A Combining Approach for 2D Face Recognition Application on IV² Database

Nefissa KHIARI HILI[1,2], Sylvie LELANDAIS[1] and Christophe MONTAGNE[1] and Kamel HAMROUNI[2]

[1] IBISC Laboratory, University of Evry Val d'Essonne, 40 Rue du Pelvoux, 91020 Evry Cedex, France

e-mail: nefissa.khiari-hili@ibisc.univ-evry.fr, s.lelandais@iut.univ-evry.fr, cmontagne@iup.univ-evry.fr

[2] LSTS Laboratory, National Engineering School of Tunis, BP-37 Le Belvédère 1002 Tunis, Tunisia

e-mail: kamel.hamrouni@enit.rnu.tn

***Abstract*—It is often difficult to deal with the problem of 2D-face recognition under unconstrained conditions. The objective of this study is to develop an original method that overcomes such obstructions. The proposed approach combines a holistic method, the Principal Component Analysis (PCA) to a local method, the Steerable Pyramid (SP). All tests were run on IV² database, with challenging variability and including 3500 to 5000 comparisons by experiment from 315 different people. The followed protocol was established in the first evaluation campaign on 2D-face images using the multimodal IV² database. Comparison with five submitted algorithms as PCA, LDA and LDA/Gabor provides satisfying results.**

***Keywords*—2D-Face recognition, Feature extraction, Method fusion, Steerable pyramid, PCA.**

I. Introduction

Human face recognition remains one of the most active areas in security and surveillance applications since it is non-invasive and requires less user cooperation. While considerable algorithmic progress has been made in controlled environments, variations remain an obstacle to deployable robust face recognition in real-life situations. Furthermore, most of the existing databases don't offer either enough variability or sufficient number of subjects. Therefore, relevant databases, like the IV² one [10], were recently developed to allow efficient evaluation taking into account real-life challenges.

Most classical approaches for face recognition are holistic appearance-based ones such as Eigenfaces and Fisherfaces [2]. On another hand, local feature-based approaches, like Gabor [6], are believed to achieve high accuracy. Both of them perform well in controlled environments; however, their performances drastically drop when variability like quality, pose and illumination occur. Therefore, new solutions are being suggested to overcome these challenges. Many of them were based on combining conventional algorithms and brought quite good results. For instance, Mellakh et al. built their method by fusing Linear Discriminant Analysis (LDA) and Gabor phase features [7]. Another combination was made by Zhang and Jia based on SP and LDA [14]. A further alternative was proposed by Su et al. allying both local and global features [12]. Another fusion method combining two different LDAs was suggested by Zuo et al. [15].

In light of the above developments, the 2D face recognition system presented in this paper was developed by combining an appearance-based method: the Principal Component Analysis (PCA) to a space-scale feature-extraction method: the Steerable Pyramid (SP). The first method commonly known as Eigenfaces was introduced by Turk and Pentland [13] and is considered as a reference face recognition method. It consists in building reduced subspaces in order to generate optimal features of facial images. On the other hand, Steerable Pyramid transform introduced by Simoncelli and Freeman in [11] is a local descriptor that associates multi-scale decompositions with differential measurements, thus able to capture both frequency and orientation information, which perfectly suits this application.

Through this paper, a promising 2D-face recognition method combining PCA and SP is introduced. The remaining part of this paper is organized as follows: Section 1 explains the theoretical concepts of PCA and SP. The database and the evaluation protocol established in the IV² project are described in Section 2. Proposed method, experimentations, and comparative results are reported in Section 3. Last section ends up with conclusions and future works.

II. Background concepts

Theoretical concepts related to the Principal Component Analysis (PCA) and the Steerable Pyramids (SP) are introduced in this section.

A. Principal Component Analysis (PCA)

The first successful face recognition system was built by Turk and Pentland by using Principal Component Analysis [13]. It is based on the decomposition of the treated image according to main directions of variation around an average image, in order to get more representative feature vectors and in a reduced size.

