Une approche par modèle déformable pour la reconstruction 3D de haute qualité d'objets photographiés

A Snake Approach for High Quality Image-based 3D Object Modeling

Carlos Hernández Esteban   Francis Schmitt

Département TSI, CNRS URA 820
Ecole Nationale Supérieure des Télécommunications, France

{carlos.hernandez,francis.schmitt}@enst.fr

Résumé

Nous présentons une nouvelle approche pour la reconstruction d'objets 3D. Elle utilise des techniques connues de la vision par ordinateur et nous permet de déterminer la géométrie et la texture d'un objet réel à partir d'une séquence d'images couleur calibrées. La méthode utilise un modèle déformable classique qui nous permet de fusionner l'information de la texture et des silhouettes pour retrouver la géométrie 3D. Nous proposons une nouvelle formulation de la contrainte des silhouettes et l'utilisation d'une approche de diffusion multirésolution du vecteur gradient pour la composante stéréo.

Mots Clefs

Reconstruction 3D, modèle déformable, diffusion multirésolution du vecteur gradient, enveloppe visuelle, silhouette, texture, multistéréo.

Abstract

In this paper we present a new approach to high quality 3D object reconstruction by using well known computer vision techniques. Starting from a calibrated sequence of color images, we are able to recover both the 3D geometry and the texture of the real object. The core of the method is based on a classical déformable model, which defines the framework where texture and silhouette information are used as external energies to recover the 3D geometry. A new formulation of the silhouette constraint is derived, and a multi-resolution gradient vector flow diffusion approach is proposed for the stereo-based energy term.

Keywords

3D reconstruction, deformable model, multigrid gradient vector flow, visual hull, texture.

1 Introduction

L'évolution très rapide des logiciels et matériels en informatique graphique nous permet de prêter davantage attention à la création ou à l'acquisition de modèles 3D de haute qualité. En particulier, un effort très important est actuellement réalisé pour la représentation numérique d'objets 3D du monde réel. L'acquisition d'objets 3D est une tâche difficile et une large bibliographie traite de ce sujet. Il existe trois approches principales au problème de la représentation d'objets 3D réels : les techniques fondées exclusivement sur le rendu par les images (*image-based rendering*), les techniques hybrides mélangeant le rendu par les images avec une reconstruction 3D partielle, et les techniques de reconstruction 3D (*3D scanning*). Les techniques fondées exclusivement sur le rendu par les images essaient de générer des vues de synthèse à partir d'un ensemble d'images originales (voir par exemple [21]). Elles n'estiment pas la vraie structure 3D à partir des images, mais elles *interpolent* directement l'ensemble des vues originales pour générer une nouvelle vue. Les méthodes hybrides [4, 20, 31, 17] estiment d'abord une approximation de la géométrie 3D puis l'utilisent avec une méthode de rendu traditionnelle afin d'améliorer la qualité des résultats. Dans ces deux types d'approches l'objectif est de générer des vues cohérentes de la scène réelle, pas d'avoir des mesures précises. En opposition a ces deux techniques, la troisième classe d'algorithmes essaie de retrouver complètement la structure 3D sous-jacente. Parmi les techniques de scanning 3D, nous pouvons distinguer deux groupes principaux : les méthodes actives et les méthodes passives. Les méthodes actives utilisent une source de lumière contrôlée tel qu'un laser ou une lumière structurée pour retrouver l'information 3D [26, 3, 16]. Les méthodes passives n'utilisent que l'information contenue dans les images de la scène [30]. Elles peuvent être classifiées par rapport au type d'information qu'elles utilisent. Une première classe est celle des méthodes de reconstruction de la forme par les silhouettes ("*shape from silhouettes*") [1, 24, 33, 23, 32]. Elles procurent une estimation initiale du modèle 3D appelée enveloppe visuelle. Elles sont rapides et robustes mais restent limitées à des objets de forme simple étant données les limitations induites par les silhouettes. Il existe des produits commerciaux basés sur cette technique. Un autre type d'approche est suivi par les méthodes de reconstruction de la forme par l'ombrage ("*shape from shading*"). Elles sont basées sur les propriétés de réflexion diffuse des surfaces Lambertiennes. Elles fonctionnent principalement pour des surfaces 2.5D (images de profondeur) et sont très dépendantes des conditions d'éclairage. Une troisième classe de méthodes utilisent l'information couleur de la scène. L'information couleur peut être utilisée différémment selon le type de scène à reconstruire. Une première utilisation est de mesurer la cohérence de la couleur entre images (photo-cohérence) pour creuser un volume numérique [28, 19]. La scène résultante est alors composée d'un ensemble de voxels, cette discrétisation spatiale rendant difficile l'extraction d'un maillage 3D de bonne qualité. Pour résoudre ce problème, les auteurs de [35] proposent d'utiliser le volume de photo-cohérence pour guider un modèle déformable. Un autre problème des algorithmes de photo-consistance est qu'ils comparent directement les valeurs de couleur, ce qui les rend très sensibles aux changements de condition d'éclairage. Une façon différente d'exploiter la couleur consiste à comparer les variations locales de la texture comme dans les méthodes de corrélation croisée [11, 25]. Il existe aussi d'autres méthodes plus spécialisées fondées sur la couleur qui utilisent en même temps d'autres informations telles que les silhouettes [19, 2, 9] ou l'albédo [7]. Même si des résultats intéressants ont pu être obtenus leur qualité reste cependant limitée, le problème principal étant la façon de fusionner efficacement les différents types de données. Certains auteurs utilisent une grille volumique pour la fusion [19, 2], d'autres utilisent un modèle déformable [7, 9]. L'algorithme que nous présentons peut être classé dans ce dernier groupe. Nous utilisons en effet un modèle déformable pour fusionner les informations liées aux silhouettes et celles liées à la texture. Les principales différences avec les méthodes citées sont dans la façon de fusionner ces informations et dans l'utilisation d'une approche multirésolution permettant d'obtenir des reconstructions de haute qualité.

2 Aperçu du système

L'objectif du système que nous proposons est une reconstruction 3D de haute qualité à partir d'une séquence d'images calibrées d'un objet réel. Pour parvenir à ce but, plusieurs types d'information contenus dans les images peuvent être exploités. Parmi ces informations, les silhouettes et la texture apparaissent les plus utiles pour la reconstruction 3D.

Le premier problème à résoudre est de décider comment mélanger les informations de silhouettes et de texture pour qu'elles coopèrent. Ceci n'est pas une tâche facile parce que ces deux informations sont de nature très différente, presque "orthogonales" de par le contenu géométrique qu'elles véhiculent.

