

Détection et reconstruction de facettes 3D par approche hiérarchique par régions à partir de couples d'images satellite en milieu urbain

Nesrine Chehata^{1,2}, Franck Jung¹, Marc Pierrot-Deseilligny¹, Georges Stamon²

¹ Institut Géographique National

2-4 Av. Pasteur 94165 St. Mandé cedex, France

Email : prénom.nom@ign.fr

² Laboratoire SIP/CRIP5, Systèmes Intelligents de Perception

UFR mathématiques et informatique, Université Paris 5.

45 Rue des Saints Pères 75006 Paris, France.

Tél : int+ 33 1 44 55 35 60

E-mail : prénom.nom@math-info.univ-paris5.fr

Résumé:

Ce travail s'insère dans un projet global d'interprétation de scènes urbaines à partir d'images satellites haute résolution et de reconstruction 3D des objets géographiques. L'interprétation porte sur deux thèmes, le bâti et les routes. Le travail présenté dans cet article traite uniquement le thème du bâti, les routes permettant d'avoir des zones de focalisation. Les algorithmes développés seront adaptés aux données images du futur satellite PLEIADE HR, le successeur de SPOT5. On disposera en entrée de couples d'images satellites panchromatiques, de haute résolution : 45-70 cm, ayant un faible rapport Base sur Hauteur B/H=0.05, 0.2. Cette étude détaille la reconstruction automatique de facettes 3D par une approche région basée sur des segmentations hiérarchiques des deux images. Dans la première partie, le processus de la segmentation est expliqué, ensuite la mise en correspondance des régions est détaillée. Dans la troisième partie, l'appariement global des deux hiérarchies est traité, ceci permettra d'aboutir aux associations finales pour la reconstruction de la scène 3D.

Mots-clés: images satellites haute résolution, stéréoscopie, reconstruction facettes 3D, système automatique, segmentation hiérarchique en régions, appariement.

1 CONTEXTE

Depuis le lancement des satellites commerciaux [Ikonos, Eros, QuickBird], on dispose de nos jours, d'images satellite de haute résolution de l'ordre de [60 cm, 1m]. Ces nouvelles données ont relancé les recherches scientifiques sur la reconstruction 3D des bâtiments et l'extraction des routes à partir d'images satellites HR. On s'intéresse dans notre étude, au futur satellite PLEIADE HR, successeur de SPOT5 et à la reconstruction 3D de bâtiments à partir de couples stéréoscopiques simples. On disposera de couples stéréoscopiques panchromatiques à [45-70 cm] de résolution et de faible rapport Base sur Hauteur B/H [0.05-0.2]. Notre objectif est de proposer un système automatique de reconstruction de bâti en contexte satellitaire. On a choisi pour cela une stratégie ascendante, basée sur l'extraction de primitives 3D fiables (segments 3D, surfaces planes) qui constitue une étape préliminaire importante pour la génération des hypothèses de bâtiments.

On s'intéresse dans ce papier, à l'extraction de facettes 3D fiables correspondant à des toits de bâtiments. Les algorithmes développés sont dédiés à un contexte satellitaire, et doivent donc être robustes au bruit et au faible rapport B/H des couples stéréoscopiques qui est le paramètre critique dans notre cas.

2 ETAT DE L'ART

Les travaux sur la reconstruction 3D des bâtiments portent essentiellement sur les images aériennes faible ou haute résolution [8cm -50 cm] avec des rapports B/H élevés pour les couples de l'ordre de [0.4-0.6]. On s'intéresse ici à la reconstruction des facettes 3D. De nombreux auteurs ont montré l'importance des surfaces planes pour décrire les scènes urbaines [17]. Certains travaux proposent d'extraire des plans 3D à partir de nuages de points 3D obtenus par focalisation sur un MNE (Modèle numérique d'élévation), soit par une approche bayésienne [3], soit en utilisant la transformée de Hough 3D pour extraire des hypothèses de plans 3D [6]. Dans [2], on recherche des demi-plans 3D pivotant autour des segments 3D extraits. D'autres travaux consistent à rechercher des rectangles à partir des images [9] ou du MNE par des processus stochastiques [8] ou bien en combinant le MNE avec des orthoimages [15] [14]. Les rectangles délimitent les bâtiments, la reconstruction 3D est obtenue en modélisant les toits par des modèles plus ou moins complexes et en récupérant les altitudes moyennes à partir du MNE. D'autres approches sont basées sur une segmentation en régions à partir des images ou du MNE [17] ou bien utilisant le MNE et une orthoimage de gradients [16]. Dans le contexte satellitaire, peu de travaux traitent le thème de reconstruction de bâti. Dans [4], l'auteur présente des évaluations qualitatives et quantitatives sur des

images IKONOS concernant les aspects radiométriques et les précisions de reconstruction de bâti avec des méthodes existantes. Dans [13], l'auteur propose une méthode de détection des bâtiments en 2D à partir d'images satellites en recherchant les directions dominantes par analyse de Fourier.

3 MÉTHODOLOGIE

Pour la reconstruction des facettes 3D, on a choisi une approche région basée sur la segmentation des images et la mise en correspondance des régions obtenues. Tout d'abord, les régions constituent des primitives plus fiables que les segments. En effet, leurs tailles étant plus grandes que la vignette de corrélation, les appariements sont plus fiables. Ceci permet d'effectuer des estimations robustes des plans 3D reconstruits et de corriger ainsi l'imprécision altimétrique due au faible rapport B/H. En outre, cette approche permet de reconstruire directement des surfaces tridimensionnelles, ce qui fournit un modèle 3-D beaucoup plus dense que ceux obtenus à partir d'une analyse en points caractéristiques ou en segments des deux images de la scène. Enfin, la segmentation en régions apporte une information topologique qui permet de propager certaines contraintes entre régions adjacentes et donc de mieux structurer la scène 3D.

