

Reconstruction de visages par stéréovision active

Face Reconstruction Through Active Stereovision

par Régis VAILLANT*, Isabelle SURIN**

* Thomson-CSF, Laboratoire central de Recherches,
Domaine de Corbeville, 91404 Orsay
e-mail : vaillant@thomson-lcr.fr

** Groupement Traitement et Simulation TTD,
Rue Guynemer BP 55, 78283 Guyancourt Cedex

Résumé

La reconnaissance automatique de visages est un problème qui suscite beaucoup d'intérêt et pour lequel divers algorithmes basés sur l'utilisation d'images acquises par des caméras CCD ont été proposés. L'information fournie aux algorithmes de discrimination mis en place dans ce type de solution est assez fruste et ces algorithmes sont peu susceptibles de fonctionner de manière robuste pour des bases comprenant un grand nombre d'individus. La surface du visage doit pouvoir être beaucoup plus discriminante. Dans cet article, nous proposons un algorithme de stéréovision active qui permet l'acquisition de surfaces de visages. Le problème de la mise en correspondance entre les éléments du motif projeté et leur image est résolu en utilisant la contrainte épipolaire et une contrainte de cohérence locale. Des résultats expérimentaux sont présentés.

Mots clés : Stéréovision Active, Reconstruction de surface, Visage.

Abstract

The automatic face recognition is a very attractive problem and several solutions can be found in the literature. Most of them rely on the analysis of the images acquired by a classical CCD camera. These images are treated and given as input to a discrimination algorithm. However, the information contained in the image is relatively poor and it is very likely that these techniques will fail in the case of a large database with a lot of people. The surface of the faces is a very discriminant information so in this article, we propose a stereovision algorithm which can be used for the acquisition of the surface of faces. The problem of matching between the pattern and its images is solved using the epipolar constraint and the local coherency constraint. Some experimental results are shown.

Key words : Active Stereovision, Surface reconstruction, Face.

1. Introduction

Dans ce travail, nous proposons un algorithme de stéréovision active qui a été tout spécialement étudié en vue de pouvoir être appliqué à la reconstruction de la surface de visages. En effet, nous nous intéressons au problème de l'identification automatique de personnes à partir de l'observation de leur visage. Une revue assez complète des travaux menés sur ce sujet peut être trouvée dans [Samal-Iyengar92]. Les grandes classes de méthodes proposées sont :

- Utilisation directe de l'image du visage : après quelques traitements relativement simples, cette image est fournie à un algorithme de discrimination qui a été préalablement entraîné avec une base de données. D'excellents taux de reconnaissance sur des bases relativement petites ont été obtenus [Turk-Pentland91a]. Récemment des résultats ont été obtenus sur des bases de plusieurs milliers de visages [Pentland-Moghaddam-Starner-Oliyide-

Turk94]. L'inconvénient majeur de ce type de méthode est la sensibilité aux conditions d'éclairage : l'image d'un visage sera sensiblement différente suivant que celui-ci est éclairé de face ou de côté. Ce problème a été souligné dans [Moses-Adini-Ullman94]. Dans [Beymer94], une technique est développée pour tenir compte des problèmes d'éclairage et d'orientation du visage. D'autre part il est probable que les taux de reconnaissance obtenus doivent diminuer assez rapidement si le nombre de visages distincts à reconnaître augmente.

- Utilisation d'une image du profil : l'image acquise est une image du profil du visage. Le contour de ce profil est extrait. Il est alors possible d'en extraire quelques traits caractéristiques qui peuvent être fournis à un algorithme de discrimination. Bien qu'a priori l'image du profil d'un visage soit moins discriminante que l'image du visage lui-même, l'avantage de cette stratégie est que les données utilisées sont indépendantes de l'éclairage puisqu'on utilise le contour, et peuvent assez facilement être normalisées pour être rendues indépendantes de la position du visage et de

son orientation dans l'image. Des taux de reconnaissance de l'ordre de 92% sont mentionnés dans [Najman-Vaillant-Pernot93], et [Harmon-Khan-Lasch-Ramig80].

- Dans la filiation directe de cette idée qui consiste à quantifier la géométrie du visage, nous avons essayé de mesurer la surface correspondant au visage en vue d'en extraire des traits caractéristiques pour permettre la reconnaissance. Une telle méthode bien qu'un peu plus difficile à mettre en œuvre que les précédentes, pourrait présenter l'avantage de permettre l'utilisation de bases comprenant beaucoup de visages distincts. Dans [Cartoux-Lapreste-Richetin89], il est montré qu'il est tout à fait envisageable d'authentifier ou de reconnaître des visages à partir d'images 3D. De même Gordon, dans [Gordon91], propose deux approches l'une basée sur la mesure du "volume" entre deux surfaces de visage après que celles-ci aient été normalisées, l'autre basée sur l'extraction d'un vecteur d'éléments caractéristiques. Dans [Lee-Milios90], il est proposé une méthode de mise en correspondance d'images de profondeur de visages basée sur l'utilisation de l'image de la surface par l'application de Gauss (Extended Gaussian Image).

La première difficulté qui se pose donc à nous est l'acquisition de la surface correspondant au visage. Plusieurs techniques sont décrites dans la littérature :

- Système de stéréovision passive : nous n'avons pas pu utiliser un algorithme de mise en correspondance des contours du type de ceux décrits dans [Ayache-Lustman87], [Robert-Faugeras91] car l'information obtenue n'aurait pas été suffisamment dense : le visage ne possède pas assez de contours. On aurait pu utiliser un algorithme de stéréovision passive basé sur des techniques de corrélation [Fua91b]. Néanmoins ce type d'algorithme ne fournit de bons résultats que si l'image observée possède suffisamment de texture. Il n'était pas sûr que cela soit le cas pour un visage. Ils sont de plus extrêmement gourmands en temps calcul. Dans [Thom88], une approche est proposée pour pallier à ce problème : un ensemble de points aléatoires est projeté sur la surface du visage. Grâce à un dispositif optique, deux images formant l'équivalent d'un couple stéréo, sont acquises. Les points éclairés sont mis en correspondance par corrélation et reconstruits en 3D par triangulation. Les travaux décrits dans [Cartoux-Lapreste-Richetin89, Cartoux89] utilisent les surfaces mesurées par ce dispositif.
- Systèmes actifs en lumière cohérente : l'inconvénient majeur est le risque que présente le faisceau laser pour les yeux.
- Systèmes de stéréovision active : le principe de base est de projeter un **motif** sur la scène en vue de créer artificiellement des contours. La correspondance entre les contours observés dans l'image et les points du motif à l'origine de ces contours est ensuite établie, ce qui permet la reconstruction tridimensionnelle par triangulation des points

observés. Des systèmes de ce type ont été proposés dans la littérature ([Wang-Liang89], [Dunn-Keizer-Yu89], [Hu-Stockman89]). En utilisant un motif suffisamment dense, un grand nombre de points de la scène peuvent être mis en correspondance ce qui permet la détermination de la surface observée avec une bonne résolution.

