HAL open science

Stéréo multi-vues : erreur de reprojection et maillages triangulaires

Amael Delaunoy, Emmanuel Prados, Pau Gargallo, Jean-Philippe Pons, Peter Sturm

To cite this version:

Amael Delaunoy, Emmanuel Prados, Pau Gargallo, Jean-Philippe Pons, Peter Sturm. Stéréo multi-vues : erreur de reprojection et maillages triangulaires. ORASIS 2009 - Congrès des jeunes chercheurs en vision par ordinateur, Jun 2009, Trégastel, France. pp.1-8. inria-00404620

HAL Id: inria-00404620

https://inria.hal.science/inria-00404620v1

Submitted on 16 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stéréo multi-vues : erreur de reprojection et maillages triangulaires

Multi-view Stereo : Reprojection Error and Triangular Meshes

Amaël Delaunoy[1*] Emmanuel Prados[1*] Pau Gargallo[1] Jean-Philippe Pons[2*] Peter Sturm[1]

[1]INRIA Rhône-Alpes / LJK
[2]CERTIS, École des Ponts

655 Avenue de l'Europe, 38330 Montbonnot St Martin, France
Prenom.Nom@inrialpes.fr

Résumé

Cet article propose une méthode variationnelle basée sur les maillages pour la reconstruction 3D de scènes (forme et radiance) à partir de plusieurs images. Notre méthode est basée sur les modèles génératifs et minimise l'erreur de reprojection (différence entre une image observée et une image obtenue à partir de la reconstruction) par une descente de gradient. Pour la première fois, nous calculons le gradient de l'erreur de reprojection pour des surfaces non lisses représentées de manière discrète par des maillages triangulés. Le gradient prend correctement en compte les changements de visibilité qui apparaissent lorsque la surface bouge durant l'évolution ; cela force les contours occultants générés par la surface à correspondre parfaitement aux contours apparents dans les images observées. Notre méthode est capable de retrouver la forme et la radiance de diverses scènes.

Mots Clef

Méthode variationnelle, reconstruction 3D, surface, descente de gradient, maillage triangulaire, stéréo multi-vues.

Abstract

This article proposes a variational multi-view stereo vision method based on meshes for recovering 3D scenes (shape and radiance) from images. Our method is based on generative models and minimizes the reprojection error (difference between the observed images and the images synthesized from the reconstruction). For the first time, we rigorously compute the gradient of the reprojection error for non smooth surfaces defined by discrete triangular meshes. The gradient correctly takes into account the visibility changes that occur when a surface moves ; this forces the contours generated by the reconstructed surface to perfectly match with the apparent contours in the input images. Our method is thus able to recover the shape and the diffuse radiance of various scenes.

Keywords

3D reconstruction, variational methods, gradient descent, surfaces, triangular meshes, multi-view stereo.

1 Introduction

La stéréo multi-vues est le problème de reconstruire une scène 3D à partir de plusieurs caméras. On souhaite reconstruire la forme $\mathcal{S}$ et éventuellement l'apparence d'un objet, étant donné une séquence d'images prises à partir de différents points de vue. Ce problème devient alors le problème inverse de la synthèse d'image, et peut donc être modélisé dans un cadre Bayesien en minimisant la différence entre les images du modèle reconstruit et celles observées, i.e. l'*erreur de reprojection*. Le problème variationnel correspondant à cette vision Bayesienne nous amène à minimiser une fonctionnelle d'énergie définie sur l'ensemble des images [6, 10, 12, 16, 20] et tient compte de la visibilité des points de la surface. Modéliser ce problème de visibilité est l'une des difficultés majeures du problème de reconstruction stéréo multi-vues. Les travaux antérieurs prennent en compte cette difficulté de manière plus ou moins élégante. La plupart du temps, la visibilité est approximée dans des pré-calculs qui viennent d'information à priori extérieure à l'algorithme de reconstruction, ou à chaque itération du processus de minimisation [1, 5, 8, 9, 12, 16, 21].

Récemment, certains auteurs [6, 20] ont réussi à prendre en compte de façon rigoureuse la visibilité dans le processus même de minimisation. Dans ces récents travaux, Yezzi et Soatto [20] (dans le cas des surfaces convexes) et Gargallo et al. [6] (dans le cas général) fournissent le gradient exact de l'erreur de reprojection. Le problème est résolu dans le cadre mathématique des surfaces continues (lisses), comme celui décrit par Solem et Overgaard [17]. Ces travaux ont montré que le calcul correct du gradient de l'erreur de reprojection faisait apparaître un nou-

*Recherche financée par l'Agence Nationale pour la Recherche, projet Flamenco - ANR-06-MDCA-007.

veau terme, qui force les contours apparents générés par la surface reconstruite à correspondre parfaitement avec les contours apparents des images observées. Les méthodes ne prenant pas en compte la visibilité dans la fonctionnelle dégage un biais de réduction de la surface et sont obligées de palier à ce problème en introduisant un terme additionnel dans la fonctionnelle. Parmi celles-ci, les "forces ballon" [18], les contraintes additionnelles par enveloppes visuelles [7, 14] ou encore les termes liés aux contours (au gradient de l'image) deviennent dorénavant inutiles.

