

CERTIS

Centre d'Enseignement
et de Recherche
en Technologies
de l'Information
et Systèmes

projets communs
INRIA
ENS Paris

Créer des mondes virtuels tridimensionnels à partir de photographies numériques : une possibilité prochainement élargie au grand public

Jean-Philippe Pons et Renaud Keriven

La généralisation des appareils photographiques et des caméras numériques, aussi bien dans l'espace public que pour les usages privés, crée de nouvelles possibilités, de nouveaux besoins de visualisation et de communication. Ainsi, la création de modèles tridimensionnels de notre environnement à partir de ces capteurs, omniprésents et d'un coût désormais dérisoire, est appelée à se généraliser dans la décennie à venir grâce à la vision par ordinateur, domaine d'expertise du CERTIS.

Les récentes avancées de cette discipline à la croisée du traitement du signal et de l'intelligence artificielle permettent de substituer un simple appareil photo numérique à l'actuel télémètre laser, coûteux, fragile, lourd et encombrant. Bientôt, ces techniques deviendront un outil privilégié pour la fabrication de mondes virtuels multi-modaux pour la réalité augmentée, l'aménagement urbain, les effets spéciaux et les jeux vidéo. Des domaines plus inattendus, tels que l'art ou l'archéologie, pourront également en bénéficier.



► [Fig. 1]

À gauche, 4 des 24 images d'entrée d'un buste en bronze. À partir de celles-ci, la méthode crée automatiquement un modèle tridimensionnel, permettant par la suite de générer à volonté des images de synthèse. À droite, avec ou sans texture.



► Estimer le relief par stéréovision

La reconstruction tridimensionnelle à partir de simples photographies est rendue possible par le principe de la stéréovision, le même principe qui est déjà largement utilisé dans le monde animal pour l'estimation du relief. Il s'agit de trouver dans différentes images numériques d'une même scène et prises de points de vue différents, des pixels correspondant au même point matériel. La position 3D de ce dernier se calcule dans un second temps, facilement par triangulation.

Malheureusement, bien que représentant une tâche triviale pour l'homme, le problème de mise en correspondance s'avère difficile à résoudre de manière fiable par un algorithme informatique. En effet, étant donnée la complexité de l'éclairément et des propriétés de réflectance dans les scènes réelles, la même partie d'un objet peut présenter des apparences très diverses dans les différentes images.

► Une approche par optimisation de forme

Le CERTIS a développé de nouvelles approches pour obtenir une reconstruction tridimensionnelle complète d'un objet à partir d'une série d'images (typiquement 10 à 50) prises tout autour de l'objet [Fig. 1].

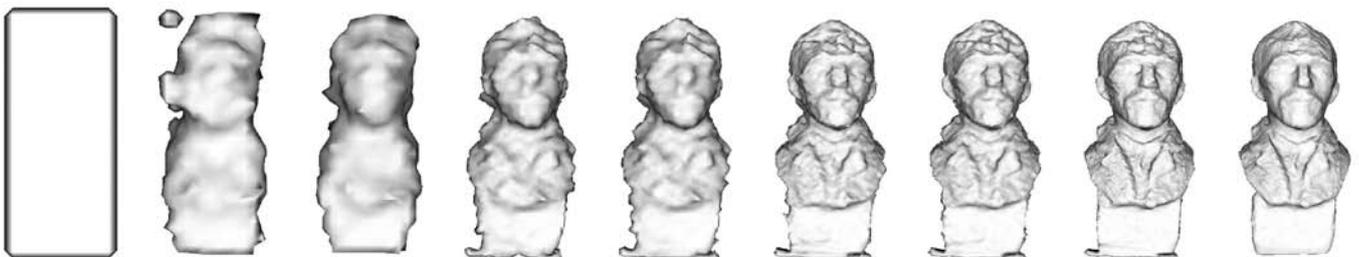
Plutôt que de s'attaquer de front au problème ambigu de mise en correspondance, il s'agit de trouver une forme qui, globalement, rende compte le mieux possible des mesures (c'est-à-dire qui soit en adéquation avec toutes les images disponibles), tout en satisfaisant des *a priori* sur la scène (comme la régularité des surfaces des objets).

