ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/228778755

Analyse d'images aériennes haute résolution pour la reconstruction de scènes urbaines

Article · January 2002

Source: OAI

CITATIONS
3

READS
163

4 authors, including:

Matthieu Cord
Sorbonne Université
272 PUBLICATIONS **5,640** CITATIONS
SEE PROFILE

Michel Jordan
Université de Cergy-Pontoise
31 PUBLICATIONS **254** CITATIONS
SEE PROFILE

Marcelo Bernardes Vieira
Federal University of Juiz de Fora
72 PUBLICATIONS **299** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Deep Metric Learning View project

Project VERSPERA View project

All content following this page was uploaded by Matthieu Cord on 27 May 2014.

The user has requested enhancement of the downloaded file.

ANALYSE D'IMAGES AERIENNES HAUTE RESOLUTION POUR LA RECONSTRUCTION DE SCENES URBAINES

Matthieu CORD, Michel JORDAN, Thomas BELLI, Marcelo BERNARDES VIEIRA

ETIS, UPRES-A CNRS 8051
ENSEA – BP 44
6, avenue du Ponceau
F 95014 CERGY-PONTOISE CEDEX
Internet : http://www-etis.ensea.fr/~image
E-mail : {cord, belli, jordan, vieira}@ensea.fr

Résumé
Nous présentons dans cet article un ensemble d'algorithmes d'analyse d'images aériennes en zone urbaine qui répond au problème de la reconstruction du relief et de la détection d'objets cartographiques tels que le bâti. Le cœur de ces algorithmes est constitué d'un module de stéréorestitution particulièrement adapté aux images haute résolution en zone urbaine, fournissant un modèle numérique d'élévation à la fois dense, précis, et respectant les discontinuités du relief. L'appariement des images couleur par stéréocorrélation est réalisé par fusion des courbes de corrélation calculées dans les trois canaux RVB. Les étapes consécutives réalisent l'estimation d'un modèle numérique de terrain, la détection des zones d'intérêt (« sur-sol ») et leur classification (bâti ou végétation). Nous proposons également une recherche interactive de régions par similarité. L'ensemble de ces algorithmes fournit des résultats ayant une précision suffisante pour des applications aussi diverses que l'étude des propagations d'onde en télécommunications ou la simulation réaliste de survols.

Mots-clefs : imagerie aérienne ; reconstruction 3D ; modèle numérique de terrain.

Abstract
This paper presents a set of algorithms for aerial image analysis in urban areas, which deals with 3D reconstruction and cartographic object detection. The core part of these algorithms is a stereorestitution process, dedicated to high resolution urban images; this process provides dense, accurate and depth-discontinuity-preserving digital elevation models (DEM). Colour image matching is performed by fusion of correlation score curves computed on each RGB channel. The following steps do the digital terrain model (DTM) estimation, the detection of regions of interest ("above-ground'" regions), and their classification ("building'' or "vegetation''). We also developed a content-based region retrieval technique. The results of these algorithms have a sufficient accuracy for various applied domains, such as telecommunications (wave propagation study) or simulation (urban tour).

Keywords: aerial imagery; 3D reconstruction; digital terrain model.

1 Introduction

Dans les dix dernières années, les besoins en informations cartographiques de haute résolution ont connu une croissance importante. Des applications variées, dans le domaine des télécommunications numériques (études de propagation d'ondes pour le placement optimal des antennes), de la planification et de l'urbanisme, du tourisme et des activités associées, *etc*., nécessitent des représentations 3D précises et régulièrement mises à jour des zones urbaines, ainsi que des modèles des objets d'intérêt, essentiellement le terrain, le bâti et la voirie, parfois la végétation.

La recherche a été très active afin d'automatiser au maximum les tâches fastidieuses de stéréorestitution et de délinéation d'objets cartographiques. Des algorithmes de mise en correspondance stéréo d'images de résolution moyenne (métrique) ont été mis au point.

Cependant, ces algorithmes sont mis en défaut en zone urbaine, où la résolution des pixels nécessaire peut descendre jusqu'à 5 à 10 cm, où les discontinuités du relief (façades, *etc*.) sont fortement marquées, où par conséquent apparaissent de grandes parties cachées dans l'une ou l'autre des images, et où aussi les zones non texturées peuvent occuper de grandes surfaces.

En zone urbaine, le problème de la reconnaissance et de la modélisation des structures artificielles (bâti et voirie) reste un sujet de recherche vivace (voir par

exemple (Gruen *et al*., 1995 ; Gruen *et al*., 1997 ; CVIU, 1998 ; Baltsavias *et al*., 2001). On trouvera également dans (Mayer, 1999) une étude bibliographique assez complète.

Avant de développer notre approche, nous rappelons brièvement ci-dessous les caractéristiques de quelques systèmes présentés ces dernières années dans la littérature et consacrés spécifiquement à la détection et à la modélisation du bâti à partir d'images aériennes.

Nevatia *et al*. (Université de Californie du Sud) ont développé une approche pour la détection et la description de bâtiments ou de groupes de bâtiments de forme polygonale à partir de plusieurs images aériennes de résolution environ 1m au sol (Kim *et al*., 2001). Les images permettent de calculer un modèle numérique d'élévations (MNE) qui est utilisé pour détecter des régions d'intérêt. Autour de ces régions d'intérêt, les segments de droite, les jonctions et les lignes parallèles sont détectées dans chaque image, puis appariées pour former des primitives 3D. Ces primitives 3D permettent de générer des hypothèses de bâtiment. La vérification de ces hypothèses s'appuie sur la présence ou non dans les images de lignes supports, de murs verticaux et de zones d'ombre portée. Des modèles de toit à pans multiples sont également testés.

Pour restituer le bâti, le projet « ASCENDER 2 » (Université du Massachussets) utilise un système à base de connaissances afin de sélectionner et combiner de manière appropriée les algorithmes de traitement d'images (Hanson *et al*., 2001). Le cœur du système réside dans un réseau hiérarchique bayésien, qui permet une combinaison efficace des résultats des algorithmes pour former des hypothèses de structures géométriques 3D. Ces hypothèses sont comparées à une bibliothèque de modèles afin de déterminer la structure de la scène complète. L'efficacité de la méthode est donc limitée par le nombre et la représentativité des modèles choisis.