2.1 Modèles déformables classiques et méthodes d'ensemble de niveaux

Les modèles déformables forment un des cadres les plus utilisés pour l'optimisation de surfaces sous les contraintes de différents types d'information. Deux approches existent selon la façon de poser le problème : les modèles déformables classiques [10] et les méthodes d'ensemble de niveaux [29]. Le principal avantage des modèles déformables classiques est leur facilité de mise en oeuvre. Leur plus gros désavantage est de conserver la topologie du modèle constante tout au long de son évolution. Les méthodes d'ensembles de niveaux ont la capacité intrinsèque de surmonter ce problème mais leurs principaux désavantages sont le temps de calcul et la difficulté de contrôler les changements de topologie. En général le seul moyen de contrôler la topologie d'une méthode d'ensembles de niveaux est le terme de régularisation. Ceci rend difficile la séparation entre la "douceur" d'une surface et sa topologie. Dans cet article nous avons choisi les modèles déformables classiques comme cadre de fusion des informations de silhouettes et de stéréo. Ceci implique que la topologie doit être connue avant l'évolution du modèle. Ce point est discuté en Section 3. Étant donné que la méthode proposée pour retrouver la bonne topologie est le concept d'enveloppe visuelle, l'exactitude de la topologie dépend des limitations intrinsèques de cette enveloppe. Ceci implique l'existence d'une classe d'objets pour lesquels les silhouettes sont incapables de retrouver la topologie correcte (cas par exemple d'un objet avec un trou n'apparaissant dans aucune silhouette) mais pour lesquels une méthode d'ensemble de niveaux pourrait correctement retrouver la bonne topologie grâce à la stéréo. Mais ce handicap n'est pas très grave en pratique car, à l'exception de quelques cas pathologiques, l'enveloppe visuelle donne la bonne topologie.

2.2 Approche des modèles déformables classiques

Le cadre des modèles déformables nous permet de définir une surface optimale qui minimise une certaine énergie $\mathscr{E}$. Dans notre cas, le problème de minimisation correspond à la détermination de la surface S de $\mathbb{R}^3$ qui minimise l'énergie $\mathscr{E}(S)$ définie de la manière suivante :

$$\mathscr{E}(S) = \mathscr{E}_{tex}(S) + \mathscr{E}_{sil}(S) + \mathscr{E}_{int}(S), \tag{1}$$

où $\mathscr{E}_{tex}$ est le terme lié à la texture de l'objet, $\mathscr{E}_{sil}$ celui lié aux silhouettes et $\mathscr{E}_{int}$ est le terme de régularisation. Minimiser Eq. 1 implique alors de trouver S_{opt} telle que :

$$\begin{aligned}\nabla\mathscr{E}(S_{opt}) &= \nabla\mathscr{E}_{tex}(S_{opt}) + \nabla\mathscr{E}_{sil}(S_{opt}) + \nabla\mathscr{E}_{int}(S_{opt}) = 0,\\ &= \mathscr{F}_{tex}(S_{opt}) + \mathscr{F}_{sil}(S_{opt}) + \mathscr{F}_{int}(S_{opt}) = 0,\end{aligned} \tag{2}$$

où les vecteurs gradients $\mathscr{F}_{tex}$, $\mathscr{F}_{sil}$ et $\mathscr{F}_{int}$ représentent les forces qui dirigent le modèle déformable. L'équation 2 établit la condition d'équilibre pour la surface optimale,

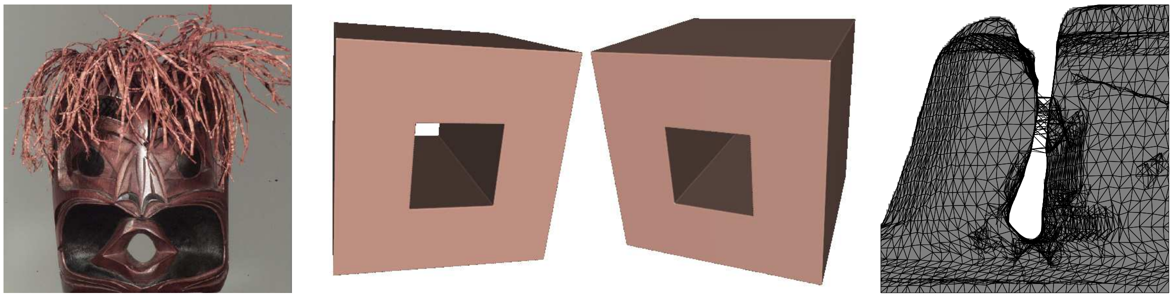

FIG. 1 – Différents problèmes topologiques. Gauche : exemple d'objet dont la topologie ne peut pas être retrouvée avec le concept d'enveloppe visuelle. Centre : exemple de problème topologique intervenant avec un nombre fini de vues. La vue de gauche est capable de retrouver la bonne topologie tandis que celle de droite ne le peut pas. Droite : topologie incorrecte résultant de la précision insuffisante de l'algorithme de reconstruction de l'enveloppe visuelle (en gris la silhouette originale ; en trait noir l'enveloppe visuelle reconstruite).

où les trois forces s'annulent mutuellement. Une solution à l'équation 2 peut être trouvée avec l'aide d'une variable temporelle t pour décrire l'évolution de la surface S, ce qui nous donne l'équation différentielle suivante :

$$S_t = \mathscr{F}_{tex}(S) + \mathscr{F}_{sil}(S) + \mathscr{F}_{int}(S). \tag{3}$$

La version discrète de cette équation devient :

$$S^{k+1} = S^k + \Delta t(\mathscr{F}_{tex}(S^k) + \mathscr{F}_{sil}(S^k) + \mathscr{F}_{int}(S^k)). \tag{4}$$

Les énergies qui dirigent le processus étant définies, il nous reste à choisir une représentation pour la surface S. Cette représentation définit comment la surface peut être déformée à chaque itération. Nous avons choisi le maillage triangulaire comme représentation du fait de sa simplicité et de ses propriétés bien connues.

Pour définir complètement le cadre de la déformation, nous avons besoin de définir une valeur initiale de S, c'est à dire une surface initiale S_0 qui va évoluer sous la contrainte des différentes énergies jusqu'à convergence.

Dans cet article nous détaillons l'initialisation du modèle dans la Section 3, la force liée à la texture de l'objet dans la Section 4, la force liée aux silhouettes dans la Section 5, et le contrôle de l'évolution du maillage dans la Section 6. Dans la Section 7 nous présentons les résultats expérimentaux.