Dans la suite de cet article, nous allons décrire le motif qui est projeté sur la scène et la technique utilisée pour calibrer le système utilisé. Nous présenterons ensuite l'algorithme que nous proposons pour mettre en correspondance les éléments du motif avec les observations réalisées puis nous montrerons quelques planches de résultats expérimentaux.

2. Le Motif – La Calibration

2.1. LE MOTIF CHOISI

Suivant le principe de la stéréovision active, on projette un motif sur la surface observée, puis on utilise des points bien choisis sur ce motif afin de réaliser la reconstruction 3D de la surface.

1. Les points doivent être en nombre suffisant pour permettre l'obtention d'une bonne représentation de la surface du visage.
2. Les points doivent être répartis régulièrement sur tout le motif.
3. Les points doivent être facilement détectables dans l'image : la première idée, qui est celle la plus fréquemment rencontrée dans la littérature (dans [Dunn-Keizer-Yu89] par exemple), est d'utiliser un quadrillage formé en projetant un ensemble de lignes. Les points utilisés pour la reconstruction sont dans ce cas les points d'intersection des lignes. Un problème se pose alors : quelle épaisseur de trait choisir? On risque de ne pas détecter un trait trop fin. Inversement, on risque de détecter deux fois un trait trop épais. De plus, l'épaisseur du trait effectivement observé dépend de la distance du projecteur à la surface sur laquelle le motif est projeté, et de la distance entre la caméra et cette même surface. Une solution à ce problème est de considérer des bandes et non des traits. On prend alors comme points d'intérêt les coins des carreaux blancs définis par les bandes. Pour éviter le problème délicat de la détection des traits, on pourra donc détecter les carreaux "blancs" formés par les bandes.
4. Les points doivent être identifiables : on aura besoin de mettre en correspondance les points du motif et ceux du motif extrait de l'image. Une façon d'identifier ces points est de les numéroter. Un moyen simple pour les numéroter est d'introduire un **repère** dans le motif comme l'illustre la figure 1. Ce repère est formé de carreaux noirs.

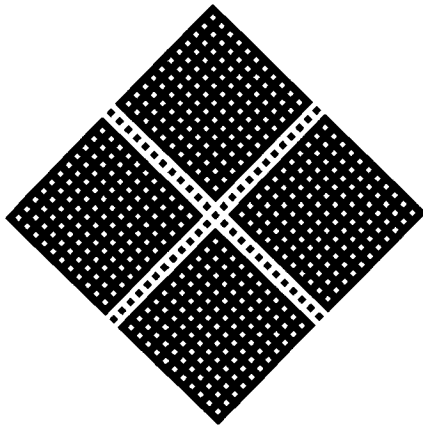


Figure 1. – Le motif.

2.2. CALIBRATION

2.2.1. Calibration des caméras

On considère que les caméras sont du type “pinhole”. L'image d'un point M de coordonnées $[X, Y, Z]^T$ dans un repère fixe indépendant de la caméra est le point $m = [u, v]^T$ tel que :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathcal{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

\mathcal{P} est la matrice de transformation perspective. Expérimentalement, la matrice \mathcal{P} est déterminée en observant un objet parfaitement connu. La technique utilisée est décrite dans [Faugeras-Toscani86] et dans [Faugeras93].

2.2.2. Calibration du projecteur

Pour chaque point m_p du motif, il nous faut déterminer la droite \mathcal{D}_p correspondant au rayon optique émis par le projecteur et associée à m_p (voir la figure 2). Dans le plan image de la caméra, on observe le point m_c , projection du point M qui est l'intersection de la surface observée (S) et du rayon optique \mathcal{D}_p . Connaissant m_c et m_p , le point M peut être déterminé en tant qu'intersection de la droite Cm_c et de la droite \mathcal{D}_p . La détermination des points M observés exige donc la connaissance de la droite \mathcal{D}_p pour chacun des points m_p du motif.

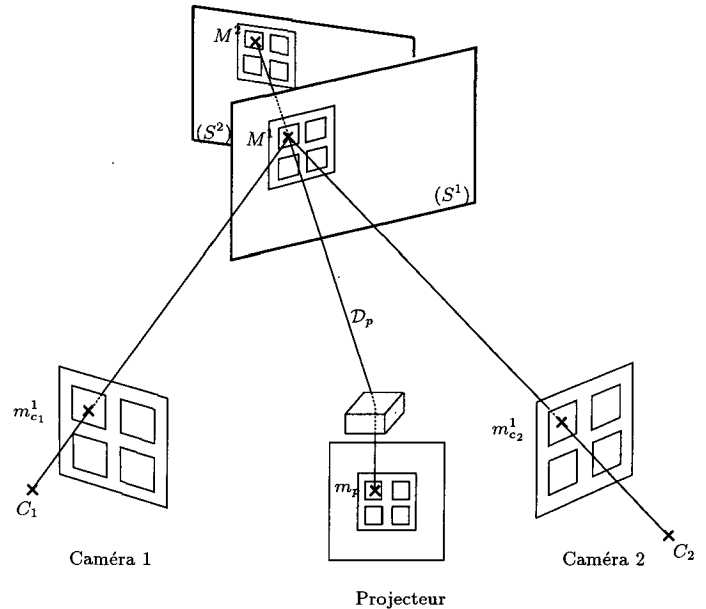


Figure 2. – Détermination de \mathcal{D}_p .

Détermination de la droite \mathcal{D}_p

Pour déterminer cette droite, il suffit d'observer au moins deux points qui sont sur le rayon optique associé à m_p . Nous projetons le motif sur deux surfaces (S^1) et (S^2) et nous observons ces projections avec deux caméras qui ont été préalablement calibrées par rapport à un même repère. Pour chacune des observations j et pour chacune des caméras, on détermine les images $m_{c_1}^j$ et $m_{c_2}^j$ des points M^j qui sont l'intersection de la surface (S^j) avec le rayon optique \mathcal{D}_p . Le point M^j est déterminé comme intersection des rayons optiques $C_1m_{c_1}^j$ et $C_2m_{c_2}^j$. C_1 et C_2 sont respectivement les centres optiques de la première et de la deuxième caméra. La droite \mathcal{D}_p est définie par M^1 et M^2 .

Pratiquement, la calibration du projecteur exige donc l'étiquetage d'images de la projection du motif sur une surface. Ce problème est similaire à celui que nous décrivons dans le paragraphe 3. Il existe néanmoins une différence majeure avec l'étiquetage réalisé lorsque le projecteur est calibré : nous n'avons pour l'instant aucune autre connaissance sur le motif que sa forme puisque nous ne connaissons ni les caractéristiques optiques du projecteur, ni la surface (S) sur laquelle le motif est projeté. Pour simplifier cette étape d'étiquetage, nous avons utilisé des surfaces (S^j) qui sont des plans. D'autre part, pour rendre plus robuste l'estimation des droites \mathcal{D}_p , nous avons utilisé plus de deux projections. Nous avons déterminé un ensemble de points $M^j, j = 1, \dots, J$ à partir desquelles la droite \mathcal{D}_p est estimée au moindre carré.

