agaian

Copyright © 2020 Hicham Zaaraoui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes the use of strings as a new local descriptor for face recognition. The face image is first divided into nonoverlapping subregions from which the strings (words) are extracted using the principle of chain code algorithm and assigned into the nearest words in a dictionary of visual words (DoVW) with the Levenshtein distance (LD) by applying the bag of visual words (BoVW) paradigm. As a result, each region is represented by a histogram of dictionary words. The histograms are then assembled as a face descriptor. Our methodology depends on the path pursued from a starting pixel and do not require a model as the other approaches from the literature. Therefore, the information of the local and global properties of an object is obtained. The recognition is performed by using the nearest neighbor classifier with the Hellinger distance (HD) as a comparison between feature vectors. The experimental results on the ORL and Yale databases demonstrate the efficiency of the proposed approach in terms of preserving information and recognition rate compared to the existing face recognition methods.

1. Introduction

Face recognition is a biometric technology to identify an individual from a picture of his face, which is a passive and nonintrusive system to verify the identity of a person, and this field has progressed significantly in recent years due to its many applications primarily in forensic sciences, driver's licenses and passport verification, missing identification, surveillance systems, social networks, access control, and location of missing persons [1, 2]. Despite the great advances in computer vision research, face identification still poses many challenges because of some typical problems as facial expressions, pose variations, lighting changes, partial occlusions, aging, and change of appearance [1, 2]. Therefore, many approaches are invented to overcome these difficulties. As a result, the topic remains active for further research.

A face recognition system consists of three critical steps: detection of a face in an image based on the position of the

eyes [3, 4], feature extraction stage, and the classification stage. Generally, a key issue in face recognition is finding an efficient facial image representation as a feature vector. Therefore, facial features approach, *holistic* or *local*, is the most vital stage to accomplish a high recognition rate.

The *holistic features approach* aims to construct a subspace representing a face image by using the entire face region; they are easy to implement, but they are sensitive to variations in face as facial expressions, pose, or illumination changes. Though the *local features approach* is appeared, the key idea is splitting the face image into small patches and extracting the local features from each subregion, and they are less affected by changes in the appearance of the face and invariant to changes in illumination and pose; therefore, they have received an increasing attention in recent years.

One of the most relevant local methods, named local binary pattern (LBP), is presented by Huang et al. [5] as a survey in the face recognition subject. The descriptor is

simple to implement, invariant to lighting changes, and efficient in terms of low computational complexity. Despite these advantages, the LBP descriptor still has several limitations such as production of long histograms, captures only information of very local texture structure and not long-range information, and limited discriminative capacity based purely on local binarized. To overcome these limitations, the LBP descriptor is the author of several local descriptors based on the pattern (rectangular or circular). For instance, full ranking (FR) named by Chan et al. [6] is a local descriptor that places the pixel positions in order according to their intensities in one of the sampling models based on the BoVW paradigm [7] and utilizes the nearest neighbor classifier (NNC) with the Chi-squared distance χ^2 to measure the similarity between two histograms.

The most important method in the computer vision research field, presented by Freeman [8], used shape representation in an image where the representation of an object is based on its form; this technique has several advantages such as the compact representation of a binary object, the comparison of objects, the calculation of any feature shape, and the compression and the conservation of topological and morphological information using in the analysis of line models in terms of speed and efficiency; these advantages allow us to apply this method in face recognition field by using strings to describe the face.

In this paper, a new local descriptor based on strings is introduced to face recognition to obtain information about both local and global properties of a face image. As opposed to the full ranking algorithm [6], the string descriptor does not require a sampling model, and the rankings are replaced by strings. The extraction of strings is based on the principle of the chain code algorithm [8]. The face description is done by bringing out strings from each pixel such that each string gives the relative positions of the successive points constituting the directional changes of the maximum intensities, as well as the minimum intensities in only four directions as in four freeman chain code (FFCC) algorithm [8], because there are more directional changes for border in four directions which ensure better border tracking such that each pixel produces two words as illustrated in Figure 1.

After dividing the face image into small regions, the words are obtained from each region as follows: a string of maximum values is extracted when the nearest pixel to the starting one (148 in this case) has the highest intensity (157 in this case), and then the string starts by a character (*R*, *L*, *U*, or *D*) that indicates one of the four directions of movement (right, left, up, or down, respectively). In this example, the new pixel (157) is in the right position, so the first letter of the sequence is *R*; the next neighbor pixel (180), is at the top, so the second letter is *U*; the same for the other characters until the new pixel has the highest value (213 in this case). We proceed in the same way to obtain the word of minimum values continuing the path of successive minima.