3 Initialisation du modèle déformable

La première étape de notre problème de minimisation est de trouver une surface initiale suffisamment *proche* de la surface optimale afin de garantir une bonne convergence de l'algorithme, *proche* devant être considéré au sens géométrique et topologique. La distance géométrique entre les surfaces initiale et optimale doit être faible afin de limiter le nombre d'itérations du processus d'évolution et réduire ainsi le temps de calcul. La topologie de la surface initiale est aussi très importante car les modèles déformables classiques gardent la topologie pendant toute l'évolution de la surface. Une bonne initialisation, comprise entre l'enveloppe convexe de l'objet et sa vraie surface, est l'enveloppe visuelle [13]. L'enveloppe visuelle peut être définie comme l'intersection de tous les cônes possibles qui contiennent l'objet. Elle peut représenter des surfaces avec n'importe quel nombre de trous. Par contre, ceci n'implique pas qu'elle puisse retrouver n'importe quelle topologie et, ce qui est pire, la topologie de l'enveloppe visuelle dépend du nombre de vues (voir Fig. 1 centre).

Calculer l'enveloppe visuelle d'une séquence d'images est un problème bien connu en vision par ordinateur [24, 23, 20]. Il existe différentes approches selon le type de sortie, le type de représentation et la fidélité à l'enveloppe visuelle théorique. Dans notre cas, nous nous intéressons aux méthodes qui produisent des maillages de bonne qualité (propriétés Euleriennes, lissé, avec un bon critère d'aspect des triangles), même si la fidélité géométrique n'est pas très grande. Les méthodes de sculpture d'un volume 3D sont un bon choix car elles fournissent des maillages de bonne qualité après une étape de maillage comme les algorithmes de type "marching cube" [18] ou "marching tetrahedron". Le degré de précision est fixé par la résolution de la grille volumique qui peut être adaptée pour une résolution donnée en sortie. Par contre, cette adaptabilité peut aussi générer de nouveaux problèmes de topologie : si la résolution de la grille est petite comparée à la taille des structures de l'enveloppe visuelle, les défauts d'échantillonnage ou *aliasing* provoqués par le sous-échantillonnage peuvent engendrer des artefacts topologiques que l'enveloppe visuelle théorique n'a pas. En conclusion, il peut y avoir jusqu'à trois types différents de déviation entre la topologie de l'objet réel et celle de l'enveloppe visuelle calculée :

- Erreurs dues à la nature de l'enveloppe visuelle (voir Fig. 1 gauche). Des objets réels peuvent avoir des trous ou tunnels qu'aucun point de vue ne permet de voir à travers. L'enveloppe visuelle ne peut alors pas représenter la bonne topologie de ce type d'objets.
- Erreurs dues à l'utilisation d'un nombre fini de vues (voir Fig. 1 centre). Elles peuvent être évitées en ayant

les points de vue qui permettent de retrouver la bonne topologie.
– Erreurs dues à la mise en oeuvre de l'algorithme (voir Fig. 1 droite). Elles sont provoquées par une précision numérique insuffisante de l'algorithme ou par un sous-échantillonnage des silhouettes. Elles peuvent être évitées en augmentant la précision numérique ou par un filtrage des silhouettes.

En pratique, nous utilisons une méthode de sculpture d'un volume *octree* suivie d'une étape de maillage par *marching tetrahedron* et d'une étape de simplification du maillage. Pour initialiser l'*octree*, une boîte englobante initiale peut être calculée analytiquement à partir des boîtes englobantes 2D des silhouettes. La rétro-projection de chaque boîte englobante 2D donne 4 demi-plans 3D. L'intersection de tous les demi-plans ainsi créés définit une enveloppe convexe dont la boîte englobante peut être calculée par une optimisation simplex de chacune des six variables qui définissent cette boîte.

4 Force définie par la texture

Dans cette section nous définissons la force due à la texture $\mathscr{F}_{tex}$ qui apparaît dans l'Eq. 2 et qui permet de retrouver la géométrie 3D pendant l'évolution du modèle déformable. Nous voulons que cette force maximise la cohérence des images qui voient la même partie d'un objet. Il existe plusieurs approches pour mesurer la cohérence d'un ensemble d'images, mais elles peuvent être classées en deux groupes selon qu'elles font une comparaison radiométrique ponctuelle (par exemple avec une mesure de photo-cohérence comme dans le cas du *voxel coloring* [28]) ou une comparaison spatiale de la distribution radiométrique relative (par exemple avec des mesures de corrélation croisée). Nous avons choisi la corrélation centrée réduite à cause de sa simplicité et de sa robustesse en présence de reflets et de changement des conditions d'éclairage.

Le critère de cohérence étant défini, nous pouvons retrouver la géométrie 3D en maximisant le critère pour un ensemble de vues. Deux types différents d'approches ont été proposés dans la littérature. Dans le premier type, la mesure de cohérence est utilisée pour évaluer le modèle courant. Si la mesure est améliorée par une déformation locale du modèle, le modèle est actualisé et le processus itéré [7, 6]. Dans le cas des ensembles de niveaux [11, 25], une bande volumique est explorée autour du modèle courant. Dans ce premier type d'approches, l'exploration reste locale et dépendante du modèle courant. Le fait de ne pas explorer toutes les configurations possibles fait que l'algorithme peut échouer à cause des maxima locaux du critère de cohérence de texture. Le second type d'approches consiste à tester toutes les configurations possibles. Ceci permet de prendre une décision plus robuste. Afin d'améliorer encore plus la robustesse, nous pouvons cumuler les valeurs du critère dans une grille 3D par une méthode de décision majoritaire [19, 22]. Nous avons choisi ce type d'approche pour sa robustesse en présence de reflets. Elle nous permet aussi de passer de l'information distribuée dans les images à une information plus utilisable du type “probabilité de trouver une surface” au voisinage d'un point 3D de l'espace.

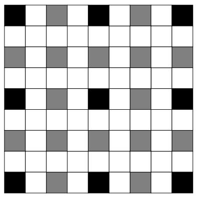

FIG. 2 – Exemple de la partition d'une image en 3 niveaux différents de résolution.

4.1 Méthode proposée de décision majoritaire

Considérons notre problème de reconstruction de la géométrie 3D à partir de la texture : nous voulons optimiser, pour un pixel donné dans une image, la cohérence de la texture au voisinage de ce point avec les autres images. Un rayon optique est défini par ce pixel. Nous cherchons le point 3D qui appartient au rayon optique et qui maximise la corrélation centrée réduite avec les autres images. Ceci peut être réalisé de façon efficace par l'échantillonnage de la projection du rayon optique dans chacune des images. En pratique, la connaissance de l'enveloppe visuelle, qui contient toujours la vraie surface, nous permet d'accélérer les calculs. L'algorithme de corrélation implanté est décrit dans [8].

