International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 4, August 2017, pp. 1915~1922
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i4.pp1915-1922 ❒ 1915

Unimodal Multi-Feature Fusion and one-dimensional Hidden Markov Models for Low-Resolution Face Recognition

Othmane El Meslouhi[1], Zineb Elgarrai[2], Mustapha Kardouchi[3], Hakim Allali[4]
[1,4]Lab.SIE FPO of Ibn Zohr University Agadir, Morocco
[2]FST of Hassan 1st University Settat, Morocco
[3]Département d'Informatique of Université de Moncton, Canada

Article Info

Article history:

Received Jan 6, 2017
Revised Mar 10, 2017
Accepted Mar 24, 2017

Keyword:

1D-HMMs
Canonical correlation analysis
Face recognition
Interpolation
Low-resolution

ABSTRACT

The objective of low-resolution face recognition is to identify faces from small size or poor quality images with varying pose, illumination, expression, etc. In this work, we propose a robust low face recognition technique based on one-dimensional Hidden Markov Models. Features of each facial image are extracted using three steps: firstly, both Gabor filters and Histogram of Oriented Gradients (HOG) descriptor are calculated. Secondly, the size of these features is reduced using the Linear Discriminant Analysis (LDA) method in order to remove redundant information. Finally, the reduced features are combined using Canonical Correlation Analysis (CCA) method. Unlike existing techniques using HMMs, in which authors consider each state to represent one facial region (eyes, nose, mouth, etc), the proposed system employs 1D-HMMs without any prior knowledge about the localization of interest regions in the facial image. Performance of the proposed method will be measured using the AR database.

Corresponding Author:

Othmane El Meslouhi,
Lab.SIE FPO of Ibn Zohr University Agadir,
Morocco.
Email: o.elmeslouhi@uiz.ac.ma

1. INTRODUCTION

Face recognition (FR) is the process of automatically identifying or verifying a person from his facial image. In real life application, such as, video surveillance cameras (placed in banks, supermarkets or in public streets), subjects are far from cameras, so face regions tend to be small. Machine face identification from this kind of images becomes an important scientific issue to solve. This issue is called low-resolution face recognition.

Like classical face recognition systems, low-resolution face recognition procedure consists of 2 major tasks: features extraction and classifier designing. Both have a significant impact on the reliability of recognition method.

However, a low-resolution face recognition system needs to consider the particular problem of dimensional mismatch between probe (with low-resloution) and gallery (generally with high-resolution) images before extracting face features. This issue can be resolved by using three main ways: (1) Applying interpolation or super-resolution to the probe facial images, (2) projecting both probe and gallery facial images into an inter-resolution space or (3) downsampling all gallery images and then perform matching in low resolution. An overview of this methods is given in [1].

For the features extraction and classification tasks, various approaches have been proposed such as: RQCr color features with nearest neighbor classifier [2], Local Frequency Descriptor with nearest neighbor classifier [3] and Support Vector Data Description with support vector machines (SVM) classifier [4].

In this paper, we use the cubic interpolation to increase the special resolution of LR images to deal with the dimensional mismatch problem. This choice is justified by the fact that it's extremely fast and it can be applied on a single image unlike super-resolution techniques which use gallery set of LR or HR images. Regarding the facial features extraction, the proposed system extract facial features by fusing two local descriptors: Gabor wavelets [5], [6] and HOG descriptors [7]. Gabor wavelets (filters) are invariant to rotation, scale and translation. In addition, they are robust to image noise. Also, Histograms of Oriented Gradients (HOG) are robust against illumination variations [8]. Their complementary nature makes them good candidate for combination. Concerning the classification tasks, we use one dimensional Hidden Markov Model to deal with the classification task. Our approach differs from existing methods using HMMs in that it doesn't need any prior knowledge about the localization of interest regions in the facial images. This advantage makes our approach fully automatic and robust even the presence of no frontal facial images.

The following section gives a detailed description of the proposed system. Experimental results are presented in section 3. Finally, we conclude this paper in section 4.

2. COMBINING GABOR AND HOG FEATURES FOR HIDDEN MARKOV MODEL RECOGNITION

This section describes the components of our face recognition system in detail: Gabor and HOG features, LDA dimensionality reduction, feature fusion using the Canonical Correlation Analysis (CCA) method and Hidden Markov Model for recognition. The stages of processing are diagrammed in Figure 1.

