Développement d'un système rapide pour le mosaïquage et la reconnaissance de visages panoramiques

Development of a fast panoramic face mosaicing and recognition system

F. Yang, M. Paindavoine, D. Ginhac, J. Dubois
Laboratoire Le2i, Aile de l'Ingénieur - Mirande Université de Bourgogne, BP 400 - 21011 Dijon cedex
Tél : 03 80 39 36 08, Fax : 03 80 39 59 10
fanyang@u-bourgogne.fr, paindav@u-bourgogne.fr

Manuscrit reçu le 24 janvier 2005

Résumé et mots clés

Dans cet article, nous présentons quelques résultats sur le développement d'un système de mosaïquage de visages panoramiques. Notre objectif est d'étudier la faisabilité de construction de visages panoramiques en temps réel. Ceci nous a conduit tout d'abord à concevoir un système d'acquisition très simple, composé de 5 caméras standards qui réalise la prise de 5 vues simultanément sous différents angles. Puis, nous avons choisi un algorithme facilement implantable sur des systèmes embarqués. Cet algorithme est basé sur des transformations linéaires successives, pour composer un visage panoramique de 150° à partir de ces 5 vues. La méthode a été testée sur une centaine de visages. Nous avons aussi effectué une étude préliminaire sur la reconnaissance de visages panoramiques dans le but de valider notre système de mosaïquage de visages. La reconnaissance est basée sur le modèle de «visages propres». Les résultats expérimentaux ont montré la faisabilité et la viabilité du système proposé permettant d'envisager une future implantation matérielle. Nous pensons aussi utiliser notre système de mosaïquage dans d'autres applications comme la reconstruction rapide de visages 3D et la catégorisation des expressions basée sur le mouvement.

Vision panoramique, mosaïquage d'images, reconnaissance de visages, analyse en Composantes Principales, FFT.

Abstract and key words

In this article, we present some development results of a system that performs mosaicing of panoramic faces. Our objective is to study the feasibility of panoramic face construction in real-time. This led us to conceive of a very simple acquisition system composed of 5 standard cameras and 5 face views taken simultaneously at different angles. Then, we chose an easily hardware-achievable algorithm: successive linear transformation, in order to compose a panoramic face of 150° from these 5 views. The method has been tested on hundreds 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 «eigenfaces» method. Experimental results obtained show the feasibility and viability of our system. This allows us to envisage later a hardware implantation. We also are considering applying our system to other applications such as human expression categorization using movement estimation and fast 3D face reconstruction.

Panoramic vision, image mosaicing, face recognition, principal Component Analysis, FFT.

Remerciements

Nous remercions Monsieur Dominique Arnoult du laboratoire LE2I pour la réalisation technique du système d'acquisition et la constitution de la base de données d'images.

1. Introduction

La biométrie est actuellement un champ de recherches très actif qui concerne plusieurs disciplines comme par exemple, le traitement d'images, la reconnaissance de formes et la vision artificielle. Le principal objectif de la biométrie est d'établir des systèmes capables d'identifier les personnes en fonction des caractéristiques intrinsèques comme le visage, les empreintes digitales, l'iris, etc. La reconnaissance de visages est un sujet beaucoup étudié dans le domaine de la biométrie par le fait que les êtres humains utilisent quotidiennement l'information faciale pour se reconnaître entre eux. Différentes techniques ont été proposées pour reconnaître automatiquement les visages. Nous pouvons citer les réseaux de neurones[1][2], les visages propres [3][4], et les champs de Markov[5]. Comme le « *vendor* test » [6][7][8] le montre, la plupart de ces modèles utilisent seulement le visage de face comme le signal d'entrée. Par conséquent, beaucoup de méthodes sont sensibles aux poses et aux conditions d'éclairage. Une des pistes pour surmonter ces limites est de combiner plusieurs modalités : couleur, profondeur, surface faciale 3D ... [9][10][11][12][13].

La plupart des systèmes d'acquisition 3D utilisent des appareils professionnels comme la caméra « *traveling* » ou le scanner 3D[11][12][14]. Typiquement, ces systèmes exigent que le sujet reste immobile pendant quelques secondes pour avoir des informations 3D. Ils ne sont pas adaptés aux applications en temps réel comme la catégorisation des expressions faciales basée sur le mouvement. En plus, leur coût est souvent prohibitif pour des applications orientées grand public. Pour simplifier ces systèmes d'acquisition 3D, certains modèles de reconnaissance de visages génèrent des visages 3D à partir de la stéréovision [15]. Dans ce cas, des calculs relativement complexes sont nécessaires afin d'assurer l'auto calibrage et l'extraction des informations 3D [16]. Une autre possibilité est d'obtenir des informations 3D à partir d'un ensemble d'images de visages 2D, sans construire la structure complète de visages 3D[9][17].

Depuis une dizaine d'années, notre laboratoire travaille sur le traitement automatique d'images de visages et a obtenu des résultats pour la détection, la reconnaissance et le suivi dynamique de visages 2D [18][19][20][21]. L'un des objectifs du projet « visages panoramiques » est d'étendre nos modèles de traitement 2D vers des traitements d'images panoramiques (3D). Dans cet article, nous décrivons le développement d'un système simple et efficace, permettant de traiter des visages 3D en temps réel. La méthode que nous proposons construit des visages panoramiques donnant des informations sur la surface 3D. Le système d'acquisition composé de 5 caméras standards permet d'obtenir 5 vues d'un visage simultanément. Ce système d'acquisition possède des avantages comme la simplicité de mise en oeuvre et le très faible coût.

Dans ce papier, nous présentons d'abord, le système d'acquisition pour les 5 vues. Ensuite, nous décrivons la méthode de création des visages panoramiques basée sur une transformation linéaire successive. Nous donnons aussi des résultats expérimentaux de la reconnaissance de visages panoramiques avant de conclure.

2. Présentation du système d'acquisition

Notre système d'acquisition est composé de 5 caméras USB logitech 4000 avec une résolution maximale de 640×480 pixels chacune. Ces caméras sont fixées sur une glissière circulaire réglable en hauteur pour pouvoir s'adapter à la taille du sujet (voir la figure 1 gauche). Les paramètres de chaque caméra (Format des images, Luminosité, Contraste ...) peuvent être ajustés indépendamment. Le programme d'acquisition (sous Visual C++) déclenche la capture de 5 images simultanément (voir la figure 1 droite). Ces images sont ensuite stockées sur le disque dur du PC avec un flot de données de $20 \times 5 = 100$ images par seconde.

Le sujet humain est placé devant le banc d'acquisition, face à la caméra centrale (caméra 3). Il possède sur son visage des marqueurs de différentes couleurs qui permettent de récupérer plus tard des points communs entre chaque vue du visage. La position de ces marqueurs correspond approximativement aux points caractéristiques du visage (coin des yeux, menton ...). La figure 2 montre la position des points choisis. Au total, il y a 10 marqueurs sur chaque visage avec au moins 3 marqueurs en commun entre chaque vue.

Avec la caméra utilisée, selon le positionnement de la personne, 1 pixel correspond à une surface qui varie entre 0.7 mm^2 et 1 mm^2. Chaque marqueur a une surface équivalente de 20 mm^2 et contient donc au moins 20 pixels. Notre méthode de localisation des marqueurs basée sur un calcul de centres de gravités (voir la section 3.1) garantit une précision minimale au pixel près, soit une erreur de la localisation inférieure à 1 mm^2 ce qui est suffisant pour notre application.

