Fast Image Mosaicing for Panoramic Face Recognition

Fan Yang, Michel Paindavoine, Hervé Abdi*, Dominique Arnoult

Laboratoire Le2i, Aile de l'Ingénieur - Mirande

Université de Bourgogne, BP 400 - 21011 DIJON cedex, FRANCE

fanyang@u-bourgogne.fr, paindav@u-bourgogne.fr

*Program in Cognition and Neurosciences, MS: Gr.4.1

The University of Texas at Dallas, Richardson TX 75083–0688 USA

herve@utdallas.edu

Abstract—**In this article, we present some development results of a system that performs mosaicing (or mosaicking) of panoramic faces. Our objective is to study the feasibility of panoramic face construction in real-time. To do so, we built a simple acquisition system composed of 5 standard cameras which, together, can take simultaneously 5 views of a face at different angles. Then, we chose an easily hardware-achievable algorithm, consisting of successive linear transformations, in order to compose a panoramic face from these 5 views. The method has been tested on a relatively large number of faces. In order to validate our system of panoramic face mosaicing, we also conducted a preliminary study on panoramic faces recognition, based on the principal component analysis method. Experimental results show the feasibility and viability of our system.**

Index Terms – **artificial vision, image mosaicking, face recognition, principal component analysis, FFT**

I. INTRODUCTION

Biometry is currently a very active area of research which comprises several sub-disciplines such as image processing, pattern recognition, and computer vision. The main goal of biometry is to build systems which can identify people from some observable characteristics such as their face, fingerprints, iris, etc. Faces seem to have a particularly strong appeal for the human users. Different techniques have been used to process faces[1][2][3]. As the recent DARPA-sponsored vendor test showed, most systems use frontal facial images as their input patterns[4]. As a consequence, most of these methods are sensitive to pose and lighting conditions. One way to override these limitations is to combine modalities (color, depth, 3D facial surface . . .)[5][6][7].

Most of 3D acquisition systems use professional devices such as a travelling camera or a 3D scanner [6][7]. Typically, these systems require that the subject remains immobile during several seconds in order to obtain a 3D scan, and therefore these systems may not be appropriate for some applications such as human expression categorization using movement estimation or real-time applications. Also, their cost can easily make these systems prohibitive for routine applications. In order to avoid using expensive and time intensive 3D acquisition devices, some face recognition systems generate 3D information from stereo-vision. Relatively complex calculations, however, are necessary in order to perform the auto-calibration and 2D projective transformation[8]. Another possible approach is to derive some 3D information from a set of face images, but without trying to reconstitute the complete 3D structure of the face[5][9][10].

The goal of the present article is to describe a system which is simple and efficient and also which can potentially process 3D faces in real-time. First, we describe our acquisition system. Then, we describe the method for creating panoramic face mosaics using successive linear transformations. Next, we present experimental results of panoramic face recognition. Finally we conclude and explore possible follow-ups and improvements.

II. ACQUISITION SYSTEM PRESENTATION

Our acquisition system is composed of five Logitech 4000 USB cameras with a maximal resolution of 640×480 pixels. Each camera is fixed on a height-adjustable sliding support in order to adapt the camera position to each individual (see Figure 1). One of its main advantages is an easy setup and very low cost. The acquisition program grabs images from the 5 cameras simultaneously. These 5 images are stored in the PC with a frame data rate of $20 \times 5 = 100$ images per second. The human subject sits in front of the acquisition system, directly facing the central camera (Camera 3). Different color markers are placed on the subject's face. These markers are used later on to define common points between different face views. The position of these color markers corresponds roughly to the face fiduciary points. Figure 2 shows the position of the chosen

Fig. 1. The Acquisition system: the top panel shows the 5 cameras and their support, the bottom panel shows the five images collected from a subject. Each image size is 240×320 pixels.

Fig. 2. Distribution of 10 markers in 5 views (clockwise from top left): Image 1, Image 2, Image 3, Image 4, and Image 5.

points. There are 10 markers on each face with at least 3 markers in common between each face view.