Facial images of size p by q pixels, represented as vectors Ii of size n $(n=p\times q)$ in high dimensional space $X = [x_1, x_2, \ldots, x_n]$, can be expressed as linear combination in an orthogonal basis Φ of a reduced dimension m $(m<<n)$:

$$x = \sum_{i=1}^{n} a_i x_i \approx \sum_{i=1}^{m} \alpha_i \phi_i \tag{1}$$

$x = [I_1, I_2, \ldots, I_L]$ is an $(n\times L)$ face matrix, L the number of faces.

The projection space defined by the orthogonal basis Φ is constructed by resolving the equation 2:

$$C\,\Phi = \Phi\,\Lambda \tag{2}$$

where C is the covariance matrix for input x:

$$C = \sum (x-\bar{x}).(x-\bar{x})^T \tag{3}$$

and $\bar{x}$ the average face:

$$\bar{x}(j) = \frac{1}{L}\sum_{i=1}^{n} I_i(j) \tag{4}$$

This operation gives the eigenvectors of size n and the eigenvalues of C which are then rearranged according to the variety directions importance. The obtained eigenvectors are called Eigenfaces and are used as the projection space for images to be treated.

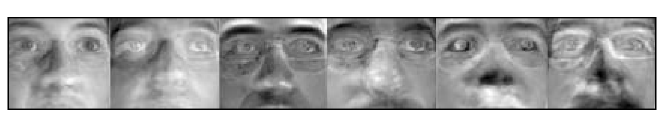

Fig. 1. Six top eigenfaces computed from the IV² training set.

Figure 1 shows an example of a few top eigenfaces from the IV² training set.

B. Steerable Pyramid (SP)

The steerable pyramid, introduced by Simoncelli & Freeman [11], is a linear multi-scale multi-orientation decomposition that provides a front-end to many image-processing applications particularly in texture analysis. The basis functions of a steerable pyramid are directional derivative operators that come in different sizes and orientations. The pyramid can be designed to produce any number of orientation bands. The representation is translation invariant (it is aliasing free) and rotation invariant (the subbands are steerable). More importantly, the transform is a tight-frame; specifically, the same filters used in the decomposition are used for the reconstruction.

The block diagram of a steerable pyramid [5] is given in figure 2 for both analysis and synthesis. In the analysis part, the image is decomposed into highpass and lowpass subbands using H_0 and L_0 filters. The lowpass band continues to break down into a collection of oriented $(n+1)$ bandpass subbands $B_0, B_1, \ldots, B_n$ and a lower lowpass subband L_1. The lower lowpass subband is subsampled by a factor of 2 in the x and y directions. This process represents the first level of decomposition of a steerable pyramid. Repeating the enclosed area on the output of subsampling provides the recursive (pyramid) structure, hence the next levels. In the synthesis part, the reconstructed image is obtained by upsampling the lower lowpass subband by a factor of 2 and adding up the collection of bandpass subbands and the highpass subband.

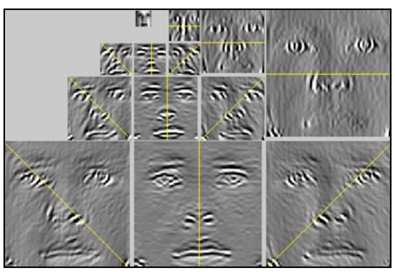

Fig. 3. Face image decomposition using a 3-level steerable pyramid with 4 orientations.

III. The IV² Database and the 2D Face Evaluation Protocol

The IV² database was designed with the aim of proposing multiple test situations to allow evaluation with regard to variability well known to be critical for the biometric systems performance, such as pose, expression, illumination and quality. The IV² database has been constructed during the Techno Vision program and has been supported by the French Research Ministry in collaboration with the French Ministry of Defense [10].

A. Database description

The IV² database contains *315* subjects with one session data where *77* of them also participated to a second session. From this database, a subset of *52* subjects, distributed as a development set, constitutes also the training set.

