Hybrid System for Robust Faces Detection

Hayet Farida Merouani and Amir Benzaoui

Abstract—**The automatic detection of faces is a very important problem. The effectiveness of biometric authentication based on face mainly depends on the method used to locate the face in the image. This paper presents a hybrid system for faces detection in unconstrained cases in which the illumination, pose, occlusion, and size of the face are uncontrolled. To do this, the new method of detection proposed in this paper is based primarily on a technique of automatic learning by using the decision of three neural networks, a technique of energy compaction by using the discrete cosine transform, and a technique of segmentation by the color of human skin. A whole of pictures (faces and no faces) are transformed to vectors of data which will be used for learning the neural networks to separate between the two classes. Discrete cosine transform is used to reduce the dimension of the vectors, to eliminate the redundancies of information, and to store only the useful information in a minimum number of coefficients while the segmentation is used to reduce the space of research in the image. The experimental results have shown that this hybridization of methods will give a very significant improvement of the rate of the recognition, quality of detection, and the time of execution.**

Index Terms—**Energy compaction, face detection, face recognition, neural networks.**

1. Introduction

The problem of the face detection is considered in any biometric system based on the face as a mean for identification. This problem is becoming a popular area of research due to its emerging applications in human computer interface, surveillance systems, secure access control, video conferencing, financial transaction, image database management system, and so on.

Several methods of face detection have appeared in the two last decades, Yang *et al.*[1] classified these methods into four classes: solutions based on knowledge, solutions based on invariant characteristics, solutions based on the mapping, and solutions based on techniques of automatic learning. For the first approach, the construction of the system is based on geometrical rules to seek candidates and each candidate is determined by the face similitude rather than the face edge characteristic. For the second approach, solutions aim at finding the face structural characteristics[2] (form, texture, color of skin, edge, and so on) which are invariant when the conditions of image such as position, point of view, etc. vary, and then using these characteristics to find the face. The third approach includes a solution that identifies faces by comparing each part of the image with a pattern of face[3]. On the contrary, the methods based on automatic learning techniques create models during the learning and then use them to find faces[4],[5]. However, the systems of the human vision have a good effectiveness because they have a solutions provided naturally by the human brain whereas the computerized solutions of vision do not present a great resemblance[6].

The objective of the proposed system aims at improving the performances and the quality of detection in a real context and fills the gap between the computerized models of the vision and their human counterparts. With this intention, we have proposed a hybrid system of faces detection in a color image, which is based primarily on a technique of automatic learning by using the decision of several neuronal networks, the learning is carried out starting from two bases of pictures (face, not face) which represent the two classes to be separated. By learning at the frequency level, this system is characterized by using the discrete cosine transform (DCT), which is a technique which has a very great effectiveness in term of energy compaction. We have profited the advantage from this transformation in order to reduce the dimension of the vectors and eliminate a very great quantity of useless information that makes it possible to gradually improve quality of the learning.

This paper includes 5 sections. After a general introduction in Section 1, we illustrate the principle of energy compaction by the DCT in Section 2. The working procedure of the system proposed is detailed in Section 3. The experimental results are exposed in Section 4. Finally,

Manuscript received April 28, 2012; revised May 24, 2012.

H. F. Merouani is with the Laboratory of Computer Research, Department of Computer Sciences, University of Badji Mokhtar, B.P. 12 Sidi Amar 23000, Annaba, Algéria (e-mail: hayet.merouani@univ-annaba.org).

A. Benzaoui is with the Laboratory of Inverses Problems, Modeling, Information and Systems, Department of Electronic and Telecommunication, University of 08 Mai 1945, P.O. box 401 Guelma 24000, Algéria (e-mail: amirbenzaoui@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://www.intl-jest.com/.

Digital Object Identifier: 10.3969/j.issn.1674-862X.2012.02.014

overall conclusion is reviewed and potential future work is discussed.

2. Energy Compaction

The time of decision and the precision for any system of pattern recognition are in direct relation with the vectors of data used to carry out learning. The dimension and the correlation of the vectors have a direct effect on quality of the learning and the time of detection. Indeed, the relevance of information is not easy to appreciate in the direct domain; therefore one seeks to carry out changes of representation which make it possible to facilitate the separation of the most relevant information. This approach is based on orthogonal transformations, which replace the origin signal by its representation according to a base of functions.

