Une modélisation probabiliste de la reconstruction 3D

Adrien Chan-Hon-Tong

Une modélisation probabiliste de la reconstruction 3D

Adrien CHAN-HON-TONG
CEA-LIST-DIASI-LVIC
adrienchanhonton@gmail.com

30 mai 2012

1 Introduction

On s'intéresse dans cet article au problème de la reconstruction 3D (l'inférence de la géométrie) d'une scène à partir de 2 photographies de la scène et de la connaissance des relations géométriques entre les points (physique ou virtuelle) de l'espace 3D et les positions dans les 2 images (ces relations constituant les calibrations des 2 photographies).

Ce problème est très étudié dans la littérature, mais les approches proposées sont souvent heuristiques. En effet, au vu de la complexité et des ambiguïtés intrinsèques du phénomène de photographie, le problème de la reconstruction 3D est intrinséquement mal posé. L'approche la plus classique consiste à faire un certain nombres d'hypothèses admises par la communauté (optique géométrique, opacité de la matière, caméras de type sténopé ou "pinhole", ...) , et, à déterminer heuristiquement des couples de points (1 dans chaque image) correspondant à des mêmes points 3D physique-matériel-plein (un point 3D par couple). Les positions de ces points 3D peuvent alors être déterminées, par les hypothèses du modèle, en calculant l'intersection des rayons correspondant aux points image associés. Malheureusement, le problème de déterminer l'ensemble maximal des couples de points image correspondant à un même point 3D est un problème encore moins bien posé que le problème de départ.

Dans cet article, on cherche à établir un modèle probabiliste simplifié du problème de la reconstruction 3D permettant de munir d'une probabilité l'ensemble des couples formés d'une géométrie et d'une observation. Ce modèle est basé sur des hypothèses relativement acceptées par la communauté de reconstruction 3D. De plus, ce modèle reste relativement proche des données observables (pas de marginalisation explicite de l'illumination de la scène...) et la détermination de la géométrie la plus probable compte tenu des observations se ramène à un problème d'optimisation. Ce problème d'optimisation est NP-complet en toute généralité, mais il reste potentiellement gérable car très proche d'un problème de flot et de plus il est simplement exprimable par un programme linéaire à

variable binaire, dans lequelle chaque variable à une signification sémantique forte.

Dans la suite de cet article, on introduit le contexte de cet article dans la section 2. Dans les section 3 et 4, on introduit notre modélisation probabiliste de la reconstruction 3D. Cette modélisation est constituée d'une partie basée sur des hypothèses admises par la communauté. Cette partie, développée dans la section 3, permet fixer les définitions, utilisées dans cet article, des concepts intrinsèques à la reconstruction 3D comme celui de géométrie. L'autre partie de notre modèle, développée dans la section 4, consiste à choisir une forme pour la probabilité qui relie les différentes grandeurs entre elle. Cette probabilité donne des valeurs quantitatives que seule l'expérience peut justifier sur des hypothèses qualitatives admises par la communauté. Dans la section 5, on montre comment le calcul de la surface la plus probable se ramène à un problème de flot harmonieux. On montre aussi que ce problème est NP-complet dans le cas général.

2 L'état de l'art en reconstruction 3D

Dans l'état de l'art, de nombreux algorithmes proposent des solutions au problème de la reconstruction 3D. Nous nous focaliserons ici sur les méthodes travaillant à partir d'images obtenues à l'aide de caméras calibrées (ie la relation géométrique entre l'espace 3D et le repère image de la caméra est connue).

Cette famille d'approches comporte deux catégories : les méthodes passives et les méthodes actives. Dans les méthodes actives, l'acquisition d'images est effectuée avec un éclairage structurant [1]. Cet éclairage permet d'assurer la présence de points de contrôle à la surface de l'objet, ce qui facilite l'étape suivante de reconstruction.