Cette approche débouche naturellement sur une formulation dite « variationnelle », c'est-à-dire sur la recherche d'une configuration d'énergie minimale, pour une fonctionnelle d'énergie à définir. Aussi fait-elle appel aux outils mathématiques de l'optimisation, notamment à l'équation d'Euler-Lagrange.

Puisque la configuration recherchée est une forme, elle fait aussi intervenir les outils de la géométrie différentielle.

La fonctionnelle d'énergie appropriée pour la stéréovision étant hautement non-convexe et l'espace de configuration de grande dimension, on ne peut qu'espérer trouver un minimum global de l'énergie.

Par conséquent, il faut se résoudre à une approche évolutive, qui part d'une estimation grossière de la forme de la scène et la déforme de façon à faire décroître l'énergie le plus rapidement possible, jusqu'à converger vers un minimum local proche [Fig. 2].



► [Fig. 2]

Aperçu de la déformation de la surface au cours de la minimisation d'énergie, à partir d'une boîte englobante positionnée par l'utilisateur, jusqu'à converger vers un minimum d'énergie répondant au problème.

CERTIS

► Progrès dans la modélisation de l'énergie

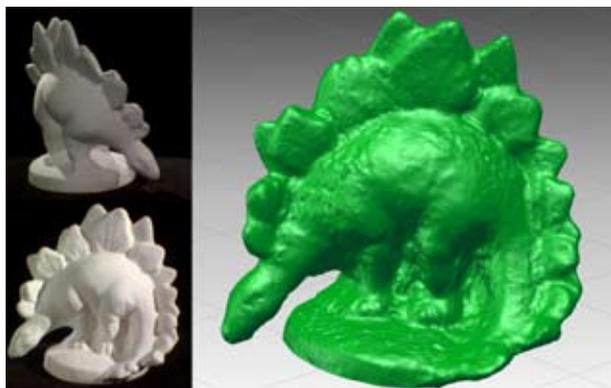
La méthodologie ci-dessus, connue sous le nom de « contours actifs » ou « modèles déformables », a fait ses preuves dans de nombreux problèmes de la vision par ordinateur, notamment celui de délimiter automatiquement les contours d'un objet dans une image.

Sa première utilisation pour la stéréovision multi-caméras a été l'œuvre de Renaud Keriven, aujourd'hui directeur du CERTIS, et d'Olivier Faugeras, en 1998.

Une reformulation récente de la question, durant la thèse de Jean-Philippe Pons, a permis de s'affranchir de plusieurs limitations partagées par la méthode initiale et par de nombreuses autres méthodes existantes.

La nouvelle approche présente l'avantage de ne nécessiter aucune hypothèse simplificatrice sur la forme des objets, alors qu'auparavant l'énergie reposait sur une approximation au mieux au premier ordre, par un plan tangent. De plus, elle s'adapte à des conditions de prise de vues très variées : caméras de réponses spectrales différentes, objets comportant des reflets et des transparences.

Enfin, elle débouche sur une implémentation à la fois plus simple et plus efficace. Notamment, tout ou partie des calculs peuvent être délocalisés sur la carte graphique. Les récentes cartes graphiques offrent en effet une puissance de calcul formidable, pour un prix minime.



► [Fig. 3]

À gauche, deux des 47 images d'entrée d'un dinosaure en plâtre, issues d'une évaluation comparative (Seitz *et al.*, 2006). À droite, aperçu d'un résultat du CERTIS.

► Confrontation des résultats au niveau international

Les résultats obtenus par les algorithmes sur des scènes complexes rivalisent avec l'état de l'art, pour des temps de calcul très inférieurs.

Afin de le démontrer, le laboratoire a participé à une évaluation comparative de différentes méthodes menée par des chercheurs de l'université de Washington.

Ils ont présenté leurs conclusions dans une des conférences majeures de vision par ordinateur, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), à New York, en juin 2006 (Seitz *et al.*, 2006).

Cette évaluation s'est poursuivie jusqu'à aujourd'hui, par le biais d'un site Internet (<http://vision.middlebury.edu/mview/>).

L'évaluation a permis de constater que les résultats initiaux du CERTIS sont toujours très bien positionnés, malgré le grand nombre de méthodes concurrentes proposées depuis [Fig. 3 et 4].



