

Research Article

Improving the Component-Based Face Recognition Using Enhanced Viola–Jones and Weighted Voting Technique

Issam Dagher  and **Hussein Al-Bazzaz**

University of Balamand, Department of Computer Engineering, El-Koura, Lebanon

Correspondence should be addressed to Issam Dagher; dagheri@balamand.edu.lb

Received 31 August 2018; Revised 17 January 2019; Accepted 27 January 2019; Published 3 April 2019

Academic Editor: Gaetano Sequenzia

Copyright © 2019 Issam Dagher and Hussein Al-Bazzaz. 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 enhances the recognition capabilities of the facial component-based techniques using the concepts of better Viola–Jones component detection and weighting facial components. Our method starts with enhanced Viola–Jones face component detection and cropping. The facial components are detected and cropped accurately during all pose-changing circumstances. The cropped components are represented by the histogram of oriented gradients (HOG). The weight of each component was determined using a validation process. Combining these weights was done by a simple voting technique. Three public databases were used: the AT&T database, the PUT database, and the AR database. Several improvements are observed using the weighted voting recognition method presented in this paper.

1. Introduction

Face recognition is a very important application of pattern recognition at which a database is used to train a classifier that tries to identify each person in it. A handful of studies concerning the face recognition problem were surveyed in [1]. Studies in cognitive science have found that local and global features can be used for face recognition [2–8]. There is enough evidence to prove that all of the holistic, configurable, and facial component information exist in the human face perception [2–15]. Additional studies in humans have concluded that some facial components are more important and useful for recognizing faces than other components. For example, the upper face is more important than the lower face [13, 16]. Researchers have approached face recognition through two methods: component-based and global-based face recognition.

1.1. Component-Based Face Recognition. This method relies on training multiple models depending on the number of components representing an image. This technique in face recognition has not been researched intensively in comparison

to the global-based technique. Therefore, they are limited in their approach [17]. Most of them use raw-pixel representation and that's what makes them less robust. Several other component-based face recognition methods have been discussed in [4, 12, 16]. The facial components used for recognition in this paper are the eye pair, the nose, and the mouth. The Viola–Jones object detection framework [18] was used to crop the facial components.

1.2. Global-Based Face Recognition. On the contrary, to the component-based concept, the global method of face recognition relies on a single array to represent a face. A comparison between the best technique in the global-based face recognition such as eigenfaces, Fisher's discriminant analysis, and kernel PCA can be found in [19, 20]. The global-based face recognition techniques have a weakness against pose changes. This technique had to include a face alignment algorithm phase or be developed to meet the standards of a component-based recognition technique [21].

The remainder of this paper is organized as follows: Section 2 explains the methods we used for component detection and cropping. The HOG features are explained in

Section 3. Section 4 presents the results summarized and compared.

1.3. 3D Face Recognition. 3D facial surface is [22] encoded into an indexed collection of radial strings emanating from the nose tip. Then, a partial matching mechanism effectively eliminated the occluding parts. Facial curves can express the deformation of the region which contains the facial curve used for detecting occluded facial areas. In [23], a novel automatic method for facial landmark localization relying on geometrical properties of the 3D facial surface working both on complete faces displaying different emotions and in presence of occlusions.

2. Component Detection and Cropping

The detection functionality is a vital process in our face recognition method. Components help to collect unique data for every person in the database. Two ways of component detection are used: Viola–Jones object detection framework [18] with geometrical approaches and Landmark detection using face alignment with an ensemble of regression trees [24]. Both facial component detection methods are used to achieve the detection of the facial components in all circumstances (changes in illumination and pose). Accurate component cropping leads to better features. The more the crop is specific to the facial component, the less the useless information is included in the representation, and therefore unique data will participate in the learning process.

2.1. Viola–Jones Object Detection Framework. Viola–Jones object detection framework is used to train a model that detects the facial components (eye pair, nose, and the mouth) needed for the recognition process. It consists of the following parts that are explained in detail in [18]: the Haar-Like features, integral image, weak classifiers and strong classifiers, AdaBoost, and the cascades.

2.2. Enhancing Viola–Jones with Geometrical Approaches. Viola–Jones is a robust object detection system. However, trained models may suffer miss detections or failures in detecting the objects. Our recognition method relies on the accurate detection of the three components (The eye pair, the nose, and the mouth). Miss detections cannot be tolerated when it comes to detecting the facial components. The component-based face recognition system needs the components to be cropped and represented accurately. The miss detection may lead to the representation of useless data (as shown in Figure 1) in the learning process and that yields a lower recognition success rate. The eye pair component is the most crucial part of the three extracted components. The eye pair carries the major unique information about a person's face. It is also the reference object used in this algorithm to detect and crop the rest of the facial component. An eye pair-location prediction model is trained to estimate

where the eye pair might be found in a face. In Figure 2, some cases where the eye pair was not found are demonstrated along with the detection result after the proposed solution. The nose and the mouth object detector might not find the component because the search area did not include the whole object, or a multiobject is detected in the search area. If the object was not detected, then the search area is expanded gradually until an object is found. The multiobject detection framework happened in the mouth area and was solved by picking the object with the maximum y coordinate.

