



HAL
open science

Reconstruction 3D du bâti à partir d'une seule image par naissances et morts multiples

Jean-Denis Durou, Xavier Descombes, Pavel Lukashevish, Alianxandr
Kraushonak

► **To cite this version:**

Jean-Denis Durou, Xavier Descombes, Pavel Lukashevish, Alianxandr Kraushonak. Reconstruction 3D du bâti à partir d'une seule image par naissances et morts multiples. GRETSI, Sep 2011, Bordeaux, France. inria-00625527

HAL Id: inria-00625527

<https://inria.hal.science/inria-00625527>

Submitted on 21 Sep 2011

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.

Reconstruction 3D du bâti à partir d'une seule image par naissances et morts multiples

Jean-Denis DUROU¹, Xavier DESCOMBES²
Pavel LUKASHEVISH³, Aliaxandr KRAUSHONAK³

¹IRIT, Université Paul Sabatier, Toulouse

²INRIA CRI-SAM, Sophia Antipolis

³UIIP, Minsk, Biélorussie

durou@irit.fr, Xavier.Descombes@inria.fr
pavelby@mail.ru, alphastorm@mail.ru

Résumé – Dans cet article, nous nous écartons de l'approche classique qui considère la reconstruction 3D comme un problème inverse et la résout en mettant en correspondance deux images d'une paire stéréoscopique. Au contraire, nous montrons qu'il est plus simple de résoudre le problème direct. Pour ce faire, nous proposons aléatoirement des configurations de bâtiments pour ne conserver que les plus pertinentes par un algorithme de type naissances et morts multiples. Nous montrons notamment que cette approche ne nécessite pas un temps de calcul prohibitif, grâce à la puissance de calcul d'OpenGL qui s'appuie sur la 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, contrary to the classical approach addressing the 3D reconstruction problem as an inverse problem and solving it by matching two images from a stereoscopic pair, we show that we can solve the direct problem in a simpler way. To do so, we randomly propose configurations of buildings while keeping only the most relevant ones, using a multiple births and deaths algorithm. Notably, we show that this approach does not imply a prohibitory computation time, thanks to the freeware OpenGL which exploits the graphic card. The first results show that the proposed approach is relevant. In particular, it allows solving ambiguities for which inverting the problem is almost impossible.

1 Introduction

Nous considérons le problème de la reconstruction 3D du bâti à partir d'images 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 [1, 2]. Les ombres ou l'ombrage peuvent également donner une information utile [3, 4].

Dans cet article, nous montrons qu'il est possible de résoudre le problème direct à partir d'une seule image. 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. Une telle approche est séduisante car elle s'affranchit des difficultés des approches inverses dues au caractère mal posé du problème. L'utilisation des ombres permet de retrouver la hauteur des bâtiments par essais/erreurs [5], alors que l'information de gradient permet de bien positionner les contours des bâtiments. Des exemples de l'approche directe, fondés sur une modélisation par processus ponctuels marqués, ont montré son efficacité pour la reconnaissance d'ob-

jets géométriques simples tels que des ellipses ou des rectangles [6]. Nous proposons ici une telle approche pour la reconstruction 3D du bâti. Néanmoins, la pertinence d'une approche directe peut être mise en doute en raison du temps de calcul. Comment générer suffisamment d'hypothèses et quel temps de calcul est nécessaire à la projection de la scène sur le plan image sont deux questions cruciales. Nous adoptons ici une démarche simple pour la génération d'hypothèses consistant en une configuration aléatoire de bâtiments suivie par une phase de sélection des objets les plus pertinents. Ce principe est décliné par un algorithme de type naissances et morts multiples, initialement proposé dans [7]. Nous montrons que cette approche, combinée à la puissance de calcul de la carte graphique et au logiciel libre OpenGL pour effectuer la synthèse des images correspondant aux différentes configurations de bâtiments, est réaliste en terme de temps de calcul.