Several transformations are used in this kind of problems; most widespread are discrete Fourier transform (DFT), Karhunen-Loève transform (KLT), and discrete cosine transform (DCT) which will be used in the design of our system. KLT is recognized by its precision and its exactitude but it presents a major disadvantage because of the complexity of its calculations. DFT is a very fast method thanks to simplicity of its algorithms but its periodicity horizontally and vertically causes its inaccuracy in certain case. DCT, a simple, fast, and effective algorithm[7], exhibits a compromise between the precision and speed; i.e., the exactitude of KLT and the speed of DFT.

DCT is a mathematical transformation which transforms a set of data from a spatial domain to a frequency spectrum[8]. In our case, we study an image, i.e., a two-dimension function with x and y indicating the spatial coordinates of a pixel and the output value (coefficient) in this pixel. The DCT of a discrete signal with two dimensions is defined as follows[9]:

$$C(u,v)=\frac{2}{N}C(u)C(v)\times \sum_{x}^{N-1}\sum_{y}^{N-1} f(x,y)\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)v\pi}{2N}$$

where

$$C(u)C(v)=\begin{cases}1/\sqrt{2}, & \text{for } u,v=0\\ 1, & \text{otherwise.}\end{cases}$$

DCT has a very great effectiveness in terms of energy compaction. It can store a maximum amount of information in a minimum amount of coefficients located low frequency (in the top left corner) and there will be a degradation of information when down towards high frequencies (in the lower right corner)[10], as shown in Fig. 1. This makes it possible to eliminate the coefficients having relatively small amplitudes without the deformation of the characteristics of the image.

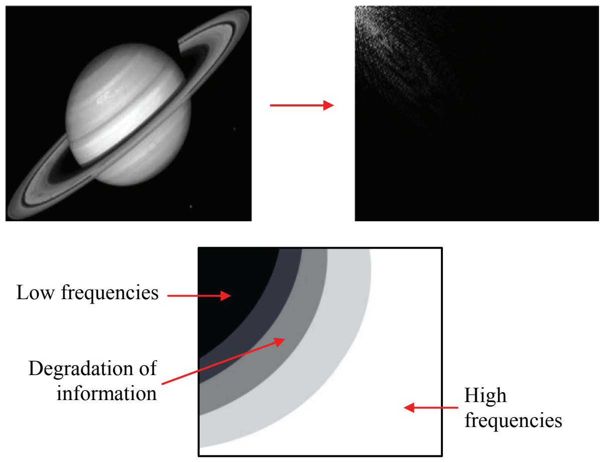


Fig. 1. Frequency distribution of DCT coefficients[10].

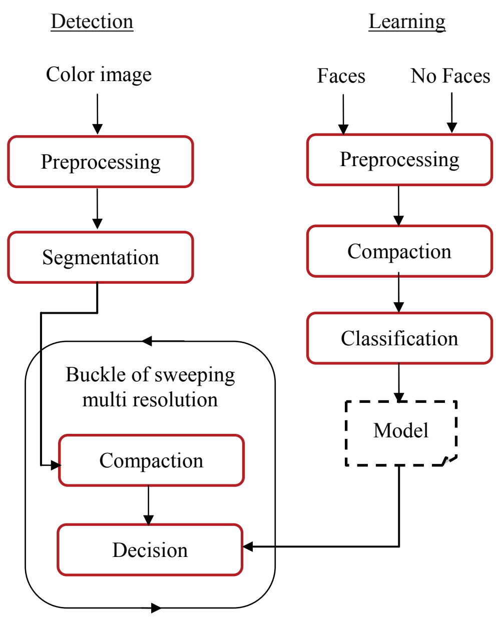


Fig. 2. Block diagram of the proposed system.

3. Structure of the Proposed System

The human face is a dynamic entity which changes constantly under the influence of several factors. In this paper, the proposed detection system takes into account all factors which contribute to the complexity of its task.

