

Rapport de Méthodes de Traitement de l'Image

Article traité:

*Extraction of buildings and trees in urban
environments de N. Haala and Brenner*

Baptiste Mougel

17 janvier 2005

Table des matières

1	Introduction	3
2	Contexte	4
2.1	Définition des Modèles 3D Urbains	4
2.2	Les applications des Modèles 3D Urbains	4
2.3	Production des modèles 3D urbains	5
3	Théories sous-jacentes	9
3.1	Les Ortho-images	9
3.2	Les MNT / MNE	11
3.3	Le cadastre	13
3.4	Les Photographies Aériennes	13
3.4.1	Acquisition de données brutes	14
3.4.2	Différents types de capteur sont disponibles	14
3.4.3	La technique d'acquisition	15
3.5	Les différents types des données	16
3.6	La radiométrie sur multicanaux	18
3.7	La détection de contours	19
3.7.1	Introduction	19
3.7.2	Quelques définitions	20
3.7.3	La méthode de Canny Deriche	21
3.7.4	Le seuillage par Hystérésis.	22
4	Méthode développée dans l'article	25
4.1	Classification	25
4.2	Reconstruction 3D	29
4.2.1	La Reconstruction 3D des arbres	29
4.2.2	La Reconstruction 3D des bâtiments	30
5	Conclusion	32

Chapitre 1

Introduction

Dans le cadre de ma formation au Master 2 IAD de l'université pierre et marie curie, j'ai participé au module de Méthodes de traitement de l'image (TIF). Afin d'évaluer, ce module nous avons dû réaliser un rapport et une soutenance sur la base d'un article en rapport avec le traitement des images.

Pour ma part j'ai travaillé sur un article intitulé Extraction of buildings and trees in urban environments (Extraction d'arbres et de bâtiments en milieu urbain) écrit en 1998 par Norbert Haala et Claus Brenner de l'Institute for Photogrammetry, University of Stuttgart Geschwister-Scholl-Strasse en Allemagne.

Cet article traite d'une méthode de reconstruction 3D de ville à partir de photographie aérienne. Cette méthode se déroule en 2 étapes dans une première étape, on va segmenter et classifier les images. Dans un second temps, on va mettre en place la reconstruction 3d photo réaliste de la ville.

Dans la suite de ce rapport, je commencerais par vous présenter le contexte de se travaille. Ensuite je vous présenterai les différentes théories sous jacente à l'article, avant de passer à la présentation de la méthode. Pour finir, je porterai une conclusion sur la méthode développée dans cet article.

Chapitre 2

Contexte

2.1 Définition des Modèles 3D Urbains

Dans notre contexte les modèles 3D sont une représentation 3D du terrain. Ils peuvent être de plusieurs formes : raster (MNS¹ ou MNT²) ou vecteur (TIN³, facettes, ou modèles complexes).

Dans le cadre de cette article, les modèles 3D urbains photoréalistes (la figure 2.1 est le produit que l'on souhaite obtenir à l'issue du traitement) .

2.2 Les applications des Modèles 3D Urbains

Les Modèles 3D Urbains sont utilisés pour de nombreuses applications, en voici quelques une mais cette liste n'est pas exhaustive :

- **La simulation des risques** est un domaine où l'on utilise de plus en plus de modèles 3D urbains. En effet, ces modèles permettent de simuler l'écoulement de l'eau lors d'inondation ou encore simuler les effets d'une avalanche sur une ville de montagne. Et ainsi, fiabiliser l'établissement de cartes des zones à risque.
- **Couverture radio** : Ici les modèles 3D permettent aux opérateurs téléphoniques de faire un état des lieux de la couverture radio et ainsi optimiser leurs réseaux d'antennes.
- **Carte de bruit & de nuisance** sont deux domaines où l'on retrouve des modèles 3D. En effet, ils permettent d'établir des cartes de nuisances fiables. De plus, grâce à eux on peut connaître les effets de l'implantation ou la destruction d'un bâtiment, sur son environnement (perte de réception ou

¹MNS : Modèle Numérique de Surface décrivant le sol + le sursol(bâtiments, arbres, ...).