Nos travaux peuvent être vus comme une extensions des précédents travaux de Gargallo et al. [6] et Yezzi et Soatto [20]. Une des limitations du gradient calculé dans le cadre mathématique des surfaces continues est que en pratique, nous travaillons sur des représentations discrètes des surfaces. Donc quelque soit la représentation utilisée, que ce soit une grille d'ensemble de niveaux (level sets) ou un maillage, il devient nécessaire de calculer une discrétisation du gradient du cas continu (voir par exemple [1, 3, 22] pour le cas de maillages triangulaires). Cependant, cette approximation introduit inévitablement des erreurs et amène des difficultés d'implémentation d'un point de vue numérique. Le calcul de gradient pour des méthodes d'évolution de maillages a été utilisé [4, 15] en tenant compte directement de la représentation. Néanmoins, ces auteurs ne minimise pas de fonctionnelles d'énergie qui dépendent de la visibilité (par exemple pour de la segmentation d'images 3D [15] ou de la correspondance de formes (shape matching [4])). Dans cet article, nous présentons le calcul exact du gradient de l'erreur de reprojection dans le cas ou la surface discrète est représentée par un maillage triangulaire.

Une autre limitation de l'algorithme présenté dans [6, 20] est due au fait que ces auteurs émettent l'hypothèse que la scène est parfaitement Lambertienne, c'est à dire que la couleur d'un point de la surface ne varie pas en fonction du point de vue. En pratique, ces objets sont rares, et donc les algorithmes de stéréo multi-vues se doivent d'être robustes aux réflexions non Lambertiennes. Pour améliorer la précision, une stratégie courante est de modifier les images d'entrée afin d'en éliminer les spots spéculaires [23]. Cependant, ces méthodes sont fortement limitées par la configuration de l'illumination. Certains auteurs considèrent le traitement des pixels des images qui ont une composante spéculaire de manière indépendante : ils traitent ces points comme des cas aberrants [1, 8] et les ignorent. Pour compenser le manque d'information, ils ont besoin davantage d'images. D'autres essayent d'être robustes aux effets non-Lambertien en incorporant directement un modèle de réflectance spéculaire dans la formulation mathématiques du problème [9, 21], néanmoins jusqu'à présent, ces formulations amènent à résoudre des problèmes d'optimisation très complexes et coûteux en calcul et qui retournent des résultats imprécis. Ce type de formulation est également limité à un type spécifique de matériaux (ceux dont le modèle de réflectance correspond parfaitement à celui utilisé dans le modèle génératif). Une idée très répandue est d'utiliser une mesure de similarité robuste entre les images [10, 12, 19]. Cependant, ces méthodes sont plus compliquées à implémenter et ne peuvent pas prendre en compte les cas ou l'erreur est grande comme par exemple des occultations. Dans cet article, nous proposons de modifier le modèle Lambertien afin de prendre en compte les variations par rapport à une illumination constante et de modéliser ces déviations à l'aide d'un terme lisse basé image. Nous ne modélisons pas explicitement les propriétés de réflectances de la scène (contrairement à [9, 21]) ou autres phénomènes photométriques complexes lié aux caméras. Malgré tout, la méthode est générique, simple, facile à mettre en oeuvre et efficace pour des effets non-Lambertiens faibles. Nous arrivons à reconstruire de façon relativement robuste la forme et la réflectance diffuse (albedo) de scènes observées par des caméras biaisées.

2 Formulation du problème

Reconstruire la forme 3D et l'information photométrique (telle que la radiance) d'une scène à partir de plusieurs images de cette scène est le problème inverse de la synthèse d'image. En particulier, si on est capable de générer des images à partir d'un modèle donné, alors la comparaison entre ces images générées et celles observées (i.e. les données d'entrée) offre une mesure naturelle de l'erreur de reprojection. En effet, cette approche correspond exactement à une formulation Bayesienne du problème qui consiste à maximiser la probabilité de vraisemblance (i.e. ici l'erreur de reprojection) et l'a priori (correspondant à un a priori sur la forme) au lieu de maximiser directement la probabilité a posteriori. La formulation en terme d'énergies correspond à minimiser la somme d'un terme d'attache aux données (l'erreur de reprojection) et d'un terme de lissage (terme d'a priori).

2.1 Modélisation de l'erreur de reprojection

Afin de comparer l'ensemble des images observées (les données) avec les images générées à partir du modèle, il est crucial de définir et modéliser le fond (arrière plan). Ceci, comme expliqué dans [6, 20], permet d'être sur que la surface estimée à partir de la zone d'intérêt de l'objet ne se limite pas à l'ensemble vide (qui est le minimum global de la plupart des fonctionnelles d'énergie décrites dans les autres travaux) et permet également aux contours générés par l'objet reconstruit de correspondre parfaitement aux contours apparents des images d'entrées. Alors que la plupart des travaux précédents supposent que le fond est connu (e.g. en le modélisant simplement comme étant de couleur uniforme, ou en exploitant les silhouettes données), ici nous estimons également les images de fond $B_i : \mathcal{I}_i \rightarrow \mathbb{R}^3$, sous l'hypothèse que ces images ont des variations lisses, de manière similaire à [20]. Nous notons par i l'index de la caméra en question et $\mathcal{I}_i$ le domaine de l'image.

Maintenant supposons que la surface $\mathcal{S}$ de la scène est Lambertienne et que l'illumination est statique. Soit $C : \mathcal{S} \rightarrow \mathbb{R}^3$, la fonction de radiance qui associe à chaque point

de la surface une couleur. Idéalement, la couleur $I_i(\mathbf{u})$ observée à un pixel $\mathbf{u}$ de l'image I_i devrait être égale à la couleur $C(\pi_{i,\mathcal{S}}^{-1}(\mathbf{u}))$ de sa rétro-projection $\pi_{i,\mathcal{S}}^{-1}(\mathbf{u})$ sur la surface ou, dans le cas où $\mathbf{u} \notin \pi_i(S)$, à la couleur $B_i(\mathbf{u})$ du même pixel sur l'image de fond (où π_i représente la projection associé à la caméra i). De plus, l'erreur de reprojection de la surface est

$$E_{data} = \frac{1}{2}\sum_i \left[\int_{\pi_i(\mathcal{S})} \left(I_i(\mathbf{u}) - C(\pi_{i,\mathcal{S}}^{-1}(\mathbf{u}))\right)^2 d\mathbf{u} + \int_{\mathcal{I}_i - \pi_i(\mathcal{S})} \left(I_i(\mathbf{u}) - B_i(\mathbf{u})\right)^2 d\mathbf{u} \right] . \quad (1)$$

Enfin, afin de rendre le problème bien posé, nous utilisons comme terme d'a priori sur $\mathcal{S}$ un terme additionnel de lissage de l'aire soit l'énergie $E_{\mathcal{S}} = \int_{\mathcal{S}} d\sigma$. On suppose également que les images de fond sont lisses en ajoutant la contrainte $E_{RB} = \sum_i \int_{\mathcal{I}_i} |\nabla B_i(\mathbf{u})|^2 d\mathbf{u}$.