The general structure of the proposed system comprises two phases: a phase of learning and a phase of detection, this structure is shown in Fig. 2.

In the learning mode, the system must have in entry two types of pictures which represent the two classes to be separated. These pictures are standardized at the size of 19×19 pixels. One type contains the pictures of human faces, which are different from each other according to each person and the position of the face (inclined, right, rotation etc.), and these pictures are taken under all possible conditions (luminosity, lighting, so on). The other type contains pictures of objects or part of faces, considered as counter examples. The system will analyze the problem, extract the data, classify the pictures, and create a model of learning (function of classification) according to the parameters which will give the best of results. Next the model will be stored for the use at the phase of detection.

In the detection mode, the system receives in entry a color image. The image will be pretreated and swept by a frame of fixed size. Then a function of decision is used to decide if the frame contains a face or not.

We will explain the working procedure of the system by presenting the most significant modules.

3.1 Preprocessing

The objective of the preprocessing module is the modification of the representation of the source image to facilitate the work of the following modules and to improve the quality of detection. The source image is a color image, it is first converted into gray scale image and then undergoes a whole of preprocessing. The preprocessing is a process of the modification of the histogram (normalization and stretching): improving the contrast, adjusting the luminosity, and decreasing the noise resulting from the original image.

Various sources of noise may exist in the input image. The fine details of the image represent high frequencies which mix up with those of noise. So low-pass filters are used to obliterate some details in the image. In this experiment, median filter is used to suppress the noise.

3.2 Compaction

The objective of this module is to recover the picture/image from the gray scale image (in matrix form) after preprocessing. And then DCT is applied in order to extract a vector of coefficients which represents the useful information of each image. This operation is carried out in several phases, as shown in Fig. 3.

The matrix of gray scale after preprocessing is converted into several blocks of $N \times N$ pixels, generally N is equal to 8, and this value is selected because it is a good compromise between the quality of the application of the DCT and the speed of calculations. The DCT is then applied for each block which ensures that the amplitudes of the lows frequencies (relevant information of the image) are higher than that of the components of highs frequencies (detail of the image). The coefficients of smaller amplitude can be eliminated and the image is represented by a small number of coefficients. Finally, the selection of the coefficients of low frequencies is carried out by a zigzag sweeping, as shown in Fig. 4. The zigzag method allows the recovery of data in an order descending from energy. It consists in traversing the elements of the left higher part of the transformed matrix, in precise order, starting from the lowest frequencies towards the highest, as shown in Fig. 4.

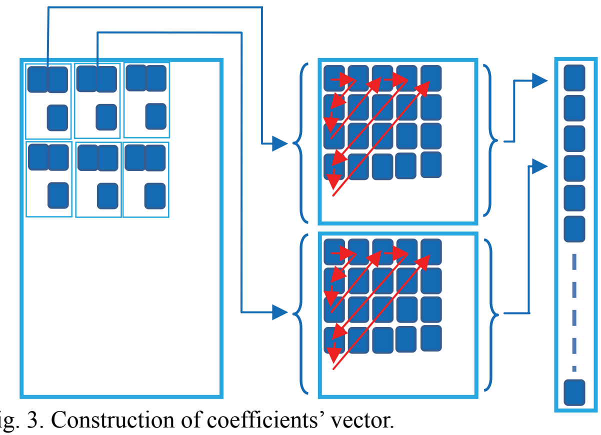


Fig. 3. Construction of coefficients' vector.

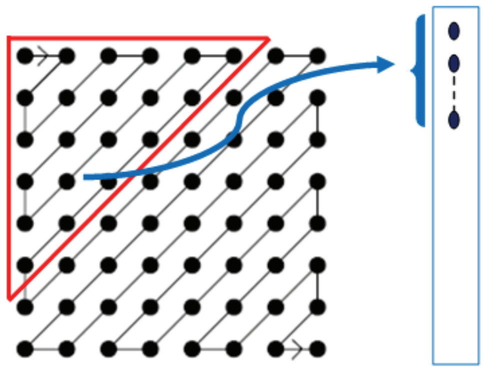


Fig. 4. Traversing in Zigzag and extraction of a block.