The face and sub-face data that are present in the IV² database are: 2D audio-video talking face sequences, 2D stereoscopic data acquired with two pairs of synchronized cameras, 3D facial data acquired with a laser scanner, and iris images acquired with a portable infrared camera; allowing monomodal and multimodal experiments. This database has been collected in several locations, by many operators. From the totality of the acquired data, are available two separate sets for development and evaluation purposes, and also an evaluation package.

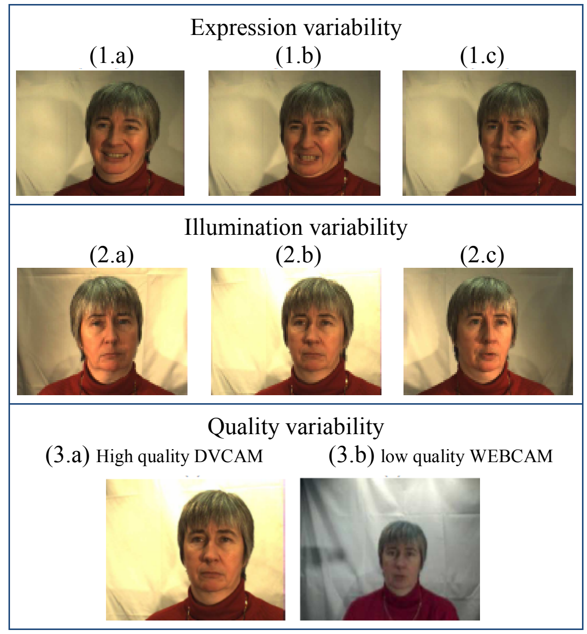

Fig. 4. Examples of variability related to (1.a-c) expression, (2.a-c) illumination and (3.a-b) quality.

The data comes from audio-visual streams (talking-face data) acquired simultaneously with high-quality digital camera (denoted as DVCAM) and low-quality Webcam (denoted as WEBCAM). The resolution of the DVCAM images is 780×576 pixels, with a mean distance of *90* pixels between the centers of the eyes. For the WEBCAM the resolution is 640×480 pixels, with a mean distance of *65* pixels between the centers of the eyes. Figure 4 presents some images with expression, illumination and quality variability. More details about the whole IV² database are available in [10].

B. The 2D face evaluation protocol

An evaluation campaign was performed at the end of the IV² project involving iris recognition, 2D and 3D-face recognition and also multimodal recognition. In the 2D-Face evaluation [9], the strategy of having "one variability" at a time was adopted in order to evaluate how challenging variability - related to illumination, expression, quality or multi-session images - can be for the biometric systems.

Table 1. Description of the images used for the 2D-Face experiments (V. means variation and N. stands for number).

Experiment	Exp1	Exp2	Exp3	Exp4
Sessions	Mono	Mono	Mono	Multi
Quality	High	High	Low	High
Expression V.	Small	No	No	No
Illumination V.	No	Yes	No	No
N. Intraclass	2595	2503	1774	1796
N. Interclass	2454	2364	1713	1796

In this evaluation campaign a set of more than *15000* images were divided into four subsets. Table 1 gives a description of the test images according to the corresponding experiment. The protocol was constructed so as to have almost the same number of client and imposters tests. This strategy allows having equivalent FAR (False Acceptance Rate) and FRR (False Rejection Rate).

Each team who proposed an algorithm conducted a list of tests with no indication of their type (intraclass or interclass). Only one still image per subject was used for enrollment and one still image for the test. A set of *52* images from *52* different subjects acquired at the development stage has been as well distributed to each team for the training phase.

Five appearance based methods were evaluated on the IV² database. The first two algorithms proposed by IBISC-1 and BioSecure were based on the Eigenfaces approach, proposed by Turk and Pentland [13]. The difference between these two algorithms was in the face space construction. IBISC-1 used the *52* images proposed in the development set while BioSecure built a face space with a set of *300* images of *30* subjects from the BANCA database [1].