2 Méthode

Nous considérons des bâtiments rectangulaires à toit bi-plan symétrique. Un bâtiment b est donc décrit par sept paramètres $b = (x, y, v_x, v_y, r, H, h) \in \mathcal{B}$, où (x, y) est le centre du bâti-

ment (qui doit se situer dans l'image), $v = (v_x, v_y)$ le vecteur représentant la demi-longueur du bâtiment, r le rapport entre largeur et longueur, H la hauteur des murs au bas du toit, et h la hauteur du toit. Pour une scène à n bâtiments, l'espace des configurations possibles est \mathcal{B}^n . L'espace complet des configurations est donc $\Omega = \emptyset \cup \bigcup_{n=1}^{+\infty} \mathcal{B}^n$. Nous définissons un processus ponctuel dans cet espace par l'intermédiaire d'une densité de Gibbs relativement à la mesure d'un processus de Poisson, ce qui nous permet de formuler la modélisation par une énergie $U(B) = U(\{b_1, \dots, b_n\})$. Cette énergie s'écrit comme la somme d'une énergie a priori $U_a(B)$, pénalisant les intersections entre bâtiments voisins, et d'une énergie d'attache aux données $U_d(B)$, calant les ombres et les contours des bâtiments de la configuration sur ceux de l'image. Nous avons alors :

$$U(B) = U_a(B) + U_d(B) = \lambda \sum_{(i,j)} \delta(b_i \cap b_j \neq \emptyset) + \sum_i \max(u_o(b_i), u_c(b_i)) \quad (1)$$

où $u_o(\cdot)$ représente l'adéquation des ombres et $u_c(\cdot)$ celle des contours. Pour définir ces énergies, nous effectuons un prétraitement pour obtenir une carte des ombres O et un champ de gradient du niveau de gris ∇I à partir de l'image initiale. Dans cette première étude, la carte des ombres est simplement obtenue par seuillage et le champ de gradient par filtrage dérivatif de type Sobel. Remarquons que le choix de l'opérateur \max dans (1) est contraignant puisqu'il impose que les deux critères, sur l'ombre et sur le gradient, soient vérifiés simultanément.

Pour définir $u_o(b)$, nous introduisons le facteur de qualité d'ombre $q_o(b)$ du bâtiment b par :

$$q_o(b) = \frac{\text{Surf}\{\text{Ombre}(b) \Delta \text{Box}_b(O)\}}{\text{Surf}\{\text{Ombre}(b)\} + \text{Surf}\{\text{Box}_b(O)\}} \quad (2)$$

où Box_b représente la boîte englobante (dans l'image) du bâtiment b et de son ombre, et Δ est la différence symétrique. Le facteur de qualité d'ombre d'un bâtiment est donc défini par la différence de son ombre avec l'ombre de l'image dans sa boîte englobante, normalisée entre 0 et 1. Moins ces deux ombres coïncident, plus ce facteur est important.

Pour définir $u_c(b)$, nous introduisons le facteur de qualité de contour $q_c(b)$ du bâtiment b . Nous considérons $n_{\parallel} + n_{\perp} = 7$ segments lorsque les deux pans de toit sont éclairés (les six côtés plus le faîte), ou $n_{\parallel} + n_{\perp} = 4$ lorsqu'un seul pan de toit est éclairé (les trois côtés du pan éclairé plus le faîte), comme montré sur la figure 1. Nous distinguons les côtés parallèles au faîte des côtés orthogonaux au faîte, que nous notons respectivement σ_i^{\parallel} et σ_i^{\perp} (N_i^{\parallel} et N_i^{\perp} désignent les normales unitaires associées). Nous définissons un score pour chaque direction comme suit :

$$q_{\parallel}(b) = \max_{i \in [1, n_{\parallel}]} \left\{ \frac{\left| \int_{\sigma_i^{\parallel}} \nabla I(s) \cdot N_i^{\parallel} ds \right|}{\left| \int_{\sigma_i^{\parallel}} ds \right|} \right\} \quad (3)$$

$$q_{\perp}(b) = \max_{i \in [1, n_{\perp}]} \left\{ \frac{\left| \int_{\sigma_i^{\perp}} \nabla I(s) \cdot N_i^{\perp} ds \right|}{\left| \int_{\sigma_i^{\perp}} ds \right|} \right\} \quad (4)$$

Finalement, nous définissons $q_c(b)$ par :

$$q_c(b) = \min(q_{\parallel}(b), q_{\perp}(b)) \quad (5)$$

Plus le gradient ∇I est élevé et plus il est orthogonal au segment considéré, plus le facteur de qualité de contour $q_{\parallel}(b)$ ou $q_{\perp}(b)$ est important. En outre, nous imposons au moins un segment à fort contraste dans chacune de ces deux directions, grâce à l'utilisation de l'opérateur \min dans (5).

