

Acquisition à faible coût de modèles 3D réalistes de visages

Emmanuel Garcia, Jean-Luc Dugelay,
Institut Eurécom, Département Multimédia,
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis,
{garciae,jld}@eurecom.fr, <http://www.eurecom.fr/~image>

Résumé

Cet article présente un procédé de reconstruction 3D de visages utilisant un rétroprojecteur classique et une caméra. La géométrie du visage est obtenue en analysant les déformations d'une grille lumineuse qui se projette dessus. On obtient la texture du visage à partir d'autres images du visage où la grille lumineuse est absente. Ce procédé nécessite une étape de calibrage du système d'acquisition après laquelle nous acquérons indépendamment la géométrie 3D puis la texture du visage. Le calibrage de l'ensemble caméra-rétroprojecteur est effectué à l'aide d'un simple objet plan sur lequel est dessiné un quadrillage dont nous prenons l'image dans au moins trois configurations différentes. La reconstruction de la géométrie complète du visage est obtenue par la fusion de quelques modèles partiels, chaque modèle partiel étant reconstruit à partir d'une seule vue du visage. Enfin, le texturage du modèle fait intervenir le recalage de l'image de texture avec le modèle 3D complet avant de la rétroprojeter sur le maillage 3D du visage.

Mots-clés

Reconstruction 3D de visage, lumière structurée, calibrage, recalage, texturage.

1 Introduction

Pour nombre d'applications multimédia (*e.g.* MPEG-4 [1]) il est important de pouvoir reconstituer les expressions faciales d'un visage sur un modèle synthétique à partir de l'analyse de signaux audio ou vidéo. Les modèles synthétiques utilisés peuvent être classés dans deux catégories : les avatars et les clones. Les avatars sont de grossières approximations d'un visage humain et sont souvent peu réalistes mais sont faciles à créer puis à animer. Les clones quant à eux doivent donner l'image d'un visage réel. Pour certaines applications, telles que la téléconférence virtuelle [2] il est nécessaire d'employer des modèles très réalistes. Mais l'acquisition de modèles réalistes nécessite souvent un matériel spécialisé tel qu'un scanner *Cyberware* [3], dont le coût est élevé et dont peu de personnes peuvent disposer. Une approximation d'un clone peut être obtenue en utilisant un maillage 3D générique de visage. Le modèle est alors personnalisé seulement en lui appliquant une texture obtenue à l'aide d'une [4] ou deux photos (face et profil). La géométrie du modèle peut également être adaptée dans une certaine mesure au sujet [5]. Sinon, il existe des méthodes classiques en vision par ordinateur pour obtenir la géométrie 3D d'un objet. Elles se rangent en deux catégories : méthodes passives et méthodes actives. Les premières sont souvent basées sur une analyse de plusieurs points de vue [6], les dernières consistent souvent en la projection d'un motif sur la personne à modéliser [7]. Dans cet article nous proposons et décrivons un cadre de travail complet pour l'acquisition et la modélisation de modèles réalistes de visages aussi bien en termes de géométrie que de texture.

2 Notre approche

Notre procédé de reconstruction utilise une seule caméra, un rétroprojecteur et un objet de calibrage plan. Un motif de quadrillage est projeté sur le visage du sujet et ses déformations sont analysées dans le plan image de la caméra pour obtenir la géométrie 3D de la partie visible

du visage. De nombreux travaux [7, 8, 9, 10, 11] ont été consacrés à ce genre de méthodes. Pour obtenir une reconstruction euclidienne le système doit néanmoins être calibré auparavant. Pour cela l'image d'un objet de calibrage connu à l'avance est analysée. Cet objet est un plan sur lequel un quadrillage est dessiné et qui doit être observé dans au moins trois positions différentes. Le calibrage par un plan, du moins en ce qui concerne le calibrage de la caméra, a également fait l'objet de nombreux travaux [14, 15, 16]. Ici nous traitons le problème du calibrage de tout le système et contrairement à [12] nous ne faisons pas l'hypothèse que la projection du motif lumineux est une projection parallèle au voisinage du visage à reconstruire. Quelques modèles construits à partir de vues différentes du visage sont utilisés pour construire un modèle complet. Ensuite une image du visage sans le motif lumineux est utilisée comme texture et plaquée sur le modèle 3D.

3 Calibrage

3.1 Modèle de calibrage

Nous adoptons ici un modèle de projection perspective sans distorsion pour la caméra et le projecteur. Dans ce cas un point de l'espace de coordonnées X dans le repère lié à un système de projection perspective se projette dans le plan image de la projection en un point de coordonnées homogènes $x \equiv AX$ où A est la matrice 3×3 des paramètres intrinsèques du système de projection, laquelle peut se mettre sous la forme classique $A = \begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ moyennant un choix convenable du repère lié au système de projection.

Le calibrage consiste à estimer les paramètres intrinsèques A_1 de la caméra, les paramètres intrinsèques A_2 du projecteur ainsi que la transformation $[R \ T]$ permettant de passer du repère $\mathcal{R}_2$ du projecteur au repère $\mathcal{R}_1$ de la caméra.