In this way, the technique of BoVW [7] is applied to reduce the number of words in a DoVW such that each subregion is represented by a histogram containing the occurrence of visual words, and these histograms are

assembled into one as a signature of the face image, where the direct extraction of LBP histograms adopted in the LBP algorithm is replaced by a DoVW. Our study focuses on the variation of the number of blocks in an image, the number of visual words in a dictionary, and the dictionary words extracted from the face images to improve the face recognition rate. The proposed algorithm is evaluated on the ORL (Olivetti Research Laboratory, Cambridge) [9] and Yale [10] databases. The results obtained are very satisfactory in terms of recognition rate and efficiency.

This paper is composed as follows. In Section 2, a brief review of related work is presented. Section 3 introduces the new descriptor based on strings. Experimental results are given in Section 4. Section 5 concludes the paper.

2. Related Work

In the face recognition system, the design of a local face descriptor remains a big challenge to three competing goals: computational efficiency, adequate discrimination, and robustness to variations between people (including changes in lighting, posture, facial expression, age, and facial occlusion). Nevertheless, the motivation of this work is to apply the bag of words model in a new local descriptor based on strings as a face descriptor.

A face image can be described by two facial extraction approaches: global or local approaches. The works on the global approaches has started since the technique invented by Turk and Pentland in [11], called principal component analysis (PCA) which is performed on training images to extract the eigenvectors which define a new reduced dimension space. The face images are then projected on this space, and the vectors obtained are used for classification. The PCA algorithm is fast to implement, simple, and popular in model identification. However, it is not optimized for discrimination (separability); for this purpose, Belhumeur et al. [12] proposed a linear discriminant analysis (LDA) named Fisherfaces which is consisted in maximizing the ratio between interclass variations (images of different individuals) and intraclass variations (images of the same individual). It only works well when many images per person are available in the same class. In contrast to the PCA and LDA that rely on supervised learning, locally linear embedding (LLE) is an unsupervised learning algorithm that computes low-dimensional of high-dimensional inputs [13]. However, LLE is only defined on learning data points, so evaluating maps for new test points remains fuzzy. This issue is tackled with the locality preserving projections (LPP) method [14], which can be basically connected to any new information point to find it in a reduced representation space, and it seeks to preserve the intrinsic geometry of the data, as well as the local structure, but considers that local dispersion in the classification phase. Whereas the consideration of local dispersion and nonlocal dispersion is treated by the unsupervised discriminant projection technique (UDP) [15] by looking for a space projection that simultaneously maximizes nonlocal dispersion and minimizes local dispersion, which makes the UDP algorithm more efficient and powerful than LPP. Although these methods

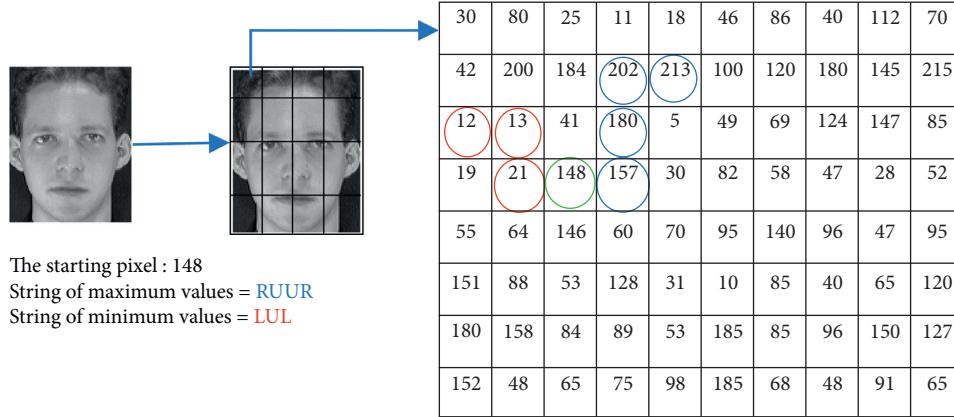


FIGURE 1: Obtaining strings from a starting pixel (148).

have been very successful, they remain sensitive to illumination change, pose, and facial expression. One way to avoid these challenges is using local approaches that are model-based and rely on a separate treatment of different divisions of the facial image. The models used are based on the knowledge that we have first at the face morphology. This category of methods manifolds advantages such as easily modelling the variations of pose, illumination, or facial expressions.

In the local methods, the face image is subdivided into small patches as in [5] where the local binary patterns (LBP) are directly extracted from each block. The idea is constructing a binary code by thresholding a neighborhood with the grey level of the central pixel. All the neighborhoods then take a value of 1 or 0 according to the sign of the difference between the neighboring pixels and the central one. As a result, the LBP code is a decimal value obtained by converting the binary code; therefore, the intensity is between 0 and 255 to describe the local texture of a region. The decimal value (LBP code) is obtained in a neighborhood of 8 pixels as in Figure 2.

