

Analyse d'images aériennes haute résolution pour l'extraction du bâti

Matthieu Cord et Michel Jordan

ETIS, URA CNRS 2235, Université de Cergy
ENSEA – 6, av. du Ponceau
95014 CERGY-PONTOISE CÉDEX, France
e-mail : cord@ensea.fr, mj@ensea.fr

RÉSUMÉ

Nous présentons dans cet article une méthode pour la détection et la localisation de bâtiments dans un couple d'images aériennes à haute résolution.

La méthode s'appuie sur un appariement stéréoscopique multi-résolutions et un groupement perceptuel de segments de droite intégrant des informations 3-D.

ABSTRACT

We present in this paper a method for building detection and localization from a high resolution aerial photographs stereo couple. This method is based on a multi-resolutions stereo matching step and a perceptual grouping of segments using 3-D informations.

1 Introduction

L'analyse d'images aériennes en vue de la modélisation du relief, et particulièrement de la reconnaissance du bâti en zone urbaine, est une question qui reçoit de plus en plus d'intérêt. Nous nous intéressons ici à la détection et à la localisation des bâtiments dans une zone urbaine de densité moyenne, à partir de couples stéréoscopiques d'images aériennes numérisées à haute résolution (pour lesquelles un pixel représente 8 cm au sol environ). L'objectif général du système est de contrôler de manière automatique les saisies manuelles des bords de bâtiments effectuées sur des images de résolution plus faible.

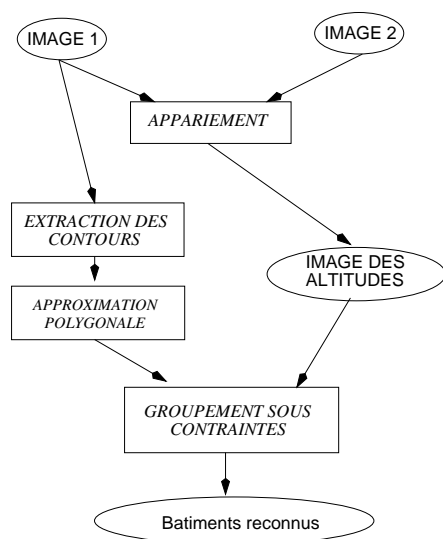


FIG. 1 — Schéma synoptique général

L'approche proposée comprend deux étapes (figure 1) :

1. une analyse stéréoscopique multi-échelles, afin d'obtenir une carte des altitudes en chaque pixel de la zone étudiée, et de détecter les structures composant le sur-sol ;
2. le groupement de segments de droite issus de la détection des contours dans une image du couple, en intégrant des informations d'altitude, afin de localiser précisément les bords des structures régulières présentes dans la zone.

2 Analyse stéréoscopique

La mise en correspondance ou *appariement* est une étape très importante de la vision par ordinateur. En stéréovision, l'objectif est d'associer chaque point d'une image à son correspondant (s'il existe) dans l'autre image, puis de calculer la position 3D du point correspondant.

2.1 Primitive et fonction de similarité

L'appariement se fait sur des objets ou *primitives* qui peuvent être les pixels eux-mêmes, des points d'intérêt ou encore des objets plus complexes (segments de contour, régions, etc.) à l'aide d'une *fonction de similarité* qui établit une mesure de ressemblance entre primitives.

De manière générale, un appariement utilisant la primitive pixel fournira un résultat plus dense mais moins précis sur des zones de discontinuité du relief qu'un appariement travaillant sur des primitives structurées. Cependant, plus les images sont riches ou *texturées*, plus les primitives structurées sont difficiles à extraire et à mettre en correspondance.

C'est pourquoi nous avons choisi une primitive peu structurée : le vecteur gradient en chaque point. Pour la primitive point, la fonction de similarité la plus utilisée est la fonction de corrélation. Dans notre cas, il faut définir une mesure de ressemblance entre vecteurs.

La méthode choisie a été proposée dans [4]. Sa grande originalité consiste à apparier non sur une simple similarité entre vecteurs mais sur des fenêtres de gradients. De plus, l'utilisation des gradients de l'image permet d'exploiter prioritairement la distribution spatiale de l'intensité.

Soient (im_E, im_O) un couple stéréoscopique d'images, et $(G_E(i, j), G_O(i, j))$ les vecteurs gradients au point (i, j) des images im_E et im_O ; sur des fenêtres rectangulaires $(2u + 1) \times (2v + 1)$, la fonction de similarité $Cgrad((i, j), \xi)$ est définie par :

$$Cgrad((i, j), \xi) = \frac{\sum_{\iota=i-u}^{i+u} \sum_{\eta=j-v}^{j+v} \| G_O(\iota, \eta) - G_E(\iota, \eta + \xi) \|}{\sum_{\iota=i-u}^{i+u} \sum_{\eta=j-v}^{j+v} [\| G_O(\iota, \eta) \| + \| G_E(\iota, \eta + \xi) \|]} \quad (1)$$

2.2 Processus multi-échelles

Les images aériennes à traiter présentent des variations de disparité importantes qui nous ont conduits à hiérarchiser le processus par des techniques multi-échelles [1] [11]. En conséquence, la mise en correspondance proposée utilise une pyramide dyadique de gradients associée à une fonction de corrélation donnée ci-dessus selon un schéma « *coarse to fine* ».