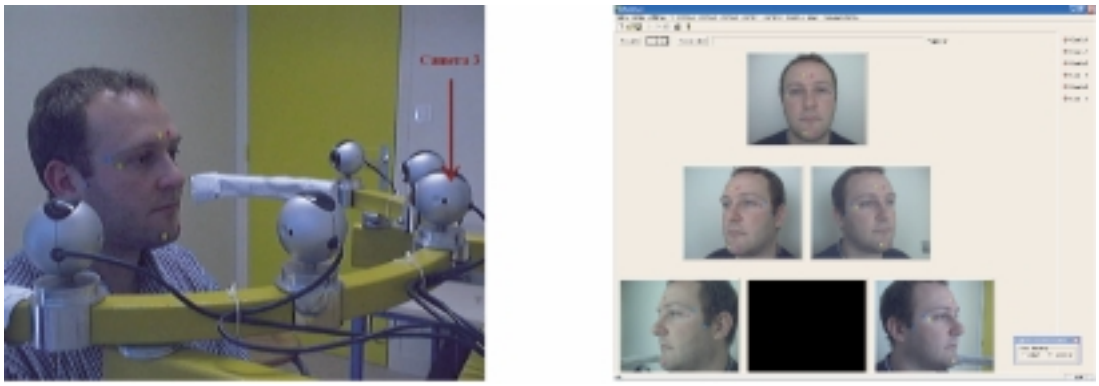

Figure 1. Système d'acquisition : la figure gauche montre les 5 caméras et leur support, la figure droite affiche les 5 images prises avec un sujet. La taille de chaque image est de 240 × 320 *pixels.*

Figure 2. Répartition des 10 marqueurs sur les 5 vues (dans le sens des aiguilles d'une montre à partir du haut à gauche) : image 1, image 2, image 3, image 4, et image 5.

3. Construction de visages panoramiques

Plusieurs algorithmes de construction de mosaïques ont été proposés dans la littérature [17] [22] [23] [24] [25] [26] [27]. A. Jain et A. Ross [25] ont développé une technique de mosaïquage d'images permettant de construire une image plus complète d'empreintes digitales à partir de deux images d'une même empreinte. Pour cela, ils alignent d'abord les 2 images en utilisant les points communs de minutie. Ensuite, une version modifiée de l'algorithme ICP (*Iterative Closest Point*) [28] est appliquée sur cet alignement afin de calculer la matrice de transformation qui définit la relation spatiale entre les deux images. À la fin du traitement, une image composite est générée grâce à la matrice de transformation qui possède 6 paramètres indépendants : 3 angles de rotation (α, β, γ) autour des axes x, y, et z et 3 composantes de translation (t_x, t_y, t_z) suivant ces 3 axes.

X. Liu et T. Chen [17] ont utilisé la géométrie faciale pour améliorer les résultats de mosaïquage de visages. Ils ont créé une projection sphérique (au lieu de la projection cylindrique) qui représente mieux le mouvement de la tête suivant les deux directions (horizontale et verticale). Un algorithme de « *matching* » géométrique a été développé pour décrire la correspondance de l'image 2D entre l'espace plane QUV et la surface sphérique $O\alpha\beta$. Le point Q se situe à la pointe du nez d'un visage, QU, QV représentent respectivement les axes horizontaux et verticaux du plan 2D QUV. Le point d'origine O, l'angle horizontal α et l'angle vertical β constituent les coordonnées de l'espace sphérique $O\alpha\beta$.

L'algorithme *Levenberg-Marquardt* a été utilisé pour trouver les deux paramètres de « *matching* » $[\Delta\alpha, \Delta\beta]^T$ afin d'accomplir la transformation depuis l'espace 2D en surface sphérique. Pour

réduire le volume de calcul, X. Liu et T. Chen approximent un visage par un ensemble de triangles en appliquant une technique de maillage.

En général, les méthodes utilisant des transformations non linéaires et des algorithmes récursifs obtiennent des résultats très corrects en terme de précision géométrique. En revanche, ces méthodes demandent des volumes de calculs relativement importants et leur réalisation matérielle en temps réel n'est pas toujours facile. Notre objectif pour ce projet « visages panoramiques » est d'aboutir à un système en temps réel. Nous avons décidé d'appliquer une transformation linéaire simple, mais efficace pour construire le mosaïquage des visages. Cet algorithme est divisé en 3 étapes (voir la figure 3) :

1. Détection des marqueurs et calcul de leurs coordonnées,
2. Estimation de la matrice de transformation et transformation de l'image,
3. Construction du visage panoramique.

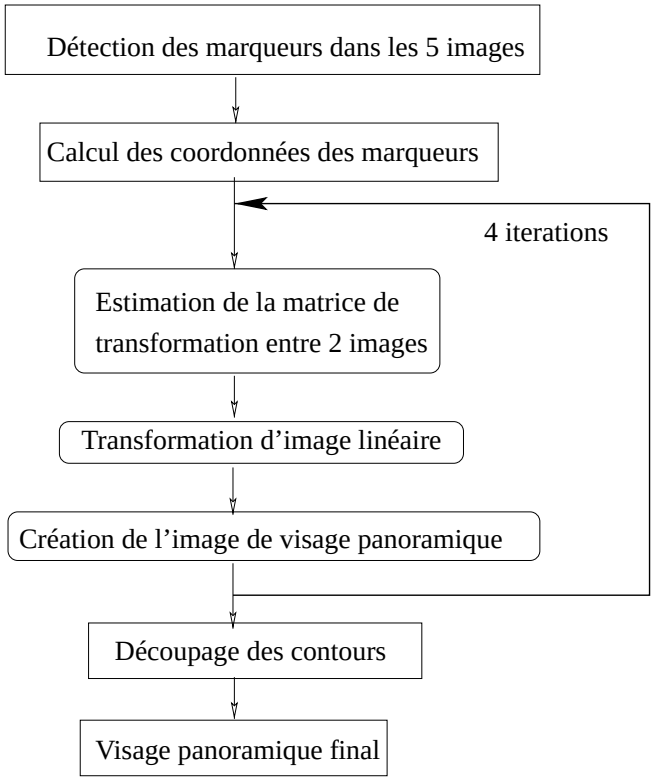

Figure 3. Construction de visages panoramiques basée sur la transformation linéaire successive.

3.1. Détection des marqueurs et calcul de leurs coordonnées

La première étape de la méthode proposée consiste à détecter les marqueurs sur le visages de sujets. Ces marqueurs sont fabriqués avec des papiers autocollants. Nous avons utilisé 3 couleurs pour les 10 marqueurs (4 bleus, 3 jaunes et 3 violets, voir la figure 2 pour une illustration). Ces trois couleurs ont été choisies car elles sont facilement identifiables par rapport à la couleur de peau. Ces marqueurs sont utilisés comme des points de référence pour coller les différentes vues d'un visage.