2.3. The Area Selection Process. The concept of the geometrical approaches is to concentrate the search for the components in the right areas. For example, the nose cannot be above the eye pair; it is located somewhere beneath the eye pair. The same concept is applied to the mouth; it has to be under the nose and the eye pair. Geometrical approaches aim to narrow search areas to where the nose and the mouth may occur [25]. The area selection algorithm (Figure 3) consists of the following steps:

- (a) The face is the first component to look for.
- (b) Eye pair detection in the cropped face image.
- (c) The area under the eye pair within the cropped face image will be the search area for the nose.
- (d) Specific area is used to detect the mouth (Figure 3). In case of multiple mouth detection, the object with the more significant y -axis value (the lowest object) is chosen to be used as the mouth component.

Several problems face the usage of Viola–Jones object detection framework for component detection. They are as follows:

- (1) Failure to detect the eye pair.
- (2) Failure to detect the nose.
- (3) Detection of multiple false mouths.

Figure 4 shows the miss detection problems and the solution of our area selection algorithm.

Figures 5 and 6 shows the miss detections and the solution of our area selection algorithm.

3. Features

Pixel patches extracted from facial images are often too large and cannot help building a robust classifier [24]. They are converted into a vector of features. A feature descriptor is an array of data that describes an image or a part of an image. It helps provide unique information about the image. It can support the recognition application for the object in that image. In this paper, we have used the histogram of oriented gradients (HOG) features [26].

3.1. Hog Features. Histogram of oriented gradients (HOG) is a feature descriptor that uses oriented gradient

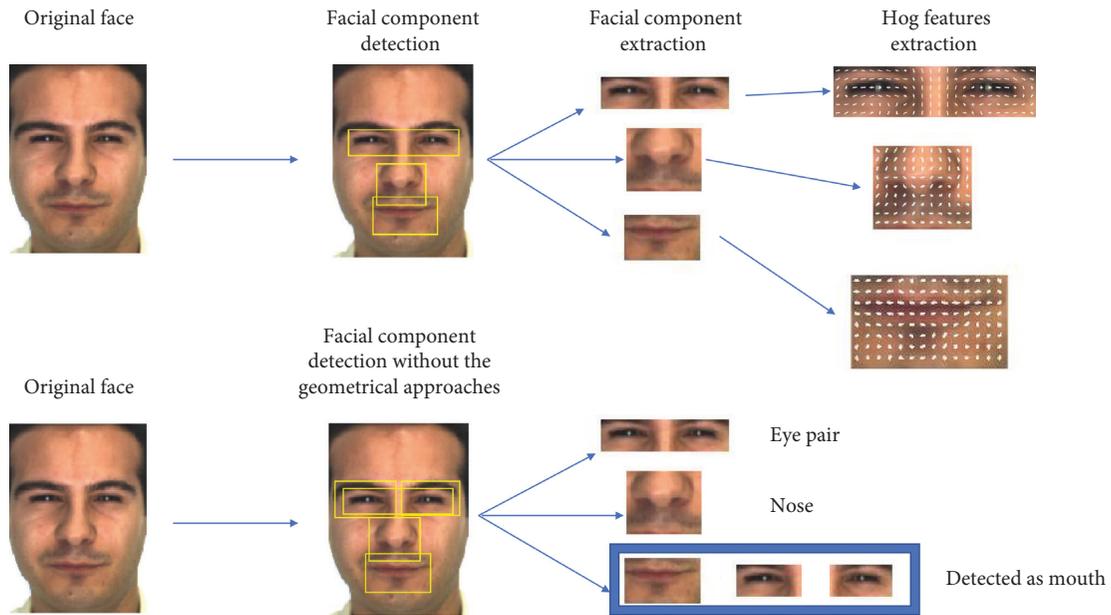


FIGURE 1: The Component Miss-Detection Risk.

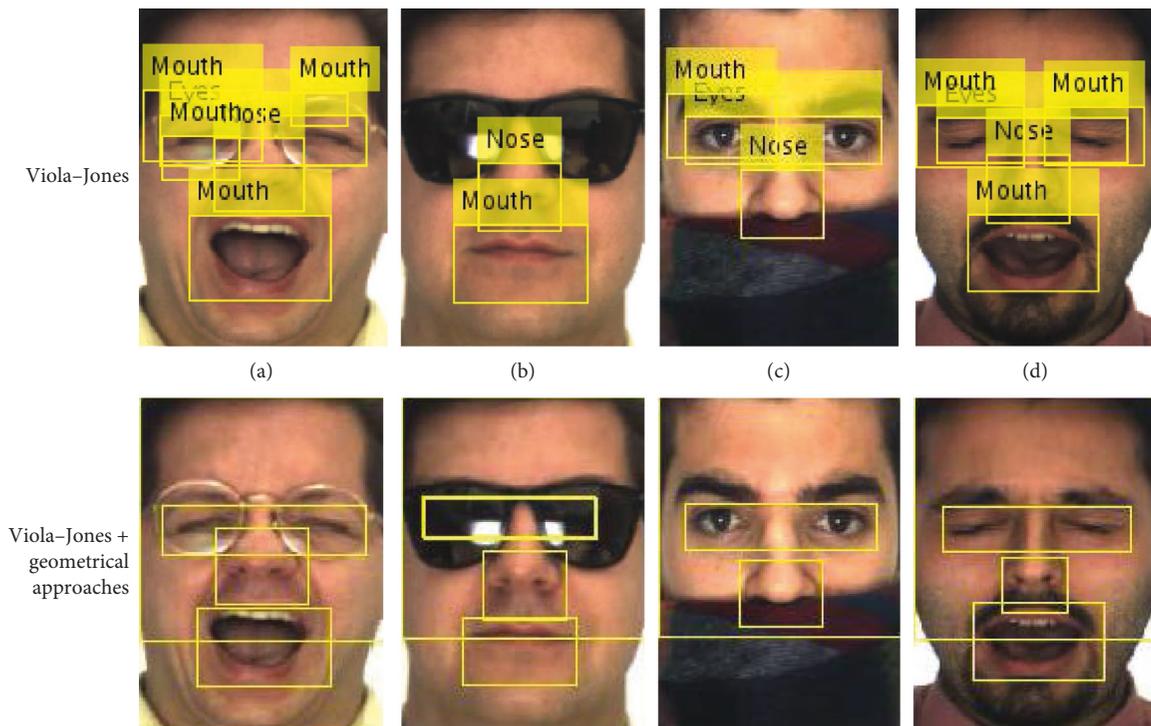


FIGURE 2: Viola-Jones vs. Viola-Jones with geometrical approaches (AR database).