Une pyramide d'images est d'abord construite par des moyennages (2×2) successifs avec sous-échantillonnage, puis un filtre de Deriche [5] est appliqué sur chaque composante image. Ensuite, les mises en correspondance débutent sur les champs de gradients de la résolution la plus grossière ; les résultats servent à l'initialisation des calculs de la résolution immédiatement supérieure, et ainsi de suite jusqu'à la résolution la plus fine.

3 Localisation des bâtiments

Le schéma que nous proposons ici s'inspire des travaux menés depuis plusieurs années en groupement perceptuel [6] [7] [8].

L'apport nouveau par rapport à ces travaux réside dans l'utilisation d'informations 3-D dans le processus de groupement. Ces informations 3-D sont obtenues à partir de l'algorithme de mise en correspondance stéréoscopique décrit au paragraphe précédent. Les disparités obtenues par cet algorithme sont transformées en altitude des points de la scène grâce aux paramètres d'orientation absolue des images dont nous disposons.

Les bâtiments que nous cherchons à reconnaître sont définis comme étant des formes régulières, à l'intérieur desquelles

l'altitude moyenne est significativement supérieure à celle du voisinage.

Nous travaillerons donc sur des segments de droite, au voisinage desquels nous rechercherons des transitions d'altitude.

3.1 Obtention des segments

Les contours sont extraits de l'image en utilisant un filtre de Deriche très fin ($\alpha = 2.5$), paramétré de manière optimale pour une bonne localisation des contours. On trouvera dans [3] l'étude théorique de ce paramétrage.

Cette première étape est suivie d'un filtrage permettant de « nettoyer » les contours correspondant à du bruit dans l'image, ainsi que les structures de faible taille, et de conserver les contours d'intérêt mais faiblement contrastés.

Le principe du filtrage est le suivant : chaque maximum du gradient donne lieu à une étude locale sur l'image initiale. Si le voisinage du point considéré est homogène, ce maximum est associé à du bruit et est rejeté. Dans le cas contraire, le maximum étudié appartient à une région de transition et il est classé point de contour. Pour apprécier l'homogénéité d'un voisinage, un filtrage d'ordre de type *c-filtre* [2] a été utilisé. De plus, ce filtrage permet de s'affranchir de l'étape de seuillage classique effectuée sur l'image de la norme du gradient.

Les contours ainsi obtenus sont polygonalisés en utilisant la méthode décrite en [9].

3.2 Groupement des segments

Il s'agit ici d'associer des segments à partir de leurs relations géométriques (colinéarité, orthogonalité) et de certaines de leurs propriétés (voisinages radiométrique et altimétrique). La principale innovation de notre travail consiste en l'insertion, parmi les contraintes d'association, d'informations sur le voisinage 3-D des segments.

De même que dans les travaux cités ci-dessus, nous avons cherché à composer des associations de formes en L, en U et en rectangle.

3.2.1 Groupement de segments deux à deux

Deux segments sont groupés s'ils respectent les deux conditions :

- cohérence du voisinage radiométrique ;
- cohérence du voisinage altimétrique.

Cela peut concerner des segments colinéaires, orthogonaux ou parallèles.

La cohérence des voisinages, tant radiométriques qu'altimétriques, est déterminée en examinant les histogrammes dans une fenêtre de taille fixe de part et d'autre des segments. La taille de cette fenêtre est un paramètre qui peut être réglé par l'utilisateur en fonction de l'*a priori* sur la scène : dimensions des bâtiments et résolution du pixel en cm au sol.

Les associations de segments testées sont limitées par un paramètre de proximité des extrémités des segments candi-

datés ; ce paramètre peut également être réglé en fonction de l'*a priori* sur la scène.

Pour chaque association possible, un taux de cohérence est calculé en fonction du degré de recouvrement des histogrammes.

3.2.2 Extraction de structures complexes

A partir des relations entre segments définies au paragraphe précédent est établi un graphe, dont les nœuds sont les segments de droite et les arcs les relations établies entre ces segments.

L'élaboration des « hypothèses de bâtiments » consiste alors à rechercher dans ce graphe des sous-graphes pouvant composer des associations selon les formes polygonales désirées [10]. En final, les meilleures associations de segments (selon les taux de cohérence définis précédemment) sont retenues.

Il est à noter que, si nous privilégions la recherche de formes polygonales fermées, nous ne l'imposons pas. Il est plus important pour nous de détecter l'ensemble des bâtiments présents, même si certains le sont incomplètement, et de prévoir des post-traitements afin de fermer les formes obtenues.

4 Résultats

Nous présentons ci-dessous (figure 2) les résultats obtenus sur un couple d'images aériennes à haute résolution (la taille d'un pixel est de 8cm au sol), représentant un quartier de la ville de Colombes, en banlieue parisienne.