2.2 Modèle correctif de réflectance

En pratique, très peu, voir aucune, scène réelle n'est vraiment Lambertienne. Dans cette section, nous présentons une modification du modèle Lambertien (1) pour prendre en compte les variations par rapport à l'hypothèse de brillance constante, sans modéliser les propriétés de réflectance de la scène de manière explicite. Nous considérons le cas d'images qui seraient produites par un modèle Lambertien, plus une composante additionnelle corrective. Ce terme additionnel contient la composante non Lambertienne de la radiance. Nous estimons donc la couleur Lambertienne C et la composante corrective $F_i : \mathcal{I}_i \to \mathbb{R}^3$. Afin de contraindre le problème, on suppose que le terme correctif a des variations lisses, en ajoutant l'énergie suivante $E_{RF} = \sum_i \int_{\mathcal{I}_i} |\nabla F_i(\mathbf{u})|^2 d\mathbf{u}$. La radiance de la surface est ainsi obtenue en utilisant l'image corrigée R_i : $\mathcal{I}_i \to \mathbb{R}^3$ tel que $R_i = I_i - F_i$. En d'autres termes, dans l'équation (1) on remplace I_i par R_i.

En plus de corriger les effets non-Lambertiens de la réflectance, les images F_i incorporent d'autres phénomènes photométriques complexes négligés dans le modèle de caméra, qui permet de retrouver de manière plus robuste la forme et la radiance diffuse C de scènes non-Lambertiennes vues par des caméras pouvant comporter de légères variations photométriques. Contrairement aux autres méthodes qui essayent d'être robustes aux effets spéculaires en modélisant la réflectance par des modèles paramétriques [21], l'algorithme que nous proposons ici est plus générique et peut être utilisé pour de nombreux matériaux sans modification. Également, nous n'avons pas besoin de mesure de similarité complexes telles que la corrélation croisée ou l'information mutuelle comme dans [12], et donc l'algorithme est plus simple et fonctionne naturellement pour des cas ou certaines parties d'images sont dégradées, comme par exemple le cas ou l'on retrouve certaines occultations dans les images des caméras.

Il est cependant important de noter qu'ici le terme de régularisation suppose des variations lisses, et qu'il serait plus judicieux de choisir une régularisation en utilisant par exemple le terme de variation totale $\sum_i \int_{\mathcal{I}_i} |\nabla F_i(\mathbf{u})| d\mathbf{u}$ permettant de préserver les discontinuités. Néanmoins ces résultats feront l'objet d'une future publication et ne seront pas traités dans cet article.

3 Minimisation de l'énergie totale

Afin de minimiser l'énergie totale $E_{total} = E_{data} + \lambda_{\mathcal{S}} E_{\mathcal{S}} + \lambda_B E_{RB} + \lambda_F E_{RF}$, on utilise une descente de gradient alternativement par rapport aux variables F_i, B_i et $\mathcal{S}$. Le calcul du gradient par rapport à F_i et B_i est classique puisque les termes sont basés images [20]. Pour une surface $\mathcal{S}$ fixe, et une couleur C fixe, on a

$$\begin{aligned} \nabla E(F) &= -\sum_i \left[(I_i - C \circ \pi_{i,\mathcal{S}}^{-1} - F_i)h + \lambda_F \Delta F_i \right] \\ \nabla E(B) &= -\sum_i \left[(I_i - B_i)(1-h) + \lambda_B \Delta B_i \right] , \end{aligned} \quad (2)$$

où h est la fonction caractéristique qui indique si $\mathbf{u}$ correspond à la projection de la surface $\mathcal{S}$ ($h(\mathbf{u}) = 1$) ou non ($h(\mathbf{u}) = 0$, $\mathbf{u}$ correspond alors au fond) ; λ est le paramètre de lissage de la surface et λ_F et λ_B sont les paramètres de lissage des termes correctifs qui sont ajustés manuellement ; $C(\mathbf{x})$ est calculé en prenant la moyenne des projections de $\mathbf{x}$ dans les images corrigées R_i où $\mathbf{x}$ est visible. En revanche, le calcul du gradient de E_{total} par rapport à $\mathcal{S}$ (plus exactement de E_{data} puisque E_{RB} et E_{RF} ne dépendent pas de $\mathcal{S}$; et $E_{R\mathcal{S}}$ est classique [2, 20]) est difficile et requière plus d'attention. En fait, le calcul exact du gradient de ce type de fonctionnelle est le résultat de [6, 20], où les calculs sont fait dans un cadre purement théorique des surfaces continues et différentiables comme celui décrit dans [17] en utilisant les dérivées de Gâteaux. Ici, contrairement à [6, 20], nous allons directement calculer le gradient exact par rapport à la surface représentée par un maillage (bien sûr, la surface n'étant plus lisse, les résultats décrits dans [6, 20] ne sont plus applicables). Comme dans [6], on réécrit tout d'abord l'énergie sous la forme d'une intégrale sur le volume visible. Ensuite dans la section 4, nous calculons le gradient exact de E_{data} par rapport au maillage *discret*. Afin de simplifier l'écriture, nous allons calculer le gradient pour une seule caméra et ainsi enlever la dépendance en i. Pour plusieurs caméras, le gradient est bien entendu la somme de tous les gradients de chaque caméra.