A third algorithm proposed by IBISC-2 was also built on Eigenfaces but as being an application of the modular PCA proposed by Pentland and Moghaddam [8] with the purpose of addressing the problem of face recognition including facial expressions, illumination and pose variability.

On the other side, IBISC-3 used LDA [2] based on image intensity where the face projection space was built using *156* face images from *52* subjects of the IV² development set.

The last algorithm developed by Intitut Télécom SudParis, combined LDA and Gabor approaches. The face image was first convoluted with a family of Gabor filters. A feature vector, with both magnitude and phase information from the filtered images, was then introduced as an input for the LDA algorithm in order to get a more discriminative representation [7].

Comparative results between these five algorithms are shown in table 5 (PCA1 to LDA/Gabor).

IV. Proposed method and experimental results

Beforehand the recognition, normalization process is performed on gray-scale images in order to extract efficiently the region of interest. As a first try, only the SP was performed on images to get features in many orientations and at different scales. Nevertheless, obtained results were relatively bad. Actually, at this level, a maximum of information is collected in number as well as in variety, but only features characterizing at best faces similarities and dissimilarities has to be kept. That's why a combination method has been developed by applying PCA on the SP-filtered image. Finally, City-block distance d_{ij} is operated on feature vectors x_i and x_j to get final scores (see equation 5).

$$d_{ij} = \sum_{k=1}^{n} \left| x_{ki} - x_{kj} \right| \qquad (5)$$

A. Normalization

Knowing a prior the eyes coordinates, normalization consists first in rotating the face so as to have a horizontal interocular axis. The face is then remapped in order to get all the face images at the same scale with a fixed interocular distance. Finally, useless regions are eliminated so that only facial region is kept and all the obtained images have the same size *(112×112)*.

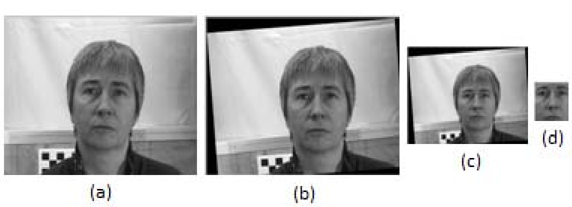

Fig. 5. Normalization process: (a) Initial image, (b) Rotated image (c) Rescaled image (d) Extracted face.

B. Characterization through SP

Many experiments have been run on how the SP has to be performed. A first experiment constructed the feature vector from the whole information at all orientations and scales provided by the entire filtered image. Another experiment was carried out by composing the feature vector of the *49 (16×16)* filtered blocs obtained from the initial image. A third try was run to build the feature vector with the *49* energy values computed from the *(16×16)* filtered blocs. Other tests have been completed to investigate how the variation of the orientations and scales numbers affects the recognition performance. Specifically an exhaustive set of experimentations has been fulfilled.

Figure 6 shows the Detection Error Trade-off (DET) curve relative to the four experiments and obtained with optimum parameters by extracting features from the entire *(112×112)* filtered image by using a four-level fifth derivative steerable pyramid (*6* orientations). The feature vector was composed of more than 10^6 intensity values.

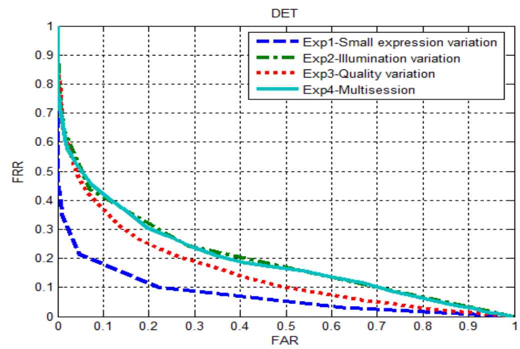

Fig. 6. DET curve of SP-based algorithm on the four IV² experiments.