Brenner *et al*. (Université de Stuttgart) présentent un système pour la constitution de « modèles de ville » à partir d'images aériennes et d'informations complémentaires (plans au sol, MNE obtenus par laser) (Brenner *et al*., 2001). A partir des plans ou de la segmentation des MNE en régions planes, le système construit des hypothèses de forme de toit pour chaque bâtiment. Ces hypothèses sont validées par une approche à base de règles de compatibilité, et l'on obtient un modèle planaire de chaque bâtiment. Afin d'aller vers des applications de « réalité augmentée », des textures réalistes sont extraites des images aériennes et d'images prises au sol, et appliquées sur toutes les faces (toit et façades) des bâtiments.

Moons *et al*. (Université Catholique de Louvain) ont présenté un système pour la restitution des toits à partir de segments 2D détectés dans 4 ou 6 images et appariés autour de régions d'intérêt désignées manuellement afin de former des segments 3D ([Frère *et al*., 1997]). Les toits sont formés de manière hiérarchique, à partir de groupements de segments 3D, eux-même composés en groupements de polygones 3D. La précision géométrique des modèles est ensuite affinée en ajustant les segments à ceux détectés dans les images.

Une approche similaire, à base de relations topologiques et géométriques afin de former des modèles polyèdres du bâti, est présentée dans (Heuel & Förstner, 2001).

De manière générale, les solutions adoptées peuvent être rassemblées en deux grandes classes. Les premières, « déductives » ou « ascendantes », cherchent à détecter et grouper des primitives de plus en plus complexes, afin d'obtenir une description de plus en plus complète et précise de la scène étudiée ; ces stratégies de groupement de primitives sont le plus souvent issues de la communauté *vision par ordinateur*. Les secondes, « inductives » ou « descendantes », principalement issues de la communauté *photogrammétrique*, cherchent à appliquer et recaler au mieux aux données images un certain nombre de modèles prédéfinis ; cela suppose généralement de détecter les régions d'intérêt, par des pré-traitements *ad hoc* ou manuellement. Des voies très diverses sont explorées : utilisation de l'information couleur, mise en correspondance de vues multiples pour le calcul des modèles numériques d'élévation (MNE), apport de contraintes géométriques *a priori*, de données externes (cadastre, *etc*.), *etc*., et les meilleurs résultats sont généralement obtenus par la combinaison de ces approches, dans des cas relativement simples : zones péri-urbaines, avec de grands bâtiments de forme simple et non accolés.

Le système que nous avons développé au laboratoire ETIS (Equipes traitement des images et du signal, Cergy) a pour objectif la détection et la modélisation du bâti en zone urbaine dense, à partir d'images aériennes haute résolution (un pixel représente moins de 10 cm au sol), et avec le degré d'automatisation le plus grand possible.

L'information altimétrique est celle que nous pensons la plus pertinente pour la détection du bâti dans ce contexte, aussi nous avons développé un algorithme de stéréocorrélation spécialement adapté, que nous présentons au paragraphe 2. La version de base de cet algorithme s'applique à un couple stéréoscopique d'images en niveaux de gris (*cf*. section 2.1), mais nous l'avons étendu au cas des images couleurs (*cf*. section 2.2).

A partir des MNE ainsi obtenus, nous effectuons une analyse fréquentielle pour estimer les MNT, et par différence détecter les objets composant le sur-sol (*cf*. section 0). Les applications présentées ici sont au nombre de deux :

– l'extraction de surfaces 3D sous-jacentes aux nuages de points 3D composant chaque objet du sur-sol : nous présentons (*cf*. section 4.1) une méthode basée sur le formalisme du « *tensor*

voting » proposé par Lee & Medioni (Medioni *et al*., 2000) ;

– une recherche interactive de régions similaires, basée sur un index calculé à partir des informations couleurs et altitudes (*cf*. section 4.2).

L'ensemble de notre chaîne de traitements est résumé en figure 1.

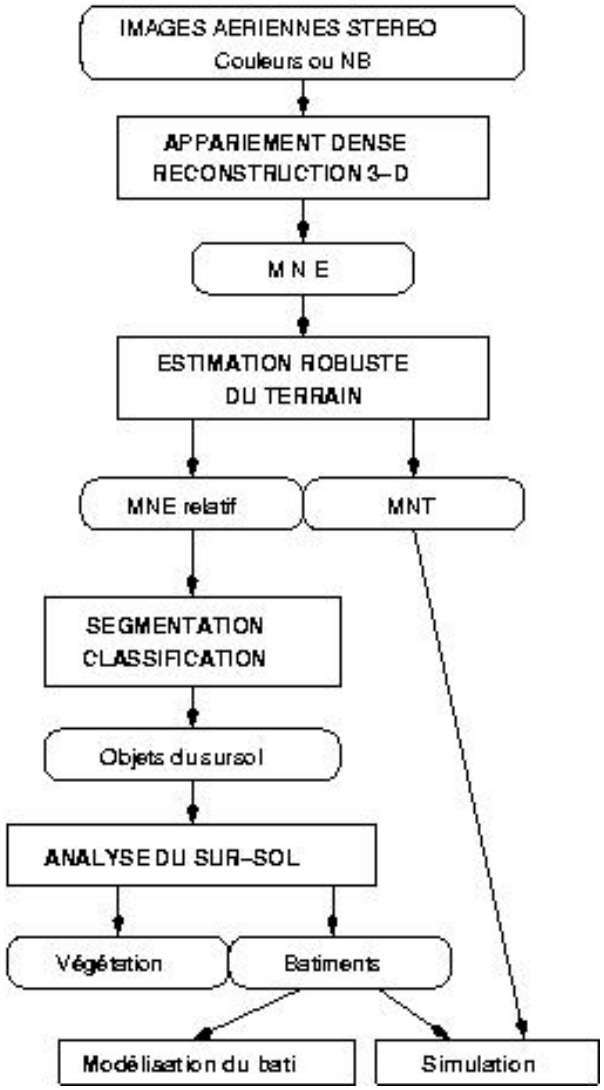

Fig. 1 – Chaîne de traitements : schéma synoptique.

Les résultats obtenus sont présentés dans chaque paragraphe et nous proposons finalement quelques conclusions et perspectives (section 5).

2 Mise en correspondance des images stéréoscopiques

L'approche de stéréo-restitution que nous proposons est basée sur un appariement des images par corrélation de voisinages de pixels, seul à même d'assurer la nécessaire densité des résultats. Cependant, ce type d'approche souffre de graves défauts dans un environnement urbain dense, en particulier au voisinage des discontinuités du relief, qui ont tendance à être lissées. Les moyens de dépasser ces défauts résident dans l'adaptation des fonctions de similarité, des supports de corrélation (fenêtres de taille et/ou de forme adaptative, ou déformées selon l'orientation locale des surfaces 3D), et l'ajout de contraintes locales sur les disparités, sur l'espace de recherche.