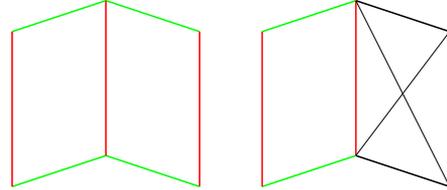


FIGURE 1 – Contours parallèles au faîte (en rouge) et orthogonaux au faîte (en vert) dans le cas d'un toit totalement éclairé (à gauche) ou d'un toit avec un pan dans l'ombre (à droite).

Pour contraindre ces facteurs de qualité à occuper l'intervalle $[-1, 1]$, nous les incorporons à des fonctions sigmoïdes, ce qui permet de définir les énergies associées comme suit :

$$u_o(b) = -1 + \frac{2}{1 + \exp\{-q_o(b)/S_o + 1\}} \quad (6)$$

$$u_c(b) = 1 - \frac{2}{1 + \exp\{-q_c(b)/S_c + 1\}} \quad (7)$$

où S_o (respectivement S_c) est un seuil en deça (respectivement au-delà) duquel l'énergie est négative, ce qui signifie que le bâtiment est favorisé.

Pour minimiser l'énergie $U(B)$, nous considérons un processus de naissances et morts multiples, initialement proposé dans [7]. Le principe de l'algorithme est d'alterner des phases de naissances, où de nouveaux objets sont ajoutés aléatoirement à la configuration, et des phases de morts, où les objets les moins pertinents, au sens de l'énergie, sont supprimés. Cette dynamique, intégrée dans un schéma de recuit simulé, converge vers le minimum global de l'énergie. L'algorithme se déroule comme suit :

Initialisation Nous initialisons la température $T = T_0$ et le taux de naissances $\delta = \delta_0$. La configuration initiale est l'ensemble vide. Nous effectuons alors alternativement et itérativement les étapes suivantes :

Naissances Nous tirons un nombre d'objets aléatoirement suivant une loi de Poisson de moyenne δ . Ces objets sont répartis de façon uniforme dans l'image, leurs paramètres sont également choisis de façon uniforme. Ces objets sont ajoutés à la configuration courante.

Tri des objets Une fois la phase de naissances effectuée, nous calculons l'attache aux données $U_d(b)$ de chaque objet b de la configuration courante. Puis, nous les classons par ordre décroissant suivant cette énergie.

Morts Pour chaque objet b de la configuration B , pris dans cet ordre, nous calculons le taux de morts :

$$d(b|B) = \frac{\delta a_T(b|B)}{1 + \delta a_T(b|B)} \quad (8)$$

où $a_T(b|B) = \exp\{-[U(B \setminus \{b\}) - U(B)]/T\}$ et supprimons l'objet b , soit $B \rightarrow B \setminus \{b\}$ avec la probabilité $d(b|B)$.

Test de convergence Si le processus n'a pas convergé, c'est-à-dire si la configuration, après l'étape de morts, a changé par rapport à l'itération précédente, nous modifions la température T et le taux de naissances δ en les multipliant par des coefficients de décroissance de l'ordre de 0,999. Nous retournons alors à l'étape de naissances pour une nouvelle itération.

3 Résultats

La figure 2 montre un exemple de reconstruction 3D à partir de la seule photographie aérienne 2-a. Rappelons que le nombre de bâtiments n'est pas connu a priori. Les ombres et les contours, estimés de façon naïve, sont représentés sur les figures 2-b,c,d. La figure 2-e montre la reconstruction 3D obtenue après 2700 itérations, pour $\lambda = 0,6$, $S_o = 0,3$ et $S_c = 15$. Elle est qualitativement bonne, dans la mesure où le nombre de bâtiments $n = 7$ est correctement retrouvé et où les bâtiments sont relativement bien positionnés et bien orientés.