III. PANORAMIC FACE CONSTRUCTION

Several panoramic image construction algorithms have been already introduced[10][11][12]. For example, A. Jain and A. Ross[12] have developed an image mosaicing technique that constructs a more complete fingerprint template using two impressions of the same finger. In the proposed algorithm, they initially aligned the two impressions using the corresponding minutiae points. Then, this alignment was used by a modified version of the Iterative Closest Point (ICP) algorithm in order to compute a transformation matrix that defines the spatial relationship between the two impressions. A resulting composite image is generated using the transformation matrix which has 6 independent parameters: three rotation angles (α, β, γ) about the x, y, and z axes, respectively, and three translation components (t_x, t_y, t_z) along the three axes.

For faces, X. Liu and T. Chen[10] have proposed using facial geometry in order to improve the face mosaicing result. They created the spherical projection (instead of the cylindrical projection), which works better with the head motion in both horizontal and vertical directions. A geometric matching algorithm has been developed in order to describe the correspondences between the 2D image plane space QUV and the spherical surface space $O\alpha\beta$. In order to accomplish this, the two matching parameters $[\Delta\alpha, \Delta\beta]^T$ are found using the Levenberg-Marquardt algorithm. In order to improve the computational efficient, X. Liu and T. Chen approximated the

mapping function with a triangular mesh, representing a face as a set of triangles.

In general, the methods using non-linear transformations and iterative algorithms obtain very correct results in terms of geometric precision. However, these methods require a large number of computations and therefore cannot be easily implemented in real-time. Because we ultimately want to be able to build a real-time system, we decided to use simple (and therefore fast) linear methods. Our panoramic face construction algorithm is performed in three stages (see Figure 3):

1) Marker detection and marker coordinate calculation.
2) Transformation matrix estimation and image linear transformation.
3) Creation of panoramic face mosaics.

A. Marker detection and marker coordinate calculation

The first step of the algorithm corresponds to the detection of the markers put on the subject's face. We used 3 colors to create 10 markers (4 blue, 3 yellow, and 3 violet ones, see Figure 2 for an illustration). These markers were used as reference points for pasting the different views of the face. In order to detect the markers, we used color segmentation based upon the hue and saturation components of each image. This procedure allows a strong color selectivity and a small sensitivity to luminosity variation. Figure 4 illustrates the steps used for the yellow marker detection process. First, color segmentation gives, from the original image (Figure 4 top left), a binary image which contains the detected markers (Figure 4 top right). Then, in order to find the marker coordinates we
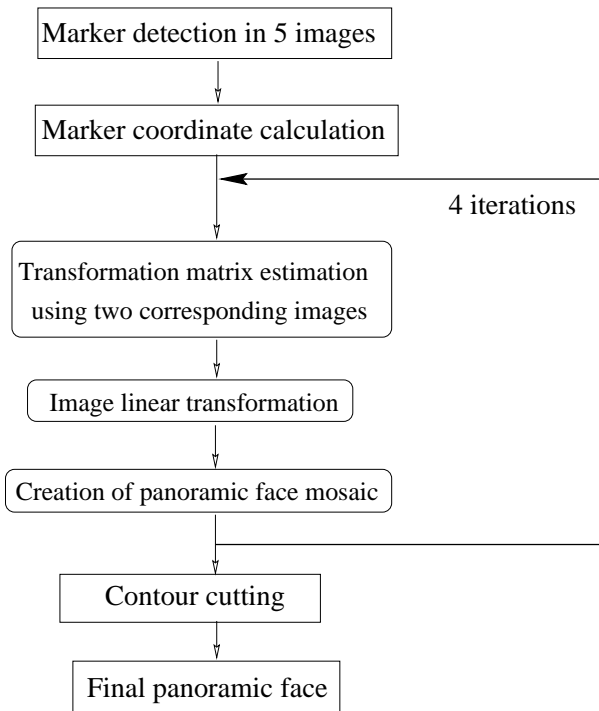
Marker detection in 5 images