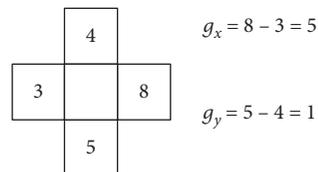
information [26]. The steps for calculating HOG are described as follows:

- (1) For each pixel $I(x, y)$, the horizontal and vertical gradient values are obtained as follows:

$$g_x(x, y) = I(x + 1, y) - I(x - 1, y),$$

$$g_y(x, y) = I(x, y + 1) - I(x, y - 1). \tag{1}$$

- (i) For example,



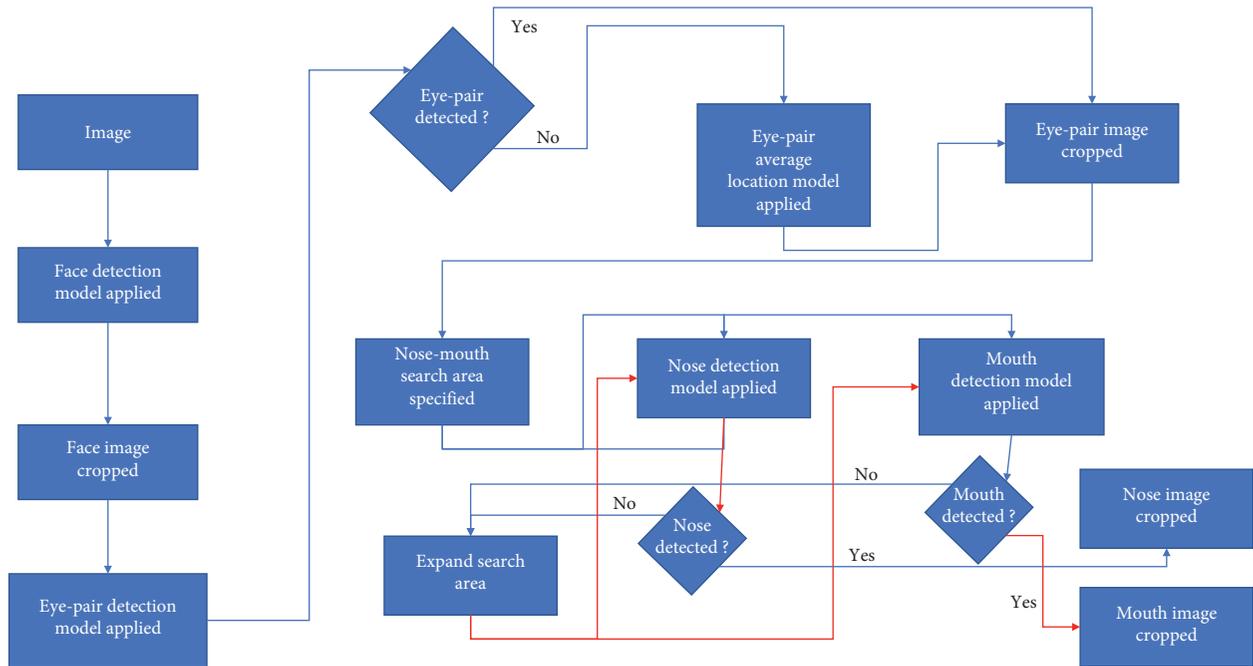


FIGURE 3: The area selection algorithm.