The eight neighboring pixels are compared with the central value (175), and each pixel is coded with 1 if the difference is greater or equal than 0 and with 0 otherwise; therefore, a binary code is obtained and converted into a decimal value (LBP code). As aforementioned, the limitations of LBP are causing a number of extensions such as multiscale block binary pattern (MB-LBP) proposed by Liao et al. [16]. The algorithm encodes not only the microstructures of image patterns but also the macrostructures because the calculation is based on the average values of the blocks instead of individual pixel as in the LBP code, where it has more efficiency than LBP. A generalization of LBP has been proposed by Tan and Triggs [17] called the local ternary pattern (LTP), where the binary code LBP is replaced by a tertiary code; this technique is more resistant to noise, but it is not strictly efficient against the change in greyscale level. In this way, Jabid et al. [18] revealed the local directional pattern (LDP), where the edge is detected at each pixel position by using Kirsch masks in eight orientations. Reducing the complexity of the LDP has motivated Perumal and Chandra Mouli [19] to develop the dimensionality reduced LDP (DR-LDP) by computing a single code for each

block by X-ORing the LDP codes retrieved in a single patch instead of eight LDP codes. Far from the comparison of pixel intensities used as a basis of local descriptors, Chan et al. [6] proposed a full ranking descriptor of a set of pixels, based on the BoVW paradigm [7] to represent face images.

3. Face Recognition by Using Strings as a Local Descriptor

A string represents the path pursued by the local descriptor, based on four freeman chain code (FFCC) [9]. Therefore, the signature of each face image is given by visual words assembled in a histogram. Generally, the feature vector (histogram) has the same dimension as the dictionary length and plays a crucial role in the feature extraction step. The nearest neighbor classifier (NNC) is used at the last step to identify subjects; Figure 3 demonstrates the complete face recognition system.

In our method, the dictionary of visual words (DoVW) is first constructed from an image chosen randomly from the training dataset, and then the strings (words) are extracted from each subregion of the face images and replaced by the nearest words in the dictionary. The assignment of words is done by a distance which gives a measure of the difference between two strings, called Levenshtein distance (LD) [20], which is equal to the minimal number of insertions, replacements, and deletions which are needed to make two strings equal. Therefore, each division is represented by a histogram. Furthermore, the face descriptor is obtained by assembling these histograms as a feature vector. However, the facial description is given on two levels of locality: the words are obtained over a small space to produce information at the regional level, and the concatenation of regional histograms provides an overall description of the face. Finally, the similarity between two feature vectors is calculated by the nearest neighbor classifier using Hellinger distance (HD) [21] in the classification stage and thus to identify a person from his face image.

3.1. Dictionary of Visual Words (DoVW). The construction of a DoVW is a critical step for the performance of face recognition systems using such representation. As

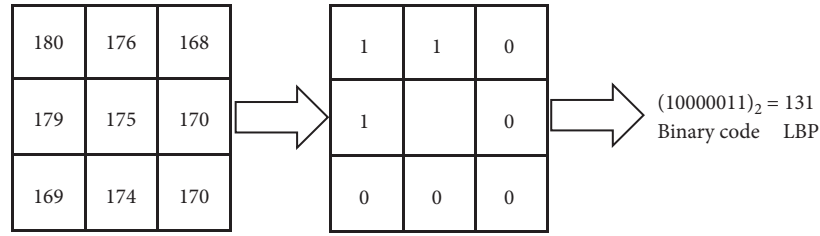


FIGURE 2: Computation of the LBP code of a given pixel (175).