²MNT : Modèle Numérique de Terrain décrivant le sol seulement

³TIN : Triangulated Irregular Networks



FIG. 2.1 – Modèle 3D Urbains

encore augmentation du bruit). Et ainsi, ils aident les urbanistes à prévenir les nuisances.

2.3 Production des modèles 3D urbains

De tels modèles peuvent être produits manuellement par photogrammétrie⁴ en utilisant la stéréoscopie⁵ sur des images aériennes. Durant les années 1970 et 1980, on a produit ce type de données (exemple BD TRAPU de l'IGN cf figure 2.2). Mais cette production est très longue et coûteuse (utilisation d'appareil de stéréorestitution, personnel qualifié,...) et elle a donc été abandonnée.

⁴photogrammétrie : Science qui regroupe les techniques qui permettent de mesurer la position, la forme et la dimension d'objet à partir d'une ou d'un ensemble d'images prises de différents points de vue.

⁵stéréoscopie : Procédé qui permet, par vision binoculaire, d'obtenir la sensation du relief d'un paysage observé à partir d'un couple d'images ou de photographies prises de points de vue différents.



FIG. 2.2 – Exemple de BD TRAPU de l'IGN sur la mairie d'amiens

De nombreux laboratoires travaillent depuis de nombreuses années sur la modélisation vectorielle 3D des bâtiments (Figure:2.3) ce qui constitue l'étape la plus importante pour la production automatique de ce type de base de données 3D urbaines.

L'approche couramment étudiée est l'utilisation d'images aériennes combinées à des algorithmes de stéréovision⁶ afin de reconstruire les bâtiments. Le problème de la stéréovision peut se résumer à la citation suivante :

"Étant donnée deux images d'une même scène obtenues simultanément de deux points de vue légèrement différents, est-il possible de reconstruire la scène en trois dimensions à partir de ces deux images ?"[Fau93]

En vision biologique, la perception du relief se fait naturellement. Le cerveau est capable d'analyser notre environnement, de trouver les homologues entre la vision de l'œil droit et de l'œil gauche et donc percevoir le relief grâce au léger décalage de ces deux points de vue.

D'un point de vue mathématique, les principes de la stéréoscopie sont bien

⁶stéréovision : Procédés utilisant le traitement d'images pour reconstruire automatiquement le relief.

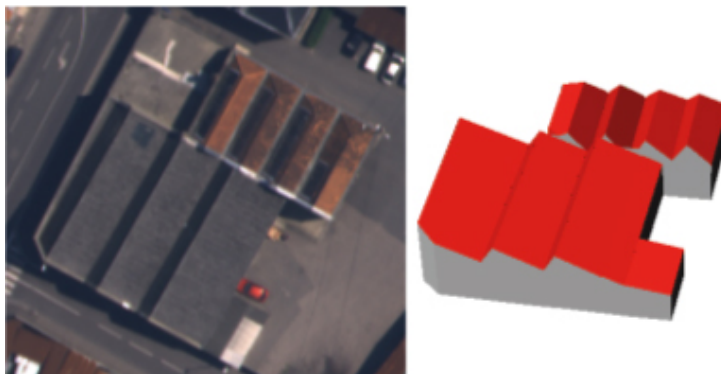


FIG. 2.3 – Reconstruction semi-automatique

connus depuis la naissance de la *photogrammétrie*⁷ à la fin du XIX^{ème} siècle. Si l'on connaît 2 points homologues i.e représentant la même partie de la scène sur deux images dont on connaît la position et l'orientation dans l'espace, on peut projeter ces deux points dans l'espace scène sous forme de faisceaux et ainsi retrouver par intersection de faisceaux le point 3D original.