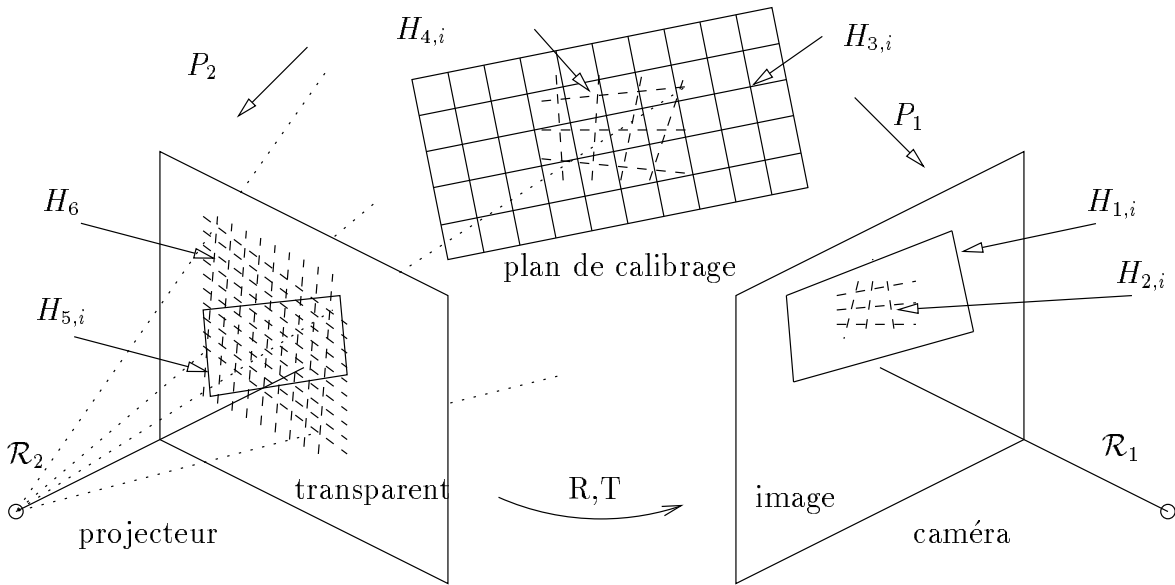
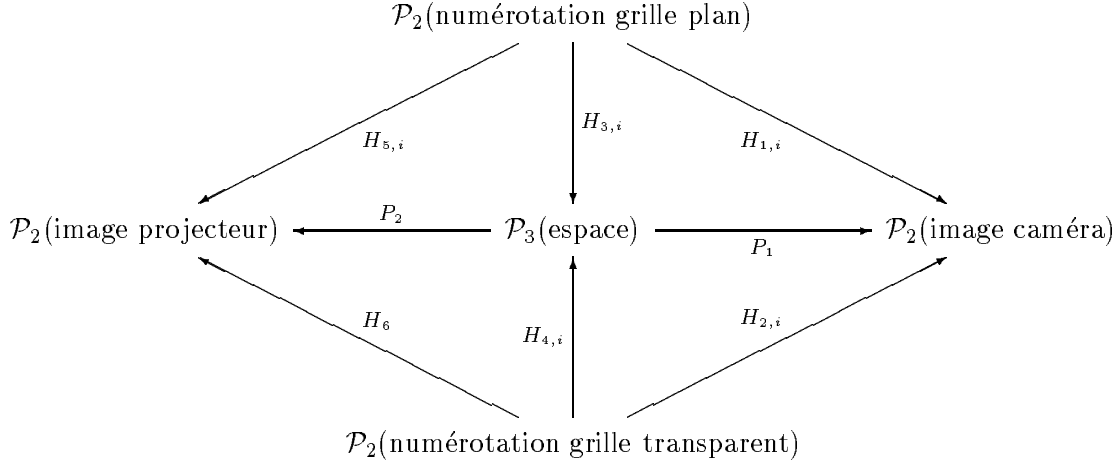


FIG. 1 – *Modèle de calibrage.*

Comme l'illustre la figure 1, notre méthode consiste à utiliser un plan sur lequel est dessiné un quadrillage carré (H_3) et sur lequel se projette le quadrillage carré dessiné sur le transparent H_6 . Un minimum de trois configurations (indiquées par i) du même plan est nécessaire pour calibrer le système.

Chacun des quadrillages peut être défini dans l'espace ou dans le plan image de la caméra ou du projecteur par une fonction homographique (désignée par la lettre H) qui associe aux indices d'un nœud d'un quadrillage ses coordonnées projectives dans l'espace ou dans l'un des deux plans image. C'est ce que résume le diagramme commutatif suivant ($\mathcal{P}_2$ et $\mathcal{P}_3$ sont les espaces

projectifs de dimensions 2 et 3 respectivement) :



Seule H_6 est indépendante de la position du plan de calibrage et on peut la supposer égale à l'identité moyennant un changement des paramètres intrinsèques du projecteur. Les homographies P_1 et P_2 sont les projections de l'espace sur le plan image de la caméra et du projecteur respectivement.