D'autre part, on a opté pour une segmentation des images et non du MNE. Compte tenu du contexte satellitaire, le MNE obtenu par des techniques de corrélation [1], présente beaucoup d'inconvénients : les bords des bâtiments ne sont pas précis, et la précision altimétrique est réduite à cause du faible rapport B/H. Par conséquent, dans notre approche, on exclut l'utilisation du MNE pour la reconstruction des plans 3D, on utilisera plutôt des segmentations sur les images. Enfin, le faible rapport B/H du couple stéréoscopique présente l'avantage de réduire les distorsions géométriques entre les images ainsi que le problème des parties cachées, ce qui facilitera l'appariement des régions entre les deux images.

L'approche globale est illustrée dans la Figure 1. Trois grandes parties se détachent : la segmentation des images, l'appariement des régions et la reconstruction des facettes 3D.

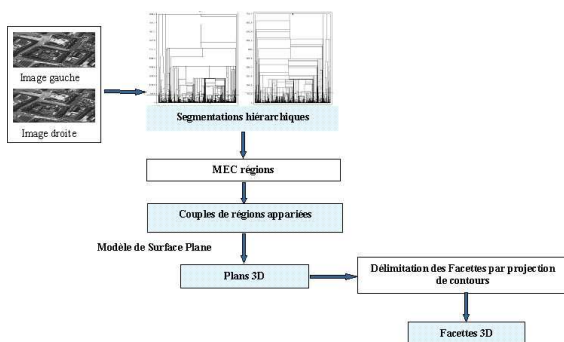


FIG. 1 – Reconstruction des facettes 3D par approche région

Tout d'abord, les deux images sont segmentées par des outils de segmentation multi-échelles que l'on décrira dans la première partie. La deuxième phase est la mise en correspondance des régions. Cette phase est elle-même composée de trois parties. Tout d'abord, les régions de chaque hiérarchie sont appariées avec les régions de la seconde hiérarchie.

Par la suite, on cherche à réaliser une coupe pour chaque hiérarchie qui renvoie le maximum de régions 2D interprétables au niveau de la scène urbaine et qui correspondent à de véritables facettes (toits de bâtiments). Un appariement global des deux coupes renvoie les meilleurs couples de régions appariées. Finalement, en supposant un modèle de surface plane, la dernière phase consiste à reconstruire les plans 3D correspondants aux couples de régions appariées et les facettes 3D finales sont obtenues par projection des contours des régions appariées.

4 SEGMENTATION MULTI-ÉCHELLES ET HIÉRARCHIE

La segmentation d'images bas-niveau pose beaucoup de difficultés du fait qu'il existe différents niveaux d'interprétation d'une image (pixel, toit, bâtiment, ville...). De plus, notre objectif étant d'apparier les deux segmentations, on préfère utiliser une segmentation hiérarchique [5], à plusieurs niveaux de segmentation, allant des régions les plus fines aux régions les plus grossières. Ainsi les appariements peuvent s'effectuer entre plusieurs niveaux afin d'obtenir les meilleures segmentations appariées dans les deux images. En vue d'apparier des segmentations, [11] propose une méthode de division récursive de l'image pilotée par la mise en correspondance des régions, le but étant de produire deux segmentations en régions aussi similaires que possible. Dans notre approche, compte tenu du faible rapport B/H du couple, les distorsions géométriques et les parties cachées entre les deux images sont réduites. Par conséquent, les deux segmentations sont assez similaires, ce qui facilitera l'appariement des régions par la suite. La mise en correspondance des régions permettra, comme on le verra dans la section 5.3 de choisir les coupes optimales dans les hiérarchies en vue de l'appariement. Donc, on commence tout d'abord par segmenter les images par un algorithme de segmentation multi-échelles, décrit dans [5]. Cet algorithme produit des segmentations de l'image $\Omega_0, \Omega_1, \dots, \Omega_n$ par une approche ascendante ou encore "Bottom-Up". En partant d'une sursegmentation initiale (par l'algorithme de partage des eaux), la hiérarchie de régions est construite par fusions successives des régions initiales suivant le modèle décrit ci-dessous.

Modèle de fusion des régions Le but est de segmenter l'image en régions correspondant à de véritables facettes 3D donc principalement des toits dans notre cas. Ces régions sont en général de forme simple et de radiométrie homogène. Pour cela, on part d'une formulation classique de la segmentation dans laquelle deux énergies sont mises en compétition : une énergie géométrique et une énergie radiométrique. L'objectif étant de trouver la partition qui minimise l'énergie globale pour un λ donné.

$$E_\lambda(P) = \sum_{R \in P} E_\lambda(R) = \sum_{R \in P} \lambda E_{\text{Géo}}(R) + E_{\text{Rad}}(R) \quad (1)$$

Le λ est un terme de régularisation : pour un fort λ , l'énergie géométrique sera prédominante conduisant à un modèle simple (peu de régions avec des formes simples), alors que si λ est faible on s'attache plus à l'énergie radiométrique et on obtient donc de plus petites régions homogènes vérifiant le modèle radiométrique (cf. Fig 2).