Modélisation du projecteur par une transformation perspective

La modélisation du projecteur précédemment décrite nous permet effectivement la réalisation des calculs qui nous intéressent : la détermination par triangulation de la position des points observés,

mais elle n'est pas très facile à manipuler. Nous avons donc envisagé d'essayer de modéliser le projecteur par une transformation perspective.

- Tout d'abord, il nous faut choisir un repère dans le plan image. Très naturellement, nous utilisons celui qui est induit par le motif. Nous pouvons attribuer des coordonnées (u_p, v_p) à chaque point m_p du motif. Un point $M = [X, Y, Z]^T$ qui est sur le rayon optique émis par le projecteur et associé au point m_p vérifie la relation :

$$\begin{pmatrix} su_p \\ sv_p \\ s \end{pmatrix} = \mathcal{P}_p \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

où \mathcal{P}_p est la matrice de transformation perspective du projecteur. Nous pouvons l'estimer en utilisant une technique similaire à celle utilisée pour les caméras. La mire de calibration est remplacée par l'ensemble des points M^j qui ont été observés en projetant le motif sur plusieurs plans distincts. Les images de ces points sont tout simplement les points m_p du motif.

Les points M^j ont été obtenus en projetant le motif sur plusieurs surfaces. Ils ne sont donc pas coplanaires, ce qui est indispensable pour pouvoir estimer la matrice \mathcal{P}_p .

- Après avoir ainsi estimé la matrice \mathcal{P}_p , il nous faut vérifier la validité de ce modèle : pour chacun des points m_p du motif, nous pouvons en utilisant la matrice \mathcal{P}_p déterminer le rayon optique \mathcal{D}_p^P et le comparer à l'estimation \mathcal{D}_p^D obtenue en ajustant une droite aux points M^j où $j = 1, \dots, J$. Pour chaque point m_p du motif nous avons effectué quelques mesures qui sont résumées dans la table 2.2.2. et que la figure 3 illustre :

- La première mesure α est l'angle entre les directions de \mathcal{D}_p^P et de \mathcal{D}_p^D .
- La deuxième mesure β est obtenue en mesurant la distance entre les points d'intersection des deux rayons optiques avec un plan. Le plan est choisi dans la zone de travail. La distance calculée indique donc l'écart entre les rayons dans la zone où nous allons les utiliser. Dans la pratique, nous choisissons un des plans sur lesquels a été projeté le motif lors de la calibration. L'équation de ce plan est estimée au moindre carré à partir des points M^j .
- La troisième mesure γ donne la distance entre le point d'intersection M_p^P du plan et de \mathcal{D}_p^P , et la projection orthogonale sur le plan de la reconstruction 3D du point m_p .
- La quatrième mesure δ donne la distance entre le point d'intersection M_p^D du plan et de \mathcal{D}_p^D , et la projection orthogonale sur le plan de la reconstruction 3D de m_p .

Nous avons porté dans la table 2.2.2. les valeurs minimum, maximum et moyenne de la mesure. Ces valeurs ont été

calculées lors de deux séries d'expériences. La première a été réalisée en utilisant 3 projections différentes. La seconde a été réalisée en utilisant 2 projections supplémentaires. Nous constatons que α est de l'ordre du dixième de degré ce qui correspond à un écart moyen inférieur au millimètre dans la zone de travail. γ et δ permettent de comparer l'écart mesuré par β avec celui qui est dû au bruit existant sur la mesure des points M^j . On trouve que γ et δ sont du même ordre de grandeur que β ce qui permet effectivement de conclure que le projecteur réalise une transformation perspective comme nous l'avions supposé. Cette propriété nous permet, ainsi que nous le verrons dans le paragraphe 3, d'utiliser la notion d'épipolaire entre l'image acquise par une caméra et l'espace dans lequel sont situés les points du motif.

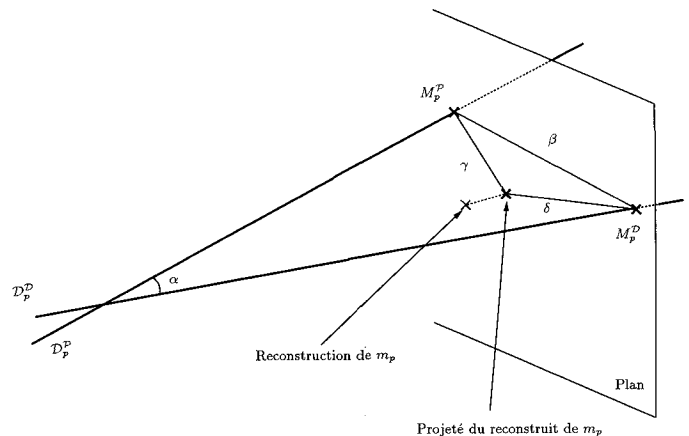


Figure 3. – Paramètres de la validité de la modélisation du projecteur par une transformation perspective.

		Min	Max	Moy
Première expérience	α (°)	0.003	0.61	0.18
	β (mm)	0.041	2.24	0.95
	γ (mm)	0.067	2.41	1.04
	δ (mm)	0.010	0.91	0.22
Seconde expérience	α (°)	0.007	0.55	0.18
	β (mm)	0.03	1.99	0.89
	γ (mm)	0.067	2.41	1.04
	δ (mm)	0.0010	1.23	0.34
Distance du plan à la caméra : 1.5 mètre				

Tableau 1. – Test de concordance avec le modèle de projection perspective

3. Etiquetage

Les carreaux observés dans l'image sont extraits en utilisant un algorithme de morphologie mathématique décrit dans l'annexe A. Dans la suite, nous appellerons, taches blanches, l'image des carreaux blancs et taches noires, l'image des carreaux noirs. Le problème est alors d'établir une correspondance entre les carreaux

de l'espace motif et les taches de l'espace image. Pour cela, on attribue à chaque tache l'étiquette que porte le carreau qui lui correspond dans l'espace motif. Au départ, une tache pourra posséder plusieurs étiquettes. Le fil conducteur de l'algorithme est l'élimination de certaines de ces étiquettes grâce à des contraintes et par propagation d'informations.

3.1. ETIQUETAGE MULTIPLE

Puisque nous avons pu modéliser le projecteur par une transformation perspective, nous nous trouvons confrontés au problème de la mise en correspondance de structures géométriques qui sont les images dans deux plans, celui de la caméra et celui du projecteur, d'un même ensemble de points tridimensionnels. Il s'agit donc du problème classique de stéréovision binoculaire. Les structures géométriques que nous souhaitons mettre en correspondance sont bien sûr les taches extraites.