Dans les méthodes passives, on effectue une acquisition d'images sous un éclairage non contrôlé. Les méthodes passives sont elles-mêmes divisibles en de nombreuses sous-familles. Parmi ces sous-familles, on trouve les méthodes qui définissent la surface par le résultat d'une minimisation notamment par la technique dite des coupes de graphe [2, 3, 4]. La théorie des coupes de graphe a été développée dans [5, 6]. D'autres méthodes cherchent à reconstruire la géométrie à partir d'éléments connus : par déformation de modèles [7] par exemple, ou par assemblage de primitives 3D simples, telles que des disques [8] ou des rectangles [9]. Les méthodes volumétriques reposent sur une idée encore différente : il s'agit de produire une surface sous la forme d'un ensemble de voxels. Ces algorithmes ont en particulier connus un essor après [10] où S. M. Seitz et al. ont introduit la méthode de coloration des voxels (voxel coloring). Plus tard, la méthode de sculpture d'espace (Space Carving) [11] a proposé une approche itérative pour affiner photo par photo une surface initiale. Dans [12], une approche demandant moins de ressource machine, basée sur la hiérarchisation des voxels, est proposée. Certains algorithmes sont le résultat de combinaisons de plusieurs types de méthodes et proposent une chaine de traitement d'images. Par exemple, dans [13] V.H. Hiep et al proposent de reconstruire la géométrie d'un objet en trois étapes. Tout d'abord, un nuage de points dense est obtenu

par couplage de points d'intérêt. Puis, une première surface est créée à partir du nuage de points par minimisation d'un critère (coupes de graphe) combinant une mesure de photoconsistance et l'erreur de visibilité. Enfin, cette première surface est raffinée par un processus d'optimisation variationnel local. Cette dernière méthode illustre l'état de l'art de la reconstruction 3D : l'algorithme est efficace et produit des objets de grande qualité, mais il reste heuristique.

Il existe aussi des méthodes non heuristiques, basées sur une modélisation mathématique. La première modélisation mathématique de la reconstruction 3D est la théorie de l'approximation d'une surface par un échantillonnage fini. La méthode du Crust [14] fut la première méthode justifiant le bien fondé de la reconstruction 3D : le Delaunay restreint de l'échantillon produit par cette méthode, est capable (après post-traitement) d'approximer précisément la surface sous-jacente si cette surface est suffisamment régulière. Cependant, cette méthode suppose que l'échantillonnage de la surface soit peu bruité. Cela pose un problème pour les nuages de points 3D provenant de mesures laser ou a fortiori d'images. Dans [15], F. Chazal propose de modéliser les surfaces par des mesures pour être capable, grâce à la théorie de la mesure, de donner un cadre formel à la reconstruction de surfaces à partir d'un nuage de points bruité. Mais, la reconstruction 3D ne peut se résumer à l'approximation de surface par un échantillonnage. En effet, l'information contenue dans n'importe quelle vidéo est plus riche que celle d'un simple nuage de point. Une autre approche est de chercher à distinguer les zones visibles de celles qui ne le sont pas. De nombreux algorithmes de l'état de l'art se basent sur une modélisation de cette contrainte de visibilité comme dans [10] (Ordinal Visibility Constraint). Enfin, une approche qui nous intéresse particulièrement consiste à voir le problème d'un point de vue statistique. Par exemple, R. Bhotika et al. [17] formulent une théorie qui introduit la probabilité qu'un voxel soit plein ou vide. Cette approche statistique qu'on retrouve aussi dans [18] semble être la seule capable de tenir compte des ambiguïtés intrinsèques de la reconstruction 3D, ce que nous développerons dans la suite de cet article.

3 Caméras, facettes, géométrie et signal

Dans cet article, la reconstruction 3D est basée sur l'hypothèse de l'optique géométrique et de la caméra sténopé. Ces deux hypothèses impliquent qu'à un point dans une image est associé un rayon dans l'espace 3D (les rayons étant disjoints et formant une partition de l'espace visible).

Dans cet article, on fait l'hypothèse supplémentaire que les valeurs des différents pixels d'une image sont indépendantes (caméra parfaite). Ainsi, une caméra induit une partition en cône, finie et naturelle de l'espace visible. Chaque cône regroupe tous les rayons associés aux points image dans le pixel considéré.

