

Reconstruction 3D du bâti par la technique des ombres chinoises

Pavel LUKASHEVISH¹, Aliaxandr KRAUSHONAK¹, Xavier DESCOMBES²
Jean-Denis DUROU³⁻⁴, Boris ZALESSKY¹, Elena ZHIZHINA⁵

¹UIIP, Minsk, Biélorussie ²EPI Ariana, INRIA Sophia-Antipolis Méditerranée

³CMLA, ENS Cachan, CNRS, Universud, Cachan

⁴IRIT, Université Paul Sabatier, Toulouse ⁵IITP, Moscou, Russie

pavel_by@mail.ru, alpha_storm@mail.ru, Xavier.Descombes@sophia.inria.fr

Jean-Denis.Durou@irit.fr, zalesky.boris.a@gmail.com, ejj@iitp.ru

Résumé – Dans cet article, nous effectuons la reconstruction 3D du bâti à partir d’images satellitaires. Nous nous écartons de l’approche classique qui consiste à considérer le problème comme un problème inverse qui doit être résolu en utilisant deux images stéréoscopiques et en les mettant en correspondance. Au contraire, nous montrons qu’il est plus simple de résoudre le problème directement par essais/erreurs comme on le fait très naturellement pour produire des ombres chinoises. Nous montrons notamment que cette approche ne nécessite pas un temps de calcul considérable, grâce à la puissance de calcul d’OpenGL qui s’appuie sur une carte graphique. Les premiers résultats obtenus montrent la pertinence de l’approche adoptée. En particulier, elle permet de résoudre des ambiguïtés pour lesquelles l’inversion du problème serait quasiment impossible.

Abstract – In this paper, we consider the 3D building reconstruction from satellite images. We avoid the classical approach consisting in considering the problem as an inverse problem which must be solved by performing a matching on a stereo pair of images. On the contrary, we show that it is easier to solve the problem directly by trial and errors as it is naturally done when using magic shadows. In this context, we show that this approach can be affordable in term of computation time thanks to the freeware libray OpenGL lying on the PC graphic card. The early obtained results show that the proposed approach is relevant. In particular, it allows solving ambiguities for which the corresponding inverse problem would be hardly solved.

1 Introduction

Nous considérons le problème de la reconstruction 3D du bâti à partir d’image(s) de télédétection. Ce problème est classiquement perçu comme un problème inverse puisqu’il s’agit de retrouver la scène qui, à travers le système optique, a produit les images. La technique la plus souvent utilisée est la stéréoscopie. Sous réserve de posséder un couple d’images sous des angles différents, il est possible de reconstruire la troisième dimension. Le problème consiste alors essentiellement à mettre en correspondance les deux images [2, 3, 4]. Les ombres ou l’ombrage peuvent également donner une information utile [5, 6].

Dans cet article, nous montrons qu’il est possible de résoudre le problème direct. L’idée consiste à proposer des configurations de bâtiments, puis à simuler les images résultant de ces configurations. Un critère est optimisé pour estimer la solution. C’est le principe des ombres chinoises qui consiste, non pas à calculer l’objet recherché à partir de l’ombre [1], mais à tester et modifier différentes configurations des mains de sorte que l’ombre produite corresponde à celle de l’objet recherché. Une telle approche est séduisante car elle s’affranchit des difficultés des approches inverses, comme par exemple dans le cas où l’ombre d’un bâtiment se projette sur un autre bâtiment. Des exemples de l’approche directe, fondés sur une modélisation par processus ponctuels marqués, ont montré son efficacité pour

la reconnaissance d’objets géométriques simples tels que des ellipses ou des rectangles [7, 8]. Néanmoins, la pertinence d’une approche directe peut être mise en doute en raison du temps de calcul. Comment générer des hypothèses et quel temps de calcul est nécessaire à la projection de la scène sur le(s) plan(s) image(s) sont deux questions cruciales. Nous adoptons ici une démarche simple pour la génération d’hypothèses consistant en un algorithme de type glouton. Un objet est modifié aléatoirement à chaque itération et ce changement est accepté si et seulement s’il améliore la configuration au sens du critère adopté. Nous montrons néanmoins qu’il est réaliste d’adopter une approche directe, en nous appuyant sur la puissance de calcul de la carte graphique et le logiciel libre OpenGL pour effectuer la synthèse des images obtenues à partir d’une configuration de bâtiments.