Nous avons effectué une segmentation des couleurs en utilisant les composantes Teinte et Saturation pour chaque image. Cette procédure permet une forte sélection des couleurs et en même temps une faible sensibilité à la variation de luminance. La figure 4 illustre les étapes de la détection des marqueurs jaunes. Tout d'abord, à partir de l'image d'origine (Figure 4 à gauche), la segmentation de couleurs nous donne une image binaire qui contient des marqueurs (Figure 4 au milieu). Ensuite, nous calculons le centre des zones détectées (Figure 4 à droite). Ces centres nous permettent de trouver les coordonnées des marqueurs dans l'image.

3.2. Estimation de la matrice de transformation et Transformation linéaire de l'image

Nous avons décidé de représenter un visage comme une mosaïque. Chaque mosaïque de visage est fabriquée par la concaténation de différentes vues collées sur une surface plane. Ainsi, pour créer un visage panoramique, les 5 vues sont combinées l'une après l'autre. Nous commençons par la vue centrale et collons une vue latérale à chaque étape (voir la figure 3). Notre méthode consiste à transformer l'image latérale et à ajuster celle-ci avec la vue centrale de manière à ce que les points communs de ces deux vues se superposent.

Nous obtenons l'image transformée par une multiplication de l'image avec la matrice de transformation. Cette matrice de transformation est estimée en fonction des coordonnées des 3 marqueurs communs entre les deux vues. $\mathbf{C}_1$ et $\mathbf{C}_2$ représentent, respectivement, les coordonnées des marqueurs de la première et de la seconde image :

$$\mathbf{C}_1 = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} \tag{1}$$

$$\mathbf{C}_2 = \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \end{bmatrix} \tag{2}$$

Nous obtenons la matrice de transformation comme suit :

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

avec

$$\mathbf{C}_2^\star = \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ 1 & 1 & 1 \end{bmatrix} \tag{4}$$

et

$$\mathbf{T} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \tag{5}$$

Figure 4. Détection des marqueurs jaunes (de gauche à droite) : image d'origine, image binaire obtenue avec le filtrage de couleurs sur les composants Teinte et Saturation et la localisation des marqueurs.

Figure 5. Image 4 avant (gauche) et après (droite) la transformation : la matrice de transformation est estimée en utilisant l'image 4 et l'image 3 (vue centrale, voir les figures 2 et 6).

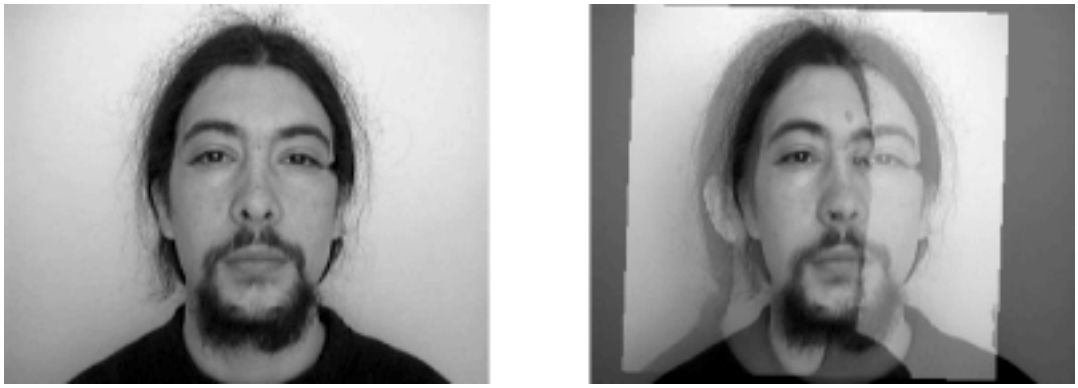

Figure 6. Superposition de l'image 3 et de l'image 4 : image 3 d'origine (gauche), et superposition de l'image 4 transformée et de l'image 3 d'origine (droite).

Puis, nous généralisons cette transformation sur l'image entière :

$$x = a_1 x' + b_1 y' + c_1$$
$$y = a_2 x' + b_2 y' + c_2 \qquad (6)$$

Cette transformation linéaire correspond à une combinaison de la rotation, de la translation et de la dilatation de l'image (voir la figure 5). La figure 6 affiche le résultat de la superposition de l'image 3 (non transformée) et de l'image 4 (transformée en utilisant les coordonnées des marqueurs jaunes).

3.3. Construction de visage panoramique

Nous commençons la construction du visage panoramique par la vue centrale (image 3 voir la figure 2). A partir de la superposition de l'image 3 d'origine et de l'image 4 transformée (voir la figure 6), nous supprimons les pixels redondants afin d'obtenir une image panoramique temporaire image 3-4 (voir la figure 7 gauche). Nous avons créé deux lignes de découpage à partir des deux marqueurs jaunes pour éliminer les pixels redondants. Cette image 3-4 devient temporairement notre image cible.

Nous répétons cette opération sur chaque vue. D'abord, l'image 2 est collée avec l'image 3-4 afin d'obtenir une nouvelle image panoramique temporaire (image 2-3-4) (voir la figure 7 droite). La matrice de transformation est générée en utilisant les 3 marqueurs violets qui sont communs avec les deux vues. Ensuite, nous calculons la matrice de transformation permettant de construire l'image 2-3-4-5 (voir la figure 8 gauche) en utilisant les 2 marqueurs bleus et le marqueur jaune le plus haut. Finalement, l'image 1 est collée avec l'image 2-3-4-5 panoramique temporaire à l'aide des 2 marqueurs bleus et le marqueur violet le plus haut (voir la figure 8 droite).

La figure 9 gauche affiche le visage panoramique final construit à partir des 5 vues. La construction préserve la forme du visage. Par exemple, le menton possède plus de courbures que les autres parties du visage, la partie basse du visage panoramique est composée avec les 5 vues : 1, 2, 3, 4 et 5. En revanche, pour composer la partie haute, du visage (front), les 3 vues suffisent (1, 3 et 5).

La figure 9 droite montre la mosaïque finale du visage obtenue après le découpage automatique des contours. Pour ce processus, nous appliquons d'abord un cercle qui englobe la zone de visage (cheveux compris) et nous effaçons la zone de fond qui se situe à l'extérieur de ce cercle. En parcourant, à l'intérieur de ce cercle, de gauche à droite et de droite à gauche, les lignes centrales, nous détectons automatiquement les points extrêmes des deux oreilles et en même temps, nous réduisons le rayon du cercle. Ensuite, nous remplaçons les segments de ce cercle par des courbes polynomiales en utilisant les coordonnées des points extrêmes localisés à l'aide de la position des marqueurs.