En outre, les deux énergies géométrique et radiométrique sont elles-mêmes composées d'une énergie de régularisation (modèle), C , et une énergie d'attache au données, D ,

$$\begin{cases} E_{Géo}(R) = W_G C_{Géo}(R) + D_{Géo}(R) \\ E_{Rad}(R) = W_R C_{Rad}(R) + D_{Rad}(R) \end{cases} \quad (2)$$

W_G et W_R constituent des termes de régularisation des deux énergies respectives. En pratique, on utilise comme modèle géométrique un coût linéaire en fonction du gradient de l'image. Pour le modèle radiométrique, on utilise un codage MDL de Gaussienne [7].

Le principe de la segmentation multi-échelles est de rechercher une famille de solutions, pour $\lambda \in \mathbb{R}^+$ qui minimise l'énergie globale et d'imposer que ces solutions forment une séquence de partitions emboîtées, modélisant un processus de description multi-échelles.

Au niveau pratique, l'idée est de construire une hiérarchie de régions d'une image à différentes échelles en augmentant à chaque fois l'échelle globale λ et en fusionnant pour cela les régions adjacentes dont la fusion diminue l'énergie globale. Pour la fusion de deux régions adjacentes R_1 et R_2 en une seule région R , on définit λ_{app} l'échelle d'apparition de la région R , qui est la valeur minimale à partir de laquelle on a :

$$E(R) < E(R_1) + E(R_2) \text{ avec } R = R_1 \cup R_2 \quad (3)$$

On obtient ainsi une hiérarchie où chaque noeud est indexé par son échelle d'apparition et correspond à une région définie par la fusion d'un ensemble de régions de base connectées. On définit aussi pour chaque noeud une échelle de disparition λ_{disp} qui correspond en fait au λ_{app} de la région mère correspondante. Ainsi, à chaque noeud correspond un intervalle $[\lambda_{app}, \lambda_{disp}]$ dans lequel la région R correspondante est visible. Autrement dit, toute coupe de la hiérarchie dans cet intervalle contiendra la région R .

La Figure 2, montre un exemple de hiérarchie. La hiérarchie est un arbre dont les caractéristiques sont décrites ci-dessous :

- Les feuilles sont les noeuds de base obtenus par sur-segmentation de l'image, la racine (i.e le sommet) est l'image entière.
- Suite au processus de fusion, les régions sont de plus en plus grandes de bas en haut.
- Une coupe horizontale suivant λ fournit une partition de l'image [5](cf.Fig 2)

5 MISE EN CORRESPONDANCE DES RÉGIONS

La mise en correspondance des régions a pour but d'apparier globalement les deux hiérarchies et est composée de trois étapes. La première étape consiste à apparier au mieux les régions de chaque hiérarchie avec la seconde hiérarchie. On disposera ainsi de correspondances 1 à N, le but étant de toujours avoir, pour une région apparierée, le bon homologue parmi la liste d'homologues possibles. Par la suite, on cherchera à réaliser des coupes dans les deux hiérarchies qui renvoient des régions correspondants à de véritables facettes 3D pour reconstruire la scène. Finalement, il faudra combiner toutes les associations afin d'aboutir aux meilleurs appariements des deux coupes au vue d'une bonne caricature de la scène.

Dans la suite de l'étude, on adoptera les notations suivantes : Ω_g et Ω_d désignent l'ensemble des régions de l'image gauche et respectivement de l'image droite. L'ensemble des homologues d'une région donnée R sera noté $H(R)$.

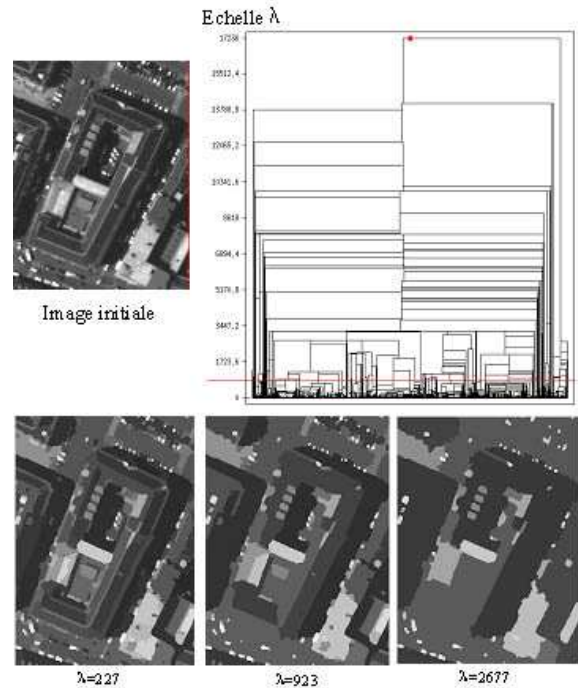


FIG. 2 – Segmentation hiérarchique

5.1 Appariement d'une région

Dans un premier temps, les régions d'une hiérarchie sont apparierées indépendamment les unes des autres avec les régions de la seconde pyramide. Le processus est réalisé sur les deux pyramides de façon symétrique. L'appariement de régions est un problème classique en vision [12][11]. Pour cette étape d'appariement, on utilise plusieurs contraintes pour réduire le nombre d'homologues à traiter ; contrainte de taille, contrainte épipolaire, contrainte altimétrique, contrainte de similarité. Ces contraintes sont détaillées ci-dessous. L'originalité de notre approche réside dans la propagation ascendante des contraintes d'appariement dans les hiérarchies. Ceci rend la recherche d'homologues plus robuste comme on le verra dans ce qui suit.

Contrainte de taille Cette contrainte est utilisée tout au long du processus pour éviter de traiter des régions jugées trop petites ou trop grandes. En fonction du niveau d'interprétation souhaité (toit, cheminées, chiens assis...) et de la résolution de l'image, on impose un seuil minimal et maximal pour la taille des régions, S_{min}^t et S_{max}^t .