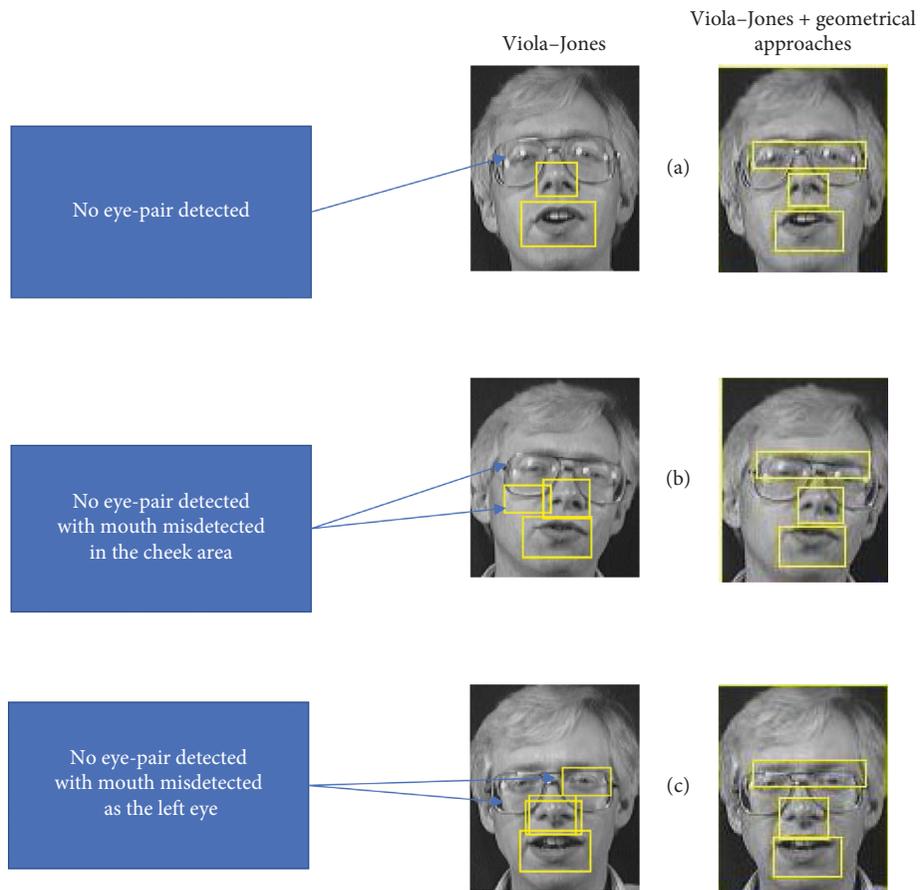


FIGURE 4: Miss detection problems and their solutions.

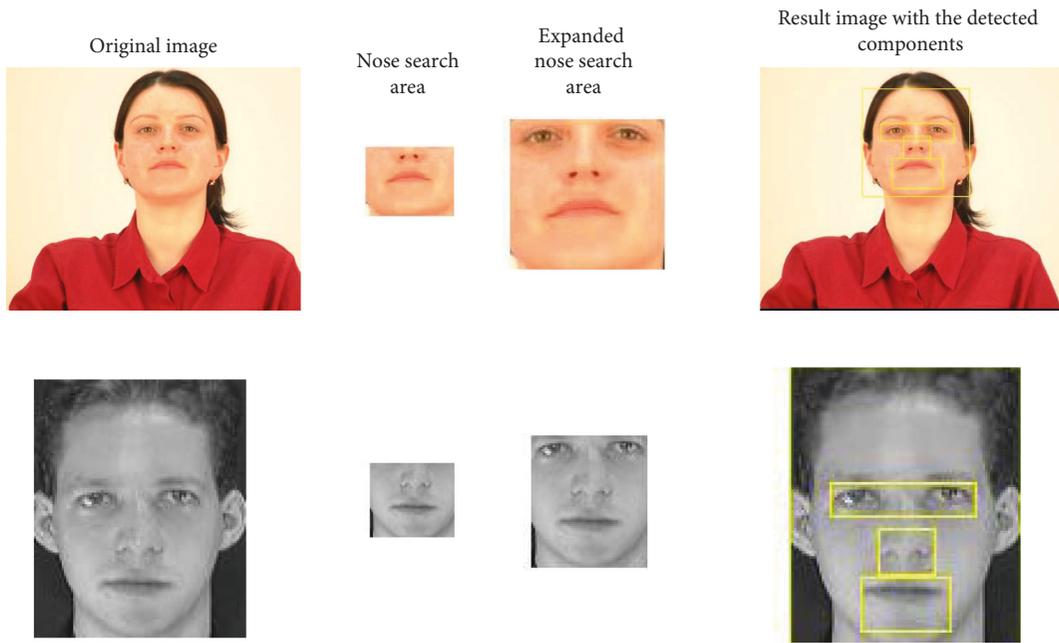


FIGURE 5: Nose miss detection problem.