Notre schéma de mise en correspondance s'appuie ainsi sur des fenêtres de forme adaptative, avec une stratégie multi-résolutions comportant une étape de validation à chaque résolution. Ce schéma est brièvement rappelé au paragraphe 2.1, on se reportera pour plus de détails à (Paparoditis *et al*., 1998) et (Cord *et al*., 2001).

2.1 Images monochromatiques

Fonction de similarité : nous utilisons la corrélation croisée normalisée centrée ou une mesure spécifique appliquée aux champs des gradients radiométriques et prenant en compte simultanément la norme et l'orientation des gradients (Crouzil *et al*., 1996). Cette dernière semble plus robuste aux changements d'illumination entre les vues.

Adaptation de la forme de la fenêtre : le défaut majeur des techniques de corrélation est perceptible au voisinage des discontinuités altimétriques, là où les fenêtres supports peuvent représenter des éléments différents de la scène (parties cachées vues dans une seule des deux images) ; il importe donc de limiter ces fenêtres afin qu'elles ne « débordent » pas au-delà d'une discontinuité du relief. Pour ce faire, nous nous appuyons sur les contours radiométriques détectés dans les deux images, en faisant l'hypothèse, très largement vérifiée, que les discontinuités altimétriques se traduisent dans nos images par des contours radiométriques.

Dans une fenêtre carrée $(2L+1)\times(2L+1)$, les seuls points pris en compte dans le calcul de la corrélation sont ceux pour lesquels il existe un chemin vers le pixel central qui ne traverse pas un contour. De plus, afin de traiter le cas de contours « ouverts », nous appliquons à chaque point P un facteur de pondération w fonction de la distance géodésique d(P,C) au pixel central C :

$$w(P,C)=\exp\left(-\frac{d(P,C)^2}{2\sigma^2}\right)$$

pour tous les pixels connectés au pixel central.

La forme de la fenêtre est calculée en chaque point de l'image de référence uniquement, contrairement à d'autres types de fenêtres adaptatives (Lotti, 1996), calculées pour chaque couple candidat. Ceci nous permet d'être quasi-équivalent à une corrélation classique en termes de rapidité de calcul, tout en restant efficace aux voisinages des discontinuités du relief.

Adaptation de la taille de la fenêtre : cette adaptation est réalisée essentiellement par le biais de notre stratégie multi-résolutions. Les deux images sont dégradées et sous-échantillonnées par un filtre moyenneur ou médian, et la mise en correspondance est réalisée à chaque résolution en utilisant comme initialisation l'appariement obtenu à la résolution précédente.

Afin de prévenir les risques de propagation d'erreurs, les pics de corrélation obtenus ne sont validés que s'ils dépassent un certain seuil ; de plus, à chaque résolution traitée, nous effectuons la corrélation en prenant comme référence successivement chacune des deux images, et seuls les appariements cohérents sont retenus, c'est-à-dire ceux qui assurent que le correspondant du correspondant d'un point est bien lui-même.

Résultats et performances : la figure 2 présente un MNE calculé sur un couple stéréoscopique d'images de très haute résolution (un pixel pour 8 cm au sol) fourni par l'IGN.

Le MNE est représenté dans la géométrie de l'image de gauche, les altitudes des pixels vont croissant de l'orangé au bleu, et les points non appariés sont en rouge, pour l'essentiel correspondant aux occultations dans l'une des images ; les contours utilisés dans les fenêtres de corrélation sont superposés à cette image.

On notera la grande densité de ce MNE, ainsi que la qualité des transitions aux bords des bâtiments, malgré l'amplitude de l'intervalle des disparités (150 pixels).

Nous avons procédé à l'évaluation de notre algorithme sur un ensemble d'images issues de la même campagne, en comparant nos résultats avec les informations extraites de la couche « bâti » de la « BD Trapu » de l'IGN, où chaque pan de toit est représenté par une face plane. Nous avons défini deux seuils de fiabilité, appliqués aux pixels appariés :

– le premier, à 1,5 m, permet de rejeter les erreurs d'appariement grossières ;
– le second, à 0,6 m, indique à la fois des erreurs d'appariement faibles, mais aussi les insuffisances de notre référence (superstructures non prises en compte, non-planéité de la surface des toits, *etc*.).

Sur plus de 80 facettes de toit (environ 310.000 points), nous obtenons les résultats suivants :

Exhaustivité	91,2 %
Fiabilité à 1,5 m	94,4 %
Fiabilité à 0,6 m	82,2 %
Ecart-type de l'erreur	0,7 pixel

2.2 Extension aux images couleurs

Il existe encore relativement peu de travaux consacrés à l'appariement d'images couleurs (Yuan & Subbarao, 1998 ; Girard *et al*., 1998), et la plupart sont des extensions de méthodes *basées intensité*, qui ne considèrent pas la spécificité de l'information couleur.

Les méthodes dites « vectorielles » proposent des fonctions de similarité travaillant non plus sur des vignettes de scalaires, mais de vecteurs couleurs.

Les méthodes « marginales » exploitent quant à elles les résultats obtenus séparément sur les trois canaux (scores de corrélation ou disparités). Le plus souvent, l'information couleur est prise en compte en fusionnant les trois MNE pour obtenir le MNE final. Ces méthodes fournissent généralement des MNE plus denses que l'appariement d'un seul canal, mais présentent deux inconvénients :

– les pixels non appariés ne présentent pas la même distribution dans les trois canaux, le schéma de fusion doit donc être différent selon que l'on a un, deux ou trois appariements à fusionner en un point ;
– il y a un risque de perte d'information en accordant trop d'importance au score de corrélation dans un canal et en négligeant d'autres appariements ayant un score de corrélation plus faible dans les deux autres canaux.

Le schéma de fusion que nous proposons pour éviter ces problèmes prend en compte l'ensemble des scores de corrélation obtenus dans les trois canaux pour un intervalle de recherche donné. L'hypothèse sur laquelle nous nous appuyons est que la valeur moyenne des scores de corrélation diminue lorsque le bruit augmente. Contrairement aux méthodes vectorielles, nous pourrons ainsi prendre en compte sans *a priori* les différences de niveaux de bruit entre les canaux RVB. Pour la fusion, nous nous appuyons sur la théorie des ensembles flous (Bloch, 1996).