Marker coordinate calculation

4 iterations

Transformation matrix estimation using two corresponding images

Image linear transformation

Creation of panoramic face mosaic

Contour cutting

Final panoramic face

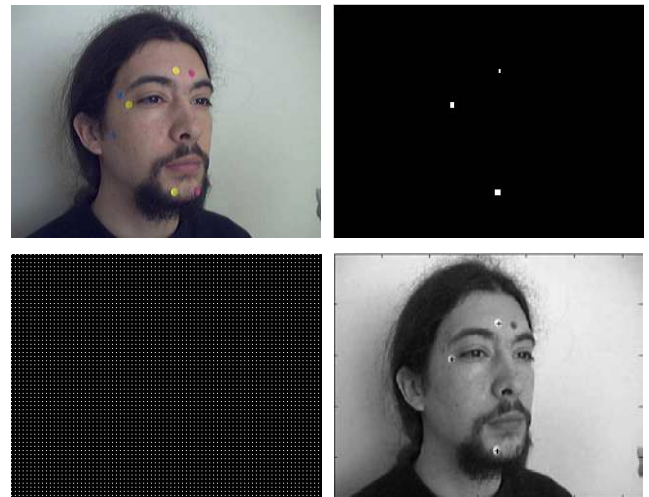Fig. 3. Block diagram of proposed panoramic face construction algorithm.

Fig. 4. Yellow marker detection (reading left to right): original image, binary image after color filtering applied on Hue and Saturation components, a grid of white pixels, and marker coordinate localization.

used a logical AND operation which was performed between the binary image and a grid including white pixels separated by a fixed distance (Figure 4 bottom left). Finally, we computed the centers of the detected zones (Figure 4 bottom right). These centers gives the coordinates of the markers in the image.

B. Transformation matrix estimation and image linear transformation

We decided to represent each face as a mosaic which is made by concatenation of the different views pasted together as if they were on a flat surface. So, in order to create a panoramic face we combine the five different views. We start with the central view and paste the lateral views one at a time (see Figure 3). Our method consists of transforming the image to be pasted in order to link common points between this image and the target image. We obtain this transformed image by multiplying it by a linear transformation matrix. This matrix is calculated as a function of the coordinates of 3 common markers between 2 images. $\mathbf{C}_1$ and $\mathbf{C}_2$ represent, respectively, coordinates of the first and second images:

$$\mathbf{C}_1 = \left[\begin{array}{ccc} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{array} \right] \qquad (1)$$

$$\mathbf{C}_2 = \left[\begin{array}{ccc} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \end{array} \right] \qquad (2)$$

We obtain the transformation matrix as follows:

$$\mathbf{T} = \mathbf{C}_1 \times (\mathbf{C}_2^\star)^{-1} \qquad (3)$$

with

$$\mathbf{C}_2^\star = \left[\begin{array}{ccc} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ 1 & 1 & 1 \end{array} \right] \qquad (4)$$

and

$$\mathbf{T} = \left[\begin{array}{ccc} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{array} \right] \qquad (5)$$

We generalize this transformation to the whole image :

$$\begin{array}{l} x = a_1 x' + b_1 y' + c_1 \\ y = a_2 x' + b_2 y' + c_2 \end{array} \qquad (6)$$

This linear transformation corresponds to a combination of image rotation, image translation, and image dilation (see Figure 5). Figure 6 displays the superposition of Image 3 (not transformed) and Image 4 (transformed using the coordinates of the yellow markers as common points).

Fig. 5. Image 4 before and after the linear transformation: the transformation matrix is computed using Image 4 and Image 3 (central view, see Figures 2 and 6).

C. Creation of panoramic face mosaics

We begin the panoramic face construction with the central view (Image 3, see Figure 2). From the superposition of the original Image 3 and transformed Image 4 (see Figure 6), we remove redundant pixels in order to obtain a temporary panoramic 3-4 image (see Figure 7 left). In order to eliminate

Fig. 6. Superposition of Images 3 and 4: original Image 3 (left), and superposition of transformed Image 4 and original Image 3 (right).