La première propriété que nous pouvons utiliser pour la mise en correspondance est l'épipolarité décrite dans [Faugeras93] et [Toscani87] qui transforme le problème bidimensionnel en un problème monodimensionnel : considérons les quatre coins d'un carreau, la tache correspondant à ce carreau se trouve dans un voisinage de la plus grande bande définie à partir des quatre épipolaires de ces coins, tracées dans l'espace image (voir figure 4). On accepte une tache comme correspondant potentiel du carreau si :

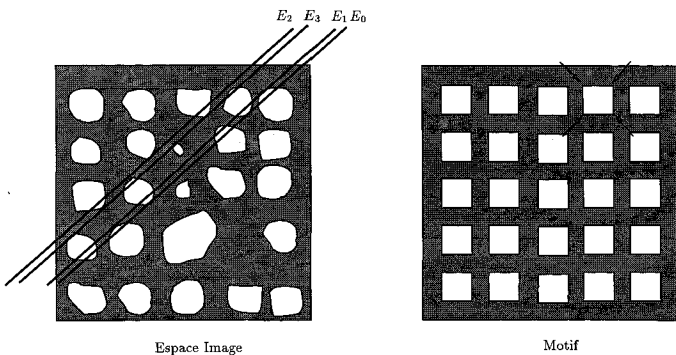


Figure 4. – Epipolaires des coins d'un carreau.

1. Sa couleur est la même que celle du carreau.
2. Son aire se situe dans la moyenne des aires des taches (on exclut les taches "hors norme").
3. Pour chaque épipolaire E_i où $i = 0, \dots, 3$:

$$\min_{j=1 \dots n} (d(E_i, P_j)) < S_e$$

où $(P_j)_{j=1, \dots, n}$ sont les pixels de la tache et S_e est un seuil qui est fixé à 3 pixels.

A chaque tache, plusieurs correspondants potentiels sont donc affectés.

3.2. DÉTERMINATION DES COINS DES TACHES

Il nous faut déterminer pour chaque tache les coins correspondants à ceux des carreaux. Le principe est d'établir pour chacun des coins qui doivent être détectés, une liste de coins potentiels C_i^p où i identifie le coin et p identifie le point candidat, et de déterminer lesquels parmi eux sont les coins effectifs en minimisant :

$$\sum_{i=0}^3 \sum_{P \in \mathcal{B}_i} dist(P, C_i^p C_{(i+1) \bmod 4}^q)$$

où \mathcal{B}_i est l'ensemble des points de la frontière de la tache situés entre C_i^p et $C_{(i+1) \bmod 4}^q$.

Un pixel P_i de la frontière est considéré comme coin potentiel pour le coin i si :

- la distance de P_i à l'épipolaire E_i est inférieure à S_e ,
- la courbure de la frontière de la tache en P_i est supérieure à un certain seuil. La courbure est donnée par l'angle que forment $P_{j-1}P_j$ et P_jP_{j+1} . Nous imposons cette contrainte sur la courbure afin de limiter le nombre de coins potentiels.

La minimisation du critère se fait en explorant les différentes solutions possibles. Le temps nécessaire est relativement faible car le nombre de coins potentiels est réduit et il est possible de mettre en place des structures de données qui réduisent le volume de calcul nécessaire.

3.3. CONTRAINTE DE REPROJECTION

Il s'agit d'une variante de la contrainte de disparité dans laquelle nous utilisons la structure des objets mis en correspondance. La surface de l'objet observé est régulière. On peut considérer en première approximation que les reconstructions respectives de deux taches voisines de l'image sont coplanaires, et que leurs projections dans l'espace motif sont voisines. On utilise cette contrainte pour vérifier que deux taches voisines dans l'image portent des étiquettes correspondant à des carreaux voisins dans le motif.

Soit B une tache, et V le barycentre d'un de ses voisins. B porte plusieurs étiquettes possibles $E_1 \dots E_n$ qui correspondent aux carreaux $C_1 \dots C_n$ de l'espace motif. Considérons E_i et C_i . La première étape consiste à déterminer le plan contenant la reconstruction 3D de B . On considère les trois coins de B les plus proches de V et on détermine le plan P_i engendré par les reconstructions 3D de ces trois points (voir figure 5). Comme V est voisin de B , on approxime la reconstruction 3D de V par sa projection sur P_i : V_{pi} . On considère maintenant la projection de V_{pi} dans l'espace motif : V_{pat} . Si B est étiqueté E_i , alors le correspondant du carreau V doit être proche de V_{pat} .

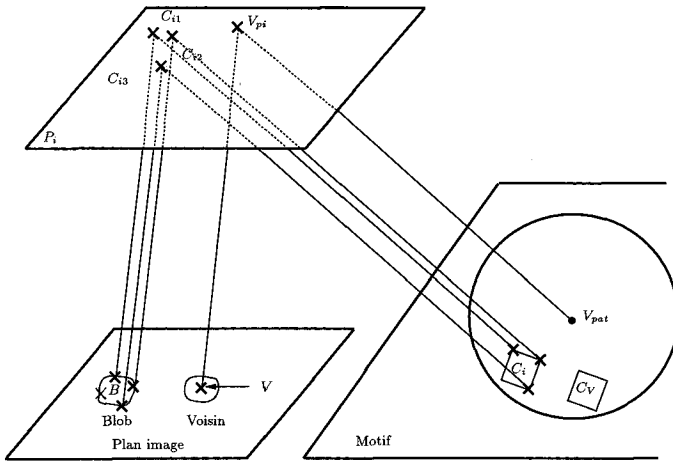


Figure 5. – Contrainte de reprojection.

Contrairement à la contrainte épipolaire, il s'agit d'une heuristique qui est basée sur l'hypothèse de continuité de la surface observée. Il n'est donc pas possible de l'appliquer systématiquement. Nous allons maintenant voir une stratégie possible pour utiliser cette contrainte en vue d'éliminer les étiquettes erronées.

3.4. VRAISEMBLANCE DES TACHES NOIRES

Soit B une tache dont les étiquettes potentielles sont E_1, \dots, E_n correspondant aux carreaux $(C_k)_{k=1, \dots, n}$. Elle possède un ensemble de voisins V^i qui portent chacun une ou plusieurs étiquettes possibles E_j^i où i identifie le voisin et j identifie l'étiquette de celui-ci. Pour un voisin donné V^i et pour une de ses étiquettes E_j^i nous appliquons la contrainte de reprojection présentée dans le paragraphe 3.3. On obtient ainsi une position possible V_{pat} pour le correspondant du barycentre de B dans l'espace motif.

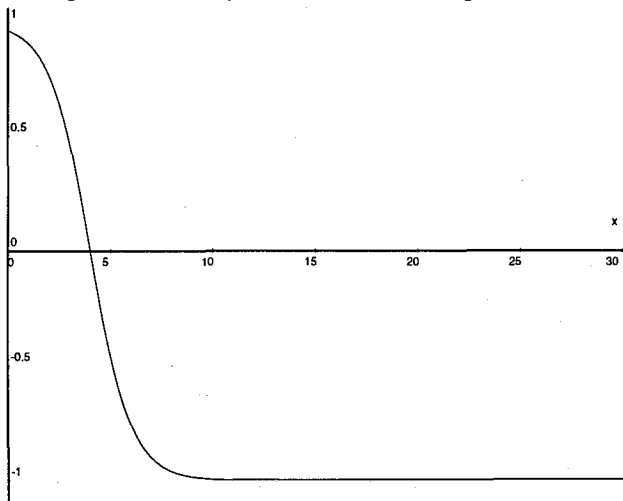


Figure 6. – Fonction f .