SP achieves better results in experiment 1 which is the easiest test (only small expression variability). But performance decreases in experiment 3 with Quality variation and becomes even worse in experiment 2 and 4 face to Illumination and multisession variations. Details about EER are given in table 5 where it can be seen that the results obtained by the SP-based algorithm are worse than the others.

C. Characterization through PCA applied to SP

While applying SP, there was a rough use of huge amount of information provided by the filtered images, which led to bad results. For this reason, PCA was considered as an alternative to encounter this problem. Therefore, PCA is used as a solution that allows reducing the feature space so as to get only relevant characteristics. Construction of the projection axes (Eigenfaces) enables the filtered images to be projected on a space of reduced size built at the training stage. Operating PCA thus keeps the entirety of information provided by the filtered image without having resort to any magnitude computation (i.e. energy) but at the same time, concentrates this information on a restricted number of axes keeping only discriminant features.

In the next, the influence of a non exhaustive list of parameters related to the combination of SP and PCA is firstly presented. Then, a comparison with the submitted algorithms at the IV² evaluation campaign is brought. The results are reported with the Equal Error Rate percentage.

A first set of tests was carried out to see whether it is better to keep the image at its entirety as an input for SP filters or to split it into *(16×16)* blocs. Table 2 shows no sensitive improvements provided by partitioning the initial image, so proceeding with the complete image was kept.

Another set of experiments focused on which subbands of the SP should be took into account as an input to PCA. Four configurations were examined. Initial one keeps the high-pass *(HP),* the band-pass

($BP_{i\,=\,0\ to\ n}$) and the last low-pass (LP_n) subbands. First modification altered to the initial configuration added the first low-pass subband (LP_0). Second modification was drove by supplementing the intermediate low-pass subbands ($LP_{i\,=\,1\ to\ n}$) to initial configuration. Finally, third alteration was realized by adding both initial and intermediate low-pass subbands ($LP_{i\,=\,0\ to\ n}$). Results illustrated in table 3 are in favor of the second modification.

Apart from examining many aspects of SP filtering, a further study concerning the impact of eigenfaces cumulative inertia has been performed to decide about the optimal number of eigenfaces to keep. Equation 6 gives the cumulative inertia $Inert\_cumul_j$ of the j^{th} first eigenfaces, with λ_i the eigenvalue related to i^{th} eigenface.

$$Inert\_cumul_j = \frac{\sum_{i=1}^{j} \lambda_j}{\sum_{i=1}^{n} \lambda_i} \qquad (6)$$

Table 2. Image splitting influence.

Application zone of SP	Exp1	Exp2	Exp3	Exp4
Entire image	**5.3 (±0.7)**	20.4 (±1.3)	**18.3 (±1.5)**	**20.0 (±1.6)**
(16x16) Sub-blocs	5.8 (±0.8)	**19.8 (±1.3)**	18.8 (±1.5)	20.2 (±1.6)

Table 3. Configuration of SP Sub-bands influence.

N° conf	Kept sub-bands	Exp1	Exp2	Exp3	Exp4
0	HP, BP_i, LP_n (i = 0..n)	5.3 (±0.7)	20.4 (±1.3)	18.28 (±1.5)	18.3 (±1.6)
1	HP, LP0, BP_i, LP_n (i = 0..n)	5.5 (±0.7)	20.0 (±1.3)	18.49 (±1.5)	18.5 (±1.6)
2	**HP, BP_i, LP_i (i =1..n)**	**5.0** (±0.7)	**19.9** (±1.3)	**17.37 (±1.5)**	**17.4** (±1.6)
3	HP, BP_i, LP_i (i = 0..n)	5.3 (±0.7)	20.0 (±1.3)	17.60 (±1.5)	17.6 (±1.6)

Table 4. Eigenfaces cumulative inertia influence (Exp1).