En vision par ordinateur, la difficulté est de trouver ces points homologues. Car, si on peut facilement à l'oeil reconnaître le coin d'un même bâtiment sur 2 prises de vues différentes, le faire automatiquement est beaucoup plus problématique. C'est pourquoi, à l'heure actuelle, la grande majorité des reconstruction 3D de bâtiments se fait toujours "à la main", c'est à dire qu'un opérateur saisit les points homologues sur un couple de photos. Comme on l'a déjà dit, ces reconstructions sont donc lentes et passablement coûteuses.

En plus des travaux sur la stéréovision, l'une des autres approches étudiées consiste à utiliser des données à balayage laser aéroporté (Figure:2.4). Ces balayages laser permettent de mesurer directement un nuage de points 3D. Ces balayages font appel à des technologies assez récentes, ils suscitent depuis plusieurs années un intérêt grandissant auprès de la communauté scientifique et métrologique mondiale. Durant ces dix dernières années, de nombreux travaux de recherche sur la modélisation 3D des bâtiments à partir de données Laser ont été réalisés [HB99], [Maa99] et [VS01].

⁷cf note 4

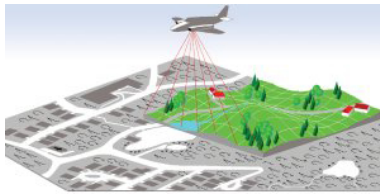


FIG. 2.4 – Acquisition Laser

Chapitre 3

Théories sous-jacentes

Afin de bien comprendre le sujet, il est nécessaire de définir quelques termes qui seront utilisés par la suite. Ainsi que les principes de bases du domaine de la reconstruction 3D.

3.1 Les Ortho-images

Toute prise de vue aérienne, présente des déformations dues à l'inclinaison de la chambre de prise de vue et au relief du terrain. De ce fait elle ne présente pas une échelle uniforme et on ne peut par conséquent pas mesurer des distances comme sur une carte (Il s'agit uniquement d'une image). Ces influences de l'inclinaison de la chambre de prise de vue et du relief du terrain peuvent cependant être neutralisées par une transformation appelée orthorectification¹ ou redressement. (Figure:3.1)

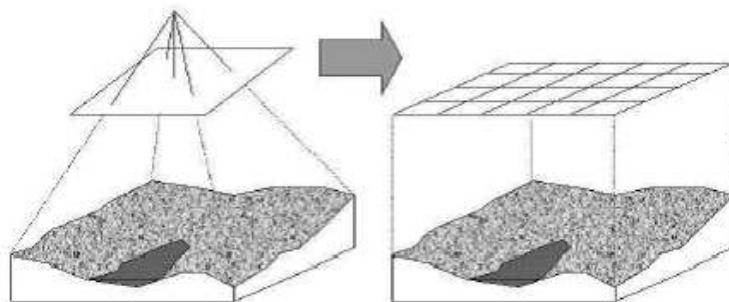


FIG. 3.1 – Transformation d'une image en ortho-image

¹Orthorectification : Traiter une image pour la corriger de toutes les déformations géométriques liées aux conditions de prises de vue et au relief du terrain.

Une Ortho-image ²(également appelée ortho-photo ou orthophoto) est une image photographique corrigée de toutes les déformations géométriques dues aux conditions de prise de vue (inclinaison de l'axe, distorsion due à l'objectif) et au relief du terrain.(Figure:3.2)