Par rapport à la méthode proposée par Liu et Chen [17], les visages panoramiques obtenus avec notre système sont moins

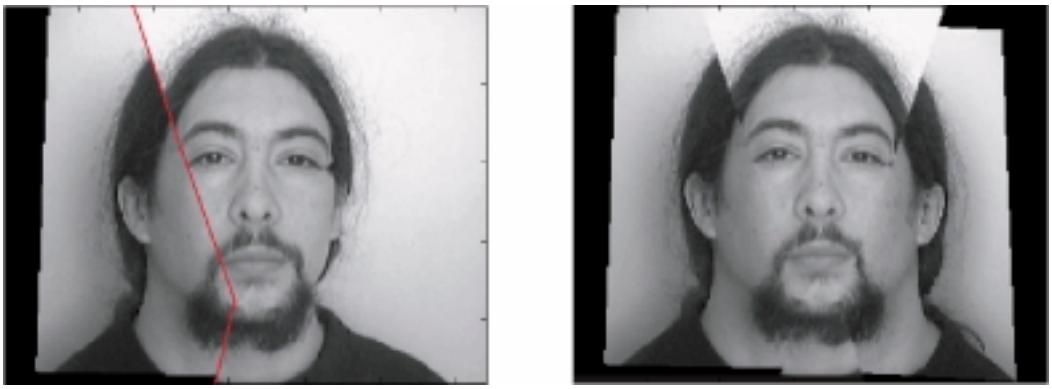

Figure 7. Résultats du mosaïquage : image 3-4 (gauche), et image 2-3-4 (droite).

Figure 8. Résultats du mosaïquage : image 2-3-4-5 (gauche), et image 1-2-3-4-5 (droite).

Figure 9. Résultats du mosaïquage : visage panoramique construit à partir des 5 vues (gauche), et mosaïque finale (droite).

précis géométriquement. Par exemple, Liu et Chen utilisent un maillage de triangles pour représenter un visage. Chaque triangle possède ses propres paramètres de transformation obtenus à l'aide d'un algorithme récursif. Dans notre système, une matrice unique de transformation est appliquée sur une image entière. Notre objectif est d'étudier un algorithme simple, mais efficace pour de futures implantations matérielles. Le système proposé crée un visage panoramique à partir de 5 vues en 1.2 seconde avec le logiciel Matlab (Pentium 4B, 2.4 GHz). Ceci nous permet d'envisager une application en temps réel sur des systèmes embarqués.

Afin de tester et de valider notre algorithme de mosaïquage de visage panoramique, nous proposons dans la section suivante, une étude de reconnaissance de visages basée sur le modèle de « visages propres » proposé par Turk et Pentland [3].

4. Reconnaissance des visages panoramiques

Nous avons créé une base de données de visages panoramiques en utilisant la méthode décrite dans la section 3. Cette base de données est composée de 12 personnes × 4 expressions × 2 sessions = 96 visages panoramiques. Les deux sessions d'acquisition ont été réalisées dans un intervalle d'un mois. Les 4 expressions de visages sont les suivantes : neutre, souriant, soucieux, endormi (voir la figure 10). Dans le but de valider notre système de mosaiquage de visages panoramiques, nous avons implémenté un algorithme de reconnaissance de visages que nous avons testé en utilisant notre base de données. Nous présentons successivement dans cette section notre algorithme de reconnaissance de visages et les résultats expérimentaux obtenus à partir de notre base de données de visages panoramiques.

4.1. Description de l'algorithme de reconnaissance de visages : ACP

Durant les 25 dernières années, plusieurs techniques de reconnaissance de visages ont été proposées en relation avec un besoin croissant d'applications liées à la sécurité mais aussi avec le développement des approches de modélisation, utilisées en sciences cognitives, pour étudier le comportement humain. L'une des techniques les plus utilisées est basée sur l'Analyse en Composantes Principales (ACP) adaptée aux images de visages [29]. Par exemple, Abdi [30], Turk et Pentland [3] ont utilisé des vecteurs propres de visages pour la détection et l'identification de visages 2D. L'Analyse en Composantes Principales est basée sur l'idée que la reconnaissance de visages peut être réalisée avec un faible nombre de données qui caractérisent d'une manière optimale un ensemble de visages connus (appris). L'application de cette méthode pour la reconnaissance de visages consiste à réaliser tout d'abord l'ACP sur un jeu d'images de visages connus (chaque personne est caractérisée par plusieurs images de visages correspondant à différentes expressions). A partir de cette analyse, un jeu de K composantes principales est obtenu. La projection de nouveaux visages sur ces composantes est réalisée dans le but de calculer les distances entre les nouveaux visages et les visages déjà appris. Ces distances permettent de vérifier si les visages en cours d'analyse sont des nouveaux visages (visages inconnus) ou des visages déjà appris. L'ACP appliquée à la reconnaissance de visages fonctionne de la manière suivante. Les visages à apprendre sont représentés par des vecteurs $\mathbf{a}_k$ où k est le numéro du visage [31]. Ces vecteurs sont obtenus en transformant l'image 2D des visages en une image 1D par la vectorisation d'une matrice. L'ensemble des images de visages transformées en vecteurs est rassemblé au sein de la matrice d'apprentissage $\mathbf{A}$ dans laquelle chaque colonne correspond à un visage.

En utilisant les propriétés de l'algèbre linéaire, cette matrice d'apprentissage peut être exprimée comme :

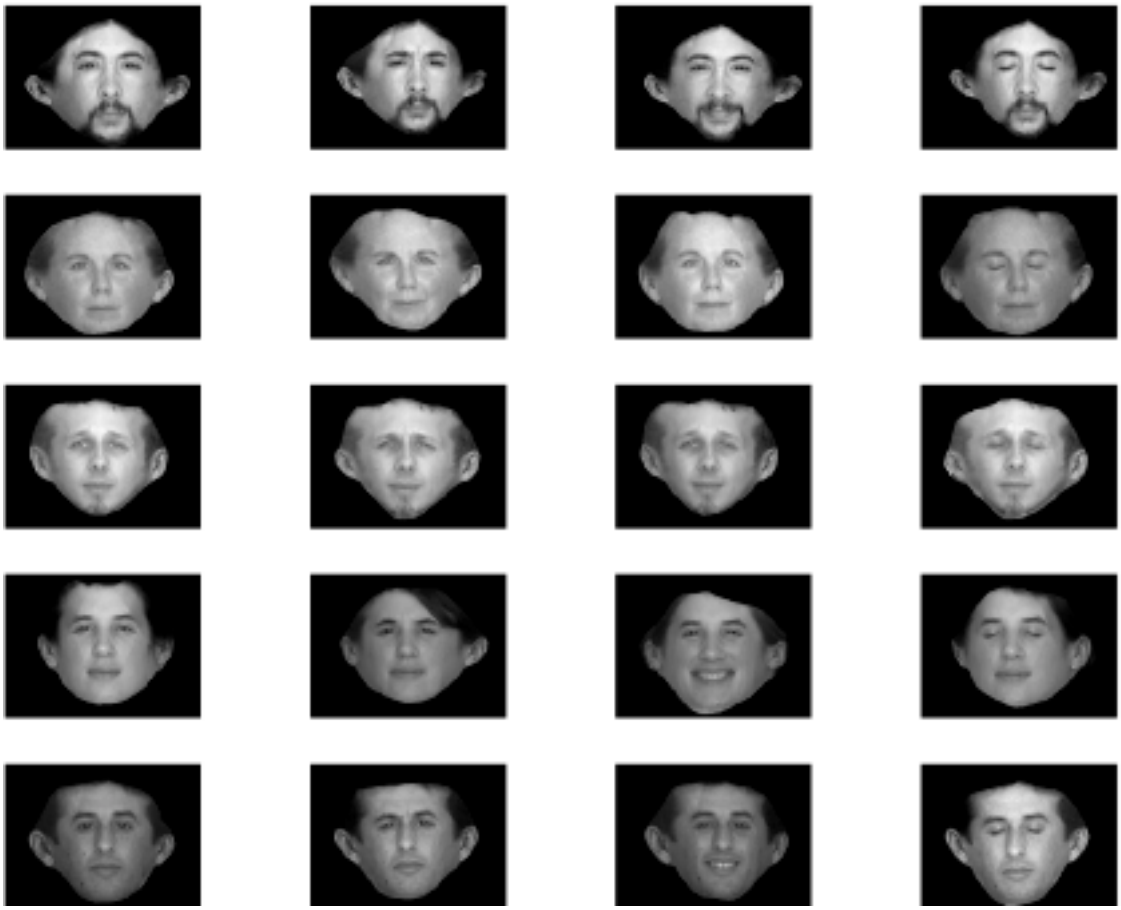

Figure 10. Exemples de visages panoramiques de la base de données obtenus lors de la première session de mesures.