Contrainte épipolaire et altimétrique Pour rechercher la liste d'homologues possibles pour une région donnée, on commence par appliquer la contrainte épipolaire sur le barycentre G de la région. Il faut que le barycentre de la région homologue appartienne à un bandeau épipolaire obtenu avec une tolérance préfixée par rapport à la ligne épipolaire correspondante à G qu'on notera E_G . L'ensemble des régions homologues, d'une région donnée R , par la contrainte épipolaire sera noté $HEpi(R)$.

Comme le montre la figure 3, afin de réduire l'espace de recherche, on utilise également une contrainte sur les al-

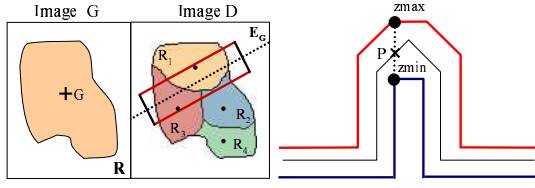


FIG. 3 – Contrainte épipolaire et altimétrique

titudes Min et Max de la région. On dispose en entrée d'un Modèle Numérique d'Élévation (obtenu par des techniques de corrélation) [1] en $2D\frac{1}{2}$. Le MNE obtenu n'est pas précis à cause du faible B/H mais robuste. Par conséquent, grâce à un volume dilaté du MNE, on dispose d'une bande de tolérance autour qui renvoie l'altitude Min et Max de chaque pixel. Cette bande de tolérance permettra de restreindre l'espace de recherche. La recherche des homologues épipolaires revient ainsi à retenir tous les noeuds de la seconde hiérarchie dont les barycentres sont contenus dans un rectangle épipolaire (cf. Fig 3).

Cependant, la contrainte épipolaire ainsi appliquée présente deux principaux inconvénients :

1. Perte de régions homologues
 - Si les régions à appairer sont segmentées différemment dans les images (cf. Figure 3).
 - Dans le cas d'un barycentre de référence G excentré, on peut perdre des régions homologues si leurs barycentres ne sont pas contenus dans le rectangle épipolaire correspondant à G .
2. La contrainte n'est pas robuste au bruit. En effet des petites régions peuvent se rajouter au cours des fusions successives et font apparaître ainsi des régions parasites parmi les homologues possibles.

Pour éviter ces problèmes, la contrainte épipolaire sera appliquée sur les noeuds de base de la hiérarchie et propagée de bas en haut de façon itérative. Un noeud père est considéré comme homologue d'un noeud de référence par la contrainte épipolaire si tous ses fils vérifient cette propriété. On évite ainsi d'une part de perdre des régions homologues et d'autre part de traiter les régions parasites dues au bruit.

Contrainte de similarité A ce niveau, chaque région R a une liste d'homologues $HEpi(R)$ dans la seconde hiérarchie. Pour réduire la combinatoire, on utilise une contrainte de similarité [12]. L'ensemble des régions homologues de R par la contrainte de similarité sera noté $HSim(R)$.

$$Sim(R_g, R_d) = \sum_{p=1}^q w_p s_p(R_g, R_d) \quad (4)$$

s_p est le score de similarité pour chaque paramètre ou attribut de la région. w_p est le poids affecté au score de similarité s_p défini par l'équation suivante. Si les attributs des régions sont similaires, le score tend vers 0.

$$s_p(R_g, R_d) = 1 - \frac{\min(A_p(R_g), A_p(R_d))}{\max(A_p(R_g), A_p(R_d))} \quad (5)$$

A_p étant un attribut de la région. Dans notre cas, on utilise trois attributs qui sont la surface, les dimensions du rectangle englobant et le moment spatial d'ordre 2. On ne prend pas en compte les caractéristiques radiométriques des

régions à cette étape. Le score de similarité doit être supérieur à un seuil préfixé S^s . Le seuil de similarité est un paramètre critique, difficile à évaluer puisqu'il dépend de la ressemblance des segmentations des deux images. En effet, si les segmentations sont très semblables, les scores de similarité tendent vers 0. Ceci est bien le cas pour des couples de faible B/H. En pratique, on choisit un seuil tolérant $S^s = 0.3$ pour éviter de perdre des régions homologues.

Contrainte de recouvrement Le principe de la contrainte de recouvrement est illustré figure 4. On commence par estimer le plan $\Pi_{(R_g, R_d)}$ correspondant au couple de régions appariées. La phase de reconstruction du plan 3D sera détaillée dans la section 5.2. En projetant les deux régions sur le plan, on obtient deux régions rectifiées R_g^π et R_d^π . Le recouvrement est égal au rapport de l'intersection des deux régions rectifiées sur leur union. Pour valider l'appariement, il faut que le recouvrement des deux régions projetées soit supérieur à un seuil S^r . Ce seuil est plus facile à interpréter que le seuil de similarité puisqu'il ne dépend pas des segmentations. On choisit un seuil strict $S^r = 60\%$.