► [Fig. 4]

À gauche, deux des 47 images d'entrée d'une réplique en plâtre du temple de Dioskouroi (en Italie), issues d'une évaluation comparative (Seitz *et al.*, 2006). À droite, aperçu d'un résultat du CERTIS.

► Perspectives

Les recherches actuelles du CERTIS s'attachent à améliorer ces techniques selon trois axes :

- La reconstruction de scènes de grande taille, notamment les scènes urbaines.

La plupart des techniques actuelles ne sont adaptées qu'aux objets compacts bien approximés par leur boîte englobante.

En effet, il s'agit souvent de méthodes volumétriques, c'est-à-dire utilisant une décomposition régulière (typiquement en cellules cubiques nommées « voxels ») du domaine d'intérêt.

L'occupation mémoire et le temps de calcul d'une telle décomposition deviennent prohibitifs pour des scènes de grande taille.

- La facilité d'utilisation pour les non-spécialistes. Cela suppose de réduire le nombre de paramètres de la méthode, mais surtout d'y intégrer une procédure d'auto-calibration, c'est-à-dire d'estimation automatique et robuste des positions, orientations relatives et des caractéristiques optiques (par exemple la distance focale) des différentes prises de vue.

- L'ajout de la dimension temporelle, afin de pouvoir reconstruire la forme et le mouvement non-rigide d'une scène à partir, non plus de plusieurs images fixes, mais de plusieurs séquences vidéo filmées simultanément, sous des points de vue différents.

Ce point fait d'ores et déjà l'objet d'une collaboration avec l'équipe PERCEPTION à l'INRIA Rhône-Alpes et d'un financement de trois ans par l'ANR.

Pour en savoir plus

www.enpc.fr/certis

Jean-Philippe Pons, Renaud Keriven et Olivier Faugeras (2007), "Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score". *The International Journal of Computer Vision*, 72(2):179-193.

Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein et Richard Szeliski (2006), "A comparison and evaluation of multi-view stereo reconstruction algorithms". *IEEE Conference on Computer Vision and Pattern Recognition*, New York, États-Unis, volume 1, pages 519-526.
<http://vision.middlebury.edu/mview/>

Patrick Labatut, Renaud Keriven et Jean-Philippe Pons (2006), "Fast level set multi-view stereo on graphics hardware". *International Symposium on 3D Data Processing, Visualization and Transmission*, Chapel Hill, États-Unis, pages 774-781.

Jean-Philippe Pons (2005), "Methodological and applied contributions to the deformable models framework". Thèse de doctorat, École des ponts.

Olivier Faugeras et Renaud Keriven (1998), "Variational principles, surface evolution, PDEs, level set methods, and the stereo problem". *IEEE Transactions on Image Processing*, 7(3):336-344.

► Le CERTIS

Le CERTIS, créé en 2004, est le laboratoire d'informatique de l'École des ponts.

Ses activités de recherche s'organisent en trois projets communs avec l'INRIA et l'ENS Paris :

- Le projet « Odyssee », commun à l'École des ponts, à l'INRIA Sophia-Antipolis et à l'ENS Paris, est consacré à l'imagerie du cerveau, à la modélisation de l'activité cérébrale et de la vision biologique et aux interfaces cerveau/machine (directeur scientifique : Olivier Faugeras).
- Le projet « Willow », commun à l'École des ponts, à l'INRIA Rocquencourt et à l'ENS Paris, a été créé en janvier 2007 ; il porte sur la reconnaissance d'objets et la compréhension de scènes à partir d'images numériques (directeur scientifique : Jean Ponce).
- Le projet « Coprin », commun à l'École des ponts et à l'INRIA Sophia-Antipolis, étudie la programmation par contraintes (directeur scientifique : Jean-Pierre Merlet).

- Personnel permanent : 19
- Doctorants et post doctorants : 37
- Personnel ITA : 3

École des ponts – CERTIS
19 rue Alfred Nobel
Cité Descartes - Champs-sur-Marne
77455 Marne la Vallée cedex 2
Tél. : 01 64 15 21 72
contact : renaud.keriven@certis.enpc.fr