Néanmoins, pour que notre méthode puisse effectivement servir à la production de MNE (Modèles Numériques d'Élévation), il est nécessaire que la hauteur des bâtiments soit estimée avec précision. En guise de vérité terrain, nous mesurons la hauteur de chaque bâtiment en détectant à la main le bord du faîte du toit et sa projection sur le bord de l'ombre : les segments rouges de la figure 2-f sont censés être parallèles, de longueurs proportionnelles aux hauteurs des bâtiments. Dans le tableau 1, nous comparons ces valeurs mesurées aux hauteurs estimées par notre méthode : la racine carrée de l'écart quadratique moyen vaut $RMSE = 3,9$ pixels.

TABLE 1 – Hauteurs mesurées et hauteurs estimées (en pixels).

Hauteurs	b_1	b_2	b_3	b_4	b_5	b_6	b_7
Mesurées	34,6	34,9	32,1	31,6	35,2	30,3	36,1
Estimées	34	36	33	30	44	32	41

4 Conclusion et Perspectives

Par rapport à notre précédent travail [5], la principale amélioration vient de ce que le nombre de bâtiments ne doit plus être connu a priori. Cela est possible grâce à la modélisation par processus ponctuel marqué et à l'algorithme de naissances et morts multiples.

Nous avons combiné deux termes d'attache aux données fondés sur les ombres et les contours. Nous envisageons d'utiliser aussi la couleur et l'ombrage, voire l'homogénéité de la couleur des toits.

L'autre principale amélioration de ce travail serait d'utiliser une carte de naissances, construite par exemple en utilisant l'ombre ou les contours, afin d'accélérer la convergence [8].

Il est également envisageable de diversifier les modèles de bâtiments. Enfin, la détection des ombres, qui est faite ici par un simple seuillage, pourrait être améliorée, ainsi que la détection des contours qui utilise un simple filtrage de Sobel.

Remerciements

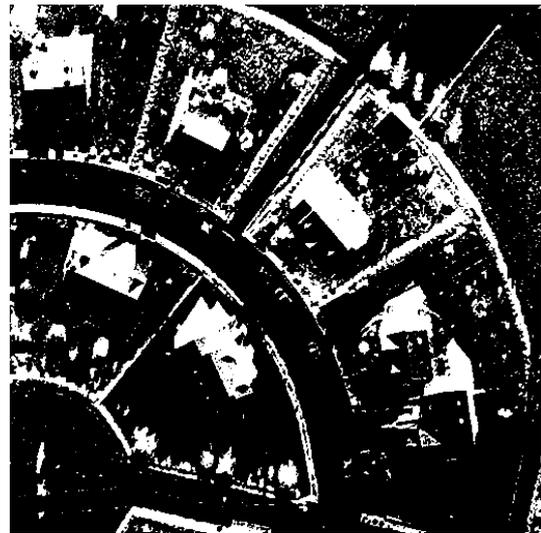
Les auteurs remercient l'INRIA pour son soutien dans le cadre de l'Équipe Associée ODESSA, et l'EGIDE pour son soutien dans le cadre du projet ECONET 18902PK.