2 Méthode

Dans cette première étude, nous considérons des bâtiments rectangulaires à toit bi-plan symétrique (voir figure 1). Un bâtiment est donc défini par sept paramètres $(x, y, \theta, L, l, H, h) \in \mathcal{B}$ où (x, y) est le centre du bâtiment (qui doit se situer dans l’image), $\theta \in [0, \pi]$ son orientation, $L \in [L_{\min}, L_{\max}]$ sa longueur, $l \in [l_{\min}, l_{\max}]$ sa largeur, $H \in [H_{\min}, H_{\max}]$ la hauteur de ses murs au bas du toit,

et $h \in [h_{\min}, h_{\max}]$ la hauteur de ses murs au faîte du toit. Dans cette première étude, nous limitons les valeurs possibles des six paramètres (x, y, L, l, H, h) à \mathbb{Z} , l'unité de longueur étant le pixel.

Nous considérons $\mathcal{I}_i, i = 1, \dots, p$, un ensemble d'images associées aux projections π_i supposées connues de la scène, et une fonction énergie, à minimiser :

$$\begin{cases} E : \mathcal{B}^N & \rightarrow \mathbb{R} \\ B = \{b_1, \dots, b_N\} & \mapsto E(B) \end{cases} \quad (1)$$

où chaque $b_i \in \mathcal{B}$ est un bâtiment et N est le nombre, supposé connu, de bâtiments.

L'énergie est composée de la somme d'un terme, dit d'attache aux données, et d'un a priori sur les bâtiments voisins :

$$E(B) = \sum_{i=1}^p D(\pi_i(B), \mathcal{I}_i) + \sum_{u \sim v} F(b_u, b_v). \quad (2)$$

Le terme d'attache aux données est la somme de distances entre les différentes projections de B , correspondant aux prises de vues des différentes images, et les données image correspondantes.

Dans ce premier travail, nous considérons un a priori hautement répulsif qui consiste à interdire tout recouvrement entre les différents bâtiments :

$$b_u \sim b_v \Leftrightarrow b_u \cap b_v \neq \emptyset, \quad F(b_u, b_v) = \infty. \quad (3)$$

En ce qui concerne l'attache aux données, nous travaillons ici à partir d'une seule image \mathcal{I} . Nous considérons deux critères différents : soit la différence symétrique des zones d'ombre de la donnée et de l'image projetée de la configuration proposée (critère C_1), soit le carré de la distance euclidienne entre la donnée et l'image projetée de la configuration proposée (critère C_2). Pour définir le critère C_1 , nous notons $O_B = \cup_{i=1, \dots, N} O(b_i)$ l'ensemble des ombres des objets de la configuration B et $O_I = \{s \in S : i(s) < o\}$, l'ensemble des pixels de S dont le niveau de gris est inférieur à o , nous avons alors :

$$D(\pi(B), \mathcal{I}) = O_B \Delta O_I = (O_B \cap O_I^c) \cup (O_I \cap O_B^c) \quad (4)$$

Pour le critère C_2 , nous définissons quatre niveaux de gris de référence, g_o, g_t, g_f et g_s , correspondant respectivement aux ombres, aux toits, aux murs et au sol. La projection $\pi(B)$ de la configuration B sur la trame discrète S est alors à valeurs dans $\{g_o, g_t, g_f, g_s\}$ et nous avons pour le critère C_2 :

$$D(\pi(B), \mathcal{I}) = \sum_{s \in S} (\pi(B)(s) - \mathcal{I}(s))^2. \quad (5)$$

Les différents paramètres radiométriques sont supposés connus dans cette première étude.

L'initialisation est effectuée en tirant aléatoirement N éléments de \mathcal{B} correspondant à N bâtiments disjoints. À chaque itération, nous tirons aléatoirement un nouvel élément de \mathcal{B} et considérons une configuration composée de la configuration courante dans laquelle un des bâtiments, choisi aléatoirement, est remplacé par ce nouvel élément. Nous conservons cette nouvelle configuration si les N bâtiments sont disjoints et si elle fait décroître le critère utilisé.