3.2 Calibrage de la caméra

Pour calibrer la caméra nous nous servons simplement de l'image de la grille dessinée sur le plan de calibrage dans au moins trois configurations différentes ($H_{1,i}$) et utilisons une méthode tout à fait similaire à celle présentée dans [14]. Pour cela nous estimons tout d'abord les homographies $H_{1,i}$ en détectant les droites dans les images de calibrage (voir plus loin). Ces homographies sont connues à une isométrie près (origine arbitraire de la numérotation, échange ou changement de sens des deux variables de numérotation) de la numérotation des nœuds des quadrillages qu'elles représentent.

Soit (t, e_1, e_2) (exprimé dans le repère $\mathcal{R}_1$) le repère du plan de calibrage tel que le nœud du plan d'indices (p, q) soit $t + pe_1 + qe_2$. En projetant dans le plan image par la matrice A_1 on obtient :

$$H_1 \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \equiv A_1 \begin{bmatrix} e_1 & e_2 & t \end{bmatrix} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (3.1)$$

d'où par identification :

$$\begin{bmatrix} e_1 & e_2 \end{bmatrix} \equiv A_1^{-1} \begin{bmatrix} c_1 & c_2 \end{bmatrix} \quad (3.2)$$

où c_1 et c_2 sont les deux premières colonnes de H_1 . Et comme (e_1, e_2) est un système de vecteurs orthonormés, on obtient les deux contraintes :

$$\begin{cases} c_1^T Q c_2 & = 0 \\ c_1^T Q c_1 - c_2^T Q c_2 & = 0 \end{cases} \quad (3.3)$$

qui sont linéaires en $Q = A_1^{-T} A^{-1}$. La matrice Q possède, comme A_1 , cinq degrés de liberté donc il faut au moins trois jeux d'équations de ce type obtenus en considérant plusieurs positions du plan de calibrage qui donnent des homographies $H_{1,i}$. On peut alors estimer Q aux moindres carrés en utilisant la décomposition en valeurs singulières de la matrice du système. La décomposition de Choleski de Q permet ensuite d'obtenir A_1^{-1} puis A_1 .

Une méthode plus poussée basée sur le même principe et prenant notamment en compte la distorsion radiale est détaillée dans [14]. Pour l'instant nous ne prenons pas en compte la

distorsion radiale car nous faisons l'hypothèse que les droites dessinées sur le plan se projettent selon des droites afin de les détecter dans l'image de la caméra. Dans notre cas, détecter des droites est en effet plus facile que détecter d'autres motifs car nous avons deux motifs superposés (celui dessiné sur le plan de calibrage et celui qui se projette dessus). Néanmoins, pour calibrer la caméra dans un premier temps, nous pourrions éteindre le rétroprojecteur, utiliser un autre motif sur le plan de calibrage, et utiliser strictement la méthode présentée par exemple dans [14] pour prendre en compte la distorsion radiale si cela était nécessaire.

3.3 Estimation de la matrice fondamentale

Pour calibrer le reste du système d'acquisition nous allons dans un premier temps essayer de tirer le maximum d'information de la seconde image de quadrillage H_2 sans utiliser l'information fournie par H_1 , (en particulier les paramètres intrinsèques de la caméra A_1), afin de limiter la propagation de l'incertitude sur cette information.

L'information de base que l'on peut tirer de H_2 est la matrice fondamentale entre les plans image de la caméra et du projecteur. En effet, les nœuds du quadrillage dessiné sur le transparent (H_6) et leurs images ($H_{2,i}$) dans la caméra sont les images du même $H_{4,i}$ par deux projections différentes et sont donc liés par la contrainte épipolaire (voir [18] pour plus de détails sur la stéréovision). Cette contrainte s'exprime par la relation $x_1^T F x_2 = 0$ où x_1 et x_2 sont les images d'un même point X de l'espace par deux projections différentes (la caméra et le rétroprojecteur dans notre cas), et où F est une matrice de rang 2 définie à un facteur près.

Le problème est que les nœuds de H_6 et des $H_{2,i}$ sont numérotés de manière indépendante d'un quadrillage à l'autre. Pour que la numérotation entre H_6 (que nous avons supposé être l'identité) et les $H_{2,i}$ soit cohérente il est nécessaire de renuméroter les nœuds des $H_{2,i}$ en

remplaçant les $H_{2,i}$ par $H_{2,i}U_i$ où $U_i = \begin{bmatrix} r_i & t_i \\ 0 & 1 \end{bmatrix}$ est une isométrie avec $r_i = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$ ou $r_i = \begin{bmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{bmatrix}$.

On peut retrouver les r_i assez simplement en supposant et en faisant en sorte lors de la mise en place du système que les droites des quadrillages $H_{2,i}$ soient réparties suivant deux directions grossièrement horizontale et verticale et en numérotant les nœuds toujours dans le même sens (vers le haut et vers la droite par exemple). Lorsque les r_i sont calculés on peut les incorporer aux $H_{2,i}$ et les ignorer (les prendre égaux à l'identité) par la suite.