Le problème de cet algorithme est son temps de calcul : pour 36 images de 2000x3000 pixels, il peut atteindre 16 heures sur une machine puissante. Ce temps peut être réduit, pratiquement sans perte, en considérant la redondance du calcul. Afin d'être capable de bénéficier des corrélations déjà calculées, les images peuvent être partitionnées en plusieurs niveaux de résolution comme indiquée en Fig.2. L'algorithme original est d'abord exécuté sur le niveau de résolution le plus bas (pixels noirs dans Fig. 2), avec l'intervalle de profondeur défini par l'enveloppe visuelle. Pour les niveaux suivants, l'intervalle de profondeur est estimé à partir des résultats du niveau précédant, un registre des résultats de corrélation étant maintenu afin de contrôler la fiabilité de l'estimation. Le gain théorique maximum que nous pouvons obtenir avec cette approche par rapport à la méthode de base est d'un facteur 16 dans le cas de 3 niveaux de résolution comme indiqué en Fig. 2. En pratique, l'amélioration est de 5 à 6 fois plus rapide pour des images bien texturées. Le cas le pire correspond à des images peu texturées où les corrélations ne sont pas du tout fiables. L'estimation de l'intervalle de profondeur échoue alors et toutes les corrélations sont calculées avec l'intervalle de profondeur maximum défini par l'enveloppe visuelle.

Une structure *octree* est également utilisée pour stocker efficacement les résultats de corrélations. Le volume *octree*

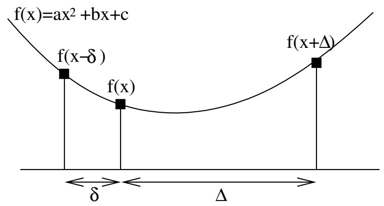

FIG. 3 – Courbe parabolique passant par 3 points.

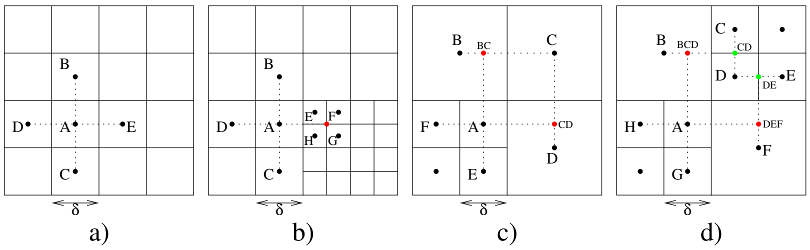

FIG. 4 – Interpolations entre valeurs (exemple 2D).

permet de stocker et cumuler les maximums de corrélations estimées en chacun des pixels. Ce volume en soi même ne peut pas être utilisé comme force pour le modèle déformable. Une possibilité serait d'utiliser le gradient du volume de corrélations comme force. Mais le problème est que cette force n'est définie que très localement au voisinage de la surface de l'objet. La solution proposée est d'utiliser un champ de vecteurs gradient (*gradient vector flow* ou GVF) pour diriger le modèle déformable.

4.2 Champ de vecteurs gradient sur grille multirésolution

Le champ de vecteurs gradient a été introduit par Xu et Prince [34] pour surmonter le problème difficile du rayon d'action limité des forces externes généralement utilisées. Ce problème est dû à une définition locale de la force et à l'absence d'un mécanisme de propagation de cette information. Pour éliminer ce problème, et pour toutes les forces dérivées du gradient d'un champ scalaire, ils proposent de générer un champ de vecteurs qui propage l'information du gradient. Le GVF d'un champ scalaire f est défini comme le champ de vecteurs $\mathbf{v} = [u, v, w]$ qui minimise la fonctionnelle $\mathscr{E}$ suivante :

$$\mathscr{E} = \int \mu ||\nabla \mathbf{v}||^2 + ||\mathbf{v} - \nabla f||^2 ||\nabla f||^2, \tag{5}$$

où μ est le poids du terme de régularisation, $\nabla \mathbf{v} = [\nabla u, , \nabla v, \nabla w]$, ∇ désignant l'opérateur gradient : $\nabla f = [f_x, f_y, f_z]$. La solution de ce problème de minimisation doit satisfaire l'équation d'Euler :

$$\mu \nabla^2 \mathbf{v} - (\mathbf{v} - \nabla f) ||\nabla f||^2 = 0, \tag{6}$$

où $\nabla^2 \mathbf{v} = [\nabla^2 u, , \nabla^2 v, \nabla^2 w]$, ∇^2 désignant l'opérateur Laplacian : $\nabla^2 f = f_{xx} + f_{yy} + f_{zz}$. Une solution numérique peut être trouvée avec l'introduction d'une variable temporelle t et la solution de l'équation différentielle suivante :

$$\mathbf{v}_t = \mu \nabla^2 \mathbf{v} - (\mathbf{v} - \nabla f) ||\nabla f||^2. \tag{7}$$

Le GVF peut être vu comme le gradient original lissé par l'action de l'opérateur Laplacien. Ce lissage nous permet en même temps d'éliminer les grandes variations du gradient et de le propager. Le degré de lissage/propagation est contrôlé par μ. Si μ est zéro, le GVF sera le gradient d'origine, si μ est très grand, le GVF sera un champ de vecteurs constants dont les composantes sont la moyenne des composantes du gradient.

Étant donné que nos données sont stockées dans une structure *octree*, le GVF doit être calculé sur une grille multirésolution. Pour ceci, nous avons besoin de définir sur cette grille :

– l'opérateur Laplacien et l'opérateur gradient,
– l'interpolation entre voxels de tailles différentes.