Références

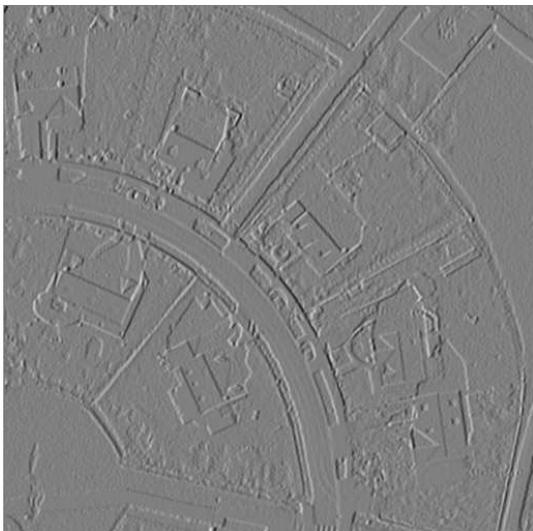
- [1] C. Baillard et A. Zisserman. *Automatic reconstruction of piecewise planar models from multiple views*. Proc. IEEE Conference on Computer Vision and Pattern Recognition, Corfou, Grèce, 1999.
- [2] F. Taillardier et R. Deriche. *Automatic Buildings Reconstruction from Aerial Images : A Generic Bayesian Framework*. Proc. ISPRS Congress, Istanbul, Turquie, 2004.
- [3] J. McGlone et J. Shufelt. *Projective and object space geometry for monocular building extraction*. Technical Report CS-94-118, Carnegie Mellon University, 1994.
- [4] C. Lin et R. Nevatia. *Building detection and description from a single intensity image*. Computer Vision and Image Understanding, 72(2), pages 101-121, 1998.
- [5] P. Lukashovich, A. Kraushonak, X. Descombes, J.-D. Durou, B. Zalesky et E. Zhizhina. *Reconstruction 3D du bâti par la technique des ombres chinoises*. Actes GRETSI, Dijon, France, 2009.
- [6] 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), pages 105-119, 2008.
- [7] 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), pages 347-359, 2009.
- [8] C. Benedek, X. Descombes et J. Zerubia. *Building Development Monitoring in Multitemporal Remotely Sensed Image Pairs with Stochastic Birth-Death Dynamics*. IEEE Transactions on Pattern Analysis and Machine Intelligence, to appear.



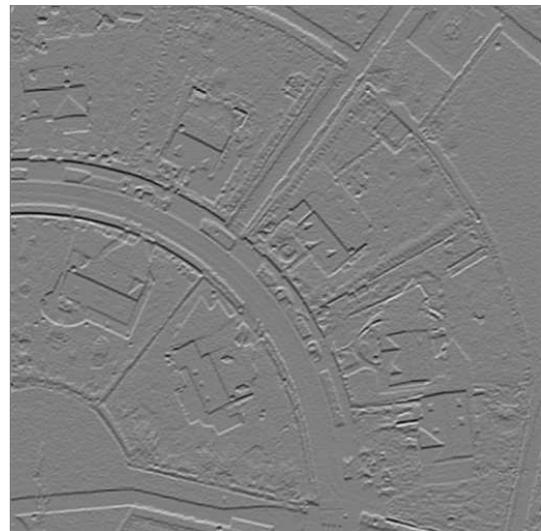
(a)



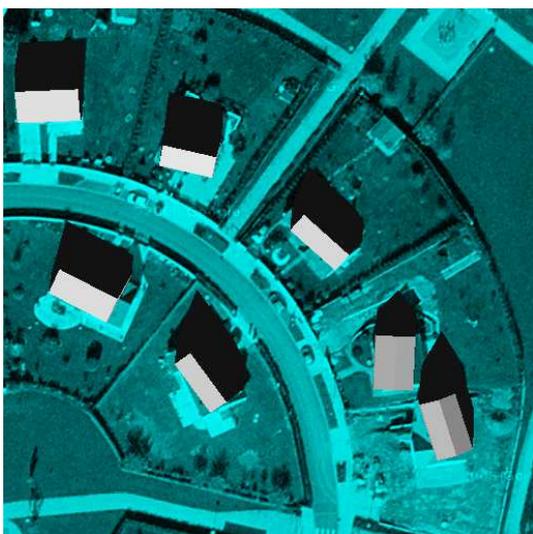
(b)



(c)



(d)



(e)



(f)

FIGURE 2 – (a) Image de la scène à reconstruire. (b) Ombres détectées par seuillage ($I \leq 75$). (c-d) Dérivée du niveau de gris obtenue par filtrage dérivatif de type Sobel (en x et en y). (e) Reconstruction 3D obtenue en 2700 itérations (25 minutes de calcul sur un processeur P4 à 2,4 GHz). (f) Pour chaque toit, détection du bord du faîte et de sa projection sur le bord de l'ombre : les segments rouges, qui sont censés être parallèles, ont des longueurs proportionnelles aux hauteurs des bâtiments.