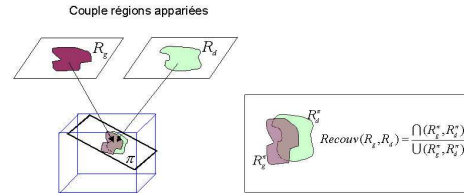


FIG. 4 – Contrainte de recouvrement

Formulation de la relation d'appariement On définit la relation d'appariement \mathcal{H} entre deux régions comme suit :

$$\forall R_g \in \Omega_g, \forall R_d \in \Omega_d \quad R_g \mathcal{H} R_d \iff$$

$$\begin{cases} 1. S_{min}^t \leq R_g \leq S_{max}^t \text{ et } S_{min}^t \leq R_d \leq S_{max}^t \\ 2. R_d \in HEpi(R_g) \\ 3. R_d \in HSim(R_g) \iff Sim(R_g, R_d) \geq S^s \\ 4. R_d \in HRecouv(R_g) \iff Recouv(R_g, R_d) \geq S^r \end{cases} \quad (6)$$

Les régions homologues doivent ainsi vérifier la contrainte de taille, la contrainte épipolaire, les contraintes de similarité et de recouvrement.

5.2 Reconstruction des plans 3D à partir de deux régions appariées

La reconstruction du plan 3D d'un couple de régions appariées est indispensable pour calculer le recouvrement des régions, pour qualifier la qualité de l'appariement comme on le verra dans la section 5.3, et bien sûr pour la reconstruction finale de la scène 3D.

La reconstruction du plan 3D peut être photométrique, en utilisant des attributs radiométriques des régions ou bien géométriques[18]. Pour notre approche, les plans 3D doivent permettre, dans un premier temps, de qualifier les appariements. Donc, la reconstruction doit être discriminante, c'est pour cette raison que l'on évite les attributs radiométriques ou géométriques qui ne permettent pas de privilégier une région mère par rapport à ses fils ou le contraire. On choisit pour notre approche de reconstruire les

plans 3D à partir de la mise en correspondance des contours des deux régions qui est dépendante de la qualité de segmentation et donc plus discriminante.

Mise en correspondance des contours des régions On rééchantillonne les contours en épipolaire de telle façon que deux points de contour homologues (u_g, v_g) et (u_d, v_d) se trouvent sur la même ligne épipolaire donc $v_g = v_d = v$. On parcourt toutes les lignes communes aux deux contours et on apparie les points de contour sur chaque ligne en respectant la contrainte d'ordre. L'appariement des points se fait par corrélation. Dans le contexte satellitaire, pour des couples à faible B/H, il est indispensable d'avoir une précision subpixelaire au niveau de l'appariement. Si on veut, par exemple avoir une précision du $\frac{1}{10}$ du pixel, on rééchantillonne le pixel homologue ; on recalcule le score de corrélation sur des fenêtres décalées à chaque fois de $\frac{1}{10}$ de pixel sur la deuxième image. Les valeurs dans la vignette sont recalculées par un interpolateur bicubique. On obtient alors une courbe de corrélation pour un pixel homologue, le pic maximum renvoie une position subpixelaire du point homologue. La corrélation subpixelaire est indispensable pour des couples à faible B/H pour améliorer la précision altimétrique de reconstruction.

Reconstruction des plans 3D L'estimation des plans 3D est réalisée, comme la mise en correspondance des contours, dans l'espace des disparités. Dans l'espace rectifié, l'équation du plan support de la face est supposée de la forme suivante :

$$au_g + bu_d + cv + d = 0 \quad (7)$$

Par conséquent, un point de l'image gauche (u_g, v) est visible aussi dans l'image droite aux coordonnées $(-\frac{au_g + cv + d}{b}, v)$. La relation est linéaire et la dépendance avec les caractéristiques du plan est simple :

$$\begin{pmatrix} u_d \\ v \end{pmatrix} = \begin{pmatrix} \frac{a}{b} & \frac{c}{b} \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} u_g \\ v \end{pmatrix} + \begin{pmatrix} \frac{d}{b} \\ 0 \end{pmatrix} \quad (8)$$

Le calcul des paramètres du plan se fait par des M-estimateurs L_p . En pratique, $p=1.2$ représente un bon compromis entre complexité et stabilité [19]. Bien que moins robustes que d'autres estimateurs (RANSAC, moindres carrés médians), les M-estimateurs donnent de bons résultats avec une bonne initialisation. Dans notre cas, l'initialisation dépend des régions appariées, donc le comportement des M-estimateurs permettra entre autres de qualifier le processus d'appariement.

L'équation 8, peut être aussi utilisée pour réaliser une reconstruction photométrique des facettes en injectant dans l'équation des attributs des deux régions et non les points de contours appariés [18]. La reconstruction photométrique peut être utilisée pour affiner les facettes 3D finales.

5.3 Recherche de la coupe optimale dans une hiérarchie

En fonction de la qualité des appariements, on cherche une coupe optimale dans une hiérarchie qui permet d'avoir la meilleure segmentation de l'image, ç.à.d des régions qui correspondent aux toits de bâtiments. Contrairement à la segmentation, l'appariement est réalisé par une approche descendante. L'appariement est en effet plus fiable sur des grandes régions, on commence donc par le niveau le plus grossier. Chaque appariement est qualifié par le score de corrélation des deux régions appariées. On utilise pour cela un volume de corrélation dans l'espace objet. Le calcul du

volume de corrélation ainsi que la qualification des appariements sont détaillés ci-dessous.

Volume de corrélation Le volume de corrélation (cf. Fig 5(a)) est calculé dans l'espace objet, en coordonnées géographiques[10]. L'emprise planimétrique de ce volume correspond à la zone de recouvrement du couple stéréoscopique. On estime ensuite un intervalle d'altitude en utilisant un modèle numérique d'élévation (MNE) calculé sur le couple, en calculant une valeur médiane et une incertitude. Une fois le volume délimité, on attribue à chacun de ses voxels un score d'attache aux données quantifiant l'appartenance de ce voxel à la scène 3D que l'on cherche à reconstruire. Pour cela, on reprojette le voxel et ses voisins horizontaux sur une fenêtre de taille fixée, et on mesure la ressemblance des images par le coefficient de corrélation normalisé centré. On a choisi comme taille de fenêtre 3x3.