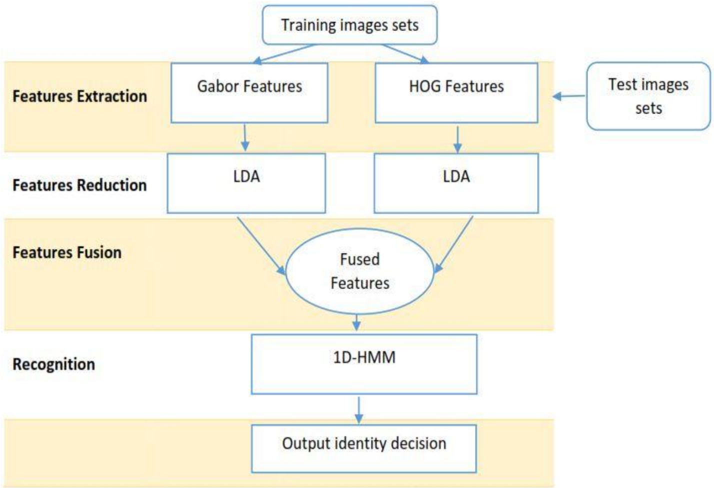

Figure 1. The overall of the proposed face recognition system.

Features extraction:

a. Gabor Features Representation: Gabor wavelet representation of face images derives desirable features gained by spatial frequency, spatial locality, and orientation selectivity. These discriminative features extracted from the Gabor filtered images could be robust to illumination and facial expression changes. A Gabor wavelet filter is a Gaussian kernel function modulated by a sinusoidal plane wave [5]:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{\frac{\|k_{u,v}\|^2 \|z\|^2}{2.\sigma^2}} . \left[e^{i.k_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

Where $z = (x, y)$ is the pixel with coordinate (x, y) in the image plan. u and v define orientation and scale of the Gabor kernels. $\|.\|$ denotes the norm operator.

A large number of local features can be generated by varying parameters associated with the position, scale, and orientation of the filters. For example, the magnitude response of the convolution of an image with 40 banks of Gabor kernels (8 orientations and 5 scales) is 40 magnitude maps in the same size as the original image, as illustrated in the Figure 2.

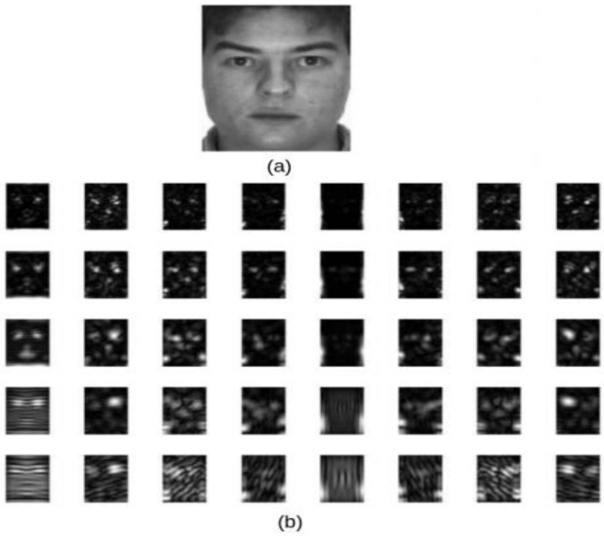

Figure 2. An example of the Gabor magnitude output: (a) the initial image (b) the magnitude output of the filtering operation with a bank of 40 Gabor filters.

b. Histogram of Oriented Gradients Features (HOG): The HOG descriptor is a local statistic of the orientations of the image gradients. It is characterized by its invariance to rotation and illumination changes. The HOG feature divides the image into many cells, in each of them a histogram counts the occurrences of pixels orientations given by their gradients. The final HOG descriptor is then built with combination of these histograms [7]. Figure 3 shows an example of facial image with extracted HOG descriptor.