Fig. 9. Mosaicing results: Panoramic face composition from 5 views (left), and final mosaic (right).

redundant pixels, we create a cutting line which goes through two yellow markers. This panoramic 3-4 image temporarily becomes our target image.

We repeat this operation for each view. First, Image 2 is pasted on the temporary panoramic Image 3-4 in order to obtain a new temporary panoramic 2-3-4 image (see Figure 7 right). The corresponding transformation matrix is generated using three common violet markers. Then, we compute the transformation matrix which constructs Image 2-3-4-5 (see Figure 8 left) using two blue markers and one yellow marker (topmost). Finally, Image 1 is pasted to the temporary panoramic Image 2-3-4-5 with the help of two blue markers and one violet marker (topmost) (see Figure 8 right).

Fig. 7. Mosaicing results: Image 3-4 (left), and Image 2-3-4 (right).

Fig. 8. Mosaicing results: Image 2-3-4-5 (left), and Image 1-2-3-4-5 (right).

Figure 9 left displays the final panoramic face composition from 5 views. This composition preserves some of the face shape. For example, the chin in a human face possesses more curvature than other parts, therefore the bottom part of the panoramic face is composed of 5 views: 1, 2, 3, 4, and 5. On the other hand, 3 views (1, 3, and 5) suffice to compose the top part.

Figure 9 right shows the final mosaic face obtained after automatically contour cutting. Note that these 10 markers allow us to link common points between 5 views. The coordinates

of the markers are computed in the marker detection process, and arranged in a table. Then, all 10 markers are erased from all 5 views using a simple image processing technique (local smoothing).

We created a panoramic face database composed of 12 persons $\times$ 4 expressions $\times$ 2 sessions = 96 panoramic faces. The two acquisition sessions were performed over an interval of one month. The four expressions were: neutral, smile, deepened eyebrows, and eyes closed (see Figure 10). We implemented a face recognition procedure using this database in order to test and validate the proposed panoramic face mosaicing algorithm.

Fig. 10. A sampling of panoramic faces from the first session's database.

IV. FACE RECOGNITION DESCRIPTION : PCA

Over the past 25 years, several face recognition techniques have been proposed, One of the most versatile approach is derived from the statistical technique called Principal Component Analysis (PCA) adapted to face images[2][13][14]. PCA is based on the idea that face recognition can be accomplished with a small set of features that best approximates the set of known facial images.

Application of PCA for face recognition proceeds by first performing a PCA on a well-defined set of images of known human faces. From this analysis, a set of K principal components is obtained and the projection of the new faces on

these components is used to compute distances between new faces and old faces. These distances, in turn, are used to make predictions about the new faces. The complete set of patterns is represented by a $I \times K$ matrix noted $\mathbf{A}$, where I represents the number of pixels of the face images and K the total number of images under consideration. Specifically, the learned matrix $\mathbf{A}$ can be expressed as:

$$\mathbf{A} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^T \qquad (7)$$

where $\mathbf{P}$ is the matrix of eigenvectors of $\mathbf{A}\mathbf{A}^T$, $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{A}^T\mathbf{A}$, and $\mathbf{\Delta}$ is the diagonal matrix of singular values of $\mathbf{A}$, $\mathbf{\Delta} = \mathbf{\Lambda}^{1/2}$ with $\mathbf{\Lambda}$ matrix of eigenvalues of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$[14].

Given the singular vectors $\mathbf{P}$, every face in the database can be represented as a weight vector in the principal component space. The weights are obtained by projecting the face image onto the left singular vectors, and this is achieved by a simple inner product operation :

$$PROJ_x = \mathbf{x}^T\mathbf{P}\mathbf{\Delta}^{-1} \qquad (8)$$

where $\mathbf{x}$ is a facial vector. It corresponds to an example face in the training process or a test face in the recognition process.

Identification of the test image is done by locating the image in the known face database whose weights have the smallest Euclidean distance from the weight of the test image.