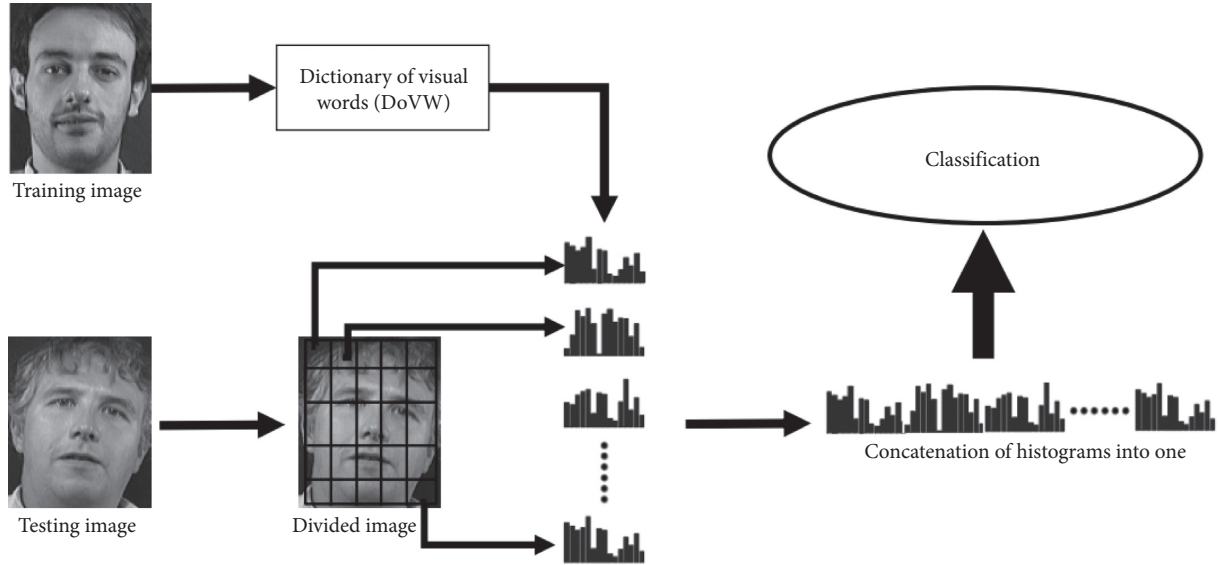


FIGURE 3: The overall face recognition system using a new local descriptor based on strings.

aforementioned, the face image can be described by a set of words since each pixel produces two words: one from the path of successive maxima and the other from the path of successive minima; consequently, the number of words is huge. However, the BoVW paradigm [8] is introduced to reduce this number in the DoVW. Algorithm 1 shows the steps to build a dictionary from a given face image.

3.2. New Local Descriptor Based on Strings. The intervention of a local descriptor in the feature extraction step is primary. For this purpose, a new local descriptor based on strings is presented in this part where the principle of freeman chain code (FCC) algorithm is applied to bring out the strings.

3.2.1. Freeman Chain Code. The chain code, introduced for the first time in 1961 by Freeman [8], aims to encode the shape of an object by a chain of coders giving the relative position of the next point of the object's outline. The properties of objects can be studied by using four-direction neighborhoods noted $I_4(x, y)$ of a pixel where x and y represent the row and column numbers, respectively. These four neighborhoods are shown in Figure 4 which corresponds to the four closest pixels to the starting one, such that all the neighborhood's points are in an equal distance from the central one; however, the distance is 1 pixel. We have

then 4 connexities (the passage of a pixel to next one is only in vertical or horizontal). Thus, using 4 directions, the coding R means that the next pixel from the contour of the object is in the right position of the current pixel, and the coding D designates the next pixel in the downward position of the current pixel.

The FFCC of an edge is determined by specifying a starting pixel $I(x, y)$ and the sequence of unit vectors obtained from left $I_L(x, y) = I(x + 1, y)$, right $I_R(x, y) = I(x - 1, y)$, up $I_U(x, y) = I(x, y - 1)$, or down $I_D(x, y) = I(x, y + 1)$ moving from a pixel to another along the contour. In our representation, the 4 connexities are chosen instead of 8 connexities [8] because directions of the border are more numerous and hence permit a better pursue of the shape. As a result, this code keeps the path of direction changes by moving from one pixel to another along an outline. The obtaining of the FCC is presented in Algorithm 2:

A coding chain gives the representation of the contour in an object, and this algorithm does not require the extraction of the border before being executed and thus has the advantage of working on a complete binary object and not only on the outline of the object.

3.2.2. Local Descriptor Based on Strings. For all the face images, the strings are extracted starting from each pixel, such as each pixel produces two different strings.

Input: Train image (matrix)

Output: Dictionary of Visual Words (DoVW)

- (1) Divide the face image into subregions.
- (2) Locate the pixels in each subregion.
- (3) Bring out the strings formed only by the four letters (R, L, U and D).
- (4) Construction of the dictionary D in each subregion i as follows:

$$D_i = [W_{i,1}^+, W_{i,2}^+, \dots, W_{i,j}^+ \dots W_{i,(m/2)}^+, W_{i,1}^-, W_{i,2}^-, \dots, W_{i,j}^- \dots W_{i,(m/2)}^-], \quad (1)$$

Where $i = 1, \dots, r$ is the number of subregions, and $j = 1, \dots, (m/2)$ is the number of words in each subregion, $w_{i,j}^+$ stands for the j th word from the i th subregion, following the path of successive maxima, and $w_{i,j}^-$ stands for the j th word from the i th subregion, following the path of successive minima.

- (5) Concatenate regional dictionaries into one: $D = [D_1, D_2, \dots, D_i \dots D_r]$ (2)

Once the DoVW is created, the words are extracted from the face image by the process pursued in the next part.

ALGORITHM 1: Building dictionary.

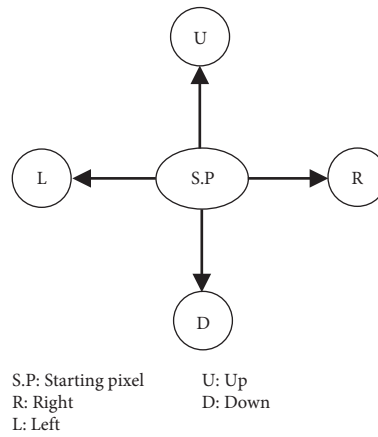


FIGURE 4: Four freeman chain code (FFCC).