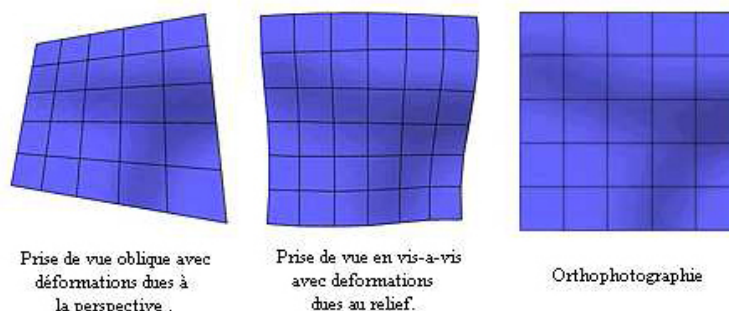


FIG. 3.2 – Déformations des images aériennes

Une ortho-image est en fait une image déformée de sorte que le centre de prise de vue soit à la verticale de chaque point (image en perspective axonométrique verticale). Elle est superposable à une carte. Chaque pixel est ainsi localisé par des coordonnées géographiques et possède une taille terrain fixe. Obtenues respectivement à partir de photographies aériennes ou d'images satellitaires rectifiées géométriquement et égalisées radiométriquement, l'orthophotographie allie la précision géométrique(cf : 3.5) de la carte à la richesse de la photographie. Elle peut être géoréférencée dans n'importe quel système de coordonnées et facilement intégrable dans un SIG (Système d'Information Géographique) car elle permet de mesurer la longueur, la surface de tout objet au sol...

Dans les ortho-images dites classiques, les défauts corrigés sont :

- Distorsion de l'objectif
- La non verticalité de l'axe de prise de vue
- Le relief du sol naturel

Cependant les objets situés en sursol (bâtiments, ponts, forêts) ne sont pas corrigés. Leurs positions et leurs tailles ne sont donc pas exactes. Ils ne peuvent donc pas être mesurés. En outre le dévers des bâtiments subsiste : les bâtiments apparaissent 'inclinés', les façades sont visibles et une partie du sol cachée. Ce problème est particulièrement sensible en milieu urbain dense , car les immeubles ne sont pas à leurs positions et peuvent complètement masquer certaines rues. La figure 3.3.a présente un exemple d'une telle ortho-image. Les 'vraies ortho-images'

²Ortho-image : Image corrigée de toutes les déformations géométriques dues aux conditions de prise de vue et au relief du terrain.

sont des ortho-images ou tous les défauts, y compris les dévers et les parties cachées ont été corrigés. Tous les objets, bâtiments compris, sont donc à leurs positions cartographiques. La figure 3.3.b présente un exemple d'une telle ortho-image.

Comme les bâtiments ont été remis à leurs places, il reste des zones non renseignées (en noir) qui correspondent aux parties cachées car on a utilisé une seule image. Ces zones seront remplies en utilisant l'information provenant d'autres images.[Bol02]. Ici on s'intéresse tout particulièrement à ces 'vraies' ortho-images appelées aussi ortho-images intégrales.

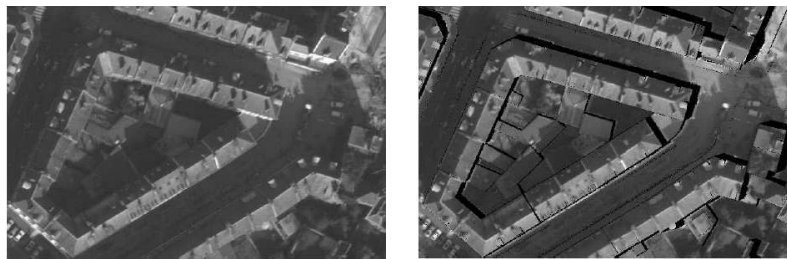


FIG. 3.3 – Ortho-image 'classique' (a) et ortho-image 'vraie' (b)

3.2 Les MNT / MNE

Les modèles numériques de terrain (MNT ou DTM en anglais) et les modèles numériques d'élévation (MNE ou DSM en anglais) fournissent une description unique de notre environnement en trois dimensions. Ce sont des fichiers altimétriques constitués par un réseau maillé régulier d'altitudes. Les MNT décrivent le profil du sol tandis que les MNE intègrent le sursol (bâti, végétation...) Quand l'information se limite aux altitudes du sol, on parle donc de modèle de terrain.(Figure:3.4). Un MNE est une modélisation numérique du sursol, incluant ou non le sol et décrit soit sous une forme 'raster', soit par facettes liées topologiquement les unes aux autres, ou encore sous forme vecteur. Les MNT et les MNE peuvent être calculés de différentes manières :