Deux caméras induisent ainsi une nouvelle partition finie de l'espace simultannément visible en supperposant les partitions individuelles.

Dans cet article, on fait l'hypothèse que les caméras sont identiques et placées dans la configuration canonique (images rectifiées). Ainsi pour tous couples de

pixels (1 dans chaque image) : soit les cônes associés aux pixels se coupent formant une intersection finie, soit ils ne se coupent pas. De plus, dans le cas où l'intersection est finie, il existe une facette polygonale à 4 points, "équivalente" à l'intersection des cônes. Typiquement, si a, b, c, d et u, v, w, x sont les 4 points correspondant aux 2 pixels (en partant du coin en haut à gauche et en tournant dans le sens direct) alors le polygone à 4 points formé des intersections des rayons associés à $(a, u), (b, v), (c, w), (d, x)$ est la facette "équivalente" au bout d'espace compris dans l'intersection des cônes. Il est important de noter que cette équivalence entre facette et intersection de cônes n'est valable que dans le cas de la configuration canonique.

Ainsi l'espace simultannément visible est équivalent à un ensemble de facette. Une géométrie peut alors se voir comme un étiquettage "plein" ou "vide" des facettes.

Néanmoins, le processus de photographie est projectif : il ne permet pas d'avoir des informations sur ce qui n'est pas visible. Dans cet article, on fait l'hypothèse qu'une facette en cache une autre (à une des 2 caméras) si et seulement si elle est pleine, elle partage un pixel avec l'autre facette et elle vient avant l'autre dans la route du cône associé au pixel. Cela revient à faire l'hypothèse qu'une facette pleine est opaque et qu'une facette vide est transparente. Cela suppose aussi que l'espace non simultannément visible est vide et transparent.

Cet modélisation n'est pas nouvelle : elle ne provient que des hypothèses d'optique géométriques de sténopé et de l'indépendance des valeurs des pixels. Mais ces hypothèses ne modélisent pas le processus de photographie lui-même. Dans cet article, on formule les hypothèses supplémentaires suivantes. La valeur d'un pixel correspond à la mesure d'un signal provenant d'une des facettes associées au rayon correspondant au pixel (précisément, de la première facette pleine et donc "visible pour cette caméra"). La mesure du signal est formée d'un spectre couleur et d'une intensité. Le spectre couleur ne dépend que la matière composant la facette. L'intensité dépend de la facette considéré et de "l'angle du cône" par rapport à la normale à la facette.

Enfin, on suppose que tout pixel rencontre dans l'espace simultannément visible une facette pleine. Cela force l'existence dans facettes pleines.

Ces hypothèses permettent de définir des objets adaptés au problème de la reconstruction 3D. Mais elle ne quantifie pas les intéractions entre ces objets.

4 Modèle probabiliste

4.1 Le modèle exprimé via des objets 3D

Pour modéliser les interactions entre les objets, on se base sur les hypothèses qualitatives suivantes, communément admises dans la littérature.

Il est peu probable que la composition spectrale des 2 signals incidents au couple de pixels associés à une facette pleine et visible (on omettra de préciser "simultannément") soient différents.

Il est peu probable que la géométrie soit irrégulière à la fois du point de

vue de la présence de matière mais aussi du point de vue de la couleur. Plus précisément, il est peu probable que la géométrie observé ait un grand nombre de bord, que les facettes pleines aient des formes très applatie, et il est peu probable que des facettes adjacentes fassent un angles fort. De plus il est peu probable que deux facettes adjacentes émettent un signal ayant un composition spectrale différente.

Enfin, il est peu probable que la distribution d'illumination soit irrégulière. Ainsi, si deux facettes sont adjacentes, il est peu probable que les illuminations mesurées soient différentes (l'illumination au niveau d'une facette pouvant être mesuré à partir de l'intensité reçu par chacun des pixels correspondant à la facette : cette intensité est une mesure de l'illumination multipliée par le cosinus de l'angle entre le cône correspondant au pixel et la normale à la facette).