$$\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}^T \tag{7}$$

où **P** est la matrice des vecteurs propres de $\mathbf{AA}^T$, **Q** est la matrice des vecteurs propres de $\mathbf{A}^T\mathbf{A}$, et Δ est la matrice diagonale des valeurs singulières de **A**, $\Delta = \Lambda^{1/2}$ avec Λ qui est la matrice des valeurs propres de $\mathbf{AA}^T$ et $\mathbf{A}^T\mathbf{A}$.

En considérant la matrice des vecteurs propres (visages propres), chaque visage de la base de données est représenté par la projection de l'image de visage sur la matrice des vecteurs propres :

$$PROJ_x = \mathbf{x}^T\mathbf{P}\Delta^{-1} \tag{8}$$

où **x** est une image de visage dépliée sous forme de vecteur. Il est à noter que l'approche des visages propres permet d'aligner et de normaliser des images pour l'apprentissage et la reconnaissance.

L'identification d'un visage est obtenue en localisant, dans la base de données, l'image d'un visage connu dont le vecteur de projection a la plus petite distance Euclidienne avec celui de l'image testée. Cet algorithme proposé par Turk and Pentland [3] est nommé «*Règle de classification des plus proches voisins*». La figure 11 représente la projection de la matrice d'apprentissage **A** sur les deuxième et troisième composantes principales. Dans cet exemple, 48 visages panoramiques sont analy-

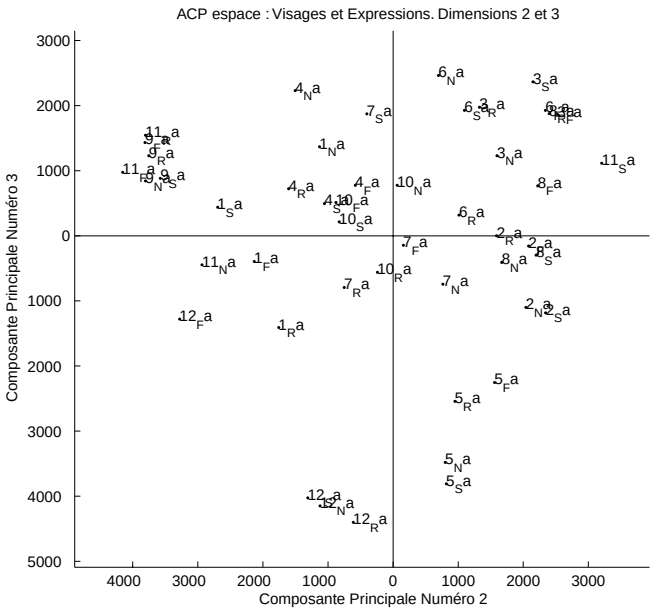

Figure 11. Représentation 2D de la projection de la matrice d'apprentissage **A** *sur les deuxième et troisième composantes principales ; chaque visage panoramique est caractérisé par son identité (numéros 1-12), son expression (N,S,R,F), et sa session d'acquisition (a, b).*

sés. Chaque visage panoramique est caractérisé par son identité (12 personnes numérotées de 1 à 12), par son expression (4 expressions différentes codées comme : N-neutre, S-sourire, R-soucieux, F-yeux fermés) et par le numéro de session d'acquisition d'images de visages (deux sessions a et b acquises à un mois d'intervalle). Dans cet exemple nous présentons uniquement le traitement des visages acquis lors de la session a.

4.2. Résultats expérimentaux de la reconnaissance de visages panoramiques

4.2.1. Représentation spatiale

Pour ces tests, nous avons utilisé des images de visages panoramiques avec un format de 240×320 pixels (représentation dans le domaine spatial). Nous avons constitué une base de données de 12 personnes, ce qui correspond à 12×4 expressions $\times$ 2 sessions $= 96$ visages panoramiques. Nous avons divisé notre base de données en deux blocs. Le premier bloc est utilisé comme base d'apprentissage et le deuxième permet de tester les performances de reconnaissance de notre système. Comme cela est illustré en figure 10, les images de visages panoramiques ont été acquises avec un fond uniforme et ceci sous un éclairage de jour.

À partir de notre base de données de 96 visages panoramiques, nous choisissons d'une manière aléatoire, pour chacune des 12 personnes, des images de visages panoramiques. Nous avons réalisé 4 expériences différentes basées sur le nombre de visages appris pour chacune des 12 personnes. Ce nombre p varie de 1 à 4 images de visages par personne et constitue le premier bloc (bloc d'apprentissage). Le reste de la base de données constitue le deuxième bloc et est utilisé comme base de visages testés en mode reconnaissance. Par exemple, quand $p = 1$, le nombre de visages pour l'apprentissage est égal à 1×12 personnes = 12 et le nombre de visages qui sont utilisés dans ce cas pour tester notre algorithme de reconnaissance est égal à $96 - 12 = 84$.

Nous présentons en Tableau 1 les résultats de reconnaissance de visages que nous avons obtenus avec notre programme Matlab et ceci pour les 4 expériences ($p = 1$ à 4). Nous avons ajouté à la phase d'apprentissage une étape d'analyse discriminante [32] permettant de déterminer le nombre de vecteurs propres utiles pour la reconnaissance de visages. On ré-ordonne les vecteurs propres non pas en fonction de leurs valeurs propres mais en fonction de leur importance pour la reconnaissance (voir [33] pour la description technique). Pour l'exemple montré à la figure 12, on voit bien que 23 vecteurs propres suffisent pour avoir le taux de reconnaissance maximum (93.75 %), les autres vecteurs propres n'apportant rien de plus à la reconnaissance.

4.2.2. Représentation fréquentielle

Nous avons testé aussi le comportement fréquentiel de notre système de reconnaissance de visages panoramiques. La figure 13 gauche représente le spectre d'amplitude d'un visage panoramique. Nous pouvons observer que ce spectre est centré autour des basses fréquences ce qui nous permet d'appliquer un filtre passe-bas dans le but de réduire la taille des donnés à traiter (voir la figure 13 droite). Ainsi, seulement 80×80 données du spectre d'amplitude sont utilisées par notre système de reconnaissance.