V. EXPERIMENTAL RESULTS OF PANORAMIC FACE RECOGNITION

A. Spatial representation

For these first tests, panoramic faces were analyzed using the original 240×320 pixel image (spatial representation) without pre-processing. The database consisted of 12 persons $\times$ 4 expressions $\times$ 2 sessions = 96 panoramic faces, and was divided into two subsets. One subset served as the face training set and the other subset provided the face testing set. From the panoramic face database, one, two, or four images were randomly chosen for each individual in order to create the training set (number of patterns for learning per individual $p = 1, 2, 4$). The rest of the panoramic faces were used in order to test the face recognition method. Several executions of our Matlab program were run for each value of p using randomly chosen training and testing sets. Then we computed mean performances, the results are presented in Table 1 where the recognition rate corresponds to correct panoramic face recognition rate. A discriminant analysis stage is added in the face recognition process in order to determine the number of necessary eigenvectors[15].

B. Frequential representation

We also tested the frequential behaviour of our recognition system. We applied the same training and testing process for FFT amplitude signal of panoramic faces as used in spatial representation. Test results are given in Table 2. We obtain a better recognition rate with the frequential representation (97.46%) than with the spatial representation (93.21%).

TABLE I

Results of panoramic face recognition with spatial representation.

Number of training examples/individual	$p = 1$	$p = 2$	$p = 4$
Number of total training examples	12	24	48
Number of eigenvectors	9	13	25
Number of tests for recognition	84	72	48
Recognition rate	70%	85.08%	93.21%

TABLE II

Results of panoramic face recognition with frequential representation.

Number of training examples/individual	$p = 1$	$p = 2$	$p = 4$
Number of total training examples	12	24	48
Number of eigenvectors	8	13	24
Number of test for recognition	84	72	48
Recognition rate	76.83%	91.26%	97.46%

This advantage of the frequential representation is due to the fact that for face images, the spectrum amplitude is less sensitive to noise (or variations) than the spectrum phase. We confirmed this interpretation by using a panoramic face image to which noise was added. Figure 11a shows a original panoramic face. Figure 11b displays the same panoramic face image with added noise. We first performed the FFT of these two images and, then their inverse FFT in the two following manners:

1) Using the spectrum amplitude of the noised image and the spectrum phase of the original image (see Figure 11c).
2) Using the spectrum phase of the noised image and the spectrum amplitude of the original image (see Figure 11d).

These results show that the face obtained with the first configuration is closer to the original face than the face obtained with the second configuration. This confirms that the spectrum amplitude is less sensitive to noise than the spectrum phase.

F. Tsalakanidou et *al.*[5] proposed a study in order to evaluate three different approaches (color, depth, combination of color and depth) for face recognition and quantity the contribution of depth. The color images were stored in portable pixmap format (ppm) with a resolution of 720×576 pixels. In addition for each person the "structured light" approach was used for capturing the 3D facial surface and thus creating the depth map. Their experimental results show significant gains of 5% with the use of depth information. They have obtained a recognition rate of 97.5% using the depth map and color frontal view. The face recognition technique used is based on the implementation of the PCA algorithm (similar to ours). By
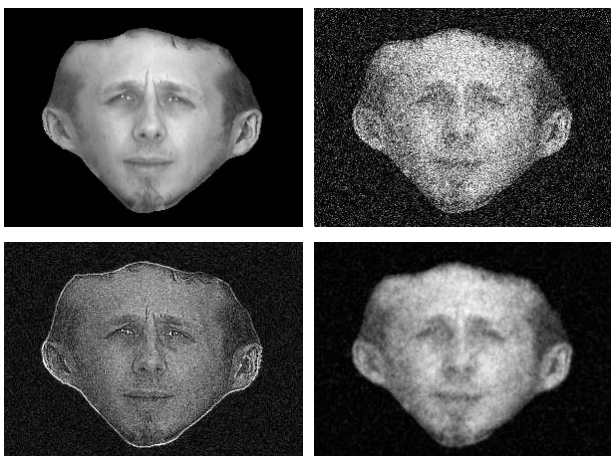
Fig. 11. Amplitude is less sensitive to noise than phase is (from left to right, top to bottom): **a).** An original panoramic face image; **b).** Original image with added Gaussian noise (mean=0 and variance=0.05); **c).** IFFT image using the spectrum amplitude of **b)** and the spectrum phase of **a)**; and **d).** IFFT image using the spectrum amplitude of **a)** and the spectrum phase of **b)**. The image **c)** is more similar to the image **a)** than the image **d)** is.