Nous allons maintenant juger la compatibilité de ce correspondant possible V_{pat} avec les $(E_k)_{k=1, \dots, n}$. Pour une étiquette donnée

E_k de B , la compatibilité $C(E_k, E_j^i)$ de E_k avec E_j^i est définie comme étant une fonction f de la distance d entre V_{pat} et le barycentre de C_k , pondérée par la distance entre le voisin V_i considéré et la tache B . f est représentée figure 6. L'abscisse est une distance mesurée dans l'espace motif dont la taille est environ 50. Elle comprend une partie dont la pente est relativement faible, pour des petites valeurs de d . En effet, la compatibilité est considérée comme étant maximale si V_{pat} appartient à un petit voisinage de C_k . Elle décroît quand d augmente, et devient négative lorsque cette distance devient grande. On applique ce processus à chaque voisin et à chacune de ses étiquettes. Il en résulte que chaque étiquette possible E_k de B se voit affecter une certaine vraisemblance $V(E_k)$ qui est la somme des compatibilités obtenues :

$$V(E_k) = \sum_i \sum_j C(E_k, E_j^i)$$

Nous venons de définir un critère de jugement sur une étiquette potentielle d'une tache. Il est tout à fait possible d'établir un critère similaire de jugement sur une tache correspondant potentiel d'un carreau, en considérant des projections dans l'espace image.

3.5. ETIQUETAGE DES TACHES NOIRES

Considérons une tache noire B qui a plusieurs étiquettes possibles. Ses voisins, noirs ou blancs, votent pour chacune de ses étiquettes. Seules celles qui ont réalisé un score honorable sont conservées.

Ainsi que décrit dans le paragraphe 3.4, on détermine la vraisemblance des étiquettes de B en utilisant ses voisins.

Soit $votemin$ (resp. $votemax$) le plus petit (resp. plus grand) score enregistré par une étiquette. Les étiquettes dont le score est inférieur à $(2 * votemax + votemin) / 3$ sont éliminées. Après cette étape, les taches noires sont, à de rares exceptions près, mono-étiquetés.

3.6. ETIQUETAGE DES TACHES BLANCHES

Il est réalisé en deux étapes :

1. Utilisation des taches noires :

Les taches noires sont en général bien étiquetés et on aimerait propager aux blanches l'information qu'elles contiennent. Pour éliminer des étiquettes dans les quadrants, dans un premier temps on fait donc voter uniquement les taches noires. On considère alors que l'ensemble des voisins d'une tache blanche est constitué par les éléments étiquetés de la croix noire.

On détermine pour chaque tache blanche la vraisemblance de chacune de ses étiquettes potentielles en utilisant la stratégie décrite dans le paragraphe 3.4. On utilise cependant une fonction de compatibilité légèrement différente car les taches noires ne sont pas voisins immédiats des taches

blanches. Le critère de reprojction ne peut s'appliquer que très approximativement. La fonction de compatibilité utilisée dans ce cas possède un plateau beaucoup plus important, et ne tient pas compte de la distance entre la tache et le voisin considéré car celle-ci est très variable. Il en résulterait une pondération trop brutale, telle que seul le vote de l'élément le plus proche jouerait un rôle.

Chaque étiquette est conservée si sa vraisemblance est supérieure à $\text{nombre de votants} / 2$ votes. Les étiquettes relatives aux mauvais quadrants sont ainsi assez systématiquement éliminées.

2. Utilisation des voisins :

On aimerait maintenant écarter les "mauvais" numéros de ligne et de colonne. Soit B une tache. Elle ne peut porter l'étiquette E_i du carreau C_i que si plusieurs de ses voisins ont parmi leurs correspondants possibles un voisin de C_i . On utilise à nouveau le critère de vraisemblance des étiquettes potentielles décrit dans le paragraphe 3.4.

D'autre part, un carreau C ne peut correspondre à une tache B que si ses voisins ont parmi leurs correspondants possibles un voisin de B . On applique la méthode de vraisemblance des étiquettes telle qu'elle est présentée dans le paragraphe 3.4 en utilisant cette fois la reprojction dans l'espace image.

Après une passe sur l'image et sur le motif, on a diminué le nombre d'étiquettes possibles. On recommence l'étape d'utilisation des voisins jusqu'à ce que l'étiquetage soit stable : on a alors exploité au maximum les informations apportées par la contrainte de continuité.

Cependant, il n'y a pas unicité de la solution. De plus, on a observé que des erreurs pouvaient se produire : on élimine parfois des étiquettes qu'il aurait fallu conserver, en particulier dans des zones perturbées. Cependant, le bilan est globalement nettement positif. On voudrait en conserver le profit, tout en remettant en cause certains résultats. Les erreurs commises étant ponctuelles, on a envie de dire qu'occasionnellement une tache peut ne plus avoir sa bonne étiquette, mais que ses voisins auront probablement encore la leur. On va donc vérifier pour chaque tache, que parmi ses étiquettes jugées erronées il n'y en a pas certaines qui seraient à cette étape déclarées vraisemblables par ses voisins. Pour les voisins on considérera donc l'ensemble réduit des étiquettes. Par contre pour pouvoir récupérer l'étiquette éventuellement perdue par la tache, on remet en jeu toutes ses étiquettes trouvées par épipolarité. On recommence ensuite la stratégie d'élimination par contrainte de voisinage jusqu'à stabilisation.

3.7. CONTRAINTE D'UNICITÉ

Toutes les taches ne sont pas mono-étiquetées. On fait maintenant intervenir la contrainte d'unicité sur les taches blanches et sur les taches noires. Une étiquette ne peut pas se trouver à plusieurs endroits, c'est-à-dire qu'un carreau ne peut pas correspondre à

différentes taches. De même, une tache ne peut pas correspondre à des carreaux différents.

- Soit B_{mono} une tache mono-étiquetée, et C_i son carreau correspondant, d'étiquette E_i . C_i ne peut pas correspondre à une autre tache que B_{mono} . On élimine en conséquence E_i comme étiquette possible de toute autre tache.
- Soit C_{mono} un carreau d'étiquette E_i , ayant un unique correspondant B_i . On élimine les étiquettes de B_i différentes de E_i , et tout carreau autre que C_{mono} ne peut correspondre à B_i .

4. Résultats expérimentaux

La mise en correspondance des carreaux du motif avec les taches observées dans l'image, permet la détermination des coordonnées tridimensionnelles de points sur la surface observée. Les points dont les coordonnées sont mesurées sont les coins des carreaux. A partir de ces points, nous construisons une représentation de la surface observée. Cette représentation est établie en réalisant une triangulation des points qui ont été mis en correspondance. La triangulation est réalisée non pas sur les points tridimensionnels mesurés mais sur les images de ces points dans le plan image.