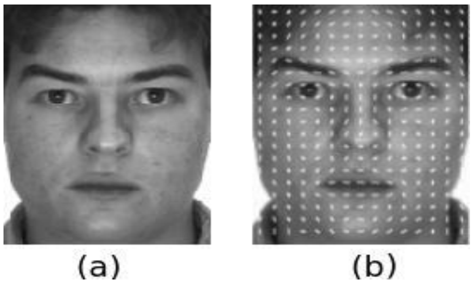

Figure 3. An example of HOG descriptor from a facial image: (a) Facial image, (b) Facial image with HOG descriptor.

c. Feature Reduction: The dimensions of Gabor and HOG feature vectors are too high, which is far too extensive for efficient processing and storage. To overcome this issue, many techniques are proposed in the literature: Principal Component Analysis (PCA) [9], Independent Component Analysis (ICA) [10], Linear Discriminant Analysis (LDA) [11-13]. In this paper, the employed dimensionality reduction technique is the Linear Discriminant Analysis (LDA). It aims at finding a feature representation by which the within-class distance is minimized and the between-class distance is maximized [14]. In order to avoid singularity issues, when computing the inverse of the within-class scatter matrix, the LDA reduction method is implemented in the Principal Component Analysis (PCA) subspace as suggested in [15].

d. Feature fusion using canonical correlation analysis: In this stage, we combine the two reduced feature vectors to obtain a single one, which is more discriminative than using only one feature modality. This is achieved by using a feature fusion technique based on Canonical Correlation Analysis (CCA) [16]. Canonical correlation analysis has been widely used to analyze associations between two sets of variables.

Given two column vectors $U = (u_1, \dots, u_n)'$ and $V = (v_1, \dots, v_n)'$ of random variables with finite second moments, one may define the cross-covariance $\sum_{UV} = cov(U, V)$ to be the $n \times m$ matrix whose (i, j) entry is the covariance $cov = (u_i, v_j)$.

Canonical-correlation analysis attempt to find vectors x and b such that the random variables a' U and b' V maximize the correlation ρ = $corr(a'$ U,b' V) = $corr(Z_1,\ Z_2)$. Z_1 and Z_2 are defined by [8]:

$$Z_1 = c' \sum_{UU}^{-1/2} U \tag{2}$$

$$Z_2 = d' \sum_{VV}^{-1/2} V \tag{3}$$

Where, $\sum_{UU} = cov(U, U)$, $\sum_{VV} = cov(V, V)$, $c = \sum_{UU}^{1/2} a$ and $d = \sum_{VV}^{1/2} b$.
After some steps of calculation, the solution is therefore:
1) a is the eigenvector with the maximum eigenvalue for the matrix $\sum_{UU}^{-1} \sum_{UV} \sum_{VV}^{-1} \sum_{VU}$.
2) b is the eigenvector with the maximum eigenvalue for the matrix $\sum_{VV}^{-1} \sum_{VU} \sum_{UU}^{-1} \sum_{UV}$.

e. Feature classification: An HMM, which was developed by Baum and Petrie [17], is a probabilistic model in which the system is assumed to be a Markov process with hidden states. HMM is used for representing probability distribution over sequences of observations. The following key elements characterize an HMM:
a. N the number of states of the model.
b. A is an $N \times N$ state transition matrix that gives the state transition probabilities between N states.
c. B is an $k \times N$ emission probability matrix while being in a particular state.
d. $\prod$ is a $1 \times N$ matrix, called initial state probability matrix, and it gives the probability of being in a particular state at the start of the process.

As mentioned above, a left to right 1D Hidden Markov Model is used in our system to achieve the recognition task. The Figure 4 shows an example of a Left-Right 1D-HMM with N states.

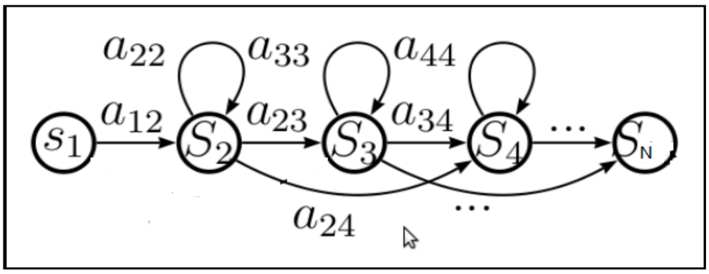

Figure 4. An example of a Left-Right 1D-HMM with N states.

In our system, the number of states N depends to the dataset subject number. The entry state 1 and the exit state N are non-emitting. The transition matrix A will have N rows and N columns. Finally we assume that each observation probability distribution is represented with single Gaussian distributions. All these parameters will be estimated during the training step using an iterative procedure, known as Baum-Welch algorithm [18]. Also, each subject will be represented with one model. Finally, the recognition step is achieved using the Viterbi algorithm [18] to find the highest likelihood score between features in the testing sets and models obtained in the training step.