L'ensemble de décision $[D_1,D_P]$ comprend P classes, correspondant aux P valeurs de disparité possibles.

Pour chaque pixel x, le score de corrélation s_i^j, calculé pour la disparité i dans le j-ième canal (image I_j), exprime le degré d'appartenance de x à la classe D_j.

Sachant que l'amplitude du bruit peut être différente selon les canaux, nous avons choisi un opérateur de fusion qui intègre les scores de corrélation respectifs dans chaque canal. De plus, nous choisissons de donner une grande importance aux disparités qui ont un score de corrélation élevée, même dans un seul canal, et de moyenner les scores de corrélation consonants.

Nous avons donc choisi l'opérateur de fusion « barycentre pondéré » (Huet & Philipp, 1998), qui répond à ces objectifs :

$$S_i^{fus} = \frac{\left(s_i^R\right)^2 + \left(s_i^V\right)^2 + \left(s_i^B\right)^2}{s_i^R + s_i^V + s_i^B}$$

Résultats : nous présentons ici les résultats obtenus sur un couple stéréoscopique fourni par l'IGN (images de la caméra numérique, environ 30 cm par pixel, ville de Rennes). On trouve en figure 3 : l'une des deux images du couple stéréoscopique ; le MNE calculé par l'algorithme section 2.1 sur la composante luminance des images ; le MNE calculé sur les trois composantes RVB des images.

On constate sur ces images la plus grande densité du MNE obtenu avec les images couleur. De plus, à partir de régions de référence levées manuellement sur la zone, nous avons réalisé une évaluation statistique qui a montré que l'écart-type des erreurs est significativement diminué par la fusion des scores de corrélation des trois canaux RVB (Belli *et al*, 2000).

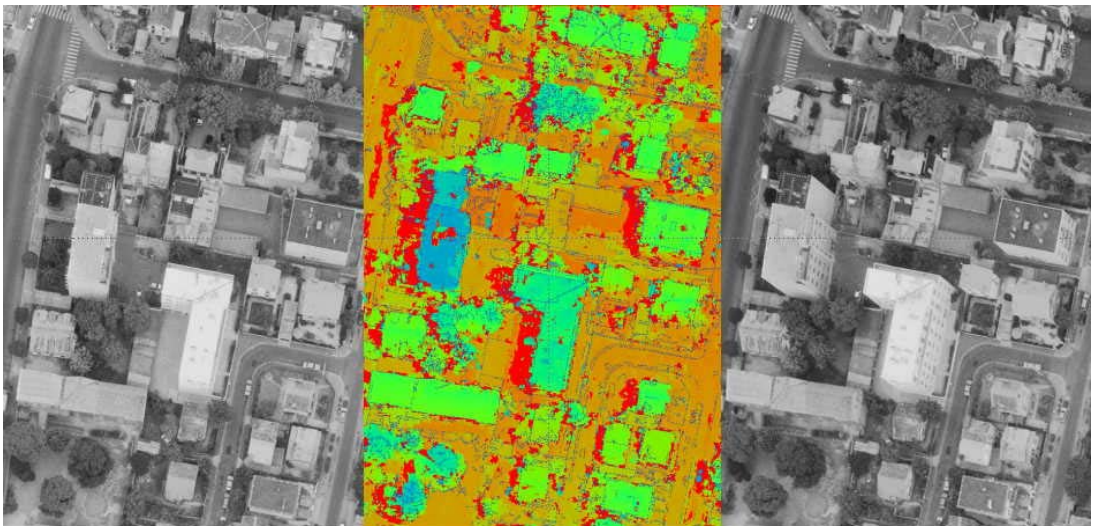

Fig.2 – MNE calculé sur un couple stéréoscopique de très haute résolution ; images de l'IGN, Colombes (image 1800 x 1500, résolution au sol 8cm par pixel).

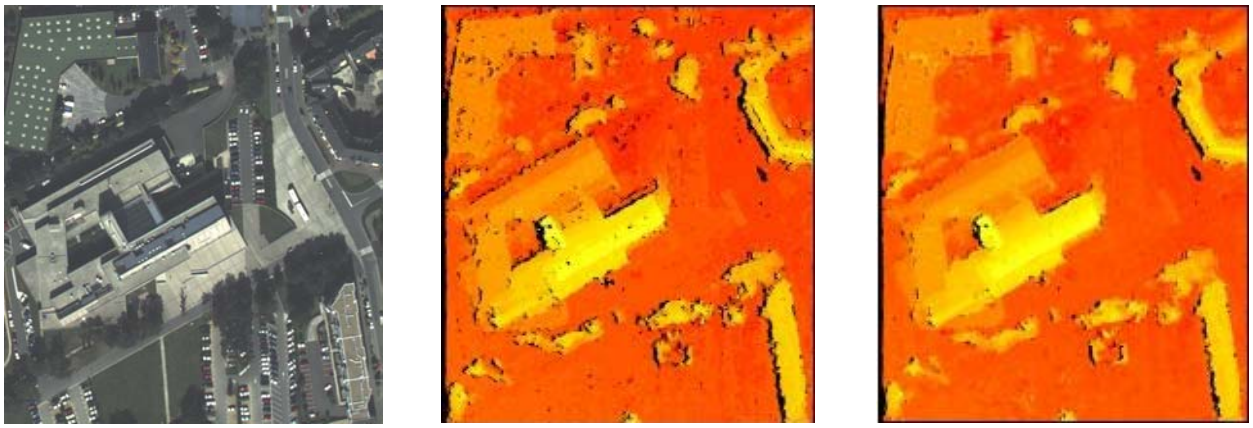

Fig.3 – MNE calculé sur un couple stéréoscopique couleurs images de l'IGN, Rennes (image 650 x 720, résolution au sol 35cm par pixel).

Nous avons également effectué des tests sur des données simulées avec un bruit d'amplitude variable selon les canaux. Il ressort de ces tests qu'il est d'autant plus intéressant d'utiliser un schéma de fusion couleur que le bruit est important. La fusion par barycentre pondéré s'avère plus performante que d'autres méthodes de fusion simples (moyenne, médian) dès lors que les bruits dans les différents canaux sont d'amplitude différente.

Cependant, des tests quantitatifs doivent encore être réalisés sur des données réelles (l'IGN doit prochainement mettre à disposition des données de référence couvrant le site d'Amiens), afin de chiffrer le gain de l'appariement couleur sur des sites urbains.