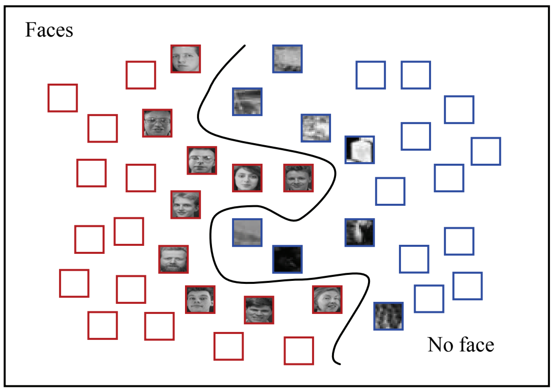


Fig. 5. Classification.

3.3 Classification and Decision

The task of the classification module is to employ the vectors of coefficients provided by the preceding stage (energy compaction) to assign the objects of interest to a category or a class (positive or negative), as shown in Fig. 5. Our system employs an artificial neurological network as a classifier. The neural network is a processor massively distributed parallel which has a natural provision to store empirical knowledge and to make it available to be used[11].

Among several types of neuronal networks, we chose the most common model of neural network which is the perceptron multi-layer (MLP), the MLP learns from the samples of entry and the values of corresponding exit by changing the synaptic weights between connections. To analyze a given problem of classification by using the MLP, it is necessary to train it; we update the weights of the neural network until the error of detection is lower than some threshold. The algorithm of learning used in this module is the back propagation of gradient. In order to render the system more effective, i.e., to reduce the number of false detections, three neuronal networks have been used to classify and choose, according to their answers, those which must be preserved or eliminated. The performances of each MLP depend on the initialization of the weights, the rate of learning, and the number of neurons of the hidden layer; this is why we have carried out several tests before preserving a good topology giving the best performances (the training is stopped when the average quadratic error is < 0.001).

After the phase of learning and the safeguard of the weights (model), a phase of evaluation comes to validate this learning. Here we test the networks by using other pictures which are not used for the learning, this test will try to classify these pictures but without changing the weights of each network. At the end, we will obtain a rate of success (rate of recognition) and starting from this rate we validate the learning by comparing it with a fixed threshold (e.g. 85%). The rate of recognition is defined as the relationship between the numbers of correctly recognized examples and the number of examples presented at the entry of the system:

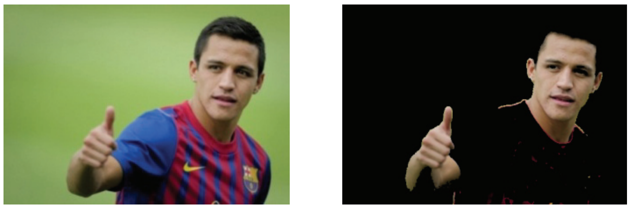
$$\text{Rate} = \frac{\text{good decision}}{\text{total number of examples}} \times 100 .$$

3.4 Segmentation

To accelerate the process of detection, we have used a technique of segmentation by the color of human skin; this technique is largely used and proved effective in several applications such as the detection of human movement and the video monitoring. It is advantageous to use this additional information to isolate the areas which contain the faces. Indeed, the elimination of the areas not having the color of the skin reduces the number of frames which are sent to the two modules (compaction and decision).

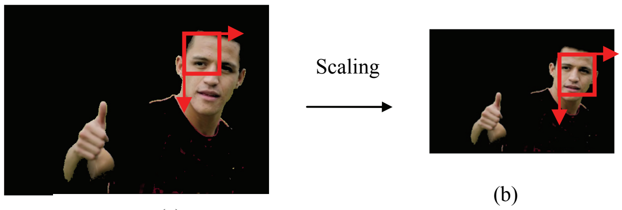
Although, the color of human skin can widely vary, recent research shows that the principal difference is in the intensity rather than in the chrominance[12]. Several spaces of color are used to label the pixels of color of human skin: RGB, RGB standardized, HSV, YCrCb, YIQ, etc.[12]. Moreover, there are methods proposed to build a model of color of human skin: methods using the tonality of pixel, non-parametric methods based on the histogram, and methods parametric using a Gaussian function[13].