Cet algorithme est glouton, mais s'avère suffisant pour montrer la pertinence de l'approche, dans la mesure où les minima locaux des critères utilisés ne constituent pas des pièges. Nous utilisons un critère d'arrêt très rudimentaire qui consiste à stopper l'algorithme lorsqu'un nombre suffisant de propositions successives ont été refusées, c'est-à-dire n'ont pas fait décroître l'énergie.

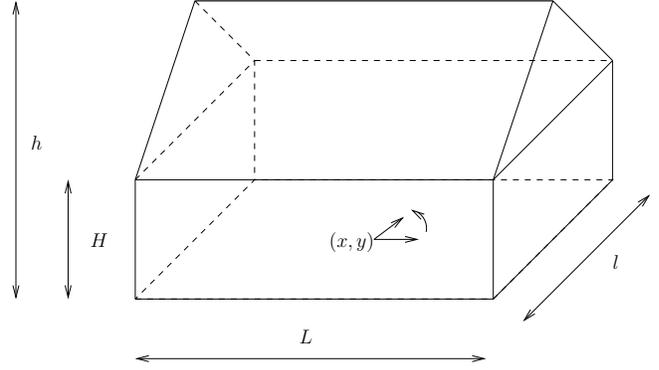


FIG. 1 – Bâtiment à toit bi-plan symétrique et ses paramètres.

3 Résultats

Nous présentons deux séries de résultats. Sur la figure 2-(a), la donnée est une image de synthèse représentant deux bâtiments situés de telle sorte que l'ombre de l'un se projette partiellement sur l'autre. Dans un tel cas de figure, il semble très difficile de reconstruire les bâtiments par résolution du problème inverse. En utilisant la technique des ombres chinoises, c'est-à-dire en essayant de minimiser le critère C_1 , nous obtenons une séquence de reconstructions. L'image de la figure 2-(b) montre la reconstruction obtenue en 61072 itérations. L'image de la figure 2-(c) montre la reconstruction obtenue en 15532 itérations avec le critère C_2 . Le résultat est manifestement meilleur en utilisant C_2 comme critère plutôt que C_1 , et qui plus est pour un nombre d'itérations moindre : 15532 au lieu de 61702. Cette dernière observation étant qualitative, nous avons jugé utile de la conforter par une évaluation quantitative des reconstructions obtenues avec les deux critères utilisés. Le tableau 1 récapitule les valeurs des sept paramètres de chaque bâtiment, pour les trois images de la figure 2. On remarque, entre autres, que le bâtiment de droite est parfaitement reconstruit en utilisant le critère C_2 . Ce résultat est prévisible puisque le critère C_2 est fondé sur une information plus riche que le critère C_1 . En effet, le critère C_1 n'utilise qu'une modélisation radiométrique des ombres alors que le critère C_2 exploite également une modélisation radiométrique du sol, des toits et des murs.

La figure 3 montre une deuxième série de résultats obtenus en utilisant comme donnée une image réelle d'une zone pavillonnaire de la région parisienne. À notre grande surprise, et bien que cela ne puisse pas être confirmé par une évaluation numérique des résultats, faute de disposer de la vérité terrain, la reconstruction obtenue en utilisant le critère C_1 est clairement meilleure que celle qui est ob-