3. EXPERIMENTAL RESULTS

The experiment was designed with the aim of validating the effectiveness of the proposed system. We performed experiments using the cropped AR face database [19]. The AR face database contains over 3,200 frontal color face images of 126 subjects (26 different images for each person), including different facial expressions, with various occlusions and under different lighting conditions. Most of the pictures were recorded in two sessions (separated by two weeks). All images were taken by the same camera under tightly controlled conditions of illumination and viewpoint. For our experiments, like in the work of [19], in the first time, 100 different subjects (50 males and 50 females) were randomly selected from this database, then, all selected images were segmented using an oval-shaped mask and finally all color images are transformed into gray images. Figure 5 shows some sample images of one subject extracted from the obtained face database.

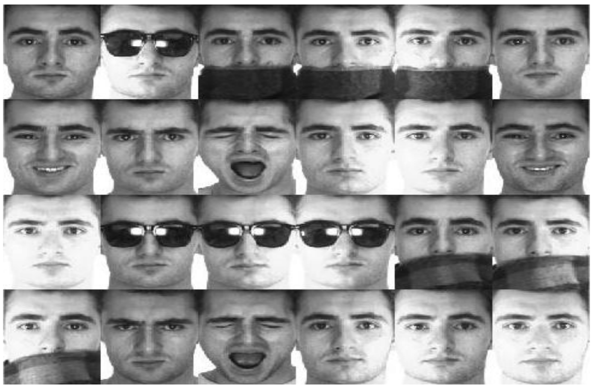

Figure 5. Images of one subject in the AR face database.

In this experiments, we investigate the face recognition as a function of the image resolution. Instead of taking images from faces which have different distances to the camera, we have tried to simulate the effect of lowering the resolution. From original pictures with the resolution of 165×120 pixels, all of pictures have been downsampled 2, 4, 8, 16 and even 32 times. It corresponds to resolutions 83×63, 42×30, 21×15, 11×8, and 6×4 respectively. Since extracting features from facial with lowest resolution would lead to serious quantization errors, we decided to upscale all images to the same resolution. After downscaling, we scaled the image up to original size (i.e 165×120) for face recognition using a cubic interpolation [20]. The results can be seen in Figure 6.

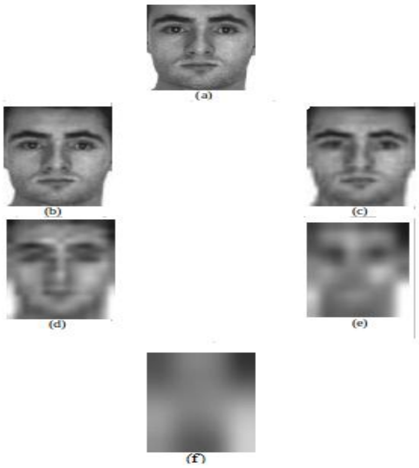

Figure 6. Samples for downsampled face image: (a) the original face image165×120. (b) downsampling the image in a by a factor of 1/2 to the size 83×63. (c) by factor of 1/4 to the size 42×30. (d) by factor of 1/8 to the size 21×15. (e) by factor of 1/16 to the size 11×8. (f) by factor of 1/32 to the size 6×4.

The experimental results presented in this section are divided into three parts. The aim of the first set of experiments is to see whether the system converges quickly or not. Recognition rates given at different epochs are plotted in Fgure 7 showing that the system converges at the end of the second epoch.

In second experiment, we make a comparison between the combined Gabor+HOG descriptor with isolated descriptors Gabor, HOG and LBP (Local Binary Patterns) for different resolution. We notice that the features size of all those descriptors is reduced using LDA method. The comparison results are shown in Table 1.

From this table, we notice that the results obtained by the combined Gabor+HOG features outperform those obtained using separated Gabor, HOG and LBP descriptors in AR database for all facial image resolution.

Finally, as shown in Table 2, Table 3, Table 4 and Table 5, we calculated recognition rate according to the size of training image set and the resolution obtained by the 1D-HMM classifier and other three classifiers: Support Vector Machine (SVM) with Radial Basic Function (RDF) Kernel, AdaBoost and Decision Tree Classifiers. From these tables, we can see that the accuracies obtained by the 1D-HMM classifier, in general, outperform the SVM, AdaBoost and Decision Tree Classifiers.