Dans cet article, on fait l'hypothèse que toutes ces probabilités s'exprime selon un modèle exponentiel. Ainsi la probabilité de la situation, $\Gamma, \chi, \phi, \rho, \varrho, \varphi, \Upsilon, \eta$ où Γ représente l'ensemble des facettes ;

χ représente le fait que la facette soit pleine ou vide (avec la contrainte que tout cône rencontre une facette) ;

ϕ représente la couleur de la facette, φ représente la ou les couleurs mesurée(s) (selon que la facette est visible ou seulement partiellement visible) ;

ρ représente l'intensité de la facette, ϱ représente la ou les intensité mesurée(s) (après les corrections dues aux angles) ;

Υ représente les couples de facettes adjacentes ;

η représente la longueur du côté commun des deux facettes ;

est donnée (à une normalisation près) par

$$\begin{aligned}
&\prod_{f\in\Gamma} \exp\left(-\chi(f)\left(\phi(f)-\varphi(f)\right)^2\right)\\
&\times\prod_{f\in\Gamma} \exp\left(-\alpha\chi(f)\left(\rho(f)-\varrho(f)\right)^2\right)\\
&\times\prod_{f,g\in\Upsilon} \exp\left(-\beta\left|\chi(f)-\chi(g)\right|\eta(f,g)\right)\\
&\times\prod_{f,g\in\Upsilon} \exp\left(-\gamma\chi(f)\chi(g)\left(\phi(f)-\phi(g)\right)^2\right)\\
&\times\prod_{f,g\in\Upsilon} \exp\left(-\delta\chi(f)\chi(g)\left(\rho(f)-\rho(g)\right)^2\right)
\end{aligned}$$

avec $\alpha, \beta, \gamma, \delta$ des paramètres inconnus. Les premiers termes représentent la probabilité associée aux bruits des mesures. Le troisième terme représente la faible probabilité d'avoir une surface possédant beaucoup de discontinuité. Le quatrième et le cinquième terme représente la faible probabilité d'avoir de brusques écarts de couleur ou de luminosités entre des facettes connexes.

4.2 Le modèle exprimé via des objets 2D

Un des intérêts de ce modèle est qu'il peut s'exprimer (via de légères approximations) en terme d'objet 2D : il y a une bijection entre les ensembles de facettes à la fois pleines et visibles et les couplages des pixels.

Soit I et J l'ensemble des pixels dans l'image 1 et 2, soit Γ le sous-ensemble de $I \times J$ des couples de pixels associé à une facette. Soit $V = \{s,t\} \cup I \cup J$ et $E = \{s\} \times I \cup J \times \{t\} \cup \Gamma$ alors un $s-t$ flot entier dans le graphe (V,E) (où toutes les arrêtes ont une capacité de 1) définit une géométrie vue comme un étiquetage "plein et visible" ou "vide ou caché". On notera F l'ensemble des flots entiers.

Ici le flot vide ne correspond pas la surface vide mais à une surface où aucun pixel n'est visible (simultannément) c'est à dire une surface où tous les pixels ne sont visibles que dans une seule des images.

Ainsi, si $(i,j) \in I \times J$ sont des pixels et les arcs $(s,i),(i,j),(j,t)$ sont saturés cela signifie que i a trouvé un couplage, j a trouvé un couplage, i et j sont couplés c'est à dire que la facette correspondante est visible et pleine. Donc cela signifie que le spectre du signal reçu par i et par j devrait être les mêmes et cela signifie que le rapport des intensités devrait se déduire des angles.

Réciproquement si les arcs $(s,i),(i,j),(j,t)$ sont vides cela signifie que i et j correspondante à 2 facettes pleines mais non simultannément visibles. Ainsi, si un arcs (s,i) ou (j,t) est vide cela représente le fait que le pixel i ou j n'est visible que dans une seule image. Donc cela signifie que les irrégularités au niveau du voxel i doivent être comptées 1 fois pour i et 1 fois pour j, alors qu'elle ne compte que pour 1 pour i et j s'ils sont couplés (c'est ce qui fait que le flot vide n'est pas optimal).