La surface obtenue peut être entachée de quelques erreurs qui relèvent principalement de deux causes :

- erreur dans les mises en correspondance entre les carreaux du motif et les taches observées. Ce type d'erreur se caractérise a priori par l'existence d'un ensemble de 4 points, les 4 coins du carreau erroné, qui se trouvent devant ou derrière les points voisins.
- imprécisions dans le calcul des coordonnées tridimensionnelles. Ces imprécisions sont liées à la manière dont les coins sont déterminés sur les taches. En effet les coins sont pris comme des pixels de la frontière de la tache, or la frontière est déterminée avec potentiellement une erreur de plus ou moins un pixel.

Nous avons donc choisi de réaliser un lissage des données mesurées. Afin d'éliminer les erreurs causées par un appariement erroné sans corrompre les autres données, nous avons choisi d'utiliser un filtre médian. Pour chaque point M , nous déterminons son voisinage : il s'agit de tous les points $(V_i)_{i=1, \dots, n}$ dont l'image est distante de l'image de M de moins de d_s pixels. Le voisinage est construit en utilisant la distance entre les points image plutôt que les points tridimensionnels car une erreur d'appariement peut introduire une grande distance entre des points dont les images sont proches. Soit C le centre optique de la caméra, nous remplaçons le point M par le point M' tel que CM et CM' soient colinéaires et que la distance CM' soit la médiane des distances $(CV_i)_{i=1, \dots, n}$.

La figure 7 montre la différence entre des données reconstruites avant et après lissage.

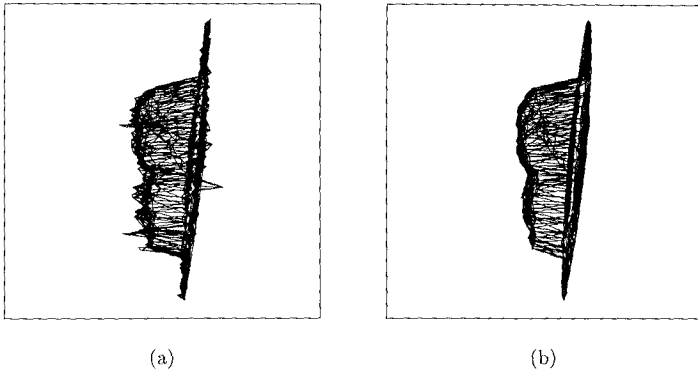


Figure 7. – Lissage de la surface obtenue : (a) Les données non lissées, (b) Les données lissées en utilisant le filtre médian.

Nous présentons des exemples de reconstructions obtenues chacune à partir de l'image d'un objet ou d'un visage, plus précisément : un masque (figure 8), deux visages de personnes différentes (figures 9 et 10).

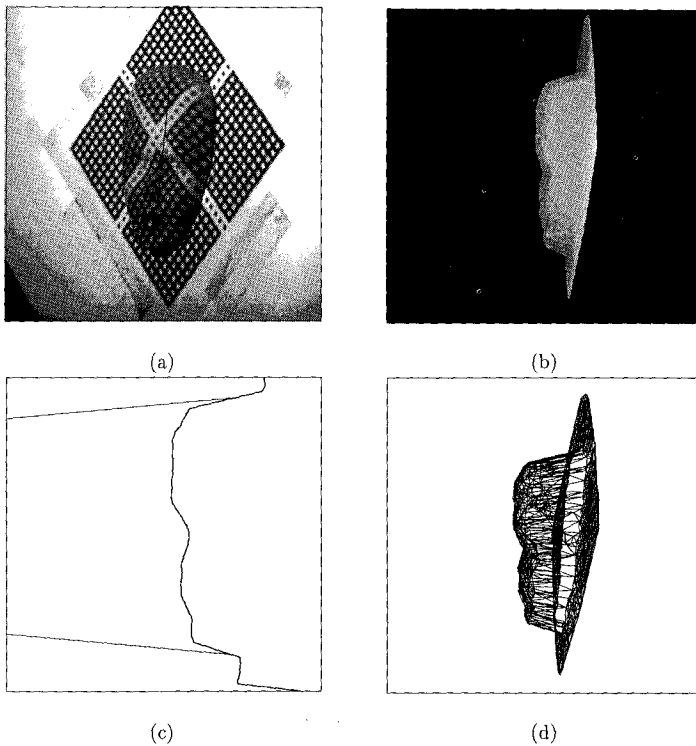


Figure 8. – Le masque.

Sur chaque planche :

- La figure (a) représente l'image de l'objet que l'on veut reconstruire, et sur lequel on a projeté le motif.
- La figure (b) est une vue en ombré de la reconstruction 3D de l'objet observé. L'intensité de chaque pixel est déterminée par la distance entre le point de vue et le point 3D qui a pour image ce pixel. Puisque le point de vue utilisé n'est pas celui de la caméra, plusieurs points de la surface reconstruite

peuvent se projeter en un seul point et l'image est donc calculée en utilisant l'algorithme de z-buffer : seul le point le plus proche est visible.

- La figure (c) est une coupe de la surface reconstruite réalisée suivant l'axe représenté figure (a). On a aussi tracé les rayons optiques correspondant aux extrémités de la ligne dessinée figure (a). On peut observer que sur les figures ?? et ?? on retrouve nettement la forme du profil.
- La figure (d) représente la triangulation de l'objet, c'est-à-dire une projection des arêtes des triangles 3D que nous avons utilisés pour reconstituer la surface.

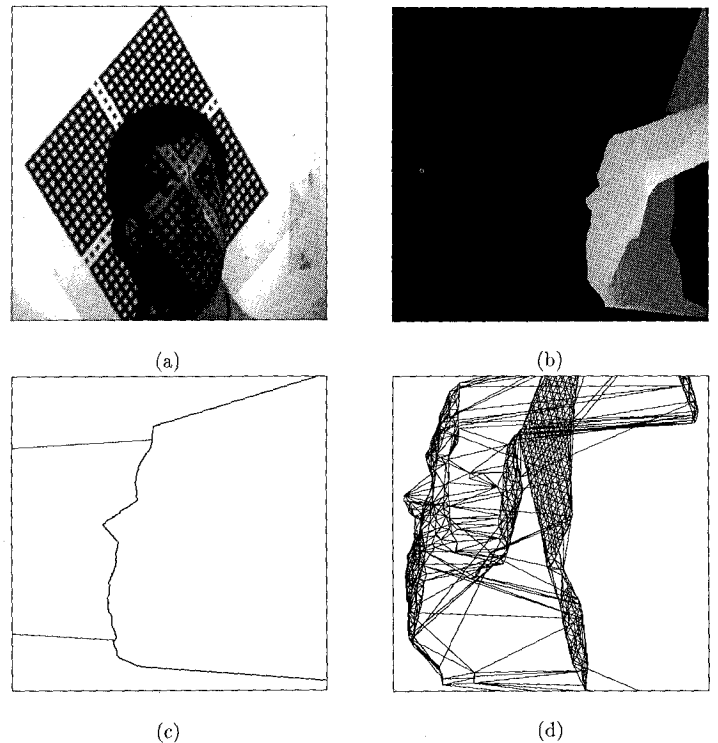


Figure 9. – Visage de la première personne.