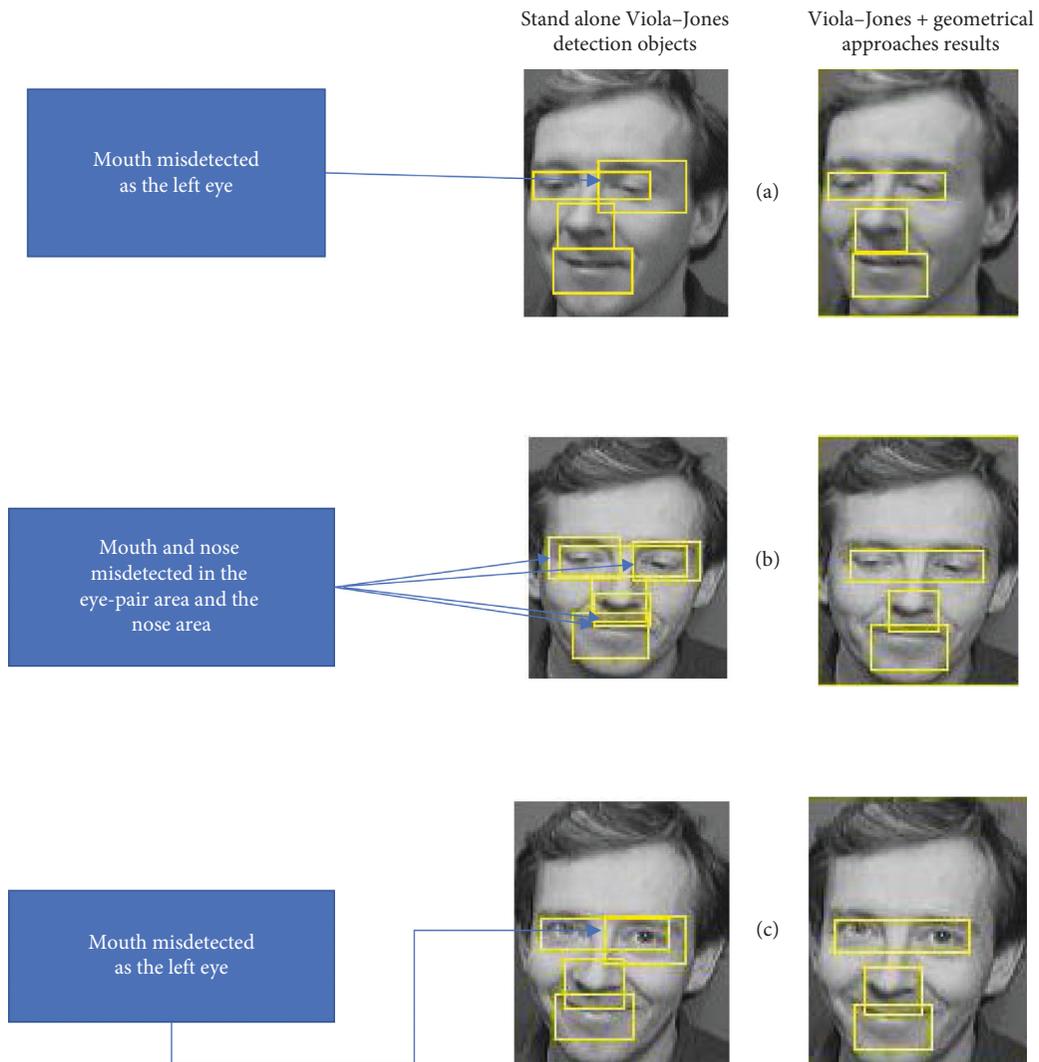


FIGURE 6: Mouth miss detection problem.

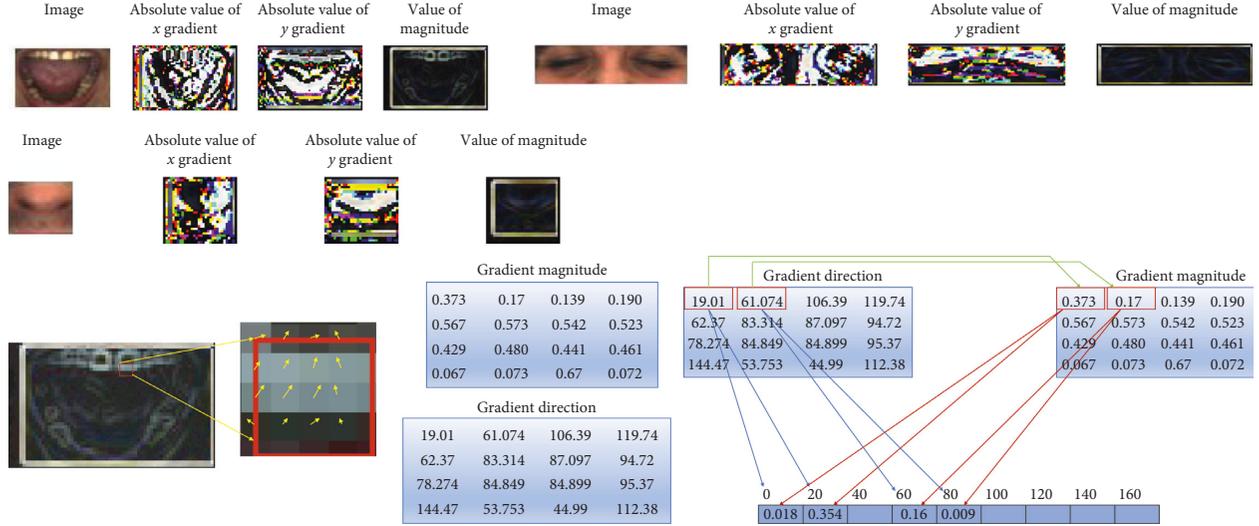


FIGURE 7: Example of the HOG feature vector.

- (ii) The gradient magnitude m and orientation θ are computed by

$$m(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2},$$

$$\theta(x, y) = \arctan \frac{g_x(x, y)}{g_y(x, y)}.$$
(2)

- (iii) The histogram is constructed based on the magnitudes accumulated by orientation.

The image is divided into several small spatial regions (cells) for each of which, a local histogram of the gradient orientations is calculated by accumulating votes into bins for each orientation. The best performance is achieved when the gradient orientation is quantized into 9 bins (0–180). On the contrary, the vote is weighted by the gradient magnitude allowing the histogram to take into consideration the importance of gradient at a given pixel. Finally, the HOG descriptor is obtained by concatenating all local histograms in a single vector.

However, it is necessary to normalize cell histograms due to the fact that the gradient can be affected by illumination variations. Figure 7 shows an example of obtaining the HOG feature vector.

4. Experimental Results

4.1. Face Databases Setups. Three databases were studied in this paper. They have been picked to test the recognition accuracy against low-resolution, missing components, and pose change circumstances. We have used the PUT [27], the AT&T [28], and the AR databases [29]. The PUT database consists of 50 people: each one has 22 colored facial images with different poses and different illumination conditions. The AT&T database consists of images of 40 persons. Each person has ten different facial images. The AR database

TABLE 1: K-Flops and their corresponding testing and learning counts (L = learn images, T = test images).