Réécrire l'énergie comme une intégrale sur le volume visible

La première étape afin de minimiser notre fonctionnelle d'énergie par rapport à la surface $\mathcal{S}$ est de réécrire l'énergie comme une intégrale sur la surface au lieu de l'image. Ce changement de variable implique d'adapter l'élément de mesure sur la surface [16] et de prendre en compte la visibilité de la surface [6, 12, 20]. Ceci est obtenu par $d\mathbf{u} = -\frac{\mathbf{x}\cdot\mathbf{n}(\mathbf{x})}{\mathbf{x}_z^3}\nu_{\mathcal{S}}(\mathbf{x})d\sigma$ où $d\sigma$ est l'élément classique

d'aire de la surface, ν_S est la fonction caractéristique de visibilité (égale à 1 si un point $\mathbf{x}$ est visible de la caméra et 0 sinon), $\mathbf{n}(\mathbf{x})$ est le vecteur unité normal à la surface $\mathcal{S}$ au point $\mathbf{x}$, et $\mathbf{x}_z$ est la profondeur de $\mathbf{x}$ liée à la caméra. En utilisant la technique de séparation du fond et de l'objet d'intérêt proposée dans [20], la fonctionnelle d'énergie devient (pour une *seule* image) :

$$\begin{aligned} E_{data}(\mathcal{S}) &= -\int_{\mathcal{S}} g_R(\mathbf{x}) \frac{\mathbf{x}\cdot\mathbf{n}(\mathbf{x})}{\mathbf{x}_z^3} \nu_{\mathcal{S}}(\mathbf{x}) d\sigma \\ &\quad + \int_{\mathcal{I}-\pi(\mathcal{S})} g_B(\mathbf{u}) d\mathbf{u}\,, \\ &= -\int_{\mathcal{S}} [g_R(\mathbf{x}) - g_B(\pi(\mathbf{x}))] \frac{\mathbf{x}\cdot\mathbf{n}(\mathbf{x})}{\mathbf{x}_z^3} \nu_{\mathcal{S}}(\mathbf{x}) d\sigma \\ &\quad + \int_{\mathcal{I}} g_B(\mathbf{u}) d\mathbf{u}\,, \end{aligned} \tag{3}$$

où $g_R(\mathbf{x})$ est $1/2\,[R(\pi(\mathbf{x})) - C(\mathbf{x})]^2$ et $g_B(\mathbf{u})$ est $1/2\,[R(\mathbf{u}) - B(\mathbf{u}))]^2$. Le terme de droite de l'équation (3) ne dépend pas de $\mathcal{S}$, donc par la suite ce terme est omit intentionnellement puisqu'il ne contribue pas à l'expression du gradient (par rapport à $\mathcal{S}$). Enfin, en notant $g(\mathbf{x}) = g_R(\mathbf{x}) - g_B(\pi(\mathbf{x}))$ pour plus de lisibilité, l'énergie à minimiser par rapport à $\mathcal{S}$ devient

$$E(\mathcal{S}) = -\int_{\mathcal{S}} g(\mathbf{x}) \frac{\mathbf{x}\cdot\mathbf{n}(\mathbf{x})}{\mathbf{x}_z^3} \nu_{\mathcal{S}}(\mathbf{x}) d\sigma. \tag{4}$$

Finalement, en utilisant le théorème de divergence de Gauss qui dit que pour tout domaine $\Omega \in \mathbb{R}^3$ et pour tout champs vectoriel $\mathbf{w} : \mathbb{R}^3 \to \mathbb{R}^3$, $\int_\Omega \nabla\cdot\mathbf{w} d\mathbf{x} = \int_{\partial\Omega} \mathbf{w}\cdot\mathbf{n} d\sigma$, la fonctionnelle de coût (4) peut être réécrite comme :

$$\begin{aligned} E(\mathcal{S}) &= -\int_{\mathbb{R}^3} \left(\nabla g(\mathbf{x}) \cdot \frac{\mathbf{x}}{\mathbf{x}_z^3}\right) \nu_{\mathcal{S}}(\mathbf{x})\, d\mathbf{x} \\ &= -\int_{\mathbb{R}^3} f(\mathbf{x})\, \nu_{\mathcal{S}}(\mathbf{x})\, d\mathbf{x}\,. \end{aligned} \tag{5}$$

Pour plus de clarté, nous notons $f(\mathbf{x}) = \nabla g(\mathbf{x}) \cdot \frac{\mathbf{x}}{\mathbf{x}_z^3}$. En particulier, ce résultat est valable pour toute surface continue par morceaux.

4 Gradient de l'erreur de reprojection pour les maillages

Dans cette section, nous calculons le gradient discret de $E(\mathcal{S})$ (et donc de E_{data}), par rapport à la surface, dans le cas où la surface $\mathcal{S}$ est représentée par un maillage triangulé $\mathbf{X}$. Les sommets de $\mathbf{X}$ sont notés $\mathbf{x}_k$ et $\mathcal{S}$ est déformée en bougeant les sommets $\mathbf{x}_k$ selon les vecteurs $\mathbf{V}_k$. La méthode pour calculer le gradient de $E(\mathcal{S})$ par rapport à $\mathbf{x}_k$ consiste à calculer la dérivée directionnelle de $E(\mathcal{S}[t])$ pour toute déformation $\mathbf{x}_k[t] = \mathbf{x}_k^0 + t\mathbf{V}_k$ et puis en l'écrivant sous la forme d'un produit scalaire de $\mathbf{V}_k$, i.e. comme $\langle \mathbf{V}_k, G\rangle$. Le vecteur obtenu G est appelé le *gradient* et l'énergie décroît nécessairement quand la surface se déforme suivant sa direction opposée $-G$. De plus, pour $\mathbf{x}_k[t] = \mathbf{x}_k^0 - tG$, on a $[E\circ\mathcal{S}]'(0) = -\langle G, G\rangle \leq 0$, voir [4].