Ainsi le modèle se réecrit en terme d'un couplage de pixel. Soit $e(i,j)$ l'erreur entre les 2 spectres et l'erreur entre les 2 intensités mesurés entre i et j. Soit $e(k)$ l'irrégularité intrinsèque lié à toute surface causant la structure autour du pixel k. Soit η donnant la longueur du coté commun au pixel générant une facette adjacente (l'ensemble des tel couples est noté Υ). La probabilité $P(f)$ du flot $f \in F$ s'écrit (à une normalisation près) :

$$\begin{aligned} &\prod_{v \in I \cup J} \exp\left(-e(i)\left(1-f(i)\right)\right) \\ &\times \prod_{(i,j)\in\Gamma} \exp\left(-e(i,j)\, f(i,j)\right) \\ \times \prod_{((i,j),(a,b))\in\Upsilon} &\exp\left(-\lambda\eta\left((i,j),(a,b)\right)\left|f(i,j)-f(a,b)\right|\right) \end{aligned}$$

où $\lambda > 0$ est un paramètre inconnu (ici $e > 0$).

5 Surface la plus probable

Étant donné le modèle que l'on se donne, la solution du problème de reconstruction 3D est un flot qui maximise la probabilité de la surface.pour résoudre l'optimisation combinatoire permetant La confiance envers cette solution est la probabilité maximale que peut avoir une surface.

Ce problème d'optimisation peut s'écrire sous la forme d'un programme linéaire en nombre entier (via $-\log(P(f))$) :

$$\arg\min_{f\in F}\left(\sum_{\alpha\in I\cup J\cup\Gamma} e(\alpha) f(\alpha) + \lambda \sum_{(\alpha,\beta)\in\Upsilon} \eta(\alpha,\beta)\left|f(\alpha) - f(\beta)\right|\right)$$

où e désigne les erreurs ou leurs opposés (ici, on peut avoir des valeurs négatives pour e ce qui fait que le flot vide n'est pas optimal) et $\lambda > 0$ est un paramètre inconnu.

Malheureusement ce problème de "flot harmonieux" est dans le cas général NP-complet. En effet, même la forme suivante est NP-complete. Soit (V, E) un graphe avec $s, t \in V$. Soit c une fonction réelle sur E. Soit $\Upsilon \subset E \times E$ Soit F l'ensemble des $s-t$ flot entier dans le graphe (V, E) quand toute les arrêtes ont une capacité de 1. Soit $\lambda \geq 0$. On cherche à résoudre

$$\arg\min_{f\in F}\left(\sum_{e\in E} c(e) f(e) + \lambda \sum_{(e,\xi)\in\Upsilon} \left|f(e) - f(\xi)\right|\right)$$

En effet, on peut ramener le problème "1 dans 3 SAT" à un problème de "flot harmonieux".

Soit une instance $v_1, ...v_n, C_1, ..., C_m$ de "1 dans 3 SAT". Soit le graphe (V, E) avec

$$V = \{s, t\} \cup \{0, 1, 2\} \times \{v_1, ..., v_n\} \cup \{0, 1, 2, 3\} \times \{C_1, C_m\}$$

$$\begin{aligned} E = \{s\} \times \{0\} \times \{v_1, ..., v_n\} &\cup \{((0, v_1), (1, v_1)), ..., ((0, v_n), (1, v_n))\} \\ &\cup \{((0, v_1), (2, v_1)), ..., ((0, v_n), (2, v_n))\} \\ &\cup \{C_1, ..., C_m\} \times \{0\} \times \{t\} \\ &\cup \{((1, C_1), (0, C_1)), ..., ((1, C_m), (0, C_m))\} \\ &\cup \{((2, C_1), (0, C_1)), ..., ((2, C_m), (0, C_m))\} \\ &\cup \{((3, C_1), (0, C_1)), ..., ((3, C_m), (0, C_m))\} \end{aligned}$$

Soit $c = -\chi_{\{C_1,...,C_m\}\times\{0\}\times\{t\}}$. Soit Υ le plus petit ensemble tel que pour toutes les clauses C si v_i est le j-ème littéral de C alors l'arc $((j, C), (0, C))$ est relié à $((0, v_i), (1, v_i))$ et si $\neg v_i$ est le j-ème littéral de C alors l'arc $((j, C), (0, C))$ est relié à $((0, v_i), (2, v_i))$.