Perspectives : il nous semble intéressant désormais de mieux exploiter les spécificités des espaces couleur dans le schéma de mise en

correspondance (Koschan, 1996), en particulier en « profitant » de la qualité colorimétrique des récentes images de l'IGN.

3 Classification, estimation du modèle numérique de terrain

Dans l'objectif de détecter des objets cartographiques particuliers (bâti, végétation, *etc.*), afin de leur appliquer des traitements spécifiques, l'étape suivante consiste à segmenter le MNE obtenu précédemment et de classer les régions en « sol » ou « sur-sol ». Dans un premier temps, nous estimerons à partir du MNE un modèle numérique de terrain (MNT), d'où sont extraites les informations 3D liées au sur-sol.

Pour cela, un simple seuillage des altitudes est le plus souvent insuffisant, car il ne permet pas de prendre en compte les pentes du terrain, qui, même en zone urbaine, peuvent être fortes. Aussi avons-nous développé une méthode originale d'estimation des MNT, à base d'analyse fréquentielle.

Modèle de terrain : pour estimer le MNT à partir du MNE obtenu précédemment, nous faisons l'hypothèse que le terrain présente des variations faibles par rapport au sur-sol. Nous supposons que la surface du sol est telle qu'elle peut être représentée par une série de Fourier, et nous l'écrivons donc sous forme de la décomposition d'ordre N sur une base de fonctions harmoniques 2-D :

$$z(x,y)=a_{0,0}+\sum_{k,l=0;k+l\neq 0}^{N}[a_{k,l}\cos(2\pi(k\nu_x x+l\nu_y y))+b_{k,l}\sin(2\pi(k\nu_x x+l\nu_y y))]$$

avec les fréquences fondamentales $\nu_x=1/T_x$ et $\nu_y=1/T_y$ pour un MNT de dimensions $T_x\times T_y$.

Estimation robuste : les MNE denses obtenus par les méthodes décrites précédemment peuvent être considérés comme des MNT corrompus par les données « aberrantes » que constituent les objets du sur-sol.

A partir de cette observation, nous allons donc utiliser les techniques « d'estimation robuste » pour calculer les $P=2(N+1)^2-1$ paramètres du modèle de l'équation précédente.

Les m points Pt_i du MNE, de coordonnées $(x_i,y_i,z(x_i,y_i))_{i=1..m}$, satisfont l'équation matricielle :

$$z=M\cdot\theta$$

où :

$$z=[z(x_i,y_i)]^T_{i=1,m}$$

$$\theta=[a_{0,0},a_{0,1},b_{0,1},\Lambda,a_{N,N},b_{N,N}]$$

$$M=\begin{pmatrix} 1 & C_{0,1}(Pt_1) & S_{0,1}(Pt_1) & \Lambda & S_{N,N}(Pt_1) \\ 1 & C_{0,1}(Pt_2) & S_{0,1}(Pt_2) & \Lambda & S_{N,N}(Pt_2) \\ M & M & M & O & M \\ M & M & M & O & M \\ 1 & C_{0,1}(Pt_m) & S_{0,1}(Pt_m) & \Lambda & S_{N,N}(Pt_m) \end{pmatrix}$$

en ayant posé $C_{k,l}(Pt_i)=\cos(2\pi(k\nu_x x_i+l\nu_y y_i))$ et $S_{k,l}(Pt_i)=\sin(2\pi(k\nu_x x_i+l\nu_y y_i))$.

Le nombre m de points 3D est bien plus grand que le nombre P de paramètres à estimer, et nous utilisons pour résoudre le système surdéterminé précédent les M-estimateurs, dont le principe consiste à réduire l'influence des points hors-modèle, ici les points du sur-sol. La norme robuste de l'erreur $\varepsilon_{i,\theta}$ sur z, à minimiser, est une fonction ρ qui minimise l'influence des erreurs les plus grandes. Par exemple, la famille ρ_c suivante, que nous utiliserons (le paramètre d'échelle c est discuté par la suite), a été proposée dans (Beaton & Tukey, 1974) :

$$\rho_c(x)=\begin{cases}\frac{c^2}{6}\left(1-\left(1-\left(\frac{x}{c}\right)^2\right)^3\right) & \text{si } |x|\leq c \\ \frac{c^2}{6} & \text{sinon}\end{cases}$$

La solution optimale est alors donnée par :

$$\theta_{opt}=\arg\min_{\theta}\sum_{i=1}^{m}\rho(\varepsilon_i,\theta)$$

La résolution numérique directe de cette équation peut s'avérer difficile dès lors que P est élevé. Dans le cadre de la théorie des M-estimateurs, une équivalence avec un schéma d'optimisation par moindres carrés repondérés itérés est établie (Beaton & Tukey, 1974) :

$$\boldsymbol{\theta}^{(k)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{m} w_{c,i}^{(k-1)} (\varepsilon_i)^2$$

$$\varepsilon_i^{(k)} = z_i - z_{\boldsymbol{\theta}^{(k)},i}$$

$$w_{c,i}^{(k)} = \frac{\rho_c'(\varepsilon_i^{(k)})}{\varepsilon_i^{(k)}}$$

Dans le cas de Tukey, on a la fonction ω_c suivante :

$$\omega_c(x) = \begin{cases} \left(1 - \left(\frac{x}{c}\right)^2\right)^2 & si\ |x| \le c \\ 0 & \text{sinon} \end{cases}$$

Ce schéma itératif est intéressant numériquement, puisque seules des estimations classiques aux moindres carrés pondérés sont exigées. Pratiquement, les valeurs des paramètres sont initialisées par une estimation classique aux moindres carrés.

Le paramètre c est alors calculé en fonction de la variance des erreurs des données au modèle initial. Il est ensuite décrémenté après chaque estimation du modèle de terrain, ce qui nous permet de construire un modèle de plus en plus sélectif, rejetant de plus en plus de points hors-modèle.

Résultats : la figure 4 présente les résultats de l'algorithme sur une scène simulée. A gauche figure le MNE, avec neuf bâtiments couvrant environ 36% de la scène, auxquels a été ajouté un bruit gaussien stationnaire ($\mu=0, \sigma=0.5$) ; à droite est représenté le MNT extrait de ces données.