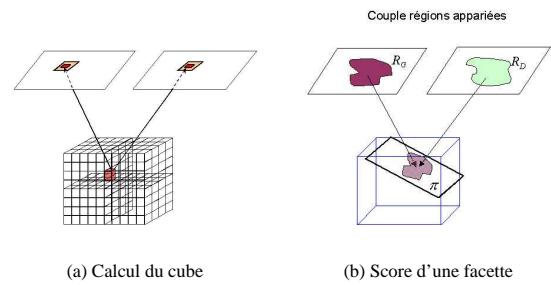


FIG. 5 – Cube de corrélation

Qualification de l'appariement de deux régions Un couple de régions appariées est qualifié par le score de corrélation moyen de la facette 3D correspondante. En effet, après la reconstruction du plan $\Pi_{(R_g, R_d)}$, on projette les deux régions R_g et R_d sur le plan 3D. L'intersection des deux régions projetées renvoie la délimitation de la facette 3D que l'on note $\mathcal{F}_{(R_g, R_d)}$. La facette est validée si et seulement si elle appartient à une bande de tolérance autour du MNE. Par la suite, pour chaque voxel de la facette 3D, on récupère le score de corrélation dans le cube précalculé (cf. Fig 5(b)). La moyenne de tous les voxels de la facette renvoie le score de corrélation qui qualifie l'appariement.

$$Q(R_g, R_d) = \frac{\sum_{\mathcal{V} \in \mathcal{F}} \text{ScoreCorrel}(\mathcal{V})}{\text{Card}(\mathcal{F}_{(R_g, R_d)})} \quad (9)$$

$\mathcal{F}_{(R_g, R_d)}$ étant considérée comme un ensemble de voxels \mathcal{V} .

L'avantage d'utiliser le cube de corrélation est le gain de temps puisqu'on calcule les scores de corrélation une seule fois pour tous les voxels du cube. Cette méthode donne de bons résultats avec des faibles B/H puisqu'il n'y a pas beaucoup de distorsions entre les images. En effet, dans notre contexte, les vignettes horizontales suffisent pour qualifier les appariements. On n'a pas besoin de rectifier les régions pour les corrélérer dans l'espace objet d'où un gain de temps important.

Calcul de la coupe optimale Le calcul de la coupe optimale est détaillé dans ce qui suit pour une hiérarchie. En pratique, le processus est appliqué aux deux hiérarchies afin

de traiter le couple de façon symétrique et d'obtenir des appariements plus fiables. On définit pour chaque région R , une qualité $Q(R)$ qui qualifie l'ensemble de ses appariements et qui correspond au maximum des scores de corrélation renvoyés avec l'ensemble de ses régions homologues.

$$\begin{cases} Q(R) = \text{Max}(Q(R, R'_i)) \quad R'_i \in H(R) \\ i = 0, \dots, \text{Card}(H(R)) \end{cases} \quad (10)$$

La recherche de la coupe optimale se fait par une approche descendante. On s'arrête à un niveau donné si les deux conditions suivantes sont réalisées.

Condition 1. La qualité des appariements de la région est supérieure ou égale à celles de tous ses fils. Ce qui se traduit par :

$$Q(R) \geq Q(R_f) \quad \forall R_f \in \text{Fils}(R) \quad (11)$$

Condition 2. Les plans 3D correspondant aux fils de la région R sont quasi-coplanaires.

On utilise l'écart angulaire entre les normales des plans pour comparer leurs pentes respectives.

Le choix d'un seuil sur la qualité des appariements d'une région étant critique, la coupe se fait par un critère relatif où l'on compare la qualité d'une région par rapport à celles de ses fils. Si la région mère a une meilleure qualité que tous ses fils, il faut vérifier que les fils renvoient des plans quasi-coplanaires ce qui veut dire qu'ils constituent une sursegmentation du toit et qu'il faut s'arrêter au niveau de la région mère.

Il faut remarquer que la qualité de la coupe dépend fortement de l'estimation des plans 3D. On peut aboutir à des coupes sursegmentées si les plans 3D sont mal estimés. En effet, pour un fils donné, une erreur d'estimation de la pente d'un plan entraîne un grand écart angulaire entre ses plans, d'où la descente à un sous-niveau de la hiérarchie et la sursegmentation. Cet inconvénient peut être résolu soit par l'étape de fusion des coupes des deux hiérarchies soit par une étape finale de fusions des plans de la scène 3D finale.

5.4 Appariement global des coupes des hiérarchies

Une fois les coupes, dans les deux hiérarchies, calculées, on cherche à combiner les résultats afin d'obtenir les meilleurs appariements. Les difficultés que l'on peut rencontrer dans l'appariement global des hiérarchies sont principalement la différence des segmentations des deux images, les occultations et les parties cachées. Ces problèmes sont toutes fois réduits dans le cas de couples à B/H faible. Tout d'abord, l'ensemble des appariements au niveau des deux coupes est fusionné entre les deux hiérarchies, ensuite un ensemble d'appariements consistant est retenu suivant le schéma de "the winner takes all" en privilégiant la qualité de l'appariement, ensuite le score de recouvrement et enfin le score de similarité.