Cette figure présente :

- les deux images (vues OUEST et EST) du couple : nous avons extrait une zone présentant deux groupes de bâtiments et des éléments de voirie et de végétation, afin de tester la faisabilité de notre méthode ;
- une carte de disparités obtenue par la méthode décrite au premier paragraphe ; les points noirs sont les points non appariés, les points les plus clairs sont ceux où l'altitude est la plus faible. La méthode opère un fort rejet des points d'occlusion par rapport à l'appariement classique (corrélation croisée normalisée et centrée, avec ou sans pyramide), les résultats sont denses et cohérents sur les régions délimitant des zones continues de l'espace ; le calcul avec la pyramide est au minimum trois à quatre fois plus rapide, selon le paramétrage inter-échelle, que celui sans pyramide.
- les contours obtenus par le filtrage de Deriche avec paramétrage optimal ; l'image présentée est celle des contours avant « nettoyage » par le filtrage d'ordre. On constate la forte densité des contours, particulièrement sur les toits des bâtiments, ce qui rend indispensable le post-traitement par filtrage d'ordre. A l'issue de ce post-traitement et de la polygonalisation des contours, 89 segments sont candidats au groupement (image non présentée).

- les groupes de segments détectés : les deux groupes de bâtiments ont été bien détectés. La superposition des segments de droite détectés à l'image d'origine montre que la méthode nous fournit des segments de droite correspondant de manière très proche aux bords des toits des bâtiments. Cependant, les limites des bâtiments détectés ne sont pas complètes, et des traitements différents doivent être envisagés pour compléter la description de la scène. Ces traitements seront simplifiés par le fait que l'essentiel des contours dus à la végétation et à la voirie ont disparus de la description.

Les images résultats sont dans la géométrie de l'image OUEST.

5 Conclusions et perspectives

Nous avons présenté dans cet article une méthode qui nous a permis de détecter de manière automatique les bâtiments présents dans une zone urbaine observée en photographie aérienne à grande échelle, et d'y identifier et localiser avec une bonne précision, bien que partiellement, les bords des bâtiments.

Cette méthode semble donner des résultats prometteurs, même si des traitements devront être mis en œuvre pour compléter la description des bâtiments obtenue, et devrait nous permettre de comparer les bords des toits des bâtiments avec ceux issus d'une saisie manuelle.

Ivana Gregorio Leòn a programmé l'algorithme de groupement de segments de droite lors d'un stage de l'université de Valencia en 1995-96.

Ce travail prend partiellement place dans une convention de recherche entre l'ETIS et l'I.G.N (service MATIS), qui nous a fourni en particulier les images de test et leurs paramètres d'orientation absolue.

Références

- [1] S.D. Cochran and G. Medioni. 3-D Surface Description from Binocular Stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(10) :981–994, 1992.
- [2] J.-P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Ed. Masson, 1995.
- [3] M. Cord, F. Huet, and S. Philipp. Optimal Adjusting of Edge Detectors to Extract Close Contours. In *SCIA*, Lappeenranta, Finland, June 1997.
- [4] A. Crouzil, L. Massip-Pailhes, and S. Castan. Mise en correspondance par corrélation des gradients. In *Reconnaissance des formes et intelligence artificielle*, pages 236–245, Rennes, janv. 1996. AFCET.
- [5] R. Deriche. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision*, 1(2) :167–187, 1987.
- [6] B. Draper. Learning Grouping Strategies for 2D and 3D Object Recognition. In *ARPA Image Understanding Workshop*, Palm Springs, CA, 1996.



Image OUEST



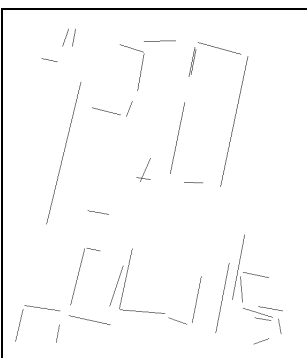
Image EST



Appariement. Calcul avec 3 niveaux de pyramide. Intervalle de recherche : 100 pixels.



Détection de contours. Filtre de Deriche : $\alpha = 2.5$



Détection des bâtiments. Voisinage de recherche pour chaque segment : 25 pixels. Taille de la fenêtre : 15 pixels.

- [7] C. Jaynes, F. Stolle, and R. Collins. Task Driven Perceptual Organization for Extraction of Rooftop Polygons. In *IEEE Workshop on Applications of Computer Vision*, Sarasota, Dec. 1994.
- [8] R. Mohan and R. Nevatia. Perceptual organization for scene segmentation and description. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(6) :616–635, June 1992.
- [9] S. Philipp and J.-P. Cocquerez. Fusion de régions sur critères géométriques. In *14ème GRETSI*, Juan-les-Pins, sept. 1993.
- [10] T. Skordas and R. Horaud. Mise en correspondance structurelle pour la vision stéréoscopique. *TSI*, 7(6) :591–608, 1988.
- [11] A. Witkin, D. Terzopoulos, and M. Kass. Signal matching through scale space. *International Journal of Computer Vision*, pages 133–144, 1987.

FIG. 2 — Détection et localisation du bâti