Pour estimer les t_i nous allons d'abord calculer une partie de la matrice fondamentale. Etant donné un nœud d'indices (p, q) dans la grille du transparent, et dont les images dans la caméra et le transparent sont respectivement x_1 et x_2 , on a $x_1^T F x_2 = 0$. De plus $F x_2$ est la droite épipolaire de x_2 , sur laquelle se trouve x_1 . Cette droite passe par e , image (épipole) du centre de projection du projecteur dans la caméra, et par x_1 , donc on a $F x_2 \equiv e \times x_1$. Comme $x_2 \equiv \begin{bmatrix} p \\ q \\ 1 \end{bmatrix}$

(car H_6 est prise égale à l'identité) et $x_1 \equiv H_2 \begin{bmatrix} p \\ q \\ 1 \end{bmatrix}$ on a par identification $F \equiv e \times H_2$.

Pour toute image i nous devons donc avoir $F \equiv e \times H_{2,i}U_i$, ce qui en notant

$H_{2,i} = [c_{1,i} \ c_{2,i} \ c_{3,i}]$, $U_i = \begin{bmatrix} I & t_i \\ 0 & 1 \end{bmatrix}$ et $t_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$ se réécrit

$F \equiv [e \times c_{1,i} \ e \times c_{2,i} \ \alpha_i e \times c_{1,i} + \beta_i e \times c_{2,i} + e \times c_{3,i}]$.

On s'aperçoit que dans cette expression de F les deux premières colonnes ne dépendent pas des t_i inconnus. On peut donc directement les estimer à un facteur d'échelle près indépendant pour chacune en posant qu'elles sont orthogonales aux $c_{1,i}$ et aux $c_{2,i}$ respectivement. Comme elles sont également orthogonales à e , leur produit vectoriel fournit une estimation de e .

La troisième colonne de F doit également être indépendante de la configuration i considérée. On lui impose d'être nulle, ce qui est possible moyennant une translation globale de tous les t_i et une modification correspondante des paramètres intrinsèques A_2 du projecteur (qui seront estimés ultérieurement). Une première estimation des t_i est donc obtenue en résolvant pour chaque i :

$$\alpha_i e \times c_{1,i} + \beta_i e \times c_{2,i} + e \times c_{3,i} = 0 \quad (3.4)$$

Une fois les t_i estimés nous pouvons les incorporer aux $H_{2,i}$ et les ignorer par la suite. Maintenant que les $H_{2,i}$ sont donc numérotés de manière cohérente entre eux et avec les nœuds du transparent, nous pouvons estimer la matrice fondamentale F par une méthode classique en stéréovision, en l'occurrence par une première estimation linéaire puis par une optimisation non linéaire visant à minimiser la somme des distances quadratiques entre l'image des nœuds observés et les droites épipolaires sur lesquelles ils devraient se trouver. L'estimation linéaire initiale consiste à résoudre aux moindres carrés le système linéaire en F des équations $x_1^T F x_2 = 0$.

Enfin, nous améliorons l'estimation des t_i à partir de cette estimation initiale de F en cherchant à minimiser le critère précédent par rapport aux t_i . Puis on estime à nouveau F avec les nouveaux t_i . On itère successivement ces deux optimisations une dizaine de fois après quoi nous n'observons plus d'amélioration sensible.

3.4 Calibrage du rétroprojecteur

Le calibrage du rétroprojecteur (paramètres intrinsèques A_2) est identique au calibrage de la caméra. Nous utilisons les images ($H_{5,i}$) dans le plan image du projecteur de la grille dessinée sur le plan de calibrage ($H_{3,i}$).

La seule difficulté est que nous ne connaissons pas ces images. Mais nous pouvons les calculer en observant le diagramme commutatif plus haut. En effet, d'après ce diagramme on a $H_5 \equiv H_6 H_2^{-1} H_1$ soit $H_5 \equiv H_2^{-1} H_1$ car on a supposé $H_6 \equiv I$. Cela ne fait intervenir que les homographies associées aux images des quadrillages observés dans la caméra mais nécessite que les différents $H_{2,i}$ soient numérotés de manière cohérente, d'où l'importance de l'étape précédente.

3.5 Estimation de la configuration spatiale du système

Après les paramètres intrinsèques associés aux deux systèmes de projection il s'agit de calculer la transformation (R, T) permettant de passer du repère du projecteur à celui de la caméra de façon à pouvoir effectuer une reconstruction euclidienne des objets observés.