For our work, a simple, effective, robust, and real time method to segment the blobs color of human skin is the good choice. For this, we chose a method of segmentation by thresholding using the space color RGB [13]. The thresholding of this space is applied to the three channels (Red, Green,



(a) (b)

Fig. 6. Example of segmentation by human skin color: (a) original image and (b) segmented image.



(a) (b)

Fig. 7. Sweeping multi resolution: (a) sweeping of the image and (b) sweeping of the reduced image.

and Blue), and the whole of the pixels of the image is classified in two classes, color skin and not-color skin, according to the rule of following classification[13]:

$$(R>95),\ (G>40),\ (B>20)$$
$$\left[\max(R,G,B)-\min(R,G,B)\right]>15$$
$$\text{ABS}(R-G)>15,\ (R>G),\ (R>B) .$$

At the end of treatment, one will obtain an image where only the pixels appear. The conditions fixed for the color of human skin are then applied. An example of human skin detection based on this method is shown in Fig. 6. We have found also that this method is very effective in detecting human skin regions.

3.5 Sweeping Multi Resolution

The required objective of the system is the detection of the faces in the image. To do this, we have defined a frame of fixed size of 19×19 pixels, which traverses in all the areas of the colors of human skin in an image to find the faces. This window has a fixed size to serve the data entering to the classifier. Also, the detection of faces will be able to take place only on images of a minimal size of 19×19 pixels. Sweeping will start on an image with its initial size, and then the image will be successively reduced to each scaling in order to be able to recover faces more or less close to this image, as shown in Fig. 7.

4. Experimental Results

To build data for training the system, we have used the data base “MIT CBCL”, witch is used for the learning and the test of the system. This database is composed of 2 parts: the pictures of the learning and the pictures of the test. The first part is composed of 6977 pictures (2429 faces, 4548 no faces) and the second part is composed of 24045 pictures (472 faces, 23573 no faces). These pictures are all on the level of gray and have a fixed size at 19×19 of pixel. Table 1 contains the best results obtained during the evaluation of the system, and that is done by successive tests carried out according to several configurations of the neural networks and using the two following modes:

1) MLP: the classifier MLP will be used only without energy compaction.

2) MLP-DCT: the classifier will be used with energy compaction by using the DCT.

Table 1: Results obtained during the evaluation

Classifier	Rate of recognition	Rate of false acceptance
MLP	70.03	7.03
MLP-DCT	86.33	2.11

According to the results obtained, we have deduced that the use of the method of energy compaction associated with the neural networks is essential to improve the performances of the detector. Because presenting raw data to the classifier without any transformation is not effective. Therefore, the realization of the classifier by the MLP-DCT method shows good and effective results.

The method proposed in this paper has several advantages compared with other methods of detection, our method is characterized by:

1) Real time detection: reduction of the space of research by using a technique of segmentation by the color of human skin.

2) Robust method (in lighting, occlusion, position, etc.): the great capacity of generalization of the neural networks.

3) Precision: improvement of the quality of the learning by the reduction and the selection of the most relevant parameters by using a technique of energy compaction.

Fig. 8. Examples of images' tested and face detection results.

Starting from the best configuration which we have obtained in the MLP-DCT mode, we also carried out some tests on images which we have taken by a personal camera and others downloaded on the Web, as shown in Fig. 8. The tests are carried out on faces which are presented in various colors and various orientations and scales.

5. Conclusions

In this paper, we have presented a hybrid system for human faces detection in unconstrained cases. The objective of our system aims to improve the performances and the qualities of detection in a real context and fill the gap between the computerized models of the vision and their human counterparts. The new proposed method in this system is primarily based on a technique of automatic learning by using the decision of three neural networks. Our method is characterized by learning at the frequency level by using the discrete cosine transform as a technique of energy compaction, which makes it possible to store a maximum of information in a minimum of coefficients located in the lows frequencies and to eliminate the coefficients having relatively small amplitudes without presenting a deformation of the characteristics of the image. Also, the technique of segmentation by the color of human skin has been used to reduce the space of research at the image. The experimental results showed that the new proposed method gave a very significant improvement at the rate of recognition, quality of detection, and the time of execution compared with other methods of detection.