Un exemple d'application sur des images à très haute résolution (images couleur fournies par le département de photogrammétrie de l'ETHZ, Zürich, résolution env. 8 cm au sol) est présenté en figure 5 : l'une des deux images du couple stéréoscopique (en haut à gauche) ; le MNE calculé par l'algorithme du paragraphe 2.2 (en haut à droite) ; le MNT extrait de ce MNE (en bas à gauche) ; un profil du MNT, en trait continu, et du MNE, le long de la colonne A (en bas à droite).

Le modèle présenté a été calculé à l'ordre $N=3$, des ordres inférieurs fournissent une approximation plus grossière du MNT ; inversement, l'utilisation d'ordres élevés peut conduire à prendre en compte dans l'estimation du modèle du sol des points appartenant au sur-sol. Sur l'ensemble des tests que nous avons réalisés, l'estimation du MNT s'est avérée fiable et robuste, pour peu que l'ordre du modèle soit correctement choisi, et que la proportion de points du sur-sol dans la scène ne soit pas trop importante. Nous avons en particulier testé notre approche sur d'autres chantiers, comme celui de Rennes fourni par l'IGN (figure 6).

L'ordre N est lié à l'étendue de la scène traitée. L'extension de la méthode à des scènes très étendues ne semble pas simple, sans passer par un découpage de la scène en blocs. D'autre part, l'utilisation de fonctions asymétriques pourrait nous permettre de mieux rejeter les données du sur-sol. Cependant, l'adaptation à des fonctions asymétriques du processus d'estimation n'est pas triviale dans la mesure où l'initialisation du modèle ne garantit pas que les erreurs initiales soient positives pour toute donnée du sur-sol.

Les MNT ainsi obtenus nous permettent de calculer des MNE relatifs, et donc de segmenter les objets ou points du sur-sol.

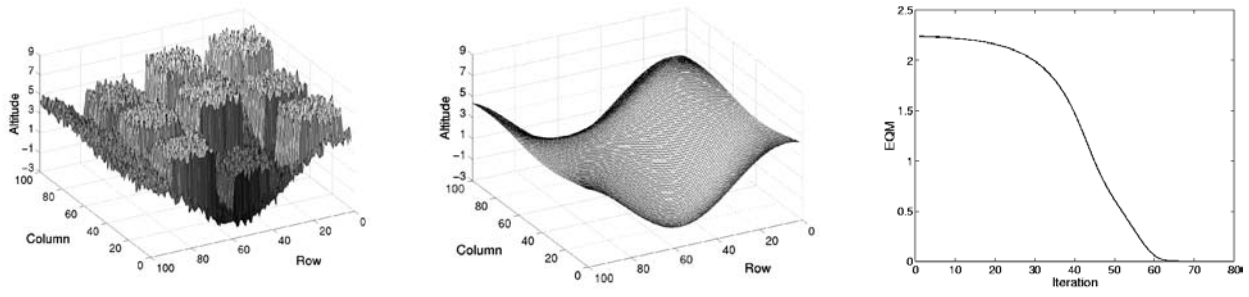

Fig. 4 – Simulation : MNE et MNT estimé , courbe de l'erreur quadratique moyenne au long des itérations.

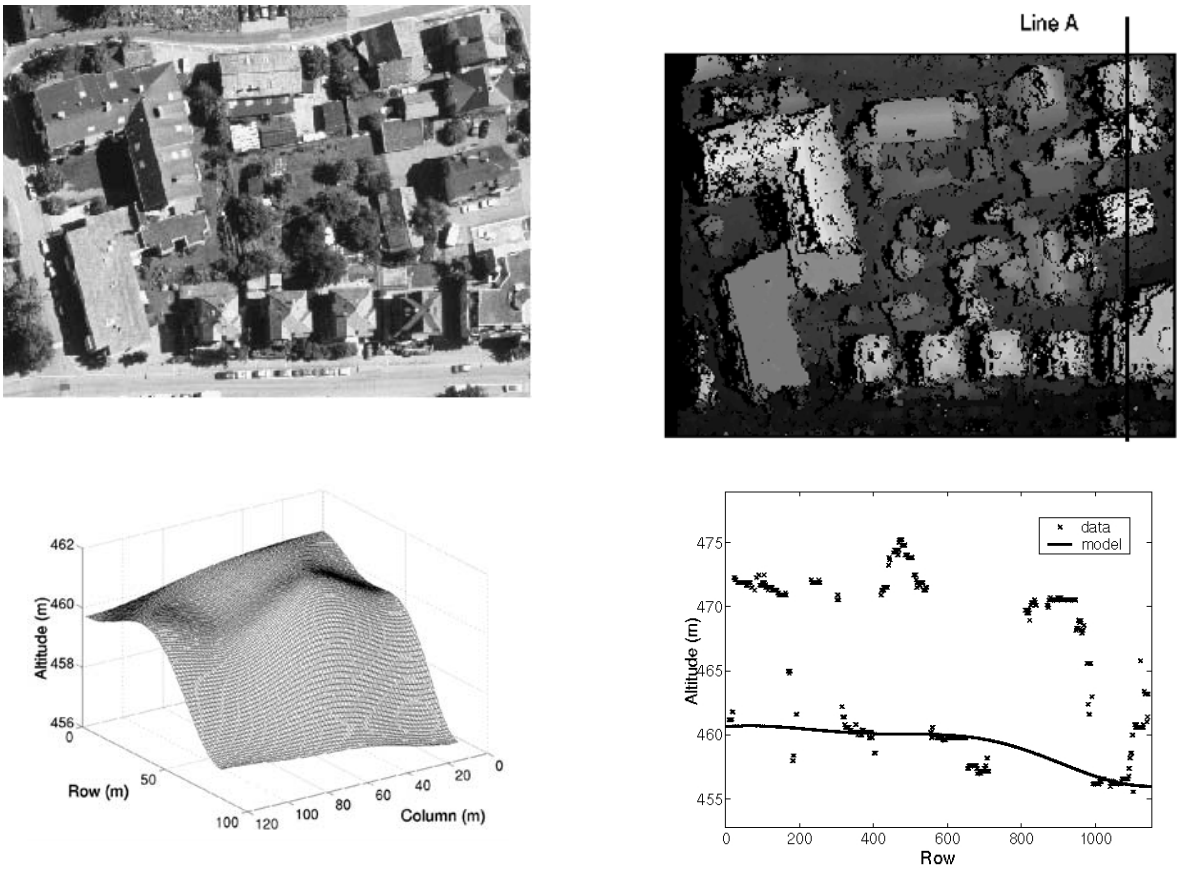

Fig. 5 – MNE et MNT estimé : données images couleurs à très haute résolution. Images de l'ETH Zürich, images 1660 x 1150, résolution 7cm au sol.