Pour cela on utilise notre connaissance de la matrice fondamentale et des paramètres intrinsèques du système, ainsi qu'une relation classique en stéréovision :

$$F \equiv A_1^{-T} R[T]_{\times} A_2^{-1} \quad (3.5)$$

où $[T]_{\times}$ représente l'opérateur linéaire correspondant au produit vectoriel par le vecteur T . D'après cette relation on a donc :

$$R[T]_{\times} \equiv A_1^T F A_2 = G \quad (3.6)$$

On trouve T , à un facteur près, comme étant dans le noyau de G . Les vecteurs colonnes de $[T]_{\times}$ engendrent le plan orthogonal à T , on connaît donc la restriction de R à ce plan (le plan image est engendré par les colonnes de G). On peut naturellement poser que l'image de T par R , est la normale au plan engendré par les colonnes de G . Connaissant ainsi l'effet de R sur un ensemble de vecteurs qui engendrent l'espace, on peut calculer R . Le résultat n'est généralement pas exactement une rotation en raison de l'incertitude sur les valeurs estimées de A_1 , A_2 et F . Pour retrouver une rotation R' proche d'une rotation approximative R , [14] suggère de décomposer R en valeurs singulières selon $R = U W V^T$ et de poser $R' = U V^T$.

Enfin, pour déterminer le facteur d'échelle à appliquer à T , nous effectuons la reconstruction 3D du plan de calibrage avec un facteur d'échelle arbitraire pour T et nous calculons le pas de la grille dessinée sur ce plan reconstruit (il est proportionnel à la norme de T). Nous ajustons alors le facteur d'échelle de T pour que ce pas soit égal à la valeur physique mesurée sur le plan réel. Il est alors possible d'effectuer une reconstruction euclidienne en unités physiques sans facteur d'échelle.

3.6 Détection des nœuds dans les images de calibrage

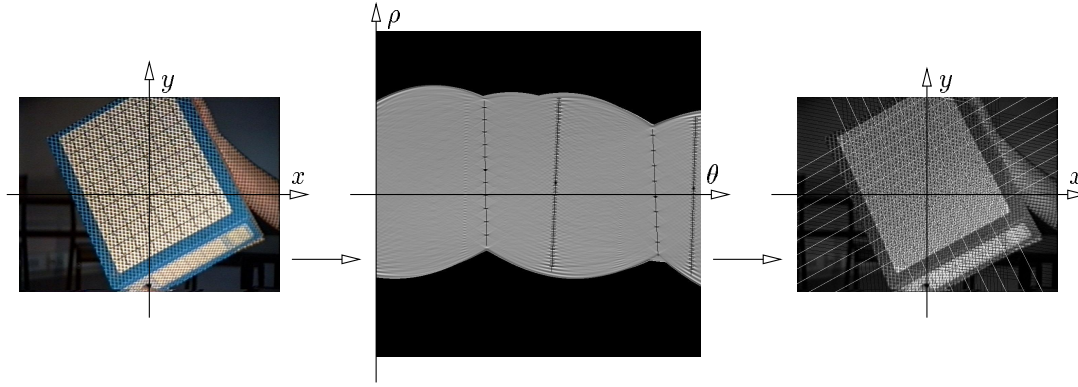


FIG. 2 – *Détection des droites dans les images de calibrage.*

Bien que les faisceaux de droites soient très denses, nous pouvons déterminer toutes les droites en utilisant la transformée de Hough. Dans la figure 2 l'intensité $I'(\rho, \theta)$ d'un point (ρ, θ) de la deuxième image est reliée à l'intensité $I(x, y)$ des points de la première image par la formule :

$$I'(\rho, \theta) = \int I(\rho \cos(\theta) - \lambda \sin(\theta), \rho \sin(\theta) + \lambda \cos(\theta)) d\lambda \quad (3.7)$$

i.e. c'est la somme des intensités des points de la première image qui se trouvent sur la droite d'équation $x \cos(\theta) + y \sin(\theta) = \rho$.

Dans l'espace de Hough chaque droite lumineuse est représentée par un point lumineux et chaque droite sombre est représentée par un point sombre. En effet, un ensemble de points d'intensité extrême situés sur une droite d'équation $x \cos(\theta) + y \sin(\theta) = \rho$ se traduisent dans l'espace de Hough par un point (ρ, θ) d'intensité extrême.

Un faisceau de droites presque parallèles est représenté dans l'espace de Hough par un ensemble de points presque alignés et régulièrement espacés. Comme nous avons quatre faisceaux de droites à détecter, nous devons chercher dans l'espace de Hough quatre ensembles de points quasiment alignés, ce qui n'est pas trop difficile si la résolution de la transformée de Hough est suffisamment élevée pour que les points à détecter soient bien séparés.

Dans la deuxième image nous avons marqué les points représentatifs des droites à détecter ainsi que leur groupement en quatre colonnes. Et dans la troisième image nous avons superposé à l'image originale les droites correspondant aux points détectés dans l'espace de Hough.

4 Reconstruction 3D

Le système étant calibré il nous est possible de reconstruire, à partir d'une image du visage, un modèle partiel du visage correspondant à la partie du visage qui est visible sous un angle d'incidence suffisamment grand à la fois depuis la caméra et le projecteur. En demandant à la personne de tourner la tête on peut prendre plusieurs vues et reconstruire ainsi autant de modèles partiels que nous fusionnons en un modèle plus complet.

4.1 Analyse des images de visage

La première étape de la construction d'un modèle partiel du visage consiste à détecter les nœuds de la grille projetée sur le visage dans l'image donnée par la caméra (Fig. 3a), et à les numéroter correctement dans leur structure de grille.