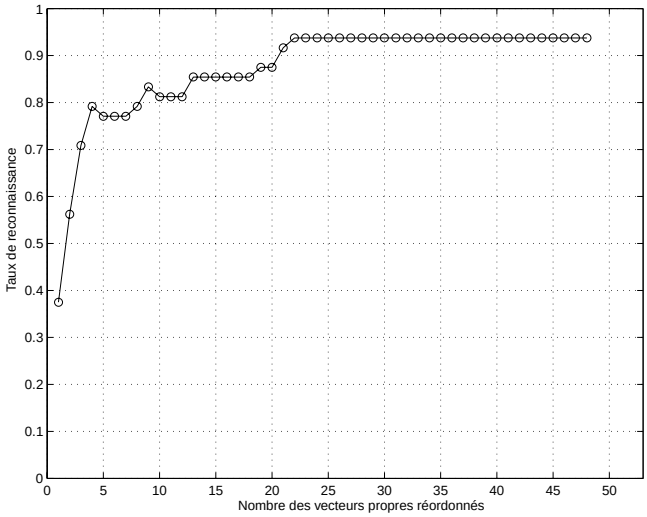

Figure 12. Résultat d'une analyse discriminante pour une base d'apprentissage de 48 visages : nombre de vecteurs propres ré-organisés en fonction de leur importance pour la reconnaissance.

Tableau 1. Résultats de la reconnaissance de visages panoramiques représentés dans le domaine spatial : le nombre de vecteurs propres utilisés correspond à la valeur moyenne obtenue pendant plusieurs exécutions avec l'analyse discriminante.

Nombre exemples appris/personne	$p = 1$	$p = 2$	$p = 3$	$p = 4$
Nombre total d'exemples appris	12	24	36	48
Nombre de vecteurs propres de visages utilisés	9	13	18	25
Nombre de tests pour la reconnaissance	84	72	60	48
Taux de reconnaissance correcte	70 %	85.08 %	90.1 %	93.21 %

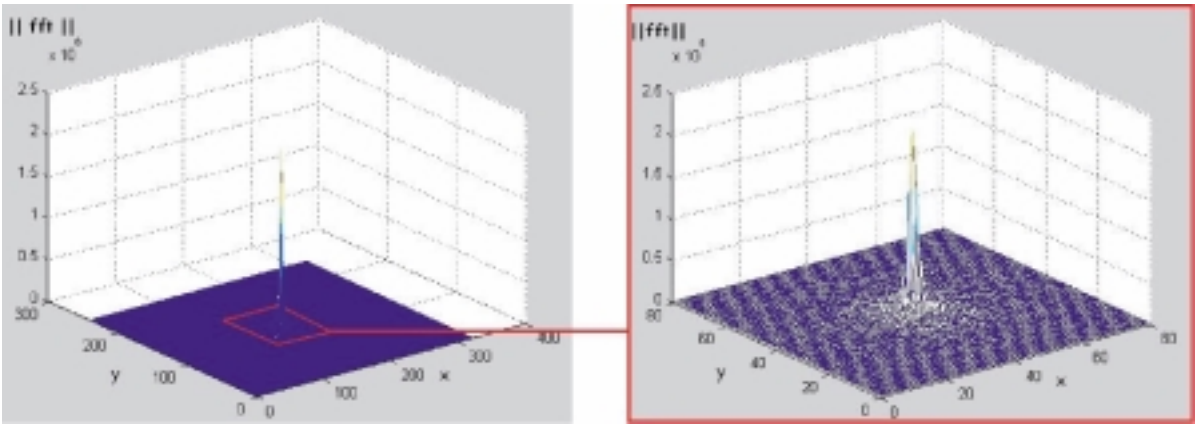

Figure 13. Représentation 3D du spectre d'amplitude original (gauche), et spectre d'amplitude avec les fréquences faibles uniquement $80 \times 80 = 6400$ *valeurs (droite) qui sont utilisées par notre système de reconnaissance.*

Tableau 2. Résultats de la reconnaissance de visages panoramiques avec la représentation fréquentielle : le nombre de vecteurs propres utilisés correspond à la valeur moyenne obtenue pendant plusieurs exécutions avec l'analyse discriminante.

Nombre exemples appris/personne	$p = 1$	$p = 2$	$p = 3$	$p = 4$
Nombre total d'exemples appris	12	24	36	48
Nombre de vecteurs propres de visages utilisés	8	13	18	24
Nombre de tests pour la reconnaissance	84	72	60	48
Taux de reconnaissance correcte	76.83 %	91.26 %	93.25 %	97.46 %

Nous avons réalisé les mêmes tests que ceux que nous avons effectué avec la représentation spatiale. Les résultats sont présentés dans la Table 2. En comparaison avec la représentation spatiale, nous obtenons un meilleur taux de reconnaissance avec la représentation fréquentielle, limitée au spectre d'amplitude (97.46 % en comparaison avec 93.21 % pour la représentation spatiale).

Ceci est dû au fait que pour les images de visages, le spectre d'amplitude est moins sensible au bruit (ou aux variations) que le spectre de phase. Pour illustrer ce phénomène, nous avons bruité volontairement une image de visage panoramique. La figure 14a représente une image d'un visage panoramique et la figure 14b représente ce même visage avec un ajout de bruit. Nous avons calculé la transformée de Fourier directe de ces deux images puis la transformée de Fourier inverse selon deux configurations différentes :

1) En utilisant le spectre d'amplitude de l'image bruitée et le spectre de phase de l'image originale (voir la figure14c),

2) En utilisant le spectre de phase de l'image bruitée et le spectre d'amplitude de l'image originale (voir la figure 14d).

Ces résultats montrent que le visage obtenu avec la première configuration est plus proche du visage original que celui obtenu avec la deuxième configuration. Ceci confirme que l'image de visage obtenue uniquement à partir du spectre d'amplitude est moins sensible au bruit que celle obtenue directement dans le domaine spatial (l'image spatiale initiale possède naturellement les spectres d'amplitude et de phase). Par ce fait les performances en reconnaissance sont supérieures avec la représentation fréquentielle limitée au spectre d'amplitude en comparaison avec les performances obtenues avec la représentation spatiale qui contient les informations en amplitude et en phase.

Tsalakanidou *et al.* [9] ont évalué trois approches pour la reconnaissance de visages en utilisant les facteurs suivants : la couleur, la profondeur et la combinaison des deux. Ils ont quantifié la contribution de la mesure de profondeur pour la reconnaissance de visages. Les images de couleurs sont stockées avec une résolution de 720×576 pixels. Pour chaque personne, ils ont créé une carte de profondeur à l'aide d'une approche « éclairage structuré » qui capte la surface 3D d'un visage. Leur expériences montrent qu'un gain significatif de 5 % sur le taux de reconnaissance a été apporté par l'information 3D. La technique de reconnaissance utilisée est basée sur l'approche ACP (similaire par rapport à la notre). Ils ont obtenu un taux de reconnaissance de 97.5 % en combinant les 2 facteurs (couleur et profondeur). Nous avons établi un système simple et efficace pour construire un visage panoramique à partir de 5 vues et les performances de reconnaissance de visages sont très proches de celles annoncées par Tsalakanidou *et al.*