Alors quelque soit $\lambda > 0$ l'existance d'un flot amenant un résultat de m est équivalent à l'existance d'une solution pour l'instance de "1 dans 3 SAT".

6 Conclusion

Dans cet article on a introduit une modélisation probabiliste du problème de la reconstruction 3D basé sur des hypothèses qualitatives communément admises dans le littérature. Cette modélisation à l'avantage de pouvoir rester très proche des données mesurées, de ramener le calcul de la surface la plus probable à un problème d'optimisation combinatoire et de fournir en plus d'une surface, une carte de probabilité sur l'ensemble des surfaces. Dans de futurs travaux nous vérifierons expérimentalement la cohérence entre le modèle et la réalité. Nous chercherons aussi une approximation polynomiale du calcul de la surface la plus probable.

Références

[1] Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, Yuya Ohta, Kazuhiro Sakashita, Ryota Zushi, Yasushi Yagi, Naoki Asada, Linear solution for oneshot active 3D reconstruction using two projectors, 3DPVT, 2010

[2] Vladimir Kolmogorov, Ramin Zabih, Computing visual correspondence with occlusions using graph cuts, ICCV, 2001

[3] G. Vogiatzis, P. Torr, R. Cippola, Multi-view Stereo via Volumetric Graph-Cuts, Proceedings of the International Conference on Computer Vision and Pattern Recognition, 2005

[4] S. Tran, L. Davis, 3D Surface Reconstruction Using Graph Cuts with Surface Constraints, Proceedings of the European Conference on Computer Vision, 2006

[5] Y. Boykov, O. Veksler, R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001

[6] V. Kolmogorov, Ramin Zabih, What Energy Functions Can Be Minimized via Graph Cuts ?, Proceedings of the European Conference on Computer Vision, 2002

[7] C. Esteban, F. Schmitt, Silhouette and Stereo Fusion for 3D Object Modeling, Computer Vision and Image Understanding, 2004

[8] M. Habbecke, L. Kobbelt, A Surface-Growing Approach to Multi-View Stereo Reconstruction, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007

[9] Y. Furukawa, J. Ponce, Accurate, Dense, and Robust Multi-View Stereopsis, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007

[10] Steven M. Seitz, Charles R. Dyer, Photorealistic Scene Reconstruction by Voxel Coloring, CVPR, 1997

[11] K. N. Kutulakos and S. M. Seitz, A Theory of Shape by Space Carving, International Journal of Computer Vision, 2000

[12] R. Guerchouche, O. Bernier, T. Zaharia, Reconstruction Volumétrique Multirésolution d'Objets 3D, RFIA, 2008

[13] Vu Hoang Hiep, Renaud Keriven, Patrick Labatut, Jean-Philippe Pons, Towards high-resolution large-scale multi-view stereo, CVPR 2009

[14] N. AMENTA,M. BERN, M. KAMVYSSELIS, A new voronoibased surface reconstruction algorithm, SIGGRAPH, 1998

[15] F. Chazal, Geometric inference for probability measures : extracting robust geometric information from noisy data, Journées STAR, 2010

[16] Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin, Bayesian Data Analysis, 1993

[17] R. Bhotika, D. J. Fleet, K. N. Kutulakos, A Probabilistic Theory of Occupancy and Emptiness, Proceedings of the European Conference on Computer Vision, 2002

[18] S. Liu, D. B. Cooper,A Complete Statistical Inverse Ray Tracing Approach to Multi-View Stereo, CVPR, 2011