K-Flops	PUT		ATT		AR	
	L	T	L	T	L	T
2	11	11	5	5	13	13
3	15	7	7	3	18	8
4	17	5	8	2	20	6
5	18	4	9	1	21	5

consists of 50 persons. Each person has 26 different colored facial images. Table 1 shows the different random training sets (k-flops). For example, for the PUT database, for $k=2$, we took 11 out of the 22 as training and 11 for testing. Images with a missing component shall substitute that particular missing component with components detected within its learning/testing set as shown in Figure 8.

The HOG features are calculated on the batch basis for each image. A batch is a part of an image cropped out to seek for its useful information, for example, the eye pair, the nose, and the mouth. The HOG features can be calculated for patches with different aspect ratios. To make the best use of these features, we have to maintain a fixed aspect ratio for all the patches within a single database. A ratio of 1 : 4, 1 : 1 and 1 : 2 was chosen for the eye pair, the nose, and the mouth, respectively (Figure 9).

4.2. The Validation Process. The purpose of this process is to figure out which model performs best for the certain database to calculate its priority. The better the score of the particular component is, the higher its priority is.

We have divided our training sets into 2 sets: training (75%) and validation (25%).

This technique uses the validation results to assign weights to each component. The higher the weight assigned to a certain component, the heavier the impact it has on the final classification result. The process is demonstrated in Figure 10.

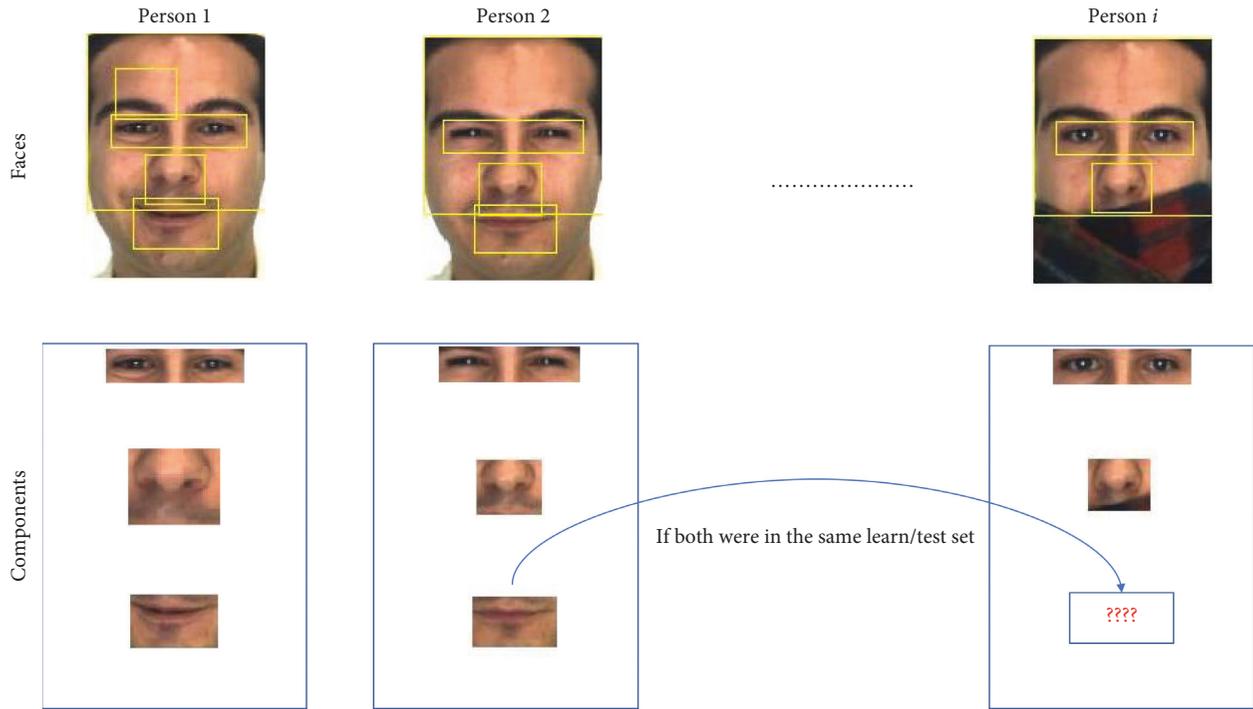


FIGURE 8: The missing component case.

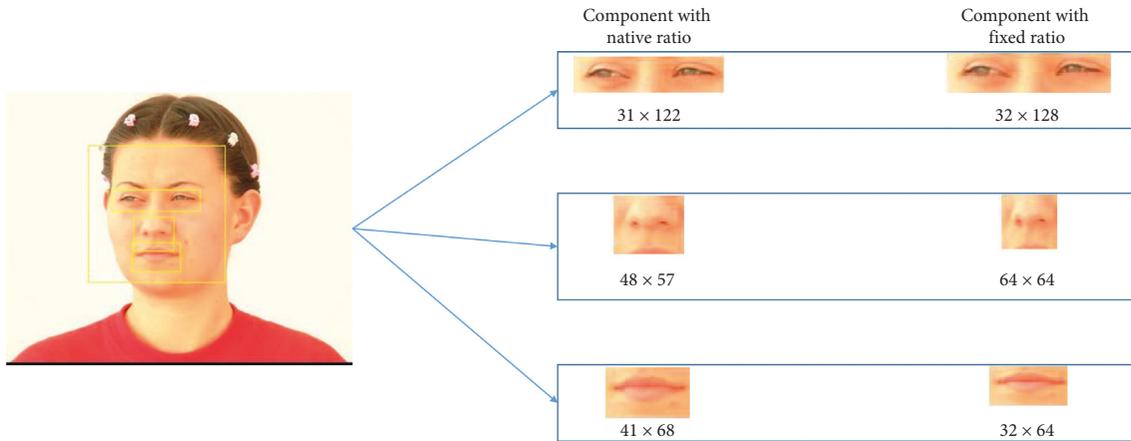


FIGURE 9: Component resizing process for the PUT database.