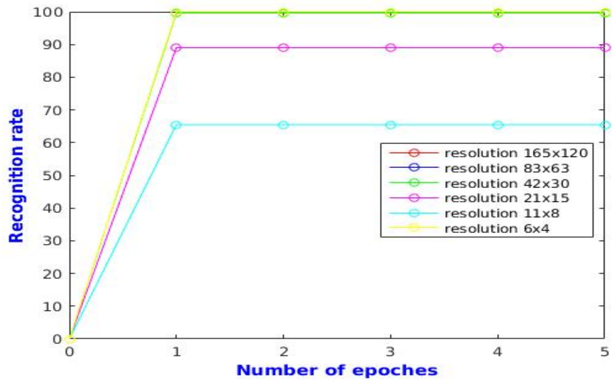

Figure 7. Recognition Rate vs. The number of training epochs.

Table 1. Best face recognition rates using different descriptors with different image resolution

	Experimental Datasets					
Descriptors	6 × 4	11 × 8	21 × 15	42 × 30	83 × 63	165 × 120
LBP	12.03%	50%	77.26%	92.23%	91.35%	95.43%
HOG	22.45%	55.14%	88.5%	98.89%	99.81%	99.81%
Gabor	13.36%	53.23%	59.37%	94.95%	99.44%	100%
Gabor+HOG	27.64%	65.49%	89.05%	99.63%	100%	100%

Table 2. Face recognition rates of four classifiers with different Image resolution using 5 training images

	Experimental Datasets					
Classifiers	6 × 4	11 × 8	21 × 15	42 × 30	83 × 63	165 × 120
SVM	0.07%	11%	45.48%	77.45%	84.83%	92.03%
AdaBoost	0.024%	3.3%	3.8%	3.5%	2.9%	3.5%
Decision Tree	0.05%	7.2%	11%	12.92%	15.9%	10.2%
1D-HMM	12.05%	19.53%	45.87%	85.03%	90.28%	93.49%

Table 3. Face recognition rates of four classifiers with different Image resolution using 10 training images.

	Experimental Datasets					
Classifiers	6 × 4	11 × 8	21 × 15	42 × 30	83 × 63	165 × 120
SVM	0.13%	20.8%	52.42%	90.3%	97.19%	98.72%
AdaBoost	0.03%	3.4%	6.1%	5.22%	9.18%	4.6%
Decision Tree	0.07%	10.58%	9.7%	16.45%	11.2%	16.83%
1D-HMM	17.86%	32%	62.37%	95.03%	98.01%	98.85%

Table 4. Face recognition rates of four classifiers with different Image resolution using 15 training images

	Experimental Datasets					
Classifiers	6 × 4	11 × 8	21 × 15	42 × 30	83 × 63	165 × 120
SVM	18.9%	37.1%	73.84%	99.44%	100%	100%
AdaBoost	5.19%	5.7%	3.33%	7.2%	6.67%	5.9%
Decision Tree	5.5%	11.13%	17.6%	14.65%	20%	23%
1D-HMM	20.59%	46.75%	80.71%	99.63%	100%	100%

Table 5. Face recognition rates of four classifiers with different Image resolution using 20 training images

	Experimental Datasets					
Classifiers	6 × 4	11 × 8	21 × 15	42 × 30	83 × 63	165 × 120
SVM	10%	31.29%	75.85%	100%	100%	100%
AdaBoost	2%	5.4%	6.46%	5.4%	8.1%	5.1%
Decision Tree	6.12%	9.18%	18.02%	19.38%	18.36%	17.6%
1D-HMM	16%	48.64%	90.48%	100%	100%	100%

4. CONCLUSION

A low face recognition system based on 1D-HMMs was presented in this paper. First, the system extract facial features using Gabor and HOG descriptors. Then it combines them using CCA fusion technique to obtain one feature vector after dimensionality reduction step using LDA method.

The standard database AR is used to evaluate the proposed system. The obtained recognition rates by the combined features and 1D-HMMs classifier outperform these obtained by isolated descriptors and some classical classifiers for all facial image resolutions.

Our future research will be focused on using the proposed system in real time context in order to integrate it in video surveillance cameras security application.