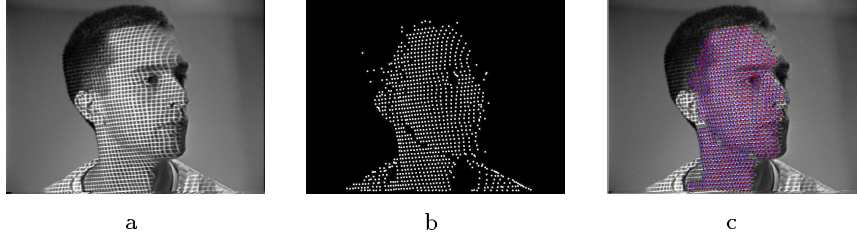


FIG. 3 – *Extraction des nœuds dans les images du visage.*

Nous utilisons ici une méthode simple qui consiste à détecter les lignes verticales puis horizontales dans l'image (on suppose et on fait en sorte que la grille se projette selon des directions grossièrement verticales et horizontales sur le visage) puis à définir les nœuds comme l'intersection de ces deux ensembles de lignes. Pour détecter les lignes verticales par exemple, nous effectuons un filtrage convolutif par la dérivée seconde d'une gaussienne dans la direction horizontale (pour faire ressortir le maximum de corrélation avec la gaussienne, que nous choisissons comme profil de section de ligne). Puis après seuillage nous effectuons quelques opérations de morphologie mathématique pour assurer la continuité des lignes détectées.

Ensuite nous numérotions les nœuds à l'aide de trois indices : numéro de composante connexe et indices de ligne et de colonne. Initialement chaque nœud appartient à une composante connexe différente et possède des indices de ligne et de colonne arbitraires. Ensuite nous ajoutons de manière incrémentale des liens entre nœuds adjacents en considérant d'abord les liens les plus probables d'après l'image. Si l'ajout d'un lien réunit deux composantes connexes jusque-là distinctes, on décale la numérotation des nœuds de l'une des composantes connexes pour que les deux nœuds juste reliés aient des indices de ligne et de colonne voisins.

Cette méthode permet généralement de détecter tous les nœuds (Fig. 3b) mais le graphe de connexité obtenu nécessite une vérification manuelle (Fig. 3c). Notamment autour des yeux ou du nez où s'il y a une discontinuité certains liens peuvent avoir été placés par erreur. Une méthode plus sophistiquée, plus précise et certainement plus robuste est présentée dans [11].

4.2 Reconstruction partielle

La reconstruction d'un modèle partiel s'effectue par triangulation comme en stéréovision car on connaît la position des nœuds du transparent (après le calibrage) ainsi que les nœuds dans l'image de la caméra (après l'étape précédente). La principale difficulté est de connaître la correspondance exacte entre les nœuds dans l'image et les nœuds dans le transparent (on suppose qu'il n'y a qu'un décalage de numérotation mais que l'orientation des numérotations est cohérente).

Soit donc $(x_{p,q})$ l'ensemble des nœuds détectés et numérotés par (p, q) dans l'image du visage. Idéalement le nœud $x_{p,q}$ correspond au nœud d'indices (p, q) et donc (voir calibrage) de coordonnées $[p \ q \ 1]^T$ du transparent. En réalité la numérotation des nœuds dans l'image du visage a été effectuée à un décalage global près, si bien que le nœud $x_{p,q}$ est le correspondant du point de coordonnées $[p + \alpha \ q + \beta \ 1]^T$ du transparent où (α, β) représente le décalage à déterminer entre les numérotations des nœuds du transparent et de l'image.

Nous estimons ces paramètres en utilisant la contrainte épipolaire et la matrice fondamentale estimée lors du calibrage. On doit en effet avoir pour chaque nœud :

$$x_{p,q}^T F \begin{bmatrix} p + \alpha \\ q + \beta \\ 1 \end{bmatrix} = 0 \quad (4.8)$$

Ces équations sont linéaires en α et β . Nous résolvons aux moindres carrés le système ainsi formé. Puis nous améliorons la solution de manière non linéaire en minimisant la somme des distances quadratiques dans l'image de la caméra entre les nœuds $x_{p,q}$ et les droites épipolaires

$F \begin{bmatrix} p + \alpha \\ q + \beta \\ 1 \end{bmatrix}$ sur lesquelles ils devraient se trouver. Une fois ces paramètres estimés nous pouvons reconstruire un modèle partiel de visage pour chaque image du visage utilisée.

4.3 Reconstruction complète

La reconstruction de la géométrie complète du visage consiste à construire plusieurs modèles partiels (Fig. 4a) obtenus en demandant au sujet de faire face à la caméra puis de présenter ses deux profils. Ces modèles doivent ensuite être recalés (par une simple transformation rigide il s'agit de reconstructions euclidiennes toutes à la même échelle) avant d'être fusionnés en un maillage unique.