Pour calculer le gradient, nous devons comprendre les variations de notre énergie lorsque $\mathcal{S}$ se déforme. Comme notre énergie $E(\mathcal{S})$ est une intégrale sur le volume visible, ses variations sont uniquement dues aux variations du volume visible. Ainsi, comme illustré Figure 1 (restreinte au cas 2D par simplicité), quand le somment $\mathbf{x}_k$ bouge selon $\mathbf{V}_k$ nous devons séparer deux cas :

1. lorsque tous les triangles adjacents à $\mathbf{x}_k$ sont visibles, la variation du volume visible est simplement la somme des tétraèdres formés par les triangles adjacents et le point déformé $\mathbf{x}_k + \mathbf{V}_k$ (voir Figure 1, partie gauche). Le calcul du gradient correspondant est détaillé dans la Section 4.1.
2. lorsque $\mathbf{x}_k$ génère un contour occultant dans les images, i.e. lorsque c'est un point d' *horizon*, son mouvement affecte la visibilité des autres points situé derrière lui (dont on appelle le premier point le *terminateur* de $\mathbf{x}_k$). Dans ce cas, la variation du volume visible correspond à la somme du cas précédent, plus l'ensemble du volume crée par les rayons crépusculaires générés par le mouvement de l'horizon (voir Figure 1, partie droite). Le calcul du gradient correspondant est détaillé dans la Section 4.2.

4.1 Terme lié au tetrahèdre formé par les triangles adjacents visibles

Par la suite, on note $\mathcal{S}_j$ le $j^{\text{ème}}$ triangle du maillage. Considérant la variation de l'énergie causée par les triangles adjacents visibles de $\mathbf{x}_k$, on a

$$E(\mathcal{S}[t]) - E(\mathcal{S}[0]) = \sum_j \varpi_{j,\mathbf{V}_k} \int_{Vol[j,\mathbf{V},t]} f(\mathbf{x})\, d\mathbf{x} + \ldots, \tag{6}$$

où $Vol[j,\mathbf{V},t]$ est le volume du tétraèdre formé par les sommets du triangle *visible* $\mathcal{S}_j$ et du point $\mathbf{x}_k + \mathbf{V}_k$. Le signe $\varpi_{j,\mathbf{V}_k}$ indique si de la matière a été ajoutée ou enlevée du volume de l'objet et est égale à $N_j \cdot \frac{\mathbf{V}_k}{|\mathbf{V}_k|}$, où N_j est la normale de la surface sur le triangle $\mathcal{S}_j$ pointant vers l'extérieure de la surface. Le terme "..." dans la partie droite de l'équation (6) est nul sauf si le sommet $\mathbf{x}_k$ considéré est un horizon ; ce terme additionnel sera détaillé dans le paragraphe suivant.

Maintenant, considérons une paramétrisation du volume $Vol[j,\mathbf{V},t]$ définie par $\mathbf{x}(u,v,w) = \mathbf{x}(u,v) + w\,t\mathbf{V}_k$, où $\mathbf{x}(u,v) = \mathbf{x}_k + u\,\overrightarrow{\mathbf{x}_k\mathbf{x}_{k1}} + v\,\overrightarrow{\mathbf{x}_k\mathbf{x}_{k2}}$ est une paramétrisation du triangle $\mathcal{S}_j$; $\mathbf{x}_{k1}$ et $\mathbf{x}_{k2}$ sont les autres sommets de $\mathcal{S}_j$; $(u,v) \in T = \{(u,v) \mid u \in [0,1] \text{ et } v \in [0, 1-u]\}$ et $w \in [0, 1-u-v]$. Par changement de variables, l'équation (6) donne

$$\sum_j \varpi_{j,\mathbf{V}_k} \int_T \int_0^{1-u-v} \Big\{ f(\mathbf{x}(u,v) + wt\mathbf{V}_k) \\ |det(\overrightarrow{\mathbf{x}_k\mathbf{x}_{k1}}, \overrightarrow{\mathbf{x}_k\mathbf{x}_{k2}}, t\mathbf{V}_k)| \Big\} dw du dv + \ldots \tag{7}$$

Soit A_i l'aire du triangle $\mathcal{S}_j$. Comme $\varpi_{j,\mathbf{V}_k} |det(\overrightarrow{\mathbf{x}_k\mathbf{x}_{k1}}, \overrightarrow{\mathbf{x}_k\mathbf{x}_{k2}}, t\mathbf{V}_k)| = -2t\, A_j \mathbf{V}_k \cdot N_j$, la