Classiquement dans un système de stéréovision où l'information tridimensionnelle est estimée par triangulation, se pose le problème de la qualité de la reconstruction obtenue. Les principaux facteurs influençant la reconstruction sont :

- la précision de l'extraction des primitives : elle est de plus ou moins un pixel pour les primitives de l'image et a priori sans erreur pour les éléments du motif qui ont été tracés en utilisant une imprimante de bonne résolution,
- la qualité de la calibration : les tests statistiques que nous avons faits pour valider le modèle du projecteur donnent un ordre de grandeur de quelques millimètres (voir table 2.2.2.),
- l'écartement entre les deux points de base utilisés pour la triangulation comparé à l'éloignement des objets : dans notre cas, l'écartement est d'environ 0.5 m pour une profondeur de 2 m au maximum. Ce ratio assure a priori une bonne reconstruction.

Compte-tenu de ces remarques et des résultats expérimentaux présentés par les planches, on pourrait envisager d'utiliser les surfaces obtenues en entrée d'un algorithme de reconnaissance de visages.

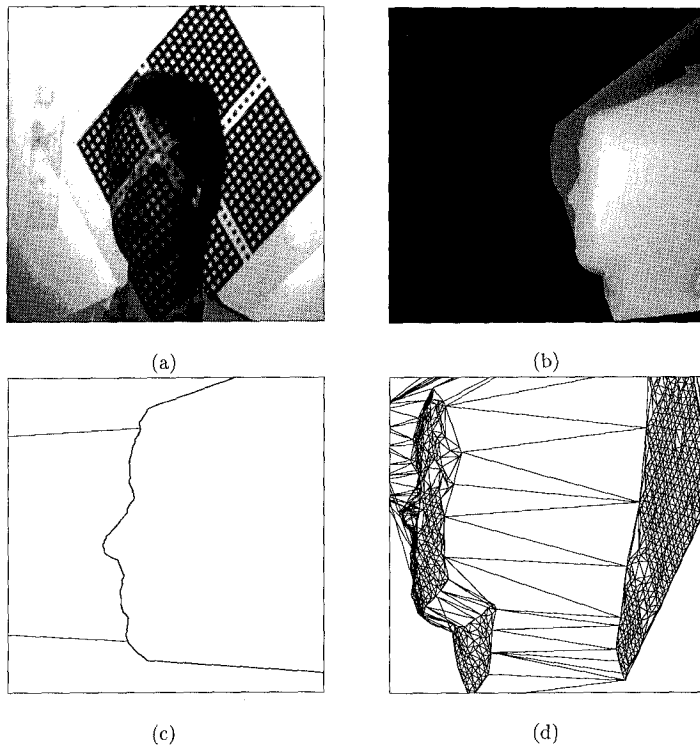


Figure 10. – Visage de la deuxième personne.

5. Conclusion

Dans cet article, nous nous sommes intéressé au problème de la mesure de la surface correspondant à un visage. Pour mesurer la surface du visage, nous avons choisi une approche de type stéréovision active. Elle repose sur la projection d'un motif sur la surface à observer. Les résultats obtenus nous paraissent satisfaisants même si l'algorithme de mise en correspondance s'est révélé parfois susceptible de produire des appariements erronés. Ils sont dus soit à l'existence de taches incorrectement détectées, soit à l'existence de zones d'ambiguïté. Ceci est notamment le cas lorsque le motif est projeté sur un plan, par exemple le mur derrière le visage observé. Dans ce cas, plusieurs solutions sont possibles pour appairer les carreaux correspondant aux zones planes. On retrouve l'un des problèmes décrits dans [Dunn-Keizer-Yu89].

Les surfaces reconstruites semblent être suffisamment précises pour que ces problèmes ne mettent pas en cause gravement la possibilité d'utiliser les surfaces mesurées pour l'identification de personnes. Il nous semble que certaines des méthodes de reconnaissance proposées dans la littérature ([Cartoux89]) pourraient être appliquées aux données obtenues par notre algorithme :

- l'observation des profils suivie de l'extraction de points caractéristiques, comme il est décrit dans [Najman-Vaillant-Pernot93], et [Harmon-Khan-Lasch-Ramig80],
- la mesure du volume entre la reconstruction d'un visage appartenant à la base de données et la reconstruction du visage à identifier,
- la méthode de vecteur d'éléments caractéristiques comme par exemple l'écartement des yeux, la longueur du nez, ...

Ces deux dernières méthodes sont décrites dans [Gordon91]

La mise en œuvre de ce type d'algorithme exige une première phase assez délicate de recalage des visages observés. Nous pouvons envisager deux solutions à ce problème :

- définir un protocole d'acquisition qui fournit directement la connaissance d'un point du visage. Nous pouvons par exemple imposer que le carreau situé à l'intersection des 2 bandes noires soit projeté exactement entre les deux yeux. Le problème de recalage serait simplifié.
- mesurer les courbures de la surface et délimiter les zones concaves et convexes. A partir de ces zones et en essayant de trouver le plan de symétrie du visage, il peut être envisageable comme dans [Gordon91] de récupérer quelques points caractéristiques sur le visage en vue de faire le recalage.

ANNEXE

A Extraction des taches par morphologie mathématique

Pour extraire les taches des images nous avons utilisé un algorithme de morphologie mathématique. La figure 11.(a) montre un exemple d'image dans laquelle on souhaite extraire les taches.

Le principe de l'algorithme est d'obtenir dans une première phase des marqueurs des taches, c'est-à-dire un ensemble de pixels connexes pour chaque tache, et de déterminer dans une deuxième phase de manière plus fine le contour de ces taches.

Extraction des marqueurs des taches blanches

L'algorithme retenu est fondé sur la notion de maxima régionaux sous contraintes. En effet, les taches claires correspondent à des maxima sur l'image. Afin de rendre robuste cette détection, nous n'avons conservé que les maxima régionaux présentant sur leurs bords un contraste suffisant, c'est-à-dire 30 niveaux de gris selon nos conditions d'acquisition des images. Une description détaillée de la procédure morphologique suivie peut être trouvée dans [Surin92].

Extraction des marqueurs des taches noires

La technique utilisée repose sur la notion de chapeau haut-de-forme. Comme nous voulons détecter les structures sombres, le

chapeau est la différence entre une fermeture et l'image originale. La fermeture doit gommer des taches compactes de diamètre au plus 15 pixels. La fermeture (algébrique) adaptée est une intersection de fermetures morphologiques par des segments de longueur 15 pixels dans diverses orientations. En effet, les taches compactes de diamètre au plus 15 pixels ne peuvent contenir un segment de longueur 15 pixels pour aucune orientation. Comme nous travaillons en trame hexagonale, nous avons pris 6 directions tous les 30 degrés. Le chapeau haut-de-forme est ensuite seuillé. Les quelques petites particules parasites sont ensuite éliminées par le filtrage "érosion de taille un puis reconstruction".

Détermination précise des contours

Le principe de la détection de contours fermés par ligne de partage des eaux est de déterminer des marqueurs (composantes connexes de pixels) correspondant à chacun des objets recherchés, trouver un marqueur du fond de l'image, puis à effectuer une ligne de partage des eaux contrainte du gradient de l'image originale. Le résultat de l'algorithme assure que chacun des objets sera entouré d'un seul lacet fermé, placé de manière optimale au sens de la ligne de partage des eaux.