REFERENCES

[1] Z. Wang, *et al.*, "Low-Resolution face Recognition: A Review," *The Visual Computer*, vol. 30, no. 4, pp. 359–386, 2014.
[2] J. Y. Choi, *et al.*, "Color face Recognition for Degraded face Images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 5, pp. 1217–1230, Oct 2009.
[3] T. Ahonen, *et al.*, *"Recognition of Blurred faces using local phase Quantization,"* Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pp. 1–4, 2008.
[4] S.-W. Lee *et al.*, "Low Resolution face Recognition based on Support Vector Data Description," *Pattern Recognition*, vol. 39, no. 9, pp. 1809 – 1812, 2006.
[5] C. Liu, *et al.*, "Gabor Feature based Classification using the Enhanced Fsher Linear Discriminant Model for face Recognition," *Trans. Img. Proc*., vol. 11, no. 4, pp. 467–476, Apr. 2002.
[6] H. Kusuma, *et al.* "Gabor-based face Recognition with Illumination Variation using Subspace-linear Discriminant Analysis", *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 10, no. 1, pp 119-128, 2012.
[7] C. Shu, *et al.*, "Histogram of the Oriented Gradient for face Recognition," *Tsinghua Science and Technology*, vol. 16, no. 2, pp. 216–224, April 2011.
[8] M. Haghighat, *et al.*, "Fully Automatic face Normalization and single Sample face Recognition in Unconstrained Environments," *Expert Systemswith Applications*, vol. 47, no. C, pp. 23–34, Apr. 2016.
[9] K. Kim, *"Face Recognition using Principle Component Analysis,"* International Conference on Computer Vision and Pattern Recognition, 1996, pp. 586–591.
[10] M. S. Bartlett, *et al.*, "Face Recognition by Independent Component Analysis," *Neural Networks, IEEE Transactions on*, vol. 13, no. 6, pp. 1450–1464, 2002.
[11] D. L. Swets, *et al.*, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.
[12] P. N. Belhumeur, *et al.*, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear Projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[13] I. Wijaya, et al. "Face Recognition Using Holistic Features and LDA Simplification", *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 10, no 4, p. 771-782, 2012.
[14] T. Savič, *et al.*, "Personal Recognition based on an Image of the Palmar Surface of the hand", *Pattern Recognition*, pp. 3152–3163, 2007.
[15] P. N. Belhumeur, *et al.*, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear Projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[16] Q.-S. Sun, *et al.*, "A new Method of Feature Fusion and its Application in Image Recognition," *Pattern Recognition*., vol. 38, no. 12, pp. 2437–2448, 2005.
[17] L. E. Baum, *et al.*, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *Annals of Mathematical Statistics*, vol. 37, pp. 1554–1563, 1966.
[18] L. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[19] A. M. Martinez, *et al.*, "Pca versus lda," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, no. 2, pp. 228–233, 2001.
[20] R. Keys, "Cubic Convolution Interpolation for Digital Image Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, Dec 1981.

BIOGRAPHIES OF AUTHORS

Othmane Elmeslouhi is an Associate Professor in the Mathematics and Computer Science Department of the Polydisciplinary Faculty Ouarzazate of Ibn Zohr University, Morocco. He has a PhD from the FST of Hassan 1st University, Morocco, in computer engineering. His research interests include : Big Data, Computer Vision, Pattern Recognition and Medical Imaging.
E-mail: o.elmeslouhi@uiz.ac.ma

Zineb Elgarrai received her engineer degree in Computer science in 2009 at the National School of computer science and systems analysis (ENSIAS) School, Rabat, Morocco. In 2012 she joined the LAVETE Laboratory of FST of Hassan 1st University, Settat, Morocco. Her actual main research interests concern Face Recognition and Computer Vision.
E-mail: elgarrai@gmail.com

Mustapha Kardouchi is a full professor at the Computer Science Departement, Moncton University, Canada. His research interests are primarily in the field of Computer Vision with a focus on Image Recognition and Video Analysis.
E-mail : mustapha.kardouchi@umoncton.ca

Hakim Allali was born in Morocco on 1966. He received the Ph.D degree from Claude Bernard Lyon I University (France) in 1993 and the “Docteur d’Etat” degree from Hassan II-Mohamedia University, Casablanca (Morocco) in 1997. He is currently Professor at Faculty of Sciences and Technologies of Hassan 1st University of Settat (Morocco) and director of LAVETE Laboratory. His research interests include technology enhanced learning, modeling, image processing, computer networking and GIS.
E-mail: hakimallali@hotmail.fr