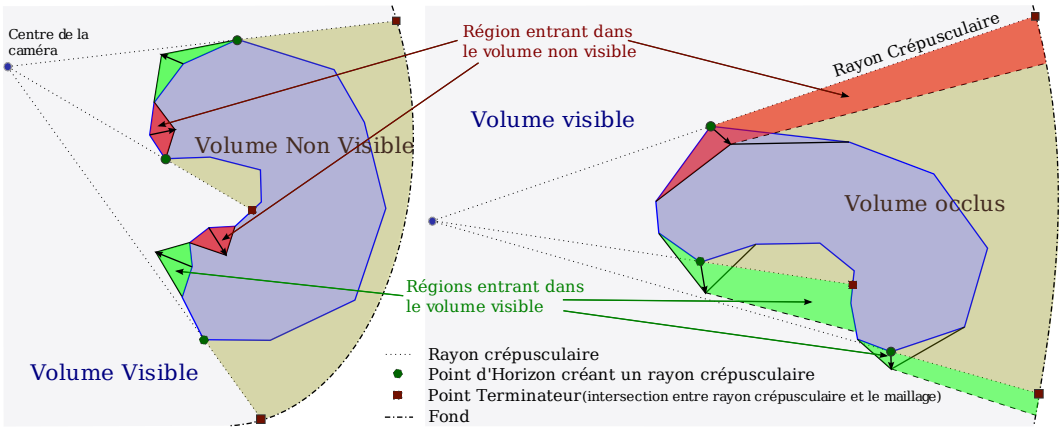

FIG. 1 – Représentation géométrique du changement de visibilité lorsque le maillage triangulé se déforme. Contrairement au terme d'intérieur (partie gauche), le mouvement de l'horizon (partie droite) affecte fortement le mouvement de l'interface de visibilité entre les volumes visibles et occultés en créant un mouvement des rayons crépusculaires.

partie gauche de l'équation (7) devient

$$-2t\,\mathbf{V}_k \cdot \sum_j A_j N_j \int_T \int_0^{1-u-v} f(\mathbf{x}(u,v) + wt\mathbf{V}_k)\, dwdudv\,. \tag{8}$$

La limite de $\frac{E(\mathcal{S}[t])-E(\mathcal{S}[0])}{t}$ quand t tend vers zéro est donc

$$\mathbf{V}_k \cdot \left[-2 \sum_j A_j N_j \int_T \nabla g(\mathbf{x}) \frac{\mathbf{x}}{\mathbf{x}_z^3} (1-u-v)\, dudv \right], \tag{9}$$

où la partie entre crochets est la première partie du gradient considéré. A noter que ce résultat reprend celui utilisé dans d'autres travaux sur les maillages [4, 15]. Considérons maintenant le cas où $\mathbf{x}_k$ génère un contour occultant et que c'est un point d'horizon.

4.2 Terme lié aux mouvements des horizons

Soit $H_{k,j}$, le vecteur tel que $[\mathbf{x}_k, \mathbf{x}_k + H_{k,j}]$ est une arête du triangle $\mathcal{S}_j$ générant des rayons crépusculaires, et où $\mathbf{x}_k$ est un point d'horizon. Le volume correspondant au mouvement de l'horizon peut être paramétré par les points $\mathbf{y}(u,v)$ du triangle $\{\mathbf{x}_k, \mathbf{x}_k + H_{k,j}, \mathbf{x}_k + v\ t\mathbf{V}_k\}$ générés par le mouvement de l'horizon. De manière plus rigoureuse, cela peut être paramétré par un ensemble de points $\mathbf{x}(u,v,r) = r\,\mathbf{y}(u,v)$ où $\mathbf{y}(u,v) = \mathbf{x}_k + u\,H_{k,j} + v\,t\mathbf{V}_k$; r correspond à la profondeur de $\mathbf{x}$ dans la direction du point de vue ; $r \in [1, \mathcal{T}_{(u,v)}]$.
Notons que $\mathbf{y}(u,v)$ dépend de t ; Soulignons cette dépendance en la notant $\mathbf{y}_t(u,v)$. Par un changement de variable, on obtient $E(\mathcal{S}[t]) - E(\mathcal{S}[0]) =$

$$\ldots + \frac{1}{2} \int_{u,v} \int_r f(r\,\mathbf{y}_t(u,v))\,(H_{k,j} \wedge t\mathbf{V}_k) \cdot \frac{\mathbf{y}_t(u,v)}{|\mathbf{y}_t(u,v)|} r^2 dudvdr\,,$$

où "..." correspond à la partie décrite dans le paragraphe précédent. A partir de maintenant, nous nous intéressons uniquement au second terme. La limite de $\frac{E(\mathcal{S}[t])-E(\mathcal{S}[0])}{t}$ lorsque t tend vers zéro est

$$\frac{1}{2} \int_{u,v} \int_r f(r\,\mathbf{y}(u))\,(H_{k,j} \wedge \mathbf{V}_k) \cdot \frac{\mathbf{y}(u)}{|\mathbf{y}(u)|} r^2 dudvdr\,, \tag{10}$$

où $\mathbf{y}(u) = \mathbf{x}_k + u\,H_{k,j}$ et $\mathcal{T}_u = \mathcal{T}_{(u,0)}$. Notons $L(u) = \int_{1,\mathcal{T}_u} f(r\ \mathbf{y}(u)) r^2 dr$, alors la partie droite de l'équation (10) peut être réécrite comme

$$\mathbf{V}_k \cdot \frac{1}{2} \int_u L(u) \left(\frac{\mathbf{y}(u)}{|\mathbf{y}(u)|} \wedge H_{k,j} \right) (1-u) du\,. \tag{11}$$

Finalement, en se rappelant que $f(\mathbf{x}) = -\nabla g(\mathbf{x}) \frac{\mathbf{x}}{\mathbf{x}_z^3}$, on a

$$L(u) = \left[g(T(\mathbf{y}(u))) - g(\mathbf{y}(u)) \right] \frac{1}{[\mathbf{y}(u)]_z^3}\,, \tag{12}$$

où $T(\mathbf{x})$ est le point terminateur de $\mathbf{x}$. Enfin, le flot complet de gradient qui minimise E_{data} (par rapport à $\mathbf{x}_k$) est le terme entre crochets de (9) obtenu précédemment, plus le terme suivant si $\mathbf{x}_k$ est un point d'horizon :

$$\sum_{\text{arêtes } H_{k,j}} \frac{1}{2} \int \left(g(T(\mathbf{y}(u))) - g(\mathbf{y}(u)) \right) \frac{\mathbf{y}(u) \wedge H_{k,j}}{|\mathbf{y}(u)|\,[\mathbf{y}(u)]_z^3} (1-u) du\,. \tag{13}$$