Dans le cas d'une grille régulière avec un espacement de $[\Delta x, \Delta y, \Delta z]$, les dérivées première et seconde f_x et f_{xx} peuvent être approximées par des différences finies centrées :

$$\begin{aligned} f_x &\approx \frac{f(x+\Delta x,y,z) - f(x-\Delta x,y,z)}{2\Delta x}, \\ f_{xx} &\approx \frac{f(x+\Delta x,y,z) - 2f(x,y,z) + f(x-\Delta x,y,z)}{\Delta x^2}. \end{aligned} \tag{8}$$

Si la grille n'est pas régulière, les différences finies ne sont pas centrées. Une façon aisée de trouver les formules équivalentes pour une grille à pas variable est d'estimer la courbe parabolique $ax^2 + bx + c$ qui passe par trois points (Fig. 3), et de calculer les dérivées de la courbe estimée [5]. Après résolution du système d'équations, nous trouvons :

$$\begin{aligned} f_x &\approx \frac{1}{(\delta+\Delta)} \left(\frac{f(x+\Delta) - f(x)}{\Delta/\delta} - \frac{f(x-\delta) - f(x)}{\delta/\Delta} \right) = 2ax + b, \\ f_{xx} &\approx \frac{2}{(\delta+\Delta)} \left(\frac{f(x+\Delta) - f(x)}{\Delta} + \frac{f(x-\delta) - f(x)}{\delta} \right) = 2a. \end{aligned} \tag{9}$$

Afin de simplifier les calculs d'interpolation, nous devons rajouter une contrainte à la topologie de la grille multirésolution : la différence de résolution dans une boule de rayon 1 en norme L1 centrée en un voxel ne doit pas être plus grande qu'un niveau. Ce n'est pas une forte contrainte car la résolution de l'*octree* doit changer doucement pour obtenir de bons résultats numériques dans le calcul du GVF.

Il existe trois configurations différentes dans l'algorithme multirésolution. La première s'obtient lorsque le voxel courant et tous ses voisins ont la même taille (voir Fig. 4.a). Dans ce cas, les calculs sont effectués comme ceux dans une grille mono-résolution. La seconde configuration correspond au cas où le voxel courant est plus grand ou égal à ses voisins (voir Fig. 4.b). Pour les voxels de même taille, les calculs sont faits de façon habituelle. Pour ceux qui sont plus petits, une moyenne est effectuée pour obtenir

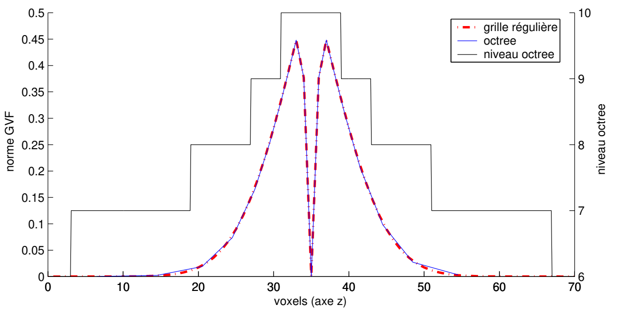

FIG. 5 – Comparaison du calcul du GVF dans une grille régulière et dans une grille *octree*.

la valeur correcte à l'échelle du voxel courant :

$$f_x(A) \approx \frac{f(EFGH) - f(D)}{2\delta}, \ f_y(A) \approx \frac{f(B) - f(C)}{2\delta}. \quad (10)$$

La troisième configuration correspond au cas où le voxel courant est plus petit ou égal à ses voisins (voir Fig. 4.c et 4.d). Nous pouvons alors distinguer deux sous-types de configuration, dans lesquels nous voulons calculer le gradient au point A. Dans le cas 4.c nous avons besoin de la valeur de la fonction f aux points E, F, BC et CD :

$$\begin{aligned} f_x(A) &\approx \frac{1}{(\delta+1.5\delta)}\left(\frac{f(CD)-f(A)}{1.5} - \frac{f(F)-f(A)}{1/1.5}\right), \\ f_y(A) &\approx \frac{1}{(\delta+1.5\delta)}\left(\frac{f(BC)-f(A)}{1.5} - \frac{f(E)-f(A)}{1/1.5}\right). \end{aligned} \quad (11)$$

Dans le cas 4.d les valeurs BCD et DEF sont obtenues par interpolation de B avec CD, et de DE avec F, respectivement. Si nous transposons ces exemples en 3D, nous avons une interpolation supplémentaire autour de la troisième dimension pour les points BC et CD dans 4.c, et BCD et DEF dans 4.d.
Dans la figure 5 nous comparons le résultat du calcul en 3D du GVF pour $\mu = 0.1$ en utilisant une grille régulière et une grille *octree*. Le champ scalaire f utilisé dans cet exemple est défini de la manière suivante :

$$f(x,y,z) = \begin{cases} 1 & \text{pour } z \in [34,36] \\ 0 & \textit{sinon} \end{cases} . \quad (12)$$

La précision du calcul en multirésolution est très comparable à celle du calcul en mono-résolution. Nous pouvons seulement distinguer une très faible différence entre les deux courbes là où la résolution est très réduite (au niveau des voxels 20 et 50). Le temps de calcul avec 10 niveaux de résolution est en valeur moyenne 2 à 3 fois plus rapide qu'avec la version mono-résolution, et l'espace de stockage mémoire nécessaire est réduit entre 10 et 15 fois.

5 Force définie par les silhouettes

La force définie par les silhouettes a pour rôle de forcer le modèle déformable à respecter les silhouettes de la séquence d'origine. Si elle est la seule force dans l'évolution, le modèle devrait converger vers l'enveloppe visuelle. Mais étant donné que nous cherchons à respecter les silhouettes en présence d'autres forces, la force des silhouettes doit tenir compte de l'auto occultation du modèle. Si une partie du modèle vérifie déjà une silhouette particulière, le reste du modèle ne doit plus être contraint par cette silhouette puisqu'elle est déjà vérifiée. Si nous comparons l'enveloppe visuelle et l'objet réel, nous constatons que l'objet réel vérifie totalement les silhouettes, mais que tous les points de l'objet ne contribuent pas à cette vérification. C'est le cas en particulier des concavités de l'objet qui sont occultées par une partie de l'objet qui les vérifient. Le principal problème est de faire la distinction entre les points qui doivent vérifier les silhouettes et ceux qui ne le doivent pas. Ceci est équivalent à trouver les générateurs de contours de l'objet. La force des silhouettes peut en fait être décomposée en deux composantes distinctes : une composante qui mesure le respect des silhouettes, et une composante qui mesure le degré d'application de la force pour un point donné du modèle déformable. La première composante est définie comme la distance *signée* à l'enveloppe visuelle. Pour un point $\mathbf{P}_M$ du maillage 3D déformable cette composante peut être calculée comme la plus petite distance euclidienne signée d_{VH} entre les contours des silhouettes S_i et la projection du point sur ces silhouettes :

$$d_{VH}(\mathbf{P}_M) = \min_i d(S_i, \mathscr{P}_i \mathbf{P}_M), \quad (13)$$

où $\mathscr{P}_i$ représente la matrice de projection dans la silhouette i. Une distance positive implique que la projection est à l'intérieur de la silhouette et une distance négative qu'elle est à l'extérieur de la silhouette. En n'utilisant que cette composante, le modèle déformable converge vers l'enveloppe visuelle.
La seconde composante mesure le degré d'occultation α d'un point du maillage $\mathbf{P}_M$ pour un point de vue donné. Le point de vue choisi c correspond au point de vue qui définit la distance minimale à l'enveloppe visuelle :

$$\begin{aligned} \alpha(\mathbf{P}_M) &= \begin{cases} 1 & \text{pour } d_{VH}(\mathbf{P}_M) \leq 0 \\ \frac{1}{(1+d(S_c snake, \mathscr{P}_c \mathbf{P}_M))^n} & \text{pour } d_{VH}(\mathbf{P}_M) > 0 \end{cases}, \\ c(\mathbf{P}_M) &= arg\min_i d(S_i, \mathscr{P}_i \mathbf{P}_M). \end{aligned} \quad (14)$$