Le recalage de deux modèles partiels est effectué en minimisant une fonction de potentiel traduisant la proximité des deux surfaces à recaler. La fonction utilisée est :

$$V = \sum_i \sum_j m_i m'_j f(x_i x'_j) u_i \cdot u'_j \text{ avec } f(x) = \begin{cases} -\frac{3}{2} + \frac{1}{2}(\frac{x}{\delta})^2 & \text{pour } x \leq \delta \\ -(\frac{x}{\delta})^{-1} & \text{pour } x \geq \delta \end{cases} \quad (4.9)$$

Dans cette formule x_i , m_i et u_i représentent respectivement le centre, la masse (en l'occurrence la surface) et la normale unitaire d'une facette d'un des modèles à recaler et x'_j , m'_j et u'_j sont analogues mais concernent le second modèle à recaler. Ici, f a la forme d'un potentiel newtonien similaire à celui engendré par une boule de densité uniforme et de rayon δ (nous avons pris $\delta = 2$ mm).

Minimiser V conduit à rapprocher les facettes les plus proches des deux surfaces (en ce sens cela ressemble à un algorithme de type *Iterative Closest Point* [17]) et à aligner les normales de ces facettes.

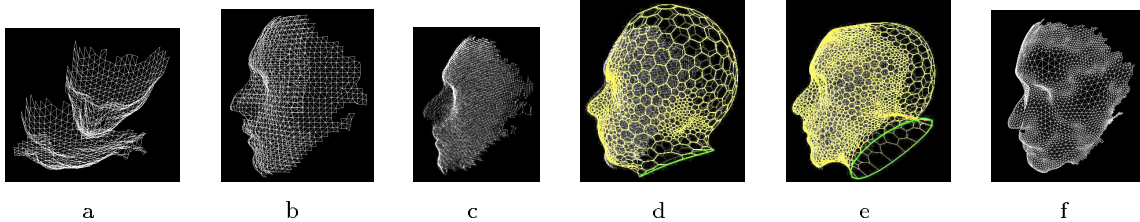


FIG. 4 – *Fusion des modèles partiels.*

Une fois les modèles partiels recalés (Fig. 4b), ils sont fusionnés en un modèle 3D complet du visage. Cette opération est réalisée en calculant des cartes de profondeur cylindriques pour chaque modèle partiel et en les moyennant en une carte de profondeur unique du modèle complet (Fig. 4c). Un maillage du modèle est alors obtenu en adaptant de manière itérative un maillage générique de visage sur la surface définie par cette carte de profondeur (Fig. 4d, 4e). Le maillage utilisé est un maillage simplexe [13] qui est converti, après adaptation à nos données, en un maillage triangulaire du visage (Fig. 4f).

5 Plaquage de la texture

Des tests préliminaires ont montré qu'il serait très difficile, voire impossible, de supprimer le motif lumineux sur les images du visage utilisées pour la reconstruction 3D tout en conservant une image de texture de haute résolution et dont les couleurs ne soient pas altérées (néanmoins l'approche exposée dans [7] semble y parvenir avec une structure de grille beaucoup plus fine).



FIG. 5 – *Plaquage d'une image de texture.*

Nous avons donc choisi de plaquer une image du visage où la grille lumineuse est absente (Fig. 5a) sur le modèle 3D global obtenu à l'étape précédente. Le problème est que cette image

est prise indépendamment des images utilisées pour la reconstruction : le sujet peut avoir bougé. Nous devons donc recalculer le modèle 3D complet avec cette image de texture. C'est-à-dire, nous devons trouver la position du modèle dans l'espace. Tel que vu par la caméra, il se projette selon les traits du visage dans l'image de texture. Pour cela nous sélectionnons manuellement des points caractéristiques correspondants sur le modèle 3D (il suffit de les sélectionner sur les images ayant servi à la reconstruction 3D pour pouvoir les localiser sur le modèle dans l'espace) et sur l'image de texture, typiquement le coin des yeux, des lèvres ou le bout du nez. Nous calculons alors la transformation rigide à appliquer au modèle 3D pour laquelle les points caractéristiques sur le modèle 3D se projettent au mieux selon les points caractéristiques correspondants dans l'image de texture. Nous rétroprojetons alors simplement l'image de texture située dans le plan image de la caméra dans l'espace et sur le modèle 3D. Ce recalage de l'image de texture et du modèle 3D peut être effectué précisément car le modèle 3D obtenu est un modèle fidèle du visage à reconstruire. La figure 5b montre le modèle texturé par l'image de la figure 5a, mais sous un angle de vue différent.



FIG. 6 – *Fusion de plusieurs images de texture.*

Enfin, quelques images de texture différentes peuvent être utilisées et moyennées en une seule image de texture, notamment pour couvrir la vue de face et les vues de profil (Fig. 6). Nous utilisons pour cela une représentation cylindrique des différentes textures qui, étant toutes recalées par rapport au modèle 3D, le sont entre elles.