Input: Train image (matrix)

Sortie: Freeman Chain Code (FCC)

- (1) Select a starting point $I(x_0, y_0)$.
- (2) Find the first coding (D, G, H or B) from the starting point.
- (3) Store the first coder.
- (4) New starting point = coding point.
- (5) **While** New starting point $\neq I(x_0, y_0)$ **do**
- (5.1) Find the next coder.
- (5.2) Store the coder.
- (6) **End while**
- (7) **Return** coding chain

ALGORITHM 2: Obtaining a FFCC.

Figure 5 shows the process of obtaining strings while pursuing the path of the maximum and minimum values in the grayscale level.

A string created by a sequence of the maxima (respectively, minima) successive can be empty if the starting pixel has the higher (respectively, lower) intensity compared to its four neighboring pixels in four directions.

The extraction of the feature vector (histogram) is the most important task in each face recognition system, and the different steps to extract this vector are presented in Algorithm 3.

The second task is to compare the histograms of testing and training sets in the classification stage.

3.3. Classification Stage. When a new face needs to be identified in the classification stage, the signature of the face image is given as a vector (histogram), so the identification is performed by finding the nearest feature vector in the database. Therefore, the design of a classifier that discriminates the feature vectors remains a decisive phase for each face

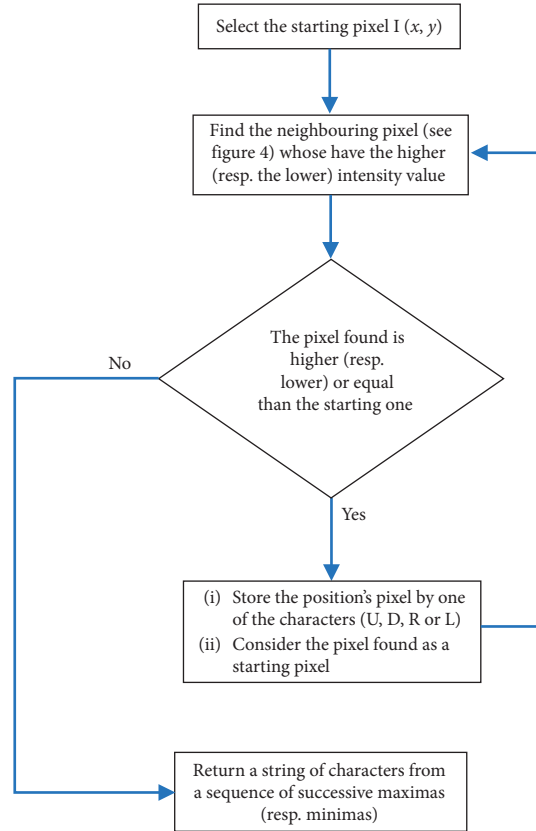


FIGURE 5: Process of finding strings from a pixel in the face image.

Input: Train an image (matrix)

Output: Feature vector (histogram)

Step 1: Divide the face image into subregions

Step 2: Extract the strings from each pixel for each region.

Step 3: Replace all strings in each subregion with the words appearing in regional dictionary based on the Levenshtein Distance, and that for all images.

Step 4: Construct a histogram of words for each image.

ALGORITHM 3: Extraction of the feature vector.

recognition system. In this work, the nearest neighbor classifier (NNC) with the Hellinger distance (HD) [21] (or coefficient of Bhattacharyya) is used to measure the dissimilarity between the histograms as:

$$D(S, M) = \sqrt{1 - \frac{\sum_{i=1}^n \sqrt{S_i M_i}}{\sqrt{\sum_{i=1}^n S_i \sum_{i=1}^n M_i}}}, \quad (1)$$

where S and M designate histograms, i is taken from their bins, and S_i and M_i are the values of the bins. Algorithm 4 shows how to classify a feature vector of a test image.

From an experimental point of view, the database is first divided into training and testing set; next, a dictionary is built from an image of training set with the steps of Algorithm 1, based on the principle of Algorithm 2; afterwards, the extraction of the feature vectors of all the images in the database is carried out using Algorithm 3, and thus each

image is described by a vector which has the same dimension as the dictionary of visual words; finally, Algorithm 4 is applied for each image from testing set to identify the face, and hence the ratio between the number of correctly recognized subjects (individuals) and the total number of testing subjects is calculated as a face recognition rate.

The experimental section treats the impact of the number of regions, when the face image is divided, the size of a dictionary, thus the number of visual words, and the changing of a dictionary by moving from an image to another on the recognition result.

4. Experimentation

To assess the performance of our approach to facial expressions, pose variations, illumination changes, partial occultations, aging, and change of appearance, we tested our

Input: Feature vector of a new face image

Output: The class of the image

Step 1: Repeat the steps in Algorithm 2 for a test image.