6 RÉSULTATS

Les premiers résultats sont présentés sur des images simulées par le CNES/QTIS* sur le site de Toulouse à 60 cm de résolution et un B/H=0.08 (Figure 7) ou sur des images aériennes avec des B/H=0.2 sur le site d'Amiens à 50 cm

(Figure 6). Ces dernières étant fournies par l'IGN †. Les résultats numériques présentés ont été obtenus sur 4 couples d'images à différentes résolutions, et différents B/H.

6.1 Les contraintes d'appariement

Contraintes	Similarité	Recouvrement
% régions appariées	0.82	0.23
Nbre Moyen homologues	3.99	2.09

TAB. 1 – Contraintes d'appariement des régions

On remarque qu'à l'issue de toutes les contraintes, uniquement 23% des régions sont appariées. La contrainte de recouvrement étant nettement plus stricte. Le nombre moyen d'homologues pour chaque région est de l'ordre de 2, ce qui réduit considérablement l'étape de fusion des appariements par la suite.

6.2 La coupe dans la hiérarchie

La recherche de la coupe permet de réduire le nombre de régions à traiter pour l'appariement global des deux hiérarchies et permet de choisir le niveau d'interprétation souhaité. D'après le tableau 2-a, on a 28% des régions appariées qui sont retenues pour la coupe dont 38% de régions fiables c.à.d à un seul homologue. Pour les noeuds à plusieurs homologues, dans 97% des cas, les homologues se trouvent sur la même branche de la hiérarchie c.à.d qu'ils ont une relation de dépendance : mère-fils, ce qui montre que l'appariement est fiable.

Régions de la coupe/ Régions appariées 28%	
a) Régions "fi ables" 1 région homologue 38%	Régions à plusieurs homologues 62%
b) Régions planes valides 75%	Régions fausses 25%

TAB. 2 – Evaluation de la coupe optimale

Le tableau 2-b montre que 75% des régions retenues correspondent à de véritables toits de bâtiments. Les toits manquants sont estimés à 14%. Pour pallier au problème de perte de régions, on peut propager les appariements par adjacence de régions afin d'obtenir des coupes denses par la suite. On peut aussi intervenir au niveau de la reconstruction finale, en récupérant des altitudes médianes à partir du MNE ou bien introduire des modèles de toits tel que les toits symétriques par exemple.

6.3 Appariement global des hiérarchies

Les figures 6 et 7 montrent les résultats de l'appariement global sur des couples d'images d'Amiens à 45 cm et de Toulouse à 60 cm. Afin de se focaliser sur les bâtiments, une classification préliminaire sol/sursol a été effectuée sur les images pour éliminer les régions du sol. Les sursegmentations obtenues par endroits sont dues aux différences de segmentations entre les deux images. En effet, si les segmentations ne sont pas similaires, les régions ne corrélaient pas bien et il faut descendre à un niveau plus bas

*Centre National d'Etudes Spatiales, Toulouse
QTIS : Qualité et Traitement des Images Spatiales.

†Institut Géographique National-St Mandé

de la hiérarchie pour retrouver des segmentations similaires d'où une sursegmentation.

Les premiers résultats sont satisfaisants, les principaux toits sont présents et la scène est bien caricaturée.

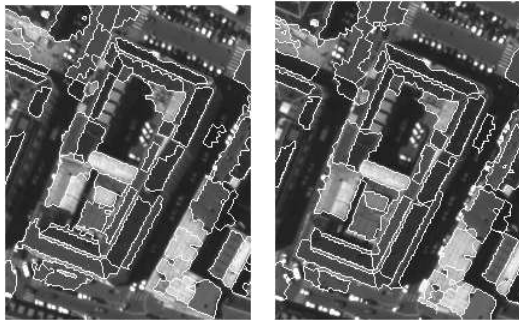


FIG. 6 – Amiens - 45cm - B/H 0.2

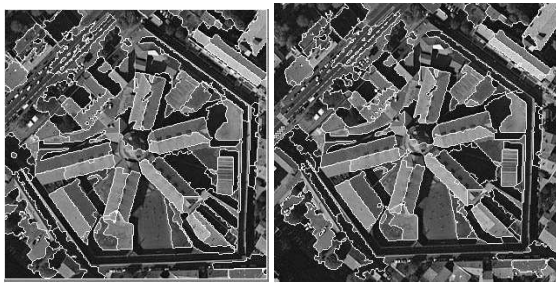


FIG. 7 – Toulouse - 60 cm - B/H 0.08

Une évaluation de l'approche ainsi que de la qualité de reconstruction des plans 3D en fonction de la résolution de l'image et du rapport B/H sont en cours d'étude.

7 CONCLUSIONS

On a présenté dans cette étude une méthode de détection et de reconstruction de facettes 3D adaptée au contexte satellitaire. L'originalité de notre approche consiste à appairer des primitives 2D entre deux images pour émettre des hypothèses de facettes 3D en milieu urbain et d'adapter chaque étape au contexte satellitaire. En effet, le couple stéréoscopique a un faible rapport B/H ce qui entraîne une perte de précision altimétrique lors de la reconstruction 3D, d'où l'utilisation des régions qui constituent des primitives plus fiables et plus adaptées à des résolutions de l'ordre de 60 cm. Cependant le couple a l'avantage de présenter de faibles distorsions entre les images d'où l'idée d'appairer deux segmentations des images qui sont initialement assez similaires.