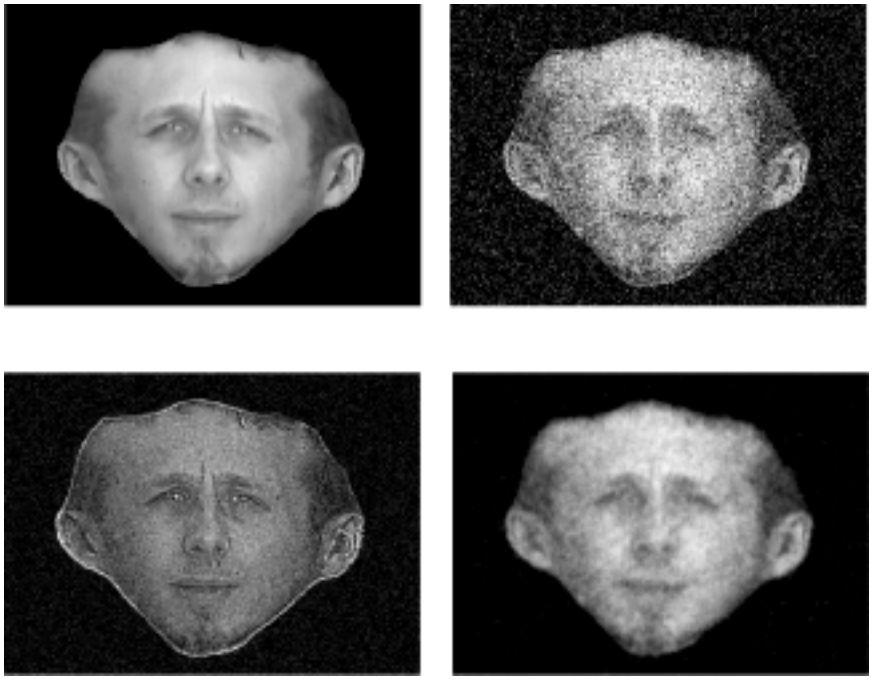

Figure 14. Le spectre d'amplitude est moins sensible au bruit que le spectre de phase : **a**) *image originale de visage panoramique (en haut à gauche) ;* **b**) *image originale avec un ajout de bruit blanc gaussien (moyenne=0 et variance=0.05) (en haut à droite) ;* **c**) *FFT inverse de l'image en utilisant le spectre d'amplitude de* **b**)*et le spectre de phase de* **a**) *(en bas à gauche) ;* **d**) *FFT inverse de l'image en utilisant le spectre d'amplitude de* **a**) *et le spectre de phase de* **b**) *(en bas à droite). Nous pouvons constater que l'image* **c**) *ressemble plus à l'image* **a**) *que l'image* **d**).

Figure 15. Visages panoramiques de quatre personnes non apprises par le système de reconnaissance.

4.2.3. Reconnaissance de visages panoramiques avec des échantillons négatifs

Dans le but de tester le comportement de notre système de reconnaissance de visages avec des personnes inconnues (non apprises), nous avons ajouté 4 nouvelles personnes dans notre base de données (voir la figure 15). Ces visages panoramiques ont été obtenus selon la méthode décrite dans la section 3.

Dans ce cas, les performances de notre système sont décrites dans la Table 3. Nous avons ajouté 4 personnes $\times$ 4 expressions $\times$ 2 sessions $= 32$ nouveaux visages panoramiques dans chaque jeu de test. Dans le but de rejeter ces visages inconnus, nous avons introduit un seuillage sur le calcul de la distance Euclidienne. Comme nous envisageons d'utiliser notre système pour des applications de contrôle d'accès pour lesquelles la confusion est plus grave que la non-reconnaissance, notre seuil de validation de visage connu est élevé et ceci dans le but d'éliminer les intrusions. Pour tous les tests, le seuil de validation est resté constant. Pour cette étude préliminaire de reconnaissance de visages panoramiques, ce seuil est imposé par l'utilisateur de façon expérimentale.

Afin de qualifier notre système, nous avons utilisé les paramètres suivants :

Taux de reconnaissance : reconnaissance correcte des visages panoramiques appris,

Taux de Non-reconnaissance : visage panoramique appris non reconnu,

Taux de Confusion : intrus confondu avec un visage panoramique appris.

Tableau 3. Résultats de la reconnaissance de visages panoramiques avec des exemples négatifs: le nombre de vecteurs propres utilisés correspond à la valeur moyenne obtenue pendant plusieurs exécutions avec l'analyse discriminante. Ces performances sont obtenues en utilisant la représentation fréquentielle.

Nombre exemples appris/personne	$p = 1$	$p = 2$	$p = 3$	$p = 4$
Nombre total d'exemples appris	12	24	36	48
Nombre de vecteurs propres de visages utilisés	8	13	18	24
Nombre de tests pour la reconnaissance	116	104	92	80
Taux de Non-reconnaissance	25.4 %	12.74 %	7.58 %	4.82 %
Taux de Confusion	5.85 %	4 %	3.5 %	2.8 %
Taux de reconnaissance	68.75 %	83.26 %	88.92 %	92.38 %

5. Conclusions et Perspectives

Dans cet article, nous avons proposé une méthode simple et rapide pour réaliser un mosaiquage d'image. Le système d'acquisition d'images est constitué d'un ensemble de cinq caméras connectées à un même PC. La fabrication des images de visages panoramiques est obtenue grâce à un algorithme simple basé sur des transformées linéaires rapides ce qui permet d'envisager des applications temps réel.

Dans le but d'étudier les performances de reconnaissance de notre système, nous avons constitué une base de données de visages panoramiques. Notre méthode de reconnaissance de visages est basée sur l'Analyse en Composantes Principales (ACP). Nous avons testé deux types de représentations d'images : spatiale et fréquentielle. La représentation fréquentielle est celle qui nous a apporté les meilleures performances en reconnaissance (97.46 % en comparaison avec 93.21 % pour la représentation spatiale). Un autre avantage de le représentation fréquentielle est la réduction du volume de données à traiter ce qui permet une accélération dans les calculs. Nous avons aussi utilisé des échantillons négatifs de visages panoramiques dans le but de tester la robustesse aux confusions de notre système. Nous avons pu ainsi montrer que dans le cas d'intrusions, le taux de reconnaissance reste supérieur à 92 %. Ainsi, les résultats expérimentaux que nous avons obtenus montrent la validité de notre système de mosaiquage pour traiter des visages dans des scènes 3D. Les performances de notre système sont proches, voire supérieures, à celles déjà décrites dans la littérature [5][1][2][9].

Notre système d'acquisition permet d'obtenir des images de visages selon 5 vues. Avec la méthode proposée, l'utilisateur dispose des 5 vues séparées, mais aussi de la recomposition panoramique des visages. Dans cet article, nous avons montré les résultats obtenus à partir de l'image panoramique qui représente des avantages comme la simplicité, l'économie de stockage et aussi la rapidité de traitement dus au fait que les 5 images sont condensées en une seule.

Avec notre dispositif, il est aussi possible d'envisager une approche mixte associant les reconnaissances à partir des 5 vues séparées et de l'image panoramique. Des travaux sont en cours actuellement dans notre équipe pour étudier, selon une approche de type fusion d'information, la combinaison de ces 2 approches et leur impact sur l'augmentation du taux de reconnaissance.