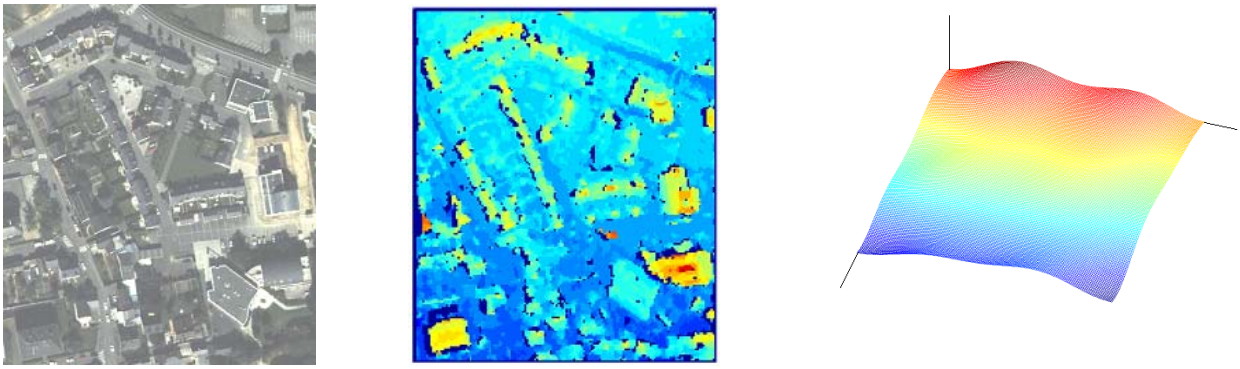

Fig. 6 – MNE et MNT (à droite, vue perspective) sur un extrait des images IGN de Rennes (résolution 35cm au sol).

4 Applications

L'estimation du MNT permet une classification binaire sol/sur-sol des points images. Cette classification est pilotée par un seuil réglé en fonction de l'altitude minimale recherchée pour une structure du sur-sol. On peut ainsi écarter du sur-sol les véhicules en stationnement ainsi que l'ensemble du mobilier urbain de faible altitude. Une simple agrégation des points sur un critère de connexité fournit ensuite un ensemble de régions du sur-sol sur la scène.

L'analyse du sur-sol consiste alors en la caractérisation de ces régions pour faire de la classification thématique (*cf.* section 4.1). Cependant, les données altimétriques corrigées peuvent également être utilisées directement dans un système de classification pixellaire multi-dimensionnel ou dans un système de recherche d'imagettes par similarité (*cf.* section 4.2).

4.1 Estimation de normales 3D par vote tensoriel

Médioni *et al.* ont proposé récemment une approche originale pour la reconstruction de surfaces 3D à partir de données bruitées (Medioni *et al.*, 2000). Le formalisme, très général, vise à fournir en tout point P de l'espace une information sur la vraisemblance de l'existence d'une surface passant par P, ainsi qu'une

estimation de la normale associée. L'approche utilise des tenseurs pour coder en chaque point l'information des normales et de l'incertitude. Des champs tensoriels sont introduits pour gérer l'influence entre les points. La phase d'accumulation utilise ces champs pour diffuser entre les points les informations.

Bien que nous visions initialement la reconstruction de surfaces, il nous a semblé intéressant d'utiliser cette approche pour caractériser les régions du sur-sol en bâti ou végétation.

En effet, à l'instar de l'information radiométrique (Dissard *et al*, 1997), l'utilisation d'informations altimétriques locales comme l'altitude et les normales peut aider à classer les différentes régions du sur-sol. Par exemple, Hug (Hug, 1997) étudie la répartition de l'angle azimutal des normales estimées en chaque point 3D de la région. Sous l'hypothèse que les régions « bâti » sont composées de quelques surfaces planes, les histogrammes doivent présenter des profils clairement différentiables selon le type de régions.

La principale difficulté réside ici dans l'estimation des normales. Nous avons utilisé l'approche tensorielle présentée ci-dessus pour effectuer une estimation fiable de la normale en chaque point. A l'instar de Hug, des histogrammes de l'angle azimutal ont été calculés. Les premiers résultats obtenus sur des régions du sur-sol (extraites de couples de Colombes (*cf.* section 2.1)) sont très encourageants (*cf*. figure 6).

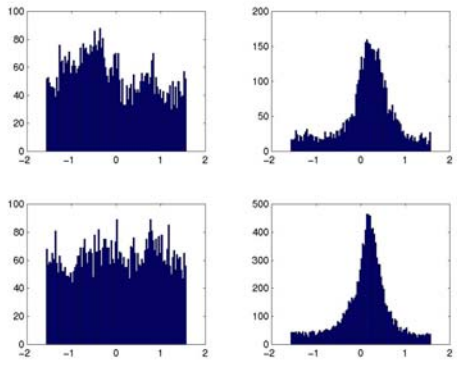

Fig. 6 – Histogrammes des angles azimutaux des normales à la surface 3D : à gauche, régions correspondant à de la végétation, à droite à du bâti (un seul pan de toit).

L'utilisation conjointe des normales et de l'information scalaire d'incertitude sur l'estimation (également disponible à la sortie du processus d'accumulation) devrait nous permettre de faire une classification très fiable des régions du sur-sol.

4.2 Recherche interactive de régions similaires

La classification d'images aériennes peut être réalisée soit sous forme de classification de régions, soit sous forme de classification pixellaire : chaque pixel reçoit une étiquette, élaborée à partir d'attributs locaux (Haala *et al*., 1998). Dans ce contexte, l'information pixellaire altimétrique brute du MNE n'est pas discriminante dès lors que l'altitude du sol varie significativement. L'utilisation de l'« altitude relative » (différence entre MNE et MNT) est en revanche pertinente.

Dans ce cadre, nous développons une application de recherche interactive de régions similaires, inspirée de systèmes utilisés pour l'indexation et la recherche d'images par le contenu (Smeulders *et al*., 2000). Le principe de cette application est le suivant :

– l'utilisateur construit une « requête » en désignant, sur une image, une région d'intérêt pour lui, correspondant à un objet particulier (bâtiment, arbre, *etc*.).
– le système présente ensuite à l'utilisateur des imagettes de régions similaires, sur la même image ou sur d'autres de la même banque d'images, et l'utilisateur peut éventuellement relancer la requête en l'affinant, par exemple en indiquant parmi les imagettes présentées celles qui sont pertinentes et celles qui ne le sont pas.