5 Expérimentations et résultats

Nous avons implémenté notre algorithme en utilisant les maillages adaptatifs avec changements de topologie proposés dans [11]. La visibilité est calculée en utilisant le Z-buffer d'OpenGL. Les horizons sont détectés en regardant le changement de signe entre les normales des facettes adjacentes et la direction du point de vue. L'évolution est effectué par une approche multi-résolution et démarre de l'enveloppe visuelle. A chaque niveau de résolution, la taille des triangles est homogène sur la surface et est fixe. L'intégration sur les triangles est effectuée de telle sorte qu'elle soit cohérente avec la taille des images.

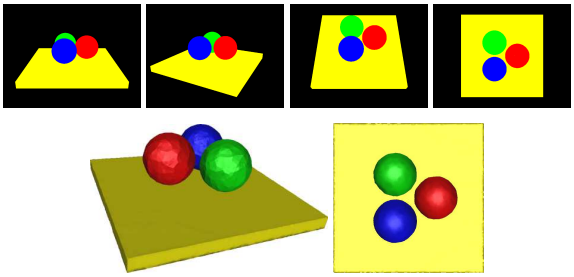

FIG. 2 – La séquence *"boules"*. 4 des 20 images d'entrée (haut) ; résultats calculés avec le terme d'horizon de la Section 4.2 vus de différents points de vues.

FIG. 3 – Séquence de synthèse *"dragon"*. De gauche à droite : 1 des 32 images d'entrée ; vérité terrain ; forme reconstruite par notre algorithme.

Par exemple si un triangle se projette sur une large zone sur l'image, le nombre d'échantillons sur le triangle sera important.

Tests sur le gradient discret

Comme dans [6], tout d'abord nous reconstruisons trois "boules" colorées de manière uniforme et placées sur un plan à partir de 20 images (640×480), voir Figure 2. De cette façon, on met en évidence l'importance du terme d'horizon puisque le gradient de la couleur est nul sur la surface sauf aux interfaces entre les objets sur les images. Si l'on utilise uniquement le terme d'intérieur (Section 4.1), alors la surface se réduit et disparaît. En utilisant uniquement le terme d'horizon (exprimé dans la Section 4.2), on arrive correctement à reconstruire et à séparer les balles, et les contours apparents se reprojettent correctement dans les images. Nous avons également testé notre algorithme sur des données de synthèse créées à partir du maillage "dragon" de Stanford (Figure 3). Ces données sont créées en considérant des hypothèses Lambertiennes et est composé de 32 images de 640×480. Les résultats montrent la reconstruction correcte de la forme, même si la texture est très lisse et certaines parties dans l'ombre ont peu de contraste et sont sombres (voir Tableau 1).

Cas de surfaces non-Lambertiennes

Afin de tester le modèle correctif introduit Section 2.2, nous modifions les images d'entrées utilisées dans les tests précédents pour simuler des problèmes photométriques liés aux caméras. 8 des 32 images d'entrée ont été modifiées en ajoutant divers gradients colorés sur les images. Notons que la forme est correctement reconstruite mais que

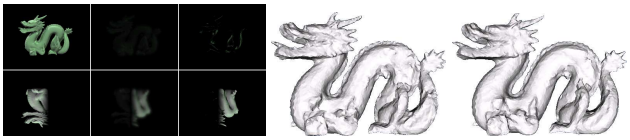

FIG. 5 – Séquence de synthèse *"dragon"* avec de la réflectance spéculaire et des occlusions. De gauche à droite : 2 des 32 images d'entrée ; composante corrective estimée ; vérité terrain des composantes correctives ; surface reconstruite sans le modèle correctif ; surface reconstruite avec le modèle correctif.

la radiance est également correctement estimée (cf. Figure 4 et Tableau 1). Les composantes correctives des images sont retrouvées à partir de la région d'intérêt sur l'image. Comme décrit Section 3, les images de fond estimées correspondent à celles des images d'entrée sur les régions complémentaires. Dans un deuxième temps, nous ajoutons un modèle de réflectance spéculaire aux images du dragon, et modifions 3 d'entre elles de telle manière que la moitié de ces images soient noires. Cela pourrait par exemple simuler des objets occultant la caméra de manière lisse comme par exemple un objet devant la caméra en dehors de la profondeur de champs. Les résultats obtenus avec la modélisation dans un cadre Lambertien (Section 2.1) sont beaucoup moins précis que ceux obtenus avec le modèle correctif de la Section 2.2. Le modèle correctif permet également de séparer et retrouver en partie les composantes diffuses et spéculaires des images (voir Figure 5 et le Tableau 1). Cependant les résultats obtenus sont limités par le modèle présenté et pourraient être améliorés en changeant le terme de régularisation pour supposer des variations lisses par morceaux.

Finalement, nous avons testé notre méthode sur les données classique de Middlebury à savoir le "Dino" (16 images de 640×480) ainsi que le "Temple" (16 images de 640×480), cf. Figure 6. Pour des résultats quantitatifs, nous référons au site web d'évaluation de Middlebury [13]. Notons que même si ces objets particuliers ont été choisis pour leur nature Lambertienne, les images ont une certaine quantité de reflets spéculaire qui peut être interprété dans les images correctives estimées. Cette composante montre une partie verdâtre sur le haut de ces images et une partie rougeâtre sur le bas, ce qui peut être du à un biais photométrique de la caméra puisque cela apparaît sur toutes les images et qu'elles ont été prises avec le même appareil.