Dans la définition de α, nous avons deux cas. Si d_{VH} est négatif, le point est à l'extérieur de l'enveloppe visuelle. Dans ce cas, la force est toujours maximale. Pour un point à l'intérieur de l'enveloppe visuelle, c correspond à la vue qui définit la distance minimale du point à l'enveloppe visuelle. $S_c snake$ est la silhouette créée par la projection du modèle déformable dans la vue c. La puissance n contrôle la vitesse de décroissance de α. Cette fonction donne une force des silhouettes maximale aux points qui constituent les contours générateurs du modèle déformable. Les autres points sont considérés comme des concavités et sont pondérés inversement à leur distance à la silhouette du modèle déformable. Ceci permet aux points du modèle de

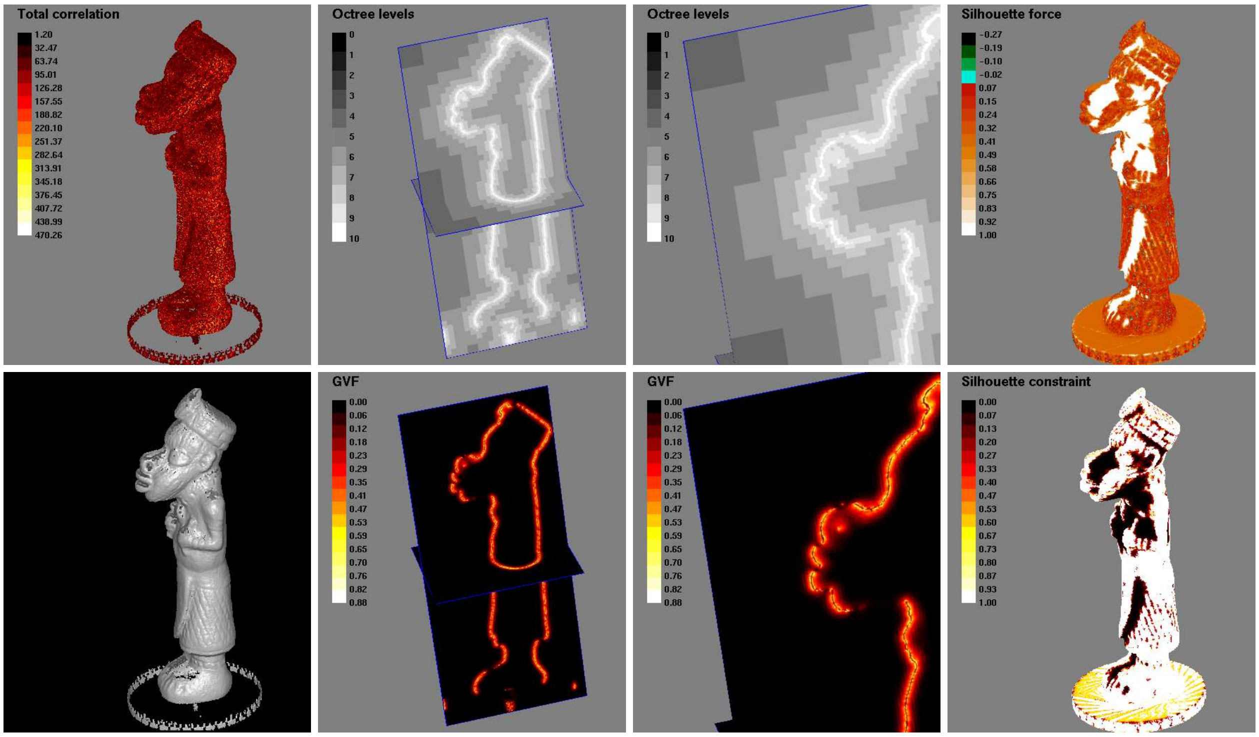

FIG. 6 – Forces externes utilisées dans la reconstruction du modèle BigHead. Haut gauche : volume des corrélations stéréo. Bas gauche : visualisation du même volume avec une estimation de la normale à la surface. Haut centre : partition *octree* utilisé dans le calcul du champ GVF (détail à droite). Bas centre : norme du champ GVF (détail à droite). Haut droite : composante d_{VH} de la force des silhouettes après convergence. Bas droite : composante α de la force des silhouettes après convergence.

se *détacher* de l'enveloppe visuelle. Une grande valeur de n permet un détachement plus facile. Mais si cette valeur est trop forte, la force devient trop locale et ne permet pas des transitions douces entre les contours et les concavités. En pratique $n = 2$ donne un bon compromis entre transitions douces et l'excavation des concavités.

La force des silhouettes pour un point $\mathbf{P}_M$ du modèle est un vecteur dont la direction est celle de la normale $\mathbf{N}_M$ à la surface du modèle et dont la norme est le produit des deux composantes :

$$\mathscr{F}_{sil}(\mathbf{P}_M) = \alpha(\mathbf{P}_M) d_{VH}(\mathbf{P}_M) \mathbf{N}_M(\mathbf{P}_M) \qquad (15)$$

6 Contrôle du maillage

Les forces externes du modèle déformable ($\mathscr{F}_{tex}$ et $\mathscr{F}_{sil}$) étant définies, il ne nous reste qu'à déterminer la force de régularisation interne $\mathscr{F}_{int}$. Nous la définissons comme la force qui attire un point donné $\mathbf{v}$ du maillage vers le centre de gravité de son 1-voisinage :

$$\mathscr{F}_{int}(\mathbf{v}) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{v}_i - \mathbf{v}, \qquad (16)$$

où $\mathbf{v}_i$ est le i^{eme} voisin de $\mathbf{v}$ et m est le nombre de ces voisins. Si la force interne est la seule force appliquée dans l'évolution du modèle déformable, le maillage s'effondrera sous l'action du filtrage barycentrique.