Step 2 (NNC): Measure the similarity between the feature vector representing the new face image and all feature vectors in the database using HD. The class (the person) is located by the vector which has the lower distance with the reference vector.

ALGORITHM 4: Classification of the feature vectors.

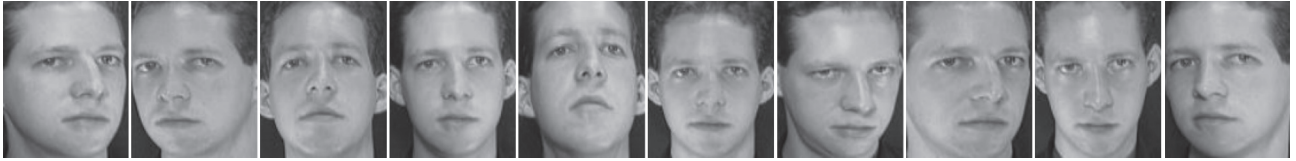


FIGURE 6: Ten views of a person from the ORL database.

TABLE 1: Impact of the divisions on recognition rates (%).

NoR	4	9	16	25	36	49	64	81	100
RR (%)	88	89,50	90	92,50	89,50	89	87,50	85,50	85

approach on two classical databases used in face recognition community: ORL and Yale.

4.1. Results on ORL Database. The ORL database is composed of 400 images of 40 distinct subjects. The images are greyscale level with a resolution of 112×92 and 10 different views for each person (see Figure 6). They were posted with lighting variations, facial expressions (happy, surprise, sadness, anger, fear, disgust, and neutral), and partial occlusions (glasses, beard, and mustache) at various times for certain individuals. The variations of poses concern the head for some people.

In our experiments, the ORL database is divided into two sets: training and testing set, selecting randomly five images of each person for the gallery set and the rest for the probe set, and the face images are taken without preprocessing (cropping, scaling, or histogram equalization) for an effective evaluation of the proposed descriptor.

Our algorithm is related to three parameters: the number of subregions, the number of visual words, in consequence the dictionary size, and the type of visual words forming a dictionary which are changed from an image to another.

4.1.1. Influence of the Number of Subregions (NoR) on Recognition Rate (RR). The number of regions (NoR) is obtained by dividing each face image from 4 (2×2) to 100 (10×10) small regions to preserve spatial information, the number of visual words in a dictionary is fixed at 500 words as in [6], and one image from the training set was chosen as a reference (here, image 3 of person 26 is chosen) to build a dictionary of visual words (DoVW). Table 1 indicates the rank one recognition rate for different NoR.

From the results in Table 1, RR increases, when the division varies from 4 to 25 subregions, and then it decreases, when the number of regions pass from 25 to 100 and that

noticed for all the other dictionaries created from each image in the gallery set. RR is high when we have divided the face image into 25 blocks because a small subregion produces short strings, which causes loss in the object's shape, while the use of the large region causes the loss of spatial information by the production of longer strings.

4.1.2. Influence of Dictionary Length on Recognition Rate (RR). To evaluate the influence of the dictionary dimension on RR, the dictionary length is changed from 100 to 1000 words, based on the number of words set in [5], while keeping the best choice of the NoF which is 25 subregions and image 3 of person 26 as a reference to build the dictionary of visual words. The results of recognition rate are obtained in Figure 7.

As it can be seen in this figure, the maximum number of estimated words for representing an image is 500 words, which makes it possible to obtain the highest recognition rate of 92.5%.

4.1.3. Influence of Changing the Visual Words Forming a Dictionary on Recognition Rate. To achieve the best recognition rate, this descriptor requires a dictionary of visual words which is constructed from each image belonging to the previous parameters which allow obtaining good results of FRR, such that each image of face is partitioned into 25 subregions, and each region is represented by 20 words in the DoVW; however, the size of the dictionary is 500 words, and we have changed the dictionary, moving from one image to another in the gallery set such that every image of each person is considered as the basis of building a DoVW. The following tables show the face recognition rates obtained from the learning images that are considered as building blocks of dictionaries.

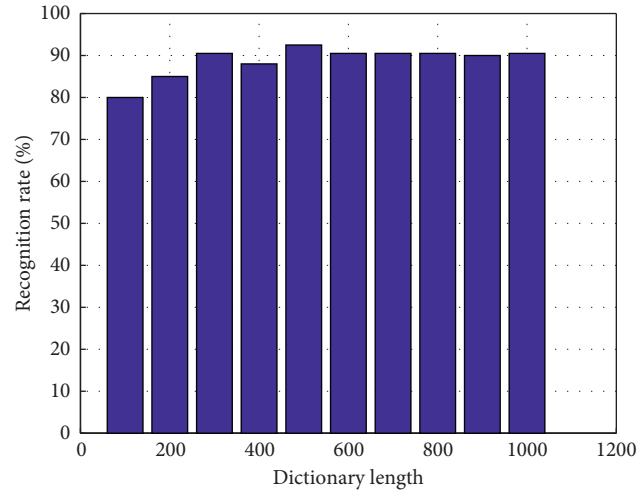


FIGURE 7: Recognition rate (%) for different dictionary's length.