Malgré la simplicité de l'énergie considérée (simple erreur quadratique entre le modèle et les données d'entrée), les résultats obtenus sont très satisfaisant et très prometteurs.

6 Conclusion

Dans cet article, nous avons présenté une nouvelle technique de modèles déformables pour résoudre le problème de stéréo multi-vues, basé sur une représentation *discrète* de la surface. Contrairement aux travaux antérieurs, pendant l'évolution nous prenons parfaitement en compte les

FIG. 4 – Séquence de synthèse *"dragon"*. De gauche à droite : 2 des 32 images originelles ; composante ajoutée pour simulé des effets photométriques sur les caméras ; images d'entrée ; composante corrective estimée (multipliée par 2 pour la visualisation) ; image de fond estimée ; surface et radiance estimées.

	Images diffuses (Fig. 3)	Images diffuses Modèle correctif (Fig. 4)	Images spéculaires Modèle Lambertien (Fig. 5)	Images spéculaires Modèle correctif (Fig. 5)
Acc. 95%	0.241mm	0.199mm	0.410mm	0.300mm
Comp. 0.5mm	98.3%	98.7%	97.2%	97.6%

TAB. 1 – Évaluation numérique de la méthode proposée sur la séquence du "dragon" qui montre la précision à 95% et la complétude à 0.5mm en utilisant la méthode décrite dans [13]. (comparée à la vérité terrain)

FIG. 6 – Séquences du "Dino" et du "Temple". De gauche à droite : 1 des 16 images d'entrée ; composante corrective non-Lambertienne estimée (multipliée par 8 pour visualisation) ; images de fond estimées (multipliée par 2 pour visualisation) ; radiance estimée ; surface reconstruite vue d'un point de vue différent.

changements de visibilité en exprimant le gradient exact de la fonctionnelle de l'erreur de reprojection. En particulier, exactement comme pour le cas continu [6], cela force les contours générés par la surface à apparaître au bon endroit sur les images, et cela supprime le biais de surface minimale présent dans la plupart des méthodes précédentes. De plus, notre méthode est robuste aux effets non-Lambertiens lisses et aux défauts photométrique des caméras (comme les erreurs d'étalonnage photométrique ou les occultations partielles), en introduisant un terme simple correctif basé image, qui a montré son efficacité dans les divers tests que nous avons fait sur des données de synthèse et des données réelles.

Références

[1] N. Birkbeck, D. Cobzas, P. Sturm, and M. Jägersand. Variational shape and reflectance estimation under changing light and viewpoints. In *ECCV*, volume 1, pages 536–549, 2006.

[2] M. Desbrun, M. Meyerm, P. Schröder, and A. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *SIGGRAPH '99*, pages 317–324, 1999.

[3] Y. Duan, L. Yang, H. Qin, and D. Samaras. Shape reconstruction from 3d and 2d data using pde-based deformable surfaces. In *ECCV*, pages 238–251, Prague, Czech Republic, 2004.

[4] I. Eckstein, J.-P. Pons, Y. Tong, C.C. Jay Kuo, and M. Desbrun. Generalized surface flows for mesh processing. In *Symp. on Geometry Processing*, pages 183–192, 2007.

[5] O.D. Faugeras and R. Keriven. Variational-principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE TIP*, 7(3) :336–344, March 1998.

[6] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *IEEE ICCV*, pages 1–8, Rio de Janeiro, Brazil, 2007.

[7] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Comput. Vis. Image Underst.*, 96(3) :367–392, 2004.

[8] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE TPAMI*, 30(3) :548–554, 2008.

[9] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *IJCV*, 63(3) :175–189, 2005.

[10] H. Jin, A.J. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *3DPVT*, pages 626–630, 2002.

[11] J.-P. Pons and J.-D. Boissonnat. Delaunay deformable models : Topology-adaptive meshes based on the restricted delaunay triangulation. In *IEEE CVPR*, Minneapolis, USA, Jun 2007.

[12] J.-P. Pons, R. Keriven, and O. Faugeras. Modelling dynamic scenes by registering multi-view image sequences. In *IEEE CVPR*, pages 822–827, San Diego, USA, Jun 2005.

[13] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE CVPR*, pages 519–528, 2006.

[14] S. N. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints : a maximum-flow formulation. In *IEEE ICCV*, pages 349–356, 2005.

[15] G. Slabaugh and G. Unal. Active polyhedron : surface evolution theory applied to deformable meshes. *IEEE CVPR*, 2 :84–91 vol. 2, 2005.

[16] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multi-view stereo. In *IEEE ICCV*, pages 974–981, 2003.

[17] J. E. Solem and N. Chr. Overgaard. A geometric formulation of gradient descent for variational problems with moving surfaces. In *Scale-Space*, pages 419–430, 2005.

[18] G. Vogiatzis, C. Hernández Esteban, P. H. S. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *PAMI*, 29(12) :2241–2246, 2007.

[19] R. Yang, M. Pollefeys, and G. Welch. Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In *IEEE ICCV*, pages 576–583, 2003.

[20] A. Yezzi and S. Soatto. Stereoscopic segmentation. *IJCV*, 53(1) :31–43, 2003.

[21] T. Yu, N. Xu, and N. Ahuja. Shape and view independent reflectance map from multiple views. *IJCV*, 73(2) :123–138, 2007.

[22] A. Zaharescu, E. Boyer, and R. P. Horaud. Transformesh : a topology-adaptive mesh-based approach to surface evolution. In *ACCV*, pages 166–175, Tokyo, Japan, Nov 2007. Springer.

[23] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. Belhumeur. Color subspaces as photometric invariants. *To appear in IJCV*, 2008.