One of the prospects for this work is to learn the system with other databases that presents strong variations in the light and the position of the head. Also, we can consider the possibility of employing it for an approach based on the characteristic elements of the face.

Acknowledgment

This work was supported by the Laboratory of Inverses Problems, Modeling, Information and Systems (PI:MIS), Department of Electronic and Telecommunication, University of 08 Mai 1945, Guelma, Algéria and the Laboratory of Computer Research (LRI), Department of Computer Sciences, University of Badji Mokhtar, Annaba, Algéria.

References

[1] X. D. Yang, P. Kort, and R. Dosselmann, "Automatically log off upon disappearance of facial image," Contract Report CR 2005-051, DRDC, Ottawa, Canada, May 2005.

[2] H.-Y. Chen, C.-L. Huang, and C.-M. Fu, "Hybrid-boost learning for multi-pose face detection and facial expression recognition," *Pattern Recognition*, vol. 41, no.3, pp. 1173–1185, 2008.

[3] S. Phimoltares, C. Lursinsap, and K. Chamnongthai, "Face detection and facial feature localization without considering

the appearance of image context," *Image and Vision Computing*, vol. 25, no. 5, pp. 741–753, 2007.

[4] H. Bae and S. Kim, "Real time face detection and recognition using hybrid information extracted from face space and facial features," *Image and Vision Computing*, vol. 23, pp. 1181–1191, Nov. 2005.

[5] I. Kukenys and B. McCane, "Support vector machines for human face detection, in *Proc. of 2008 New Zealand Computer Science Research Student Conf.*, Christchurch, 2008, pp. 226–229.

[6] Z. Li Stan and A. K. Jain, *Handbook of Face Recognition*, 1st ed. New York: Springer, 2004.

[7] S. A. Khayem, "The discete cosine transform: theory and application," Tutorial Report: WAVES-TR-ECE802.602, Michigan State University, March 2003.

[8] Ken Cabeen and Peter Gent. Image compressing and the discrete cosine transform [Online]. Available: http://online.redwoods.cc.ca.us/instruct/darnold/laproj/fall98/pken/dct.pdf

[9] R. B. Dubey and R. Gupta. High quality image compression [Online]. Available: http://www.webmedcentral.com/wmcpdf/Article_WMC001673.pdf

[10] A. M. Vijaya Prakash and K. S. Gurumurthy, "A novel VLSI architecture for digital image compression using discrete cosine transform and quantization," *Int. Journal of Computer Science and Network Security*, vol. 10, no. 9, pp. 175–182, 2010.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New York: Macmillan, 1994.

[12] T. P. Nguyen, "Détection de visage dans images de couleur en utilisant une caractéristique invariante," presented at the 3rd Int. Conf. on Sciences of Electronic, Technologies of Information and Telecommunications, Sousa, Tunisie, 2005.

[13] B. Chemesse-Ennehar, B. Abd El halim, and T. Hicham, "Segmentation de la Couleur de Peau par Seuillage Selon Différents Espaces de Couleur," in *Proc. of 3èmes Journées Internationales sur l'Informatique Graphique*, Constantine, 2007, pp. 207–211.

Hayet Farida Merouani received her Eng. degree from Annaba University, Algeria in 1984, and the Ph.D. degree from Robert Gordon University, Aberdeen, UK. She is currently an associate professor with Badji Mokhtar University, Annaba. She also leads Research Group of Pattern Recognition. Her current work focuses on the computer vision, medical imaging, and biometry.

Amir Benzaoui was born in Annaba, Algéria in 1988. He received his Master degree in image processing and artificial vision from the Department of Computer Sciences, Badji Mokhtar University, Annaba, Algeria in 2011. He is currently pursuing the doctorate degree with the Department of Electronic and Telecommunication, 08 Mai 1945 University, Guelma, Algeria. His research interests include images and signals processing, biometric and security of individuals, pattern recognition, and inverses problems.