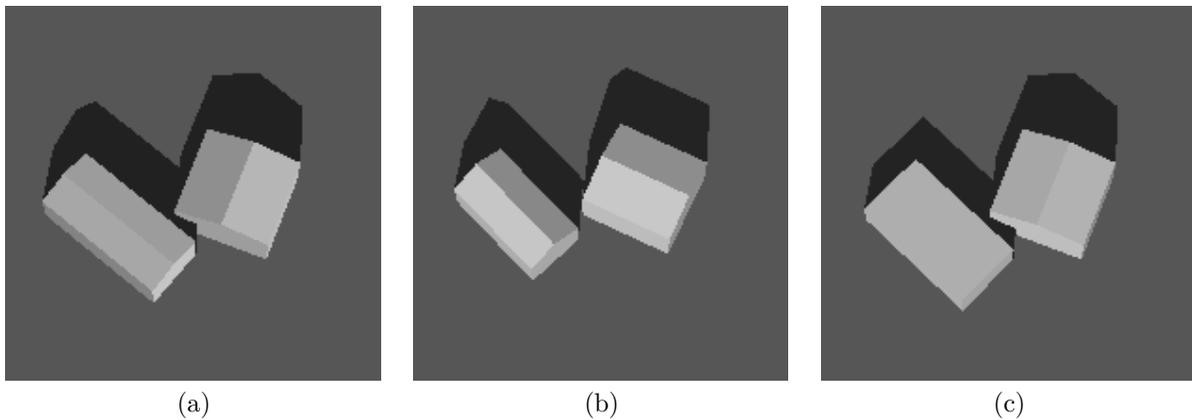


FIG. 2 – (a) Donnée de synthèse. (b) Reconstruction obtenue en 61702 itérations avec le critère C_1 . (c) Reconstruction obtenue en 15532 itérations avec le critère C_2 . Pour des données de synthèse, nous constatons empiriquement que le critère C_2 converge plus vite que C_1 et qu’il converge vers une meilleure reconstruction.

TAB. 1 – Paramètres des bâtiments de gauche (indice g) et de droite (indice d) correspondant aux trois images de la figure 2.

	x_g	y_g	θ_g	L_g	l_g	H_g	h_g	x_d	y_d	θ_d	L_d	l_d	H_d	h_d
Vérité terrain	60	76	0,875	18	39	27	33	124	95	2,774	27	25	29	36
Critère C_1	54	84	0,752	17	31	24	34	124	98	1,124	24	26	30	40
Critère C_2	62	73	0,805	20	37	22	22	124	95	2,774	27	25	29	36

tenue en utilisant le critère C_2 , et ce pour des nombres d’itérations comparables. Une validation plus approfondie de ce constat semble nécessaire, mais on peut d’ores et déjà avancer l’explication suivante : dans la mesure où les zones d’ombre sont radiométriquement très homogènes, il est normal que le critère C_1 soit plus robuste à toutes les sources d’erreur provenant des écarts au modèle élémentaire de bâtiment que nous avons utilisé. Le modèle radiométrique constant pour le sol est trop rudimentaire et pénalise le critère C_2 . Il pourrait donc s’avérer que la technique des ombres chinoises, bien que naïve, soit une piste très prometteuse pour la reconstruction 3D de bâtiments à partir d’image(s) de télédétection. Une énergie d’attache aux données locale, fondées sur les caractéristiques radiométriques des objets, s’avère plus robuste qu’un terme global, fondé sur une métrique entre deux images. Dans cet article, nous avons uniquement utilisé l’ombre comme caractéristique locale liée à l’objet “bâtiment”. Il est tout à fait envisageable d’étoffer le modèle par d’autres caractéristiques locales, comme par exemple la modélisation d’un fort gradient perpendiculairement aux arêtes des toits.

4 Conclusion

L’objectif de cet article est de montrer qu’il n’est pas illusoire de traiter le problème de la reconstruction 3D de bâti comme un problème direct. En effet, l’utilisation intensive d’une carte graphique performante via des outils tels que OpenGL rend cet objectif atteignable. Même si nous traitons ici d’image(s) très petites, nous parions qu’il est possible de traiter de scènes beaucoup plus gran-

dea, en un temps raisonnable. Pour cela, il faut bien évidemment proposer une approche d’optimisation plus efficace que l’algorithme glouton étudié ici. De récents développements montrent que de telles méthodes d’optimisation sont accessibles.

Notons que l’approche directe possède de nombreux avantages par rapport aux classiques problèmes inverses. En effet, le caractère mal posé de la plupart des problèmes inverses rend l’algorithmie nécessaire à leur résolution complexe. Dans notre cas, la résolution par problème inverse de la reconstruction 3D de bâti à partir de la seule ombre, est impossible dès que les ombres de bâtiments se projettent sur d’autres bâtiments ou sur d’autres ombres. Bien entendu, dans le cas d’un couple stéréoscopique, le problème est moins ardu. Néanmoins, les occultations sont un problème majeur. Une boucle analyse/synthèse consistant à générer des hypothèses, à synthétiser les données qu’auraient produites ces hypothèses pour les comparer aux données réelles semble donc une approche séduisante. Ce travail préliminaire a pour ambition de montrer la faisabilité d’une telle approche.