using a simple and easily built system, we obtain some 3D face information. Obtained performances of panoramic face recognition are very close to those of F. Tsalakanidou et *al.*

C. Panoramic face recognition with negative samples

In order to test the behavior of our system for unknown people, we added 4 people to the test database (see Figure 12). These panoramic faces were obtained as described in Section III.

Fig. 12. Panoramic faces of four unknown persons.

Table 3 displays performance of different tests. We added 4 persons × 4 expressions × 2 sessions = 32 panoramic faces in each test set. In order to reject these unknown faces, we established a threshold of Euclidean distance. Because we are working on applications of typical access control, where confusion is more harmful than non-recognition, we decided to use a severe acceptance threshold in order to reject intruders. Note that the acceptance threshold is constant for all tests. Efficiency is defined as follows:

Recognition: correct recognition of a panoramic face,
Non-recognition: a panoramic face has not been recognized,
Confusion: a panoramic face is confused with an intruder.

These performances were obtained using the frequential representation. Performance declined in comparison with tests without negatives samples.

TABLE III
Results of panoramic face recognition with negative samples.

Number of training examples/individual	$p = 1$	$p = 2$	$p = 4$
Number of total training examples	12	24	48
Number of eigenfaces	8	13	24
Number of test for recognition	116	104	80
Non-recognition rate	25.4%	12.74%	4.82%
Confusion rate	5.85%	4%	2.8%
Recognition rate	68.75%	83.26%	92.38%

VI. CONCLUSIONS AND PERSPECTIVES

In this article, we proposed a fast and simple method for panoramic face mosaicing. The acquisition system consists of several cameras followed by a series of fast linear transformations of the images. The simplicity of the computations makes it possible to envisage real-time applications.

In order to test the recognition performance of our system, we used the panoramic faces as input to a recognition system based on principal component analysis. We tested two panoramic face representations: spatial and frequential. We found that a frequential representation gives the best performance with a correct recognition rate of 97.46% versus 93.21% for spatial representation. We used negative samples for the panoramic face recognition system, and the correct recognition rate was 92.38%.

Experimental results show that our fast mosaicing system provides relevant 3D facial surface information for recognition application. Obtained performance is very close or superior to published levels[1][5].

In the future, we plan to simplify our acquisition system by replacing the markers by a structured light. We also hope to use our system without markers. For this, we will detect control points on faces (corners and maximum curvature ...). Another line of development is to improve the geometry quality of our panoramic face mosaic construction[10]. For this, we will use realistic human face models. We are also exploring processing panoramic face recognition using others classifiers with more variable conditions. Then, 3D face applications are envisaged such as real time human expression categorization using movement estimation and fast 3D facial modeling for compression and synthesis such as in video-conferencing.

REFERENCES

[1] A.J. Howell and H. Buxton, *Learning identity with radial basis function networks*, Neurocomputing, Vol.20, pp.15-34, 1998.
[2] M. Turk and A. Pentland, *Eigenfaces for recognition*, Journal Cognitive neuroscience, Vol.3, pp.71-86, 1991.
[3] H. Abdi, D. Valentin and A. O'Toole, *A generalized auto-associator model for face semantic process*, in *Optimization and neural network*, edited by D.Levine (Erlbaum, Hillsdale), 1997.
[4] P.J. Phillips, P. Grother and *al.*, *Face recognition Vendor Test 2002*, IEEE International workshop on Analysis and Modeling of Faces and Gestures (AMFG), 2003.

[5] F. Tsalakanidou, D. Tzovaras and M.G. Strintzis, *Use of depth and colour eigenfaces for face recognition*, Pattern recognition Letters, Vol.24, pp.1427-1435, 2003.