TABLE 2: Recognition rates (%) for each dictionary built from an image of the training set.

	Image 1	Image 2	Image 3	Image 4	Image 5
Person 1	89,50	89,50	90,50	90,50	88,50
Person 2	88,00	89,50	87,50	89,00	90,50
Person 3	89,00	88,00	89,00	87,50	91,00
Person 4	90,50	90,50	89,50	89,50	89,50
Person 5	90,50	91,00	90,00	90,50	88,50
Person 6	90,00	90,00	89,50	89,00	90,50
Person 7	90,50	90,00	91,50	89,00	90,50
Person 8	88,50	89,00	90,00	89,50	89,00
Person 9	90,00	90,00	89,50	90,00	90,00
Person 10	90,50	89,00	90,00	89,00	88,50
Person 11	88,50	89,50	89,50	90,50	91,00
Person 12	89,50	89,00	89,50	90,50	91,00
Person 13	89,50	89,00	89,50	87,50	90,00
Person 14	89,50	89,50	89,50	88,50	89,00
Person 15	91,50	90,00	89,50	90,50	88,50
Person 16	90,00	88,50	91,00	90,00	89,50
Person 17	89,00	90,50	89,50	90,00	87,50
Person 18	88,50	90,50	88,50	90,50	90,50
Person 19	91,00	90,50	90,50	88,50	90,00
Person 20	90,50	90,50	88,00	90,00	89,00
Person 21	91,00	88,50	89,50	91,50	89,50
Person 22	89,50	91,50	88,50	90,00	91,00
Person 23	90,50	88,50	88,50	88,00	89,00
Person 24	91,50	90,00	91,50	90,00	91,50
Person 25	91,00	90,00	88,50	90,50	88,00
Person 26	88,50	91,00	92,50	89,50	91,00
Person 27	89,50	90,00	88,00	90,00	88,50
Person 28	91,50	91,50	89,50	91,00	89,00
Person 29	89,50	89,50	90,00	90,50	89,50
Person 30	88,50	89,00	90,00	89,00	91,50
Person 31	87,50	90,50	90,50	88,50	91,00
Person 32	88,00	91,00	88,50	88,00	89,00
Person 33	89,50	92,00	90,50	91,00	89,50
Person 34	89,50	88,50	92,00	90,00	89,00
Person 35	90,00	89,50	91,00	91,50	89,50
Person 36	89,50	91,00	88,50	90,50	88,50
Person 37	90,50	91,00	90,50	92,00	89,50
Person 38	89,00	89,00	90,50	89,00	92,00
Person 39	92,00	89,00	90,00	90,50	89,50
Person 40	92,00	90,50	90,50	90,50	87,50

TABLE 3: Recognition rates (%) of the proposed method and different approaches on the ORL database.

Recognition method	Recognition rate (%)
PCA [22]	90,50
LDA [23]	90,35
LPP [23]	90,80
FR [6]	92,00
Our method	92,50



FIGURE 8: Eleven images of a person extracted from the Yale database.

TABLE 4: Recognition rates (%) of the proposed method and different approaches on the Yale database.

Recognition method	Recognition rate (%)
PCA [23]	83,22
LDA [23]	88,00
LPP [23]	88,22
FR [6]	89,33
Our method	90,66

According to Table 2, the choice of visual words is the most important parameter to achieve a height recognition rate, and the lowest and the highest recognition rates are 87.5% and 92.5%, respectively, which shows that the change of the dictionary influences the results of recognition rate with a shift up to 5%. In addition, the calculation of the average recognition rate for all the images of each person shows that the change of the dictionary from one image to another of the same person is not too much affected by the change of lighting, facial expression, or pose since the recognition results are very close to the average with a gap not exceeding 2%.

The performance of our method is compared with PCA, LDA, LPP, and FR. Therefore, the same parameters for our method and the FR descriptor can be used for a fairly comparison: the number of subregions is set to 25 (from Table 1) so each face image from the ORL database is divided into 5×5 blocks, and the number of visual words forming the dictionary is fixed to 500 words (from Figure 7). Moreover, the circular model (8, 1) is used as a mask for FR, and the experiment is repeated two hundred times (see Table 2) taking a new dictionary each time (extracted from a new face image) for the two descriptors, and the best recognition rate (%) is taken into account in Table 3.

From the recognition results reported in Table 3, we note that our technique is more efficient than the classical facial feature extraction methods such as PCA, LDA, and LPP and comparable or even outperforms the Full Ranking (FR) descriptor as one of the recent local methods based on the BoVW paradigm with a little gap (0, 50%). Therefore, the recognition rate of the proposed method confirms the good

performance of our method compared with the class of descriptors based on the BoVW model.