Pour aller plus loin, deux voies restent donc à investiguer. La première concerne l’énergie. Un a priori plus structuré pourra être pris en compte en fonction du type de scène (urbaine, pavillonnaire, ...). Le terme d’attache aux données peut également être affiné, par exemple par la prise en compte d’un terme de gradient sur les bords des toits. Pour finir, le schéma d’optimisation utilisé pour cet article est très simple, mais peu efficace. En outre, dans un cadre opérationnel, le nombre de bâtiments n’est pas connu. L’étape suivante va donc consister à adapter la nouvelle dynamique de naissances et morts multiples,

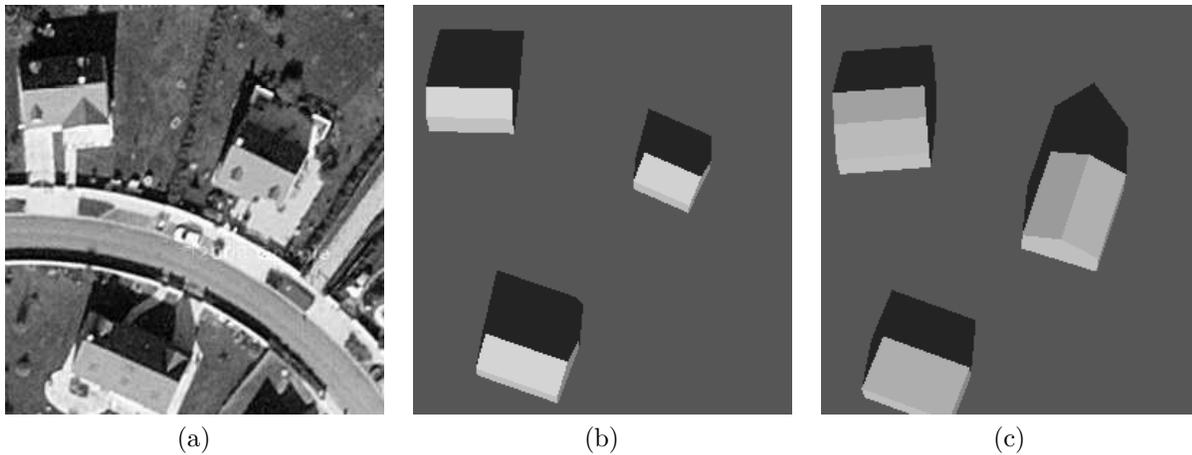


FIG. 3 – (a) Donnée réelle. (b) Reconstruction obtenue en 28873 itérations (critère C_1). (c) Reconstruction obtenue en 25822 itérations (critère C_2). Pour des données réelles, en revanche, nous constatons que le critère C_1 converge vers une meilleure reconstruction que le critère C_2 , en des temps de calcul comparables.

proposée dans [9], à notre contexte.

Références

- [1] X. Huang et L.K.. Kwoh. *3D Building Reconstruction and Visualization for Single High Resolution Satellite Image*. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Barcelone, Espagne, juillet 2007, pages 5009–5012.
- [2] C. Baillard et A. Zisserman. *Automatic reconstruction of piecewise planar models from multiple views*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Corfou, Grèce, juin 1999, pages 559–565.
- [3] F. Taillandier et R. Deriche. *Automatic Buildings Reconstruction from Aerial Images : a Generic Bayesian Framework*. Proceedings of the ISPRS Congress, Istanbul, Turquie, juillet 2004.
- [4] F. Lafarge, X. Descombes, J. Zerubia et M. Pierrot-Deseilligny. *Building reconstruction from a single DEM*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, États-Unis, juin 2008.
- [5] J. McGlone et J. Shufelt. *Projective and object space geometry for monocular building extraction*. Rapport technique, Carnegie Mellon University, mars 1994.
- [6] C. Lin et R. Nevatia. *Building detection and description from a single intensity image*. Computer Vision and Image Understanding 72(2) :101-121, 1998.
- [7] M. Ortner, X. Descombes et J. Zerubia. *A marked point process of rectangles and segments for automatic analysis of Digital Elevation Models*. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(1) :105-119, janvier 2008.
- [8] S. Descamps, X. Descombes, A. Béchet et J. Zerubia. *Automatic flamingo detection using a multiple birth and death process*. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, États-Unis, mars-avril 2008, pages 1113-1116.
- [9] X. Descombes, R. Minlos et E. Zhizhina. *Object Extraction Using a Stochastic Birth-and-Death Dynamics in Continuum*. Journal of Mathematical Imaging and Vision 33(3) :347-359, mars 2009.