Cumulative inertia	Eigenvalues Nbr	% EER	% FRR (FAR= 0.001)
100%	**52**	**5.28 (±0.7)**	**23.6**
95%	32	5.39 (±0.7)	26.0
90%	23	5.56 (±0.8)	32.8
85%	17	5.79 (±0.8)	31.0
80%	12	5.72 (±0.8)	41.0

Results presented in table 4 are relative to experiment 1, and were also confirmed by the other experiments. They state that keeping the totality of projection axis perform better. Moreover, operating a PCA with a training set of *52* images implies a total number of *52* Eigenfaces. That is to say that the feature vector size has been drastically reduced from more than 10^6, when applying only SP, to *52* elements when operating PCA on filtered images.

The method having the optimum parameters was to filter the entire *(112×112)* image by a *6*-orientation and *4*-scale SP. Besides the high-pass *(HP)* and the band-pass ($BP_{i=\,0\ to\ n}$) subbands, the intermediate low-pass bands ($LP_{i=\,1\ to\ n}$) were also retained at each level to get more information. Then, PCA was applied on filtered image using all the eigenfaces found out at the training step i.e. *100%* of the cumulative inertia.

Table 5 illustrates evaluation of the proposed method put side by side with the other ones. It can be seen that combining PCA to SP improves considerably the performance of SP. On another hand, a comparison to PCA1 [3] is essential as it is the same PCA that was operated in proposed method. Enhancements are obvious in the three first experiments, whereas almost identical performances are achieved in the fourth one.

Regarding PCA2, it has to be underlined that the training set on which the face space has been constructed isn't the same as indicated in the protocol. In fact, it is built using *300* images from BANCA database (*30* subjects, *10* images per subject) with *3* different quality images [9]. While proposed method strictly followed the protocol using only *52* images of *52* individuals acquired under quite good conditions, which is not the case of the test subsets where many variations are present. The small number of trained images besides the different acquisition conditions between training and test subsets represents an additional challenge, which explains the results obtained in experiments 3 and 4. Despites, proposed method outperforms PCA2 in experiments 1 and 2.

Compared to LDA, except for the first controlled scenario, proposed method achieves higher performance in the other more challenging ones. But it still remains less robust than LDA/Gabor which is also a combining approach of a projection-based method (LDA) and a space-scale feature-extraction method (Gabor).

Table 5. Comparative results between proposed algorithms (two last lines) and IV² first evaluation campaign ones (five first lines).

Participants	Exp1	Exp2	Exp3	Exp4
PCA1	6.7 (±0.8)	20.7 (±1.3)	22.2 (±1.6)	20.1 (±1.6)
PCA2	7.3 (±0.8)	21.6 (±1.4)	16.3 (±1.4)	13.6 (±1.4)
mod PCA	5.3 (±0.7)	20.7 (±1.4)	20.5 (±1.5)	19.5 (±1.6)
LDA	**3.7** (±0.6)	22.5 (±1.4)	19.7 (±1.5)	21.7 (±1.7)
LDA/Gabor	4.2 (±0.6)	**12.0** (±1.1)	**11.3** (±1.2)	**8.3** (±1.1)
SP	14.8(±1.2)	26.2 (±1.5)	23.0 (±1.7)	26.2 (±1.7)
PCA/SP	5.0 (±0.7)	20.0 (±1.4)	17.4 (±1.5)	20.2 (±1.7)

Unlike the hereby suggested method, which focuses on the verification aspect of face recognition, other combining studies related to face identification by using SP exist in literature. El Aroussi et al. [3]fused SP to LDA. However the recognition accuracy they claim (*98.75%* on ORL, *100%* on Yale and *91.72%* on subset of FERET database) were achieved on much smaller (*40, 15* and *145*-subject) databases and with bigger training sets, which makes the results extremely dependent on the subjects present in that subsets. On the other side, Zhang et al. [14] conceived a face identification approach combining SP to LDA. Databases employed in their work were very relevant in number as well as in variability (*295* subjects in XM2VTS and *429* subjects in FERET database). Furthermore, pertinent statistics based on Cumulative Match Scores were specified to show the robustness of the method and its superiority to other standard ones.