En perspectives, nous avons aussi l'intention de simplifier notre système d'acquisition. Nous sommes en train de mettre en place un système composé de 2 rétro-projecteurs pour projeter des marqueurs virtuels de différentes couleurs sur le visage au moment de l'acquisition. Ceci apportera un caractère dynamique à notre dispositif et améliorera l'aspect ergonomique. Nous souhaitons aussi utiliser notre système sans marqueurs. Pour cela, nous allons utiliser des algorithmes de détection de coins et de courbures maximales dans le but de détecter les points caractéristiques des visages. Enfin nous envisageons d'utiliser notre système de reconnaissance de visages panoramiques pour la catégorisation temps réel des expressions humaines, aussi bien dans le cadre d'applications liées à la modélisation en sciences cognitives (étude de l'expression des visages en relation avec le goût des aliments par exemple) que dans le cas d'applications en vidéoconférence.

Références

[1] A.J. HOWELL and H. BUXTON, *Learning identity with radial basis function networks*, Neurocomputing, Vol.20, pp.15-34, 1998.

[2] T. SIM, R. SUKTHANKAR *and al.*, *Memory-Based face recognition for visitor Identification*, 4th IEEE International Conf. On automatic face and gesture recognition, Grenoble, France, 26-30 March 2000.

[3] M. TURK and A. PENTLAND, *Eigenfaces for recognition*, Journal Cognitive neuroscience, Vol.3, pp.71-86, 1991.

[4] 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.

[5] M. SLIMANE, T. BROUARD *and al.*, *Unsupervised learning of pictures by genetic hibrydization of hidden Markov chain*, Signal Processing, Vol.16, No.6, pp.461-475, 1999.

[6] 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.

[7] http://www.frvt.org/FRVT2002/default.htm

[8] http://www.cs.colostate.edu/evalfacerec/

[9] F. TSALAKANIDOU, D. TZOVARAS and M.G. STRINTZIS, *Use of delth and colour eigenfaces for face recognition*, Pattern recognition Letters, Vol.24, pp.1427-1435, 2003.

[10] C. BEUMIER, M. ACHEROY, *Face verification from 3D and grey level clues*, Pattern recognition letters, Vol.22, pp.1321-1329, 2001.

[11] C. HEHSER, A. SRIVASTAVA and G. ERLEBACHER, *A nouvel technique for face recognition using range imaging*, 7th International Sympoosium on Siganl Processing and its Apllications (ISSPA), 2003.

[12] X. LU, D. COLBRY and A.K. JAIN, *Three-Dimensional model based face recognition*, Proc. International Conference on Pattern Recognition, Cambridge, UK, August, 2004.

[13] K.W. BOWYER, K. CHANG and P. FLYNN, *A survey of 3D and multi-modal 3D+2D face recognition*, International Conference on Pattern Recognition (ICPR), 2004.

[14] V. BLANZ and T. VETTER, *Face recognition based on fitting a 3D morphable model*, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.25, pp.1063-1074, September, 2003

[15] J.G. WANG, R. VENKATESWARLU and E.T. LIM, *Face tracking and recognition from stereo sequence*, Computer Science, Vol. 2688, pp.145-153, 2003.

[16] R. HARTLY and A. ZISSERMAN, *Multiple View Geometry in Computer vision*, Cambridge University Press, Second Edition, 2003.

[17] 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.

[18] F. YANG et M. PAINDAVOINE, *Détection et Localisation de Visages dans une Scène : Implantation Parallèle sur un Réseau de DSPs*, Revue Traitement du signal. Vol.17, N° 2, 2000.

[19] F. YANG, M. PAINDAVOINE, N. MALASNÉ, *Localisation et Reconnaissance de Visages en temps Réel avec un Réseau de Neurones RBF : Algorithme et Architecture*, Revue Traitement du signal. Volume 20, No.4, 2003.

[20] F. YANG, M. PAINDAVOINE, H. ABDI and J. MITÉRAN, *A new filtering technique combining a wavelet transform with a linear neural network: application to face recognition*, Optical Engineering SPIE. Vol.39, No.11, 2000.

[21] F. YANG, M. PAINDAVOINE, *Implementation of a RBF neural network on embedded systems: Real time face tracking and identity verification*, IEEE Transactions on Neural Networks, Vol.14, No.5, pp.1162-1175, September 2003.

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

[23] Y. ZHOU, H. XUE and M. WAN, *Inverse image alignment methode for image mosaicing and video stabilization in fundus indocyanine green angiography under confocal scanning laser ophthalmoscope*, Computerized Medical Imaging and Graphics, Vol.27, pp.513-523, 2003.

[24] P.F. MCLAUCHLAN and A. JAENICKE, *Image mosaicing using sequential bundle adjustment*, Image and Vision computing, Vol.20, pp.751-759, 2002.

[25] A.K. JAIN and A. ROSS, *Fingerprint Mosaicking*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, Florida, May, 2002.

[26] D.N. WOOD, A. FINKELSTEIN, J.F. HUGHES, C.E. THAYER and D.H. SALESIN, *Multiperspective Panoramas for Cel Animation*, Proceedings of SIGGRAPH 97, pp.243-250, august, 1997.

[27] P. RADEMACHER and G. BISHOP, *Multiple-Center-of-Projection Images*, Proceedings of SIGGRAPH, pp.199-206, 1998.

[28] C. LANGIS, M. GREENSPAN and G. GODIN, *The Parallel Iterative Closest Point Algorithm*, Proceedings of the Third International Conference on 3D Digital Imaging and Modeling (3DIM), Québec, Canada, 2001.

[29] 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.

[30] 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).

[31] H. ABDI, D. VALENTIN and B. EDELMAN, *Neural Networks*, Sage, Thousand Oaks, 1999.

[23] 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.

[33] 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**

Fan Yang est maître de conférences de l'Université de Bourgogne depuis 2000. Elle enseigne actuellement l'informatique industrielle, l'automatisme et la robotique à l'IUT de Dijon. Ses travaux de recherche au sein du groupe Architecture du LE2I-CNRS UMR 5158 s'intéressent aux méthodes de traitement automatique d'images de visages 2D et 3D, et aux systèmes embarqués en temps réel.

Michel **Paindavoine**

Michel Paindavoine est professeur à l'Université de Bourgogne. Il enseigne le traitement du signal et des images à l'Ecole d'Ingenieurs ESIREM et à l'IUP Electronique et Image. Il est le directeur du laboratoire LE2I-CNRS UMR 5158 (Laboratoire d'Electronique, d'Informatique et d'Image). Il effectue ses travaux de recherche dans le domaine de l'Adéquation Algorithmes Architectures en traitement d'images.

Dominique **Ginhac**

Dominique Ginhac est maître de conférences à l'Université de Bourgogne depuis 2000. Il effectue ses travaux de recherches dans le domaine de la conception micro-électronique de capteurs d'images dédiés intégrant des traitements analogiques et/ou numériques.

Julien **Dubois**

Julien Dubois a rejoint le Laboratoire Electronique, Informatique et Image (LE2I) depuis 2003, où il occupe les fonctions de Maître de Conférence. Ses thématiques principales sont la conception et le développement de caméra rapide et intelligente. A travers ces développements matériels, il s'intéresse plus particulièrement à l'intégration de nombreux algorithmes de traitements d'images, tels que la compression d'images ou la reconnaissance de visage, sur des cibles de type FPGA et DSP (co-design).