4.2. Results on the Yale Database. The Yale database is composed of 165 images of 15 people such that each individual contains 11 images (Figure 8). These images are taken under different illumination conditions (left, right, and center), several facial expressions (surprised, sad, happy, sleepy, wink, and normal), and facial details with and without glasses. To facilitate processing, the original images with 320 pixels in width and 243 pixels in length are reduced to a resolution of 112×92 , which is the same resolution of the images in the ORL database.

To examine the effectiveness of our descriptor, we tested it on the Yale database too. As previously, the partition of the face image (5×5) and the dictionary size (500 words) are preserved, and the number of samples in the training set database is fixed at 75 images, by randomly choosing 5 images per person and the remainder for the test. We repeated the experiment 75 times, by changing the image to build the dictionary each time and reporting the best recognition rate obtained in Table 4 with the recognition rates of the other approaches (PCA, LDA, LPP, and FR).

As shown in Table 4, our descriptor outperforms the traditional methods (PCA, LDA, and LPP) on the Yale face database and comparable with FR descriptor because the highest recognition rate of our approach is 1.33% which is higher than that of FR. Therefore, the proposed method confirms the reliability of our descriptor against the high variations of illumination, pose, facial expression, as well as facial detail.

5. Conclusion

In this paper, we use strings as a new local descriptor and the nearest neighbor classifier with the Hellinger distance (HD) to identify people from their images. The string descriptor ensures an efficient description of edges and shapes in image analysis and easy to implement; however, to build a face feature vector, we have divided it into nonoverlapping blocks, where a description of each region is given by a set of words, which are assigned to the dictionary's words using the Levenshtein distance (LD). The proposed descriptor was evaluated on two standard databases: ORL and Yale. Experimental results have demonstrated the robustness of our approach in terms of recognition rates with suitable parameters and even surpass one of the best methods, named FR, which is also based on the bag of visual words model.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] A. J. S. Li, *Handbook of Face Recognition*, Springer-Verlag, Berlin, Germany, 2nd edition, 2011.
- [3] S. El Kaddouhi, A. Saaïdi, and M. Abarkan, "Eye detection based on the Viola-Jones method and corners points," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 23077–23097, 2017.
- [4] S. El Kaddouhi, A. Saaïdi, and M. Abarkan, "Eye detection based on Viola & Jones Detector, skin color, and eye template," *International Journal of Control and Automation*, vol. 11, no. 5, pp. 59–72, 2018.
- [5] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765–781, 2011.
- [6] C. H. Chan, F. Yan, J. Kittler, and K. Mikolajczyk, "Full ranking as local descriptor for visual recognition: a comparison of distance metrics on sn," *Pattern Recognition*, vol. 48, no. 4, pp. 1324–1336, 2015.
- [7] J. Sivic and A. Zisserman, "Video Google: a text retrieval approach to object matching in videos," *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 2, pp. 1470–1477, 2003.
- [8] H. Freeman, "On the encoding of arbitrary geometric configurations," *IEEE Transactions on Electronic Computers*, vol. EC-10, no. 2, pp. 260–268, 1961.
- [9] AT&T Laboratories Cambridge, "ORL database of faces," 2005, <http://www.cam-orl.co.uk/facedatabase.html>.
- [10] Y. University, "Yale Database," 2009, <http://vision.ucsd.edu/content/yale-face-database>.
- [11] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [14] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 37–44, 2004.
- [15] J. Yang, J. Y. Yang, D. Zhang, and B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 650–664, 2007.
- [16] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, *Learning Multi-Scale Block Local Binary Patterns For Face Recognition*, Springer-Verlag, Berlin, Germany, 2007.
- [17] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [18] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (LDP) for face recognition," in *Proceedings of the Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, vol. 8, no. 4, pp. 2423–2437, Las Vegas, NV, USA, January 2010.
- [19] S. R. Perumal and P. V. S. S. R. Chandra Mouli, "Dimensionality reduced local directional pattern (DR-LDP) for face recognition," *Expert Systems with Applications*, vol. 63, pp. 66–73, 2016.
- [20] L. Yujian and L. Bo, "A normalized Levenshtein distance metric," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1091–1095, 2007.
- [21] S. D. Bharkad and M. Kokare, "Performance evaluation of distance metrics: application to fingerprint recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 6, pp. 777–806, 2011.
- [22] C. Zhou, L. Wang, Q. Zhang, and X. Wei, "Face recognition based on PCA image reconstruction and LDA," *Optik*, vol. 124, no. 22, pp. 5599–5603, 2013.
- [23] Y. Wen, L. Zhang, K. M. Von Deneen, and L. He, "Face recognition using discriminative locality preserving vectors," *Digital Signal Processing*, vol. 50, pp. 103–113, 2016.