V. Conclusion

The main contribution of this work is to investigate a new approach combining Steerable Pyramids and PCA 2D-face recognition from still images. It has been proved that joining a projective method to the SP brought significant enhancements. Future works are intended to consider the effectiveness of LDA opposed to PCA so as to compare it afterward to Zhang et al. study. Further research will investigate non linear classifiers which are reputed to be more efficient.

ACKNOWLEDGMENT

The authors are grateful to Anis Chaari from IBISC laboratory in the University of Evry Val d'Essonne for his implementation in MATLAB of the PCA algorithm and to Joseph Colineau from Thales Research & Technology for providing the IV² evaluation package.

REFERENCES

[1] Bailly-Bailliere E., Bengio S., Bimbot F., Hamouz M. & al., 2003. The BANCA database and evaluation protocol. *In Inter. Conf. on Audio- and Video-based Biometric Person Authentification (AVBPA),* 2688(4):625-638, Guilford, UK.

[2] Belhumeur P.N., Hespanha J.P. and Kriegman D.J., 1997. Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI),* 19(7):711–720.

[3] Chaari A., Lelandais S. and Ben Ahmed M., 2009. A pruning approach improving face identification system. *In Advanced Video and Signal Based Surveillance (AVSS)*, Italy.

[4] El Aroussi M., El Hassouni M., Ghouzali S., Rziza1 M. and Aboutajdine D., 2009. Novel face recognition approach based on steerable pyramid feature extraction. *In IEEE Inter. Conf. on Image Processing (ICIP),* 4165 – 4168, Cairo, Egypt.

[5] Krasaridis A. and Simoncelli E.P., 1996. A Filter Design Technique for Steerable Pyramid Image Transforms. *In IEEE, Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP),* 4:2387-2390, Atlanta, USA.

[6] Li S. Z. and Jain A. K., 2005. *Handbook of Face Recognition*. Springer.

[7]Mellakh A., Chaari A., Guerfi S. & all., 2009. 2D Face Recognition in the IV² Evaluation Campaign. *In Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Bordeaux, France.

[8] Pentland A.P., Moghaddam B. and Starner T.E., 1994. View-based and modular eigenspaces for face recognition. . *In Conf. on Computer Vision and Pattern Recognition (CVPR)*, 84–91, Washington, USA.

[9] Petrovska-Delacrètaz D., Chollet G. and Dorizzi B., 2009. *Guide to Biometric Reference Systems and Performance Evaluation*. Springer.

[10] Petrovska-Delacretaz D., Lelandais S., Colineau J. & al., 2008. The IV2 Multimodal Biometric Database (Including Iris, 2D, 3D, Stereoscopic, and Talking Face Data), and the IV2- 2007 Evaluation Campaign. *In 2nd IEEE Inter. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, USA.

[11] Simoncelli E.P. and Freeman W.T., 1995. The Steerable Pyramid: A Flexible Architecture For Multi-scale Derivative Computation. *In IEEE Inter. Conf. on Image Processing (ICIP),* 3(2):444-447, USA.

[12] Su Y., Shan S.G., Chen X.L., and Gao W., 2009. Hierarchical ensemble of global and local classifers for face recognition. *Journal of IEEE Trans. on Image Processing,* 18(8): 1885-1896.

[13] Turk M. and Pentland A.P., 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

[14] Zhang X. and Jia Y., 2005. Face Recognition Based on Steerable Feature and Random Subspace LDA. *In Inter. workshop on Analysis and Modelling of Faces and Gestures (AMFG),* 3723(2):170-183, Beijing, China.

[15] Zuo W., Wang K., Zhang D. and Zhang H., 2007. Combination of two novel LDA-based methods for face recognition. *Journal of Neurocomputing, Elsevier,* 70 (4-6): 735-742.