Etant donné un point de coordonnées x dans le repère des images de texture cylindrique, on note $C_i(x)$ sa couleur (vecteur des composantes rouge, verte et bleue) dans l'image de texture cylindrique numéro i . Cette couleur n'est définie, pour une image de texture cylindrique donnée, que dans un domaine D_i (correspondant à la partie du visage visible dans l'image utilisée pour la construction de la texture cylindrique). Une texture cylindrique moyenne et plus complète C est construite en posant pour chaque point x :

$$\left(\sum_{i|x \in D_i} d(x, F_i) \right) C(x) = \sum_{i|x \in D_i} (d(x, F_i) C_i(x)) \quad (5.10)$$

où F_i représente la frontière de D_i et $d(x, F_i)$ la distance de x à F_i (nous avons utilisé la distance induite par la 4-connexité des pixels). Cette moyenne pondérée des $C_i(x)$ par $d(x, F_i)$ permet d'éviter les discontinuités dues aux changements du nombre de termes intervenant dans la moyenne $C(x)$ aux frontières F_i .

6 Conclusions

Nous avons présenté un protocole complet de reconstruction 3D de visage utilisant peu de moyens matériels. Ce protocole comporte en particulier une étape de calibrage du système, une étape de reconstruction de modèles partiels du visage et leur fusion puis une étape de texturage du modèle complet.

Nous nous sommes focalisés sur la mise en œuvre du protocole complet mais chaque étape nécessite maintenant d'être revue séparément, évaluée quantitativement et si possible améliorée, notamment du calibrage qui semble assez sensible.

Il reste que nous avons pu utiliser un modèle construit par notre procédé dans nos expériences de suivi de visages [19], lesquelles font intervenir une boucle d'analyse/synthèse et utilisaient originellement un modèle *Cyberware*. Notre prochain objectif est la construction automatique de modèles animés fidèles à partir de simples images ou vidéos.

7 Remerciements

Ces travaux sont en partie supportés par France Télécom Recherche et l'INRIA Sophia Antipolis (COLOR 2000). Nous remercions en particulier Hervé Delingette de l'INRIA pour avoir mis à notre disposition son logiciel de manipulation de maillages simplexes [13].

Références

- [1] *Overview of the MPEG-4 standard*, Technical Report ISO/IEC/JTC1/SC29/WG11 N2725, Int. Organization for Standardization, Séoul, Corée, mars 1999.
- [2] J.-L. Dugelay, S. Valente, K. Fintzel, *Synthetic/Natural Hybrid Video Processings for Virtual Teleconferencing Systems*, Proc. Picture Coding Symp., avril 1999.
- [3] CYBERWARE Home Page - <http://www.cyberware.com>.
- [4] A.C. Andrés del Valle, J. Ostermann, *Design of Quickly Adaptable Models for Talking Heads to be used in Interactive Internet Applications*, Internal Report, AT&T Labs research, 2000.
- [5] L. Yin, A. Basu, *MPEG-4 Face Modeling Using Fiducial Points*, International Conference on Image Processing, vol. 1, 1997, pp. 109-112.
- [6] European Project ACTS-092 PANORAMA - <http://www.tnt.uni-hannover.de/project/eu/panorama/>.
- [7] M. Proesmans, L. Van Gool, *One shot 3D shape and Texture Acquisition of Facial Data*, Audio- and Video-based Biometric Person Authentication, Crans-Montana, Suisse, mars 1997, pp. 411-418.
- [8] L. Guisser, R. Payrissat, S. Castan, *PGSD: an accurate 3D vision system using a projected grid for surface descriptions*, Image and Vision Computing, 2000, vol. 18, pp. 463-491.
- [9] C. Beumier, M. Acheroy, *3D Facial Surface Acquisition by Structured Light*, International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorin, Grèce, 15-17 septembre, 1999, pp. 103-106.
- [10] G. Hu, G. Stockman, *3-D surface solution using structured light and constraint propagation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11 (4), avril 1989, pp. 390-402.
- [11] M. Proesmans, L. Van Gool, A. Oosterlinck, *Active Acquisition of 3D Shape for Moving Objects*, IEEE International Conference on Image Processing (3), 1996, pp. 647-650.
- [12] M. Proesmans, L. Van Gool, A. Oosterlinck, *One-Shot Active 3D Shape Acquisition*, Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, 1996, pp. 336-340.
- [13] H. Delingette, *General Object Reconstruction Based on Simplex Meshes*, Int. Journal of Computer Vision, 1999, vol. 32 (2), pp. 111-146.
- [14] Z. Zhang, *A Flexible New Technique for Camera Calibration*, Microsoft Research, Technical Report MSR-TR-98-71.
- [15] M. Devy, V. Garric, J.-J. Orteu, *Camera calibration from multiple views of a 2D object, using a global non linear minimization method*, Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, 1997, pp. 1583-1589.
- [16] P. F. Sturm, S. J. Maybank, *On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 1999, pp. 432-437.
- [17] P.J. Besl, H.D. McKay, *A method for registration of 3-D shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14(2), février 1992, pp. 239-256.
- [18] O. Faugeras, G. Toscani *The calibration problem for stereo*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1986, pp. 15-20.
- [19] S. Valente, J.-L. Dugelay, *Face Tracking and Realistic Animations for Telecommunicant Clones*, IEEE Multimedia Computing and Systems Magazine, janvier-mars 2000, pp. 34-43.