- Soit à partir de courbes de niveaux vectorisées.
- Soit à partir de couples d'images radar ou laser.
- Soit à partir de couples stéréoscopiques d'images aériennes ou satellitales, par corrélation automatique ou par restitution.
- Soit à partir de données laser (dans cette article c'est cette méthodes qui est utilisé)

Ils sont indispensables pour tout calcul d'intervisibilité, d'écoulement, de pente, d'ombre portée, pour toute modélisation de propagation d'ondes, de bassins ver-

sants, d'érosion... Les MNT sont également nécessaires pour la rectification des images aériennes ou satellitaires et donc la production d'orthophotographies. Quant aux MNE, ils sont d'un grand intérêt pour toutes les applications urbaines, en particulier pour les télécommunications (couverture de réseau de téléphonie mobile). Pour la création d'ortho-image on utilise les MNE pour les ortho-images intégrales.(Figure:3.5) et les MNT pour la construction d'ortho-images classiques de base.

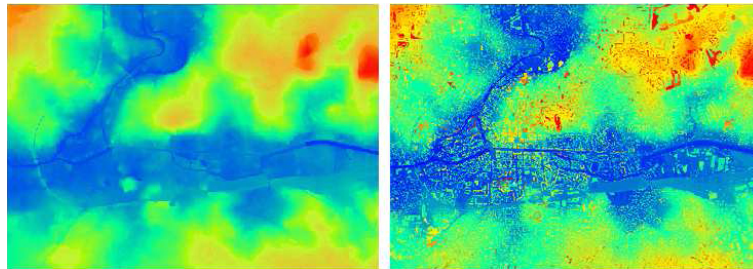


FIG. 3.4 – MNT et MNE sur la ville de Rennes (2001)



FIG. 3.5 – Modèle Numérique d'Élévation sur le centre ville d'Amiens.

3.3 Le cadastre

Le cadastre est un ensemble de documents établis à la suite de relevés topographiques et d'opérations administratives, et destinés à permettre la détermination des propriétés foncières d'un territoire, la constatation de la nature de leurs produits et l'évaluation de leur revenu. Les documents du cadastre sont réalisés puis archivés en premier lieu sur papier et sont ensuite scannés c'est à dire numérisés. En France le cadastre est exclusivement fiscal, on l'utilise pour la rédaction de divers actes relatifs à la propriétés foncières (actes de ventes, déclaration de successions, contrats de mariage, acte de partage ou d'échange, de location...). La parcelle est l'unité cadastrale, c'est une portion de terrain non divisée par des séparations matérielles, elle présente une même nature de culture ou une même affectation et appartient en général à un même propriétaire. Le sol d'un bâtiment et la cour attenante ou le jardin contigu ne forment qu'une parcelle. (Figure:3.6)