La méthode de reconstruction de facettes 3D proposée se base sur trois grandes parties classiques qui tiennent compte à chaque étape du faible B/H du couple et qui traitent de façon symétrique le couple d'images. Tout d'abord, la segmentation hiérarchique où l'on utilise des critères de fusion adaptés pour obtenir des régions correspondants aux toits de bâtiments et qui permet surtout de choisir le niveau d'interprétation souhaité grâce à des coupes dans la hiérarchie. Ensuite on traite le problème de l'appariement

des régions. La nouveauté, dans notre approche, est la propagation des contraintes dans la hiérarchie qui rend l'algorithme plus robuste au bruit, aux sursegmentations des régions et réduit considérablement le temps de calcul. Ensuite les étapes d'appariement sont adaptées au contexte satellitaire : au niveau du calcul du volume de corrélation initial et de l'appariement subpixelaire des contours des régions pour la reconstruction des facettes 3D et finalement l'estimation robuste des plans 3D pour la scène 3D finale.

8 PERSPECTIVES

Les premiers résultats obtenus sont satisfaisants. Des améliorations sont en cours d'étude. On envisage pour rendre les coupes des hiérarchies plus robustes, de réaliser une remontée dans la hiérarchie pour valider la coupe.

Des évaluations restent à réaliser en fonction de la résolution des images, du rapport B/H, des différents seuils utilisés et de la qualité du MNE en entrée.

Pour la reconstruction finale des facettes 3D, on envisage de réaliser une mise en correspondance fine des contours des régions, continue par des méthodes de programmation dynamique subpixelaire. Les facettes 3D obtenues serviront à émettre des hypothèses de bâtiments.

REMERCIEMENTS

Les auteurs tiennent à remercier l'IGN et le CNES pour nous avoir fourni les images aériennes de la ville d'Amiens et les simulations des images satellites haute résolution réalisées sur Toulouse.

RÉFÉRENCES

- [1] C. Baillard. *Analyse d'images aériennes stéréoscopiques pour la restitution 3-D des milieux urbains*. PhD thesis, ENST, 1997.
- [2] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. *CVPR*, pages 559–565, June 1999.
- [3] M. Cord, M. Jordan, J-P. Cocquerez, and N. Paparoditis. Automatic extraction and modeling of urban buildings from high resolution aerial images. *ISPRS*, September 1999. Munich, Germany.
- [4] C.S. Fraser, E. Baltsavias, and A.Grün. 3-d building reconstruction from high resolution ikonos stereo imagery. In L.Van Gool ETH Zurich Switzerland E.P.Baltsavias, A.Gruen, editor, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, pages 331–344, Centro Stefano Franscini, Monte Verità,Ascona, June 2001.
- [5] L. Guigues, H. Le-Men, and J-P. Cocquerez. Scale-sets image analysis. *ICIP'03*, pages 299–306, October 2003. Barcelone,Espagne.
- [6] H. Jibrini, M-P. Deseilligny, N. Paparoditis, and H. Maître. Reconstruction de modèles de bâtiments à partir de données cadastrales vectorisées et d'images aériennes. *RFIA*, III :299–306, 2000. Paris,France.
- [7] T. Kanungo, B. Dom, W. Niblack, and D. Steele J. Sheinvald. Mdl-based multi-band image segmentation using a fast region merging scheme. Research Report RJ 9960, IBM, May 1995.
- [8] X.Descombes M.Ortner and J.Zerubia. Un nouveau modèle pour l'extraction de caricatures de bâtiments sur des modèles numériques d'élévation. In

- TAIMA 03, pages 213–218, Hammamet, Tunisia, October 2001.
- [9] S. Noronha and R. Nevatia. Detection and modeling of buildings from multiple aerial images. *PAMI*, 23(5) :501–518, May 2001.
 - [10] N. Paparoditis, C. Thom, and H. Jibrini. Surface reconstruction of urban area from highly overlapping digital aerial images. *International Archives of Photogrammetry and Remote Sensing, XXXIII ISPRS*, Geo Information for all, July 2000. Amsterdam, Netherlands.
 - [11] S. Randriamasy and A. Gagalowicz. Region-based top-down segmentation controlled by stereo matching. *IAPR :MVA1992*, 1992. Japan.
 - [12] P. Sander, L. Vinet, L. Cohen, and A. Gagalowicz. Hierarchical region based stereo matching. *SCIA, In Proc of the 6th Scandivian Conference on Image Analysis*, pages 71–78, June 1989.
 - [13] G. Sohn and I.J.Dowman. Extraction of buildings from high-resolution satellite data. In L.Van Gool ETH Zurich Switzerland E.P.Baltsavias, A.Gruen, editor, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, pages 345–355, Centro Stefano Franscini, Monte Verità, Ascona, June 2001.
 - [14] S.Vinson and L-D.Cohen. Multiple rectangle model for buildings segmentation and 3d scene reconstruction. In *ICPR, 16th Interbational Conference on Pattern Recognition*, volume I, pages 125–134, Québec, Canada, August 2002.
 - [15] L-D.Cohen S.Vinson and F.Perlant. Extraction of rectangular buildings using dem and orthoimage. In *SCIA, 12th Scandinavian Conference on Image Analysis*, Bergen, Norway, June 2001.
 - [16] F. Taillandier, L. Guigues, and R. Deriche. A framework for constrained multi-scale range image segmentation. In *ICIP'03*, October 2003. Barcelone, Espagne.
 - [17] C. Vestri and F. Devernay. Using robust methods for automatic extraction of buildings. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition, IEEE*, 2001. Hawaii.
 - [18] J-M. Vézien and J-P. Tarel. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandivian Conference on Image Analysis*, pages 1061–1070, 1995. Uppsala, Sweden.
 - [19] Z.Zhang and O.Faugeras. *3D scene analysis :a stereo based approach*. Springer, 1992.