4.3. Results. The results for the three databases are shown in the following subsections.

4.3.1. The PUT Database Recognition Results. Using our validation process, Table 2 shows the priority of each component for the PUT database. Combining these priorities with a voting technique reached a 100% accuracy success rate for $k=5$ (Table 3).

4.3.2. The AT&T Recognition Results. Table 4 shows the priority of each component for the AT&T database. The voting recognition success rate reached 96% accuracy success rate for $k=5$ (Table 5).

4.3.3. The AR Database Recognition Results. Table 6 shows the priority of each component in the AR database. The voting criteria improved the recognition success range from 73% to 87% for $k=2$ and from 84% to 94% for $k=5$ (Table 7).

4.3.4. Summary of Results. Three public databases were used:

- AT&T with 40 subjects and 400 images.
- PUT database with 50 subjects and 1100 images.
- AR database with 50 subjects and 1300 images.

Our method has the following advantages:

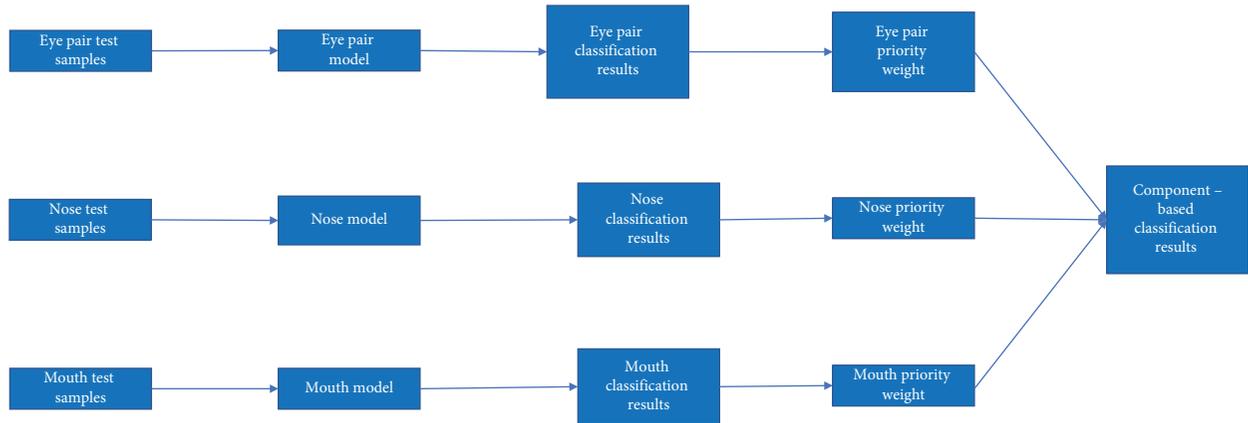


FIGURE 10: The priority voting process.

TABLE 2: Validation recognition success rates for each component of the PUT database.

	PUT database			
	2	3	4	5
Eye pair	0.89091	0.90286	0.968	0.97
Eye pair priority	0.94727	0.97714	0.992	0.99
Nose	0.91075	0.91379	0.936	0.96
Nose priority	0.95455	0.98286	0.992	0.99
Mouth	0.96182	0.97714	0.968	0.985
Mouth priority	0.97636	0.98571	0.992	1

TABLE 3: Recognition success rates improvement using our approach (PUT database).

Face recognition method	PUT database			
	2	3	4	5
Average KNN success rate	0.92	0.93	0.95	0.97
Facial component priority voting	0.97	0.98	0.99	1

TABLE 4: Validation recognition success rates for each component of the AT&T database.

	AT&T database			
	2	3	4	5
Eye pair	0.8	0.775	0.7625	0.8625
Eye pair priority	0.9125	0.875	0.9125	0.925
Nose	0.8875	0.93333	0.8875	0.975
Nose priority	0.925	0.925	0.875	0.9375
Mouth	0.8375	0.88333	0.9125	0.9375
Mouth priority	0.9375	0.94167	0.9375	0.9625

TABLE 5: Recognition success rates improvement using our approach (AT&T database).

Face recognition method	AT&T database			
	2	3	4	5
Average KNN success rate	0.84	0.86	0.85	0.92
Facial component priority voting	0.93	0.94	0.93	0.96

- (i) Excellent accuracy in detecting facial components during all pose-changing circumstances.
- (ii) Improved recognition accuracy by combining multiple classifications using majority voting.

TABLE 6: Validation recognition success rates for each component of the AR database.

	AR database			
	2	3	4	5
Eye pair	0.78308	0.8275	0.82	0.868
Eye pair priority	0.87077	0.9225	0.92	0.944
Nose	0.69385	0.785	0.71333	0.804
Nose priority	0.84154	0.8975	0.88	0.928
Mouth	0.73692	0.7925	0.76667	0.852
Mouth priority	0.84462	0.915	0.85667	0.928

TABLE 7: Recognition success rates improvement using our approach (AR database).