Etant donné que la force de texture $\mathscr{F}_{tex}$ peut parfois devenir tangente à surface du maillage au cours de l'évolution du modèle nous n'utilisons pas cette force directement mais sa projection $\mathscr{F}_{tex}^N$ sur la normale à la surface :

$$\mathscr{F}_{tex}^N(\mathbf{v}) = (\mathscr{F}_{tex}(\mathbf{v}) \cdot \mathbf{N}(\mathbf{v}))\mathbf{N}(\mathbf{v}). \qquad (17)$$

Ceci permet d'éliminer des problèmes de cohérence qui peuvent intervenir entre les forces des points voisins et d'aider la force interne à conserver une forme régulière au maillage.

Le processus d'évolution du modèle déformable à la $k^{ème}$ itération (Eq. 4) peut être décrit comme l'évolution de tous les points du maillage $\mathbf{v}_i$:

$$\mathbf{v}_i^{k+1} = \mathbf{v}_i^k + \Delta t(\mathscr{F}_{tex}^N(\mathbf{v}_i^k) + \beta \mathscr{F}_{sil}(\mathbf{v}_i^k) + \gamma \mathscr{F}_{int}(\mathbf{v}_i^k)) \quad (18)$$

où Δt est le pas temporel, et β et γ sont les poids relatifs de la force des silhouettes et de la force interne de régularisation par rapport à la force de texture. L'équation 18 est itérée jusqu'à équilibre de tous les points du maillage. Le pas temporel Δt doit être choisi selon un compromis entre stabilité et temps de calcul. Une étape additionnelle de remaillage du modèle déformable est effectuée à la fin de chaque itération afin de préserver des distances minimum et maximum entre points voisins. Elle est réalisée avec une décimation et un raffinement contrôlé

FIG. 7 – Modèles Oceania et BigHead après convergence (45843 et 114496 points respectivement). Gauche : visualisation de la surface 3D après convergence. Droite : même vues avec placage de texture.

du maillage. La décimation met en oeuvre l'opérateur de suppression d'arête et le raffinement celui de subdivision en $\sqrt{3}$ des triangles [12].

7 Résultats

Dans cette section nous présentons quelques résultats obtenus avec la méthode proposée. Ils intègrent aussi un placage de texture obtenu avec une méthode similaire à [27, 15]. Nous avons amélioré la qualité de la texture avec un filtrage des reflets. Ceci est rendu possible grâce à l'existence de plusieurs images dans lesquelles un même triangle est visible.

Toutes les reconstructions présentées dans cet article ont été obtenues, par objet, à partir d'une seule séquence en rotation de 36 images, chaque image ayant 2008x3040 pixels. La séquence est géométriquement calibrée par la méthode exposée en [14]. Les valeurs de β et γ sont les mêmes pour toutes les reconstructions : $\beta = 0.2$, $\gamma = 0.15$. La valeur de β ne dépend que du rapport entre la taille des images et la taille de l'*octree* car l'itération est toujours faite dans le domaine de l'*octree* GVF. Des valeurs typiques de γ sont comprises entre 0.1 et 0.25, selon le lissage requis.

Les temps de calcul sont dominés par l'étape de corrélation

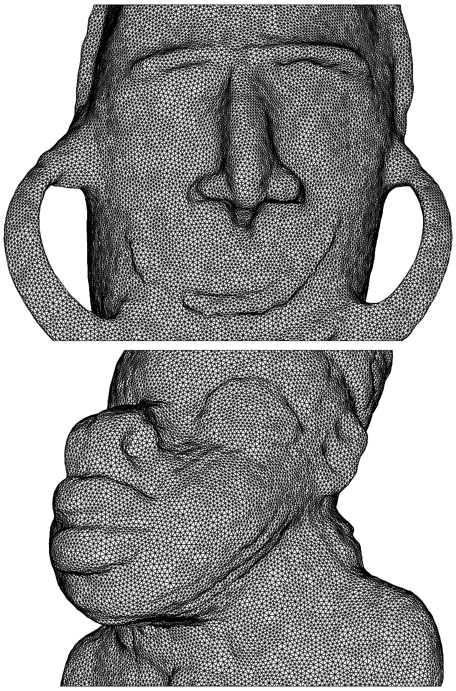

FIG. 8 – Détail du maillage des modèles Oceania et BigHead.

stéréo : une valeur moyenne du temps de calcul pour 36 images de 6 M pixels est de 3 heures sur une machine avec un Pentium 4 à 1.4 GHz.

La Fig. 6 nous présentons les différentes forces utilisées dans le modèle déformable. À gauche nous montrons une visualisation du volume de corrélations. Nous pouvons y voir que le socle n'a corrélé qu'autour des marqueurs d'angles qui fournissent suffisamment de détail pour l'algorithme de corrélation. Nous avons utilisé 10 niveaux d'*octree* (haut centre), ce qui permet une très bonne précision dans le calcul du GVF (bas centre). À la fin du processus d'itération, toutes les concavités sont automatiquement détectées et correspondent aux zones sombres (bas droite). Nous constatons qu'en l'absence de corrélation, et donc de force de texture sur le socle, celui-ci est entièrement reconstruit grâce à la force des silhouettes (en fait, étant donné que le modèle initial était l'enveloppe visuelle, le socle n'a pas évolué).

Plusieurs exemples de reconstruction sont montrés en Fig. 7, Fig. 8 et Fig. 9. Nous pouvons apprécier en Fig. 8 la haute qualité des modèles reconstruits.

8 Conclusions

Nous avons présenté une nouvelle approche pour la reconstruction d'objets 3D fondée sur la fusion des informations de texture et de silhouettes. Nos deux principales contributions sont 1) la définition de la force des silhouettes et son intégration dans le cadre d'un modèle déformable et 2) une approche système complète et robuste de la reconstruction 3D d'un objet à partir d'une séquence d'images dans laquelle plusieurs techniques existantes sont utilisées

FIG. 9 – Quelques exemples de reconstruction. Gauche : initialisation du modèle déformable avec l'enveloppe visuelle. Centre : Visualisation Gouraud des modèles reconstruits (haut : 159534 points ; bas : 60189 points). Droite : modèles texturés.

et améliorées pour obtenir des résultats de haute qualité.