Les figures 11.(b) et 11.(c) montrent respectivement les taches blanches et noires extraites de l'image 11.(a).

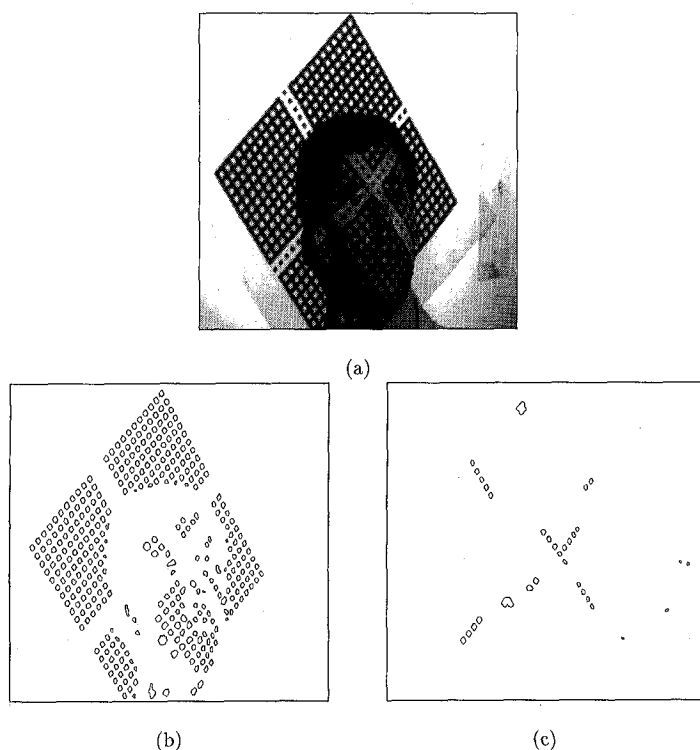


Figure 11. – Image analysée et les taches extraites.

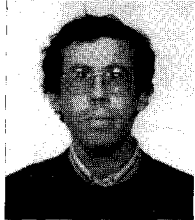
BIBLIOGRAPHIE

- [1] Nicolas Ayache and Francis Lustman. Fast and Realible Passive Trinocular Stereovision. In *First International Conference on Computer Vision*, June 1987.
- [2] D. Beymer. Face Recognition under Varying Pose. In *Computer and Pattern Recognition*, June 1994.
- [3] J.Y. Cartoux. *Formes dans les images de Profondeur: Application à la Reconnaissance et à l'authentification de Visages*. PhD thesis, Université Blaise Pascal de Clermont-Ferrand, October 1989.
- [4] J.Y. Cartoux, J.T. Lapreste and M. Richetin. Face Authentication by Profile Extraction from Range Images. In *Proceedings of IEEE Workshop on Interpretation of 3D Scenes*, pages 194-100, November 1989.
- [5] Stanley M. Dunn, Richard L. Keizer and Jongdaw Yu. Measuring the Area and Volume of the Human Body with Structured Light. *IEEE Transactions on Systems Man and Cybernetics*, 19 :1350-1364, November/December 1989.
- [6] Oliver Faugeras and Giorgio Toscani. The Calibration Problem for Stereo. In *Proceedings of CVPR '86, Miami Beach Florida*, pages 15-20, 1986.
- [7] Olivier D. Faugeras. *Three-dimensional Computer Vision : a geometric viewpoint*. MIT Press, 1993.
- [8] Pascal Fua. Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserves Discontinuities. In *12th International Joint Conference on Artificial Interlligence*, pages 1292-1298, August 1991.
- [9] Gordon Gaile Gibson. *Face Recognition from Depth and Curvature*. PhD thesis, Harward University, Cambridge, Massachusetts, September 1991.
- [10] L.D. Harmon, M.K. Khan, Richard Lasch and P.F. Ramig. Machine Identification of Human Faces. *Pattern Recognition*, 13 : 2 :97-110, 1980.
- [11] Gongzhu Hu and George Stockman. 3-D Surface Solution Using Structured Light and Constraint Propagation. *Pattern Analysis and Machine Intelligence*, 11(4) :390-402, April 1989.
- [12] John C. Lee and Evangelos Milios. Matching Range Images of Human Faces. In *Third International Conference on Computer Vision*, 1990.
- [13] Yael Moses, Yael Adini and Shimon Ullman. Face Recognition : the Problem of Compensating for Changes in Illumination Direction. In *European Conference on Computer Vision*, volume 1, pages 286-296, Stockholm, May 1994.
- [14] Laurent Najman, Régis Vaillant and Etienne Pernot. From Face Sideviews to Identification. In Gianni Vemazza, editor, *Image Processing : Theory and Applications*, San Remo, Italy, 1993. Elsevier.
- [15] A. Pentland, B. Moghaddam, O. Starner, T. Oliyide and M. Turk. View-Based and Modular Eigenspaces for Face Recognition. In *Computer Vision and Pattern Recognition*, June 1994.
- [16] Luc Robert and Olivier Faugeras. Curve-Based Stereo : Figural Continuity and Curvature. In *Computer Vision and Pattern Recognition*, 1991.
- [17] Ashok Samal and Prasana Iyengar. Automatic recognition and analysis of human faces and facial expressions : a survey. *Pattern Recognition*, 25(1) : 65-77, 1992.
- [18] Isabelle Surin. Reconstruction de Visages par stéréovision Active. Technical Report ASRF-92-5, Thomson-CSF, L.C.R., September 1992. Rapport de stage du DEA Robotique et Vision par Ordinateur de l'Université de Nice Sophia-Antipolis.
- [19] Thom. Production Automatique de Modèles Numériques Tridimensionnels d'Objets en Temps Quasi-Réel à l'aide d'une caméra CCD Vidéo. Research report. IGN, November 1988.
- [20] Giorgio Toscani. *Système de Calibration et Perception du Mouvement en vision artificielle*. Ph.D thesis, Université de Paris-Sud Centre d'Orsay, December 1987.
- [21] Mathew A. Turk and Alex P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1) :72-86, 1991.
- [22] Y.F. Wang and P. Liang. A New Method Computing Intrinsic Surface Properties. In *Conference on Computer Vision and Pattern Recognition*, pages 235-240, 1989.

Manuscrit reçu le 5 mai 1994

LES AUTEURS

Régis VAILLANT



Régis Vaillant : École Polytechnique 1986; Doctorat en Sciences de l'Informatique 1990. Ingénieur au Laboratoire Perception du Laboratoire Central de Recherches de THOMSON-CSF où il mène des activités de recherche en vision tridimensionnelle.

Isabelle SURIN



Isabelle Surin : Maîtrise de mathématiques (Université de Nice Sophia-Antipolis, 1991), DEA "Théorie et Applications de la Vision Artificielle et Robotique" (Université de Nice Sophia-Antipolis, 1992). Ingénieur d'études en algorithme de traitement d'images dans le cadre d'une thèse de doctorat (convention CIFRE).