L'indice de similarité entre deux régions est établi sur les informations « couleurs » et « altitude relative » de chaque pixel, et calculé sur des blocs carrés de taille fixe.

L'information « altitude » est bien sûr prépondérante pour la reconnaissance du bâti, mais l'introduction de l'information couleur nous permet d'éliminer par exemple la végétation, de détecter les bâtiments d'aspect colorimétrique et de hauteur semblables. L'interaction est au cœur de ce genre de systèmes qui vise principalement à aider l'utilisateur à extraire dynamiquement des informations pertinentes dans un vaste ensemble de données hétérogènes.

5 Conclusion

Nous avons présenté ici un système d'analyse d'images aériennes à haute résolution, particulièrement adapté au traitement de zones urbaines.

A partir d'un couple stéréoscopique de photographies aériennes noir et blanc ou couleurs, nous calculons un modèle numérique d'élévations dense et fiable, préservant les discontinuités altimétriques. Une extension de ces travaux nous conduira à traiter des vues multiples de la même scène, afin de réduire le plus possible les parties cachées de la scène, sur lesquelles aucune information 3D ne peut être reconstruite.

De ce modèle, nous déclinons à la fois un modèle numérique de terrain – nous présentons ici un algorithme original à base d'analyse fréquentielle et d'estimation statistique robuste – et une classification des pixels ou des régions de l'image, de manière à séparer les objets de la scène (essentiellement bâti et végétation) du terrain nu.

La modélisation indépendante du terrain et de chaque bâtiment nous permet aujourd'hui d'envisager

des applications aussi diverses que la simulation de « survol virtuel » de zones urbaines (pour l'urbanisme et le tourisme), ou la constitution de bases vectorielles d'objets cartographiques pour les télécommunications, la cartographie, *etc*.

Références

Baltsavias, E.P., Gruen, A. & Van Gool, L. (eds.) (2001). Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). A.A. Balkema Publishers.

Beaton, A.E., & Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16:147-185.

Belli, T., Cord, M., & Philipp-Foliguet, S. (2000). Colour Contribution for Stereo Image Matching. *Proc. of Int. Conf. on Colour in Graphics and Image Processing, CGIP'2000*, pp.317-322, Saint-Etienne, France.

Bloch, I. (1996). Information Combination Operators for Data Fusion: A Comparative Review with Classification. *IEEE Trans. on Systems, Man and Cybernetics*, 26(1):52-67.

Brenner, C., Haala, N. & Fritsch, D. (2001). Towards fully automated 3D city model generation. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III),* A.A. Balkema Publishers, pp.47-57.

Cord, M., Jordan, M., & Cocquerez, J.-P. (2001). Accurate Building Structure Recovery from High Resolution Aerial Images. *CVIU journal*, 82(2):138-173.

Crouzil, A., Massip-Pailhès, L., & Castan, S. (1996). Mise en correspondance par corrélation des gradients. *Congrès RFIA*, AFCET, pp.236-245, Rennes.

Grün *et al*., editors (1998). Special issue on Automatic Building Extraction from Aerial Images. *Computer Vision and Image Understanding*, vol. 72, no.2.

Dissard, C., Baillard, C, Maître, H., & Jamet, O. (1997). Above-ground objects in urban scenes from medium scale aerial imagery. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp 183-192, Basel. Birkhaüser Verlag.

Frère, D., Hendrickx, M., Vandekerckhove, J., Moons, T. & Van Gool, L. (1997). On the reconstruction of urban house roofs from aerial images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II),* Birkhaüser Verlag, pp.87-96.

Girard, S., Guérin, P., Maître, H., & Roux, M. (1998). Building Detection from High Resolution Colour Images. *International Symposium on Remote Sensing, EUROPTO'98*, Barcelona.

Grün, A., Kübler, O., & Aggouris, P., editors (1995). *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Basel. Birkhaüser Verlag.

Grün, A., Baltsavias, E., & Henricsson, O., editors (1997). *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Basel. Birkhaüser Verlag.

Haala, N., Brenner, C., & Stätter, C. An Integrated System for Urban Model Generation. *Int. Archives of Photogrammetry and Remote Sensing*, vol. XXXII-2, pp.96-103.

Hanson, A.R., Marengoni, M., Schultz, H., Stolle, F. & Riseman, E.M. (2001). Ascender II: a framework for reconstruction of scenes from aerial images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III),* A.A. Balkema Publishers, pp.25-34.

Heuel, S. & Förstner, W. (2001). Topological and geometrical models for building reconstruction from multiple images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III),* A.A. Balkema Publishers, pp.13-24.

Huet, F., & Philipp, S. Fusion of Images Interpreted by a Multi-Scale Fuzzy Classification. *Pattern Analysis and Applications*, 1:231-247.

Hug, C. (1997). Extracting artificial surface objects from airborne laser scanner data. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Basel. Birkhaüser Verlag.

Kim, Z., Huertas, A. & Nevatia, R. (2001). A model-based approach for multi-view complex building description. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III),* A.A. Balkema Publishers, pp.181-193.

Koschan, A. (1996). Using perceptual attributes to obtain dense depth maps. *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, pp.155-159, Texas.

Lotti, J.-L. (1996). Mise en correspondance stéréo par fenêtres adaptatives en imagerie aérienne haute résolution. *Thèse de doctorat*, *Université de Nice Sophia-Antipolis*.

Mayer, H. (1999) Automatic Object Extraction from Aerial Imagery - A Survey Focusing on Buildings. *Computer Vision and Image Understanding*, 74(2):138-149.

Medioni, G., Lee, M.-S., & Tang, C.-K. (2000). A Computational Framework for Segmentation and Grouping. *Elsevier Science B.V.*, 1st edition.

Paparoditis, N., Cord, M., Jordan, M., & Cocquerez, J.-P. (1998). Building Detection and Reconstruction from Mid and High Resolution Aerial Images. *Computer Vision and Image Understanding*, 72(2):122-142.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content-based Image Retrieval at the End of the Early Years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1-27.

Yuan, T. and Subbarao, M. (1998). Integration of multiple-baseline color stereo vision with focus and defocus analysis for 3D shape measurement. *Proc. of SPIE Conf. on 3D Imaging, Optical Metrology and Inspection*, vol.3520, pp.44--51. SPIE.

View publication stats