Les deux limitations les plus importantes de cette approche sont aussi ses deux sources de robustes : le volume de corrélations multi-stéréo par décision majoritaire et l'utilisation d'un modèle déformable avec une topologie constante. Le cumul des corrélations permet d'avoir de bonnes reconstructions en présence de reflets, mais le fait qu'il soit effectué dans un volume discret limite la résolution maximum du modèle 3D. Cette limitation pourrait être surmontée en introduisant le modèle final dans une autre évolution où l'énergie liée à la texture prendrait en compte la surface courante du modèle 3D (corrélation fondée sur le plan tangent ou l'approximation par une quadrique). Étant donné que le modèle est déjà très près de la solution, nous n'aurions besoin que de très peu d'itérations. La seconde limitation est celle de la topologie constante. Elle permet de garantir la topologie du modèle final mais est aussi une limitation pour les objets dont la topologie ne peut pas être captée par l'enveloppe visuelle. Une solution possible consisterait à détecter les auto collisions du maillage et de démarrer une étape fondée sur les ensembles de niveaux pour retrouver la bonne topologie. D'autres améliorations pourraient comprendre : i) l'auto calibrage de la séquence d'images à partir des silhouettes ou d'autres approches classiques, ii) une amélioration de la stratégie de convergence afin d'accélérer l'évolution du modèle dans les régions de l'espace vides de corrélation, iii) l'utilisation de la courbure pour permettre une évolution multirésolution du maillage, iv) des développements supplémentaires au niveau de la génération et du placage de la texture.

Remerciements

Ces travaux ont été partiellement supportés par le projet Européen SCULPTEUR IST-2001-35372. Nous remercions le musée Thomas Henry à Cherbourg pour les séquence d'images correspondantes aux Fig.7 et 9.

URL : www.tsi.enst.fr/3dmodels/

Références

[1] B. G. Baumgart. *Geometric Modelling for Computer Vision*. PhD thesis, Standford University, 1974.

[2] G. Cross and A. Zisserman. Surface reconstruction from multiple views using apparent contours and surface texture. In *NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Ljubljana, Slovenia*, pages 25–47, 2000.

[3] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH '96*, pages 303–312, 1996.

[4] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs : A hybrid geometry and image-based approach. In *SIGGRAPH '96*, pages 11–20, 1996.

[5] B. Fornberg. Generation of finite difference formulas on arbitraily spaced grids. *Mathematics of Computation*, 51 :699–706, 1988.

[6] P. Fua. From multiple stereo views to multiple 3d surfaces. *International Journal of Computer Vision*, 24 :19–35, 1997.

[7] P. Fua and Y.G. Leclerc. Object-centered surface reconstruction : Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16 :35–56, September 1995.

[8] C. Hernández and F. Schmitt. Multi-stereo 3d object reconstruction. In *3DPVT '02*, pages 159–166, 2002.

[9] J. Isidoro and S. Sclaroff. Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *Proc. ICCV*, pages 1335 –1342, 2003.

[10] M. Kass, A. Witkin, and D. Terzopoulos. Snakes : Active contour models. *International Journal of Computer Vision*, 1 :321–332, 1988.

[11] R. Keriven and O. Faugeras. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3) :336–344, 1998.

[12] L. Kobbelt. $\sqrt{3}$-subdivision. In *SIGGRAPH 2000*, pages 103–112, 2000.

[13] A. Laurentini. The visual hull concept for silhouette based image understanding. *IEEE Trans. on PAMI*, 16(2) :150–162, 1994.

[14] J. M. Lavest, M. Viala, and M. Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration ? In *Proc. ECCV*, volume 1, pages 158–174, 1998. Germany.

[15] H. Lensch, W. Heidrich, and H. P. Seidel. A silhouette-based algorithm for texture registration and stitching. *Journal of Graphical Models*, pages 245–262, July 2001.

[16] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project : 3d scanning of large statues. In *SIGGRAPH 2000*, pages 131–144, 2000.

[17] M. Li, H. Schirmacher, M. Magnor, and H. Seidel. Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes. In *Proceedings of IEEE 2002 Workshop on Multimedia and Signal Processing*, pages 9–12, 2002.

[18] W. E. Lorensen and H.E. Cline. Marching cubes : A high resolution 3d surface construction algorithm. In *Proceedings of SIGGRAPH '87*, volume 21, pages 163–169, 1987.

[19] Y. Matsumoto, K. Fujimura, and T. Kitamura. Shape-from-silhouette/stereo and its application to 3-d digitizer. In *Proceedings of Discrete Geometry for Computing Imagery*, pages 177–190, 1999.

[20] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. *SIGGRAPH 2000*, pages 369–374, July 2000.

[21] L. McMillan and G. Bishop. Plenoptic modeling : An image-based rendering system. In *SIGGRAPH '95*, pages 39–46, 1995.

[22] G. Medioni, M. Lee, and C. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.

[23] W. Niem and J. Wingbermuhle. Automatic reconstruction of 3d objects using a mobile monoscopic camera. In *Int. Conf. on Recent Advances in 3D Imaging and Modeling*, pages 173–181, 1997.

[24] M. Potmesil. Generating octree models of 3d objects from their silhouettes in a sequence of images. *CVGIP*, 40 :1–29, 1987.

[25] A. Sarti and S. Tubaro. Image based multiresolution implicit object modeling. *EURASIP Journal on Applied Signal Processing*, 2002(10) :1053–1066, 2002.

[26] F. Schmitt, B. Barsky, and W. Du. An adaptative subdivision method for surface-fitting from sampled data. In *SIGGRAPH '86*, pages 179–188, 1986.

[27] F. Schmitt and Y. Yemez. 3d color object reconstruction from 2d image sequences. In *IEEE Interational Conference on Image Processing*, volume 3, pages 65–69, 1999.

[28] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 38(3) :197–216, 2000.

[29] J. A. Sethian. *Level Set Methods : Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences*. Cambridge University Press, 1996.

[30] G. Slabaugh, W. B. Culbertson, T. Malzbender, and R. Shafer. A survey of methods for volumetric scene reconstruction from photographs. In *International Workshop on Volume Graphics 2001*, 2001.

[31] G. Slabaugh, R. Schafer, and M. Hans. Image-based photo hulls. In *3DPVT '02*, pages 704–862, 2002.

[32] S. Sullivan and J. Ponce. Automatic model construction, pose estimation, and object recognition from photographs using triangular splines. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(10) :1091–1096, 1998.

[33] R. Vaillant and O. Faugeras. Using extremal boundaries for 3d object modelling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2) :157–173, 1992.

[34] C. Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, pages 359–369, 1998.

[35] A. Yezzi, G. Slabaugh, R. Cipolla, and R. Schafer. A surface evolution approach of probabilistic space carving. In *3DPVT '02*, pages 618–621, 2002.