[6] C. Hehser, A. Srivastava and G. Erlebacher, *A novel technique for face recognition using range imaging*, 7th International Symposium on Signal Processing and its Applications (ISSPA), 2003.

[7] X. Lu, D. Colbry and A.K. Jain, *Three-Dimensional model based face recognition*, Proc. International Conference on Pattern Recognition, Cambridge, UK, August, 2004.

[8] R. Hartly and A. Zisserman, *Multiple View Geometry in Computer vision*, Cambridge University Press, Second Edition, 2003.

[9] F. Yang, M. Paindavoine and H. Abdi, *Fast panoramic face mosaicing and recognition*, IEEE International Conference on Image and Signal Processing and Analysis (ISPA), Zagreb, Croatia, September, 2005.

[10] X. Liu and T. Chen, *Geometry-assisted statistical modeling for face mosaicing*, IEEE International Conference on Image Processing (ICIP), Vol.2, pp.883-886, Barcelona, Spain, 2003.

[11] Y. Kanazawa and K. Kanatani, *Image mosaicing by stratified matching*, Image and Vision computing, Vol.22, pp.93-103, 2004.

[12] A.K. Jain and A. Ross, *Fingerprint Mosaicing*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, Florida, May, 2002.

[13] D. Valentin, H. Abdi, A.J. O'Toole and G.W. Cottrell, *Connectionist models of face processing: A survey*, Pattern Recognition, Vol.27, 1208-1230, 1994.

[14] H. Abdi, *A generalized approach for connectionist auto-associative memories: interpretation, implications and illustration for face processing*. In J. Demongeot (Ed.), *Artificial Intelligence and Cognitive Sciences.* Manchester: Manchester University Press, (1988).

[15] A.J. O'Toole, F. Jiang, H. Abdi and J.V. Haxby, *Partially distributed representations of objects and faces in ventral temporal cortex*, Journal of Cognitive Neuroscience, Vol.17, (in press, 2005).

Fan Yang received a B.S degree in eletrical engineering from the University of Lanzhou (China) in 1982. She was a Scientific Assistant at the Departement of electronics in the University of Lanzhou. She received an M.S. (D.E.A.) degree in computer science and a Ph.D. degree in image processing from the University of Burgundy (France), respectively in 1994 and in 1998. She is currently full "Maitre de Conferences" and member of LE2I CNRS-UMR (Laboratory of Electronic, Computing and Imaging Sicences). Her research interests are in the areas of patterns recognition, neural network, motion estimation based on spatio-temporal Gabor filters, parallelism and real-time implementation, and more specifically, automatic face image processing : algorithms and architectures.

Michel Paindavoine received the PHD in Electronics and Signal Processing from the Montpellier University, France 1982. He was with Fairchild CCD company for two years as engineer specialized on CCD sensors. He joined the Burgundy University in 1985 as "Maitre de Conferences" and is currently full professor and director of LE2I UMR-CNRS, Laboratory of Electronic, Computing and Imaging Sicences, Burgundy University, France. His main research topics is image acquisition and real time image processing. He is also a member of ISIS (a resarch group in Signal and Image Processing of the French National Scientific Research Committee).

Hervé Abdi received an M.S. in Psychology from the University of Franche-Comté (France) in 1975, an M.S. (D.E.A.) in Economics from the University of Clermond-Ferrand (France) in 1976, an M.S. (D.E.A.) in Neurology from the University Louis Pasteur in Strasbourg (France) in 1977, and a Ph.D. in Mathematical Psychology from the University of Aix-en-Provence (France) in 1980. He was an assistant professor in the University of Franche-Comté (France) in 1979, an associate professor in the University of Bourgogne at Dijon (France) in 1983, a full professor in the University of Bourgogne at Dijon (France) in 1988. He is currently a full professor in the School of Human Development at the University of Texas at Dallas. He was a visiting associate professor of Cognitive and Linguistic Sciences at Brown University in 1986 and 1987 and a Fulbright scholar. His interest includes neural network and cognitive modeling, experimental design and statistics.