Face recognition method	AR database			
	2	3	4	5
Average KNN success rate	0.73	0.80	0.76	0.84
Facial component priority voting	0.87	0.92	0.92	0.94

5. Conclusion

Enhancing the recognition capabilities of the facial component-based techniques was the objective of this paper. This was done by using the concepts of better Viola–Jones component detection and weighting facial components. Each component was given a certain weight using a validation process. We used a voting technique which incorporates all of these weights. The component-weighted technique supplied the opportunity to involve multiple features into the success rate, granting that chance to use a particular feature's strength to suppress other feature's weakness. The improvement of the weighted voting method is demonstrated for the databases that we have used. The voting technique has boosted the recognition success rate. The boost in the success rate within the voting technique distributes the weight importance among the facial components without settling for one major facial component.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–741, 1995.
- [2] J. L. Bradshaw and G. Wallacei, "Models for the processing and identification of faces," *Perception and Psychophysics*, vol. 9, no. 5, pp. 443–448, 1971.
- [3] J. Sergent, "An investigation into component and configural processes underlying face perception," *British Journal of Psychology*, vol. 75, no. 2, pp. 221–242, 2011.
- [4] A. Schwaninger, S. Schumacher, H. Bülthoff, and C. Wallraven, "Using 3D computer graphics for perception: the role of local and global information in face processing," in *Proceedings of 4th Symposium on Applied Perception in Graphics and Visualization*, pp. 19–26, Tübingen, Germany, July 2007.
- [5] A. Schwaninger, C. Wallraven, D. W. Cunningham, and S. D. Chiller-Glaus, "Processing of facial identity and expression: a psychophysical, physiological, and computational perspective," *Understanding Emotions*, vol. 156, pp. 321–343, 2006.
- [6] D. Maurer, R. L. Grand, and C. J. Mondloch, "The many faces of configural processing," *Trends in Cognitive Sciences*, vol. 6, no. 6, pp. 255–260, 2002.
- [7] N. Sagiv and S. Bentin, "Structural encoding of human and schematic faces: holistic and part-based processes," *Journal of Cognitive Neuroscience*, vol. 13, no. 7, pp. 937–951, 2001.
- [8] M. L. Matthews, "Discrimination of identikit constructions of faces: evidence for a dual processing strategy," *Perception and Psychophysics*, vol. 23, no. 2, pp. 153–161, 1978.
- [9] E. E. Smith and G. D. Nielsen, "Representations and retrieval processes in short-term memory: recognition and recall of faces," *Journal of Experimental Psychology*, vol. 85, no. 3, pp. 397–405, 1970.
- [10] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 177–190, 2013.
- [11] R. Dewi Agushinta, A. Suhendra, and Y. Hanum, "Facial feature distance extraction as a face recognition system component," *ICSITT*, vol. 2007, p. 239, 2007.
- [12] G. Davies, H. Ellis, and J. Shepherd, "Cue saliency in faces as assessed by the "photofit" technique," *Perception*, vol. 6, no. 3, pp. 263–269, 2016.
- [13] M. J. Farah, K. D. Wilson, M. Drain, and J. N. Tanaka, "What is "special" about face perception?," *Psychological Review*, vol. 105, no. 3, pp. 482–498, 1998.
- [14] J. J. Richler, O. S. Cheung, and I. Gauthier, "Holistic processing predicts face recognition," *Psychological Science*, vol. 22, no. 4, pp. 464–471, 2011.
- [15] J. M. Gold, P. J. Mundy, and B. S. Tjan, "The perception of a face is no more than the sum of its parts," *Psychological Science*, vol. 23, no. 4, pp. 427–434, 2012.
- [16] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.
- [17] B. Heisele, T. Serre, and T. Poggio, "A component-based framework for face detection and identification," *International Journal of Computer Vision*, vol. 74, no. 2, pp. 167–181, 2006.
- [18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, December 2001.
- [19] A. M. Martinez and A. C. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [20] M. H. Yang, "Face recognition using kernel methods," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1457–1464, Vancouver, Canada, December 2002.
- [21] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, Seattle, WA, USA, June 1994.
- [22] O. N. Dagnes, E. Vezzetti, F. Marcolin, and S. Tornincasa, "Occlusion detection and restoration techniques for 3D face recognition: a literature review," *Machine Vision and Applications*, vol. 29, no. 5, pp. 789–813, 2018.
- [23] F. Marcolin, F. Marcolin, S. Tornincasa, L. Ulrich, and N. Dagnes, "3D geometry-based automatic landmark localization in presence of facial occlusions," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 14177–14205, 2017.
- [24] V. Kazemi and S. Josephine, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 1867–1874, Columbus, USA, June 2014.
- [25] A. ElMaghraby, M. Abdalla, O. Enany, and M. Y. E. Nahas, "Detect and analyze face parts information using viola- Jones and geometric approaches," *International Journal of Computer Applications*, vol. 101, no. 3, pp. 23–28, 2014.
- [26] G. Tsai, *Histogram of Oriented Gradients*, University of Michigan, Vol. 1, University of Michigan, Ann Arbor, MI, USA, 2010.
- [27] A. Kasinski, A. Florek, and A. Schmidt, "The PUT face database," *Image Processing and Communications*, vol. 13, no. 3-4, pp. 59–64, 2008.
- [28] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of Second IEEE Workshop on Applications of Computer Vision*, pp. 138–142, Sarasota, FL, USA, December 1994.
- [29] A. M. Martinez, "The AR face database. CVC technical report," Tech. Rep. 24, Centre de Visió per Computador (CVC), Barcelona, Spain, 1998.



Hindawi

Submit your manuscripts at
www.hindawi.com