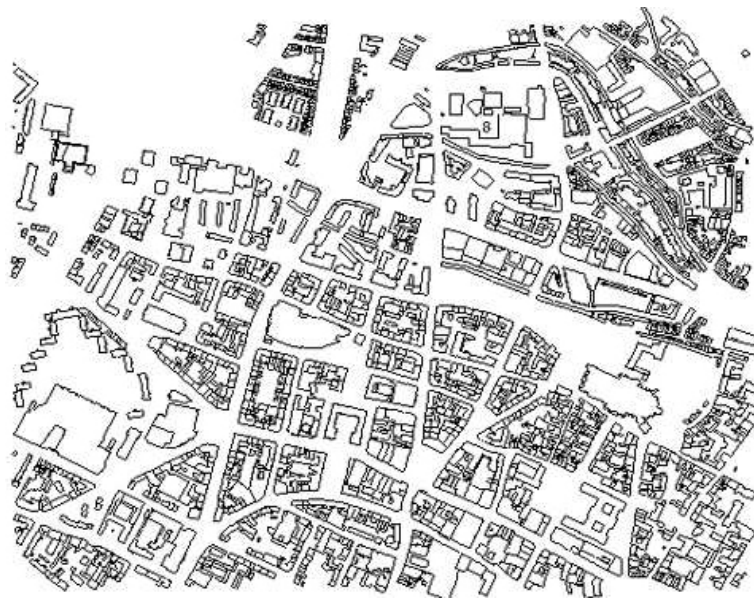


FIG. 3.6 – Cadastre vectorisé sur le centre ville d'Amiens

3.4 Les Photographies Aériennes

La photographie aérienne permet d'obtenir une image objective et esthétique du territoire. La photographie aérienne est en plus une base idéale pour tout projet d'aménagement rural, d'urbanisme opérationnel ou de cadastre.

3.4.1 Acquisition de données brutes

Tous les vols photogrammétriques sont effectués avec GPS embarqué et d'une centrale inertielle (INS) ce qui permet de connaître exactement la trajectographie de l'avion et ainsi d'améliorer la précision des informations recueillies. De plus, il est fortement recommandé, de disposer d'au moins une station GPS au sol, à proximité de la zone de vol (30km maximum), pour améliorer la précision du positionnement géographique du capteur (Figure:3.7)

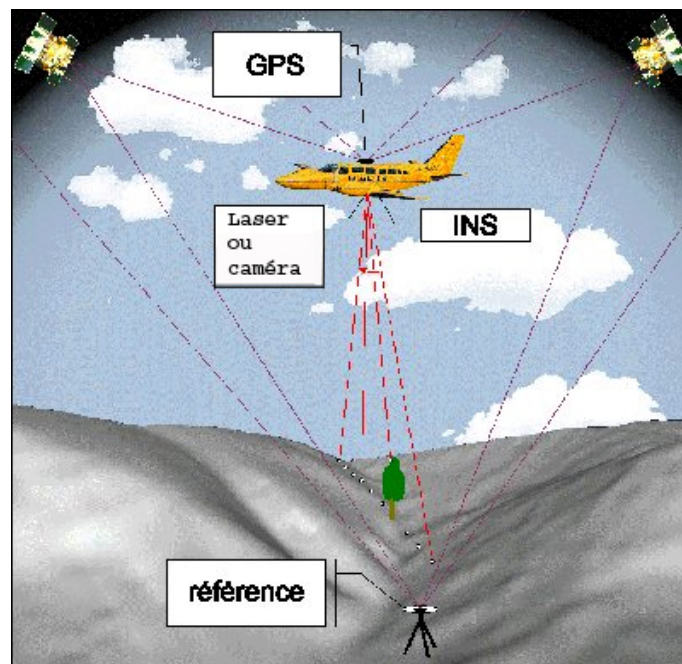


FIG. 3.7 – Principe d'acquisition des mesures laser

3.4.2 Différents types de capteur sont disponibles

- le capteur radar analogique ou numérique qui présente de bons résultats dans les zones nuageuses ou pour des études de texture géologique.
- le capteur optique analogique ou numérique avec différents choix pour l'émulsion : Infra Rouge noir et blanc, Infra Rouge couleur, panchromatique, couleurs naturelles.

Les caméras à capteur optique dispose d'une perspective conique, elles sont équipées d'un système de compensation de filé, du positionnement par GPS et de disques durs à haut débit permettant une fréquence de prise de vues élevée.

Elles assurent :

- une parfaite stabilité géométrique
- un très haut rendu dynamique (vision dans les zones sur et sous-exposées comme les zones d'ombre)
- des images avec un faible dévers grâce à l'utilisation en standard de longues focales.
- une taille compatible avec les optiques du commerce.

3.4.3 La technique d'acquisition

Sur la figure 3.8 on peut voir la technique d'acquisition des images stéréoscopiques. Les points noirs représentent les sommets de prises de vues et les rectangles les emprises des premiers clichés. Le recouvrement entre les images permet d'obtenir l'effet stéréoscopique, et donc de reconstruire le relief (par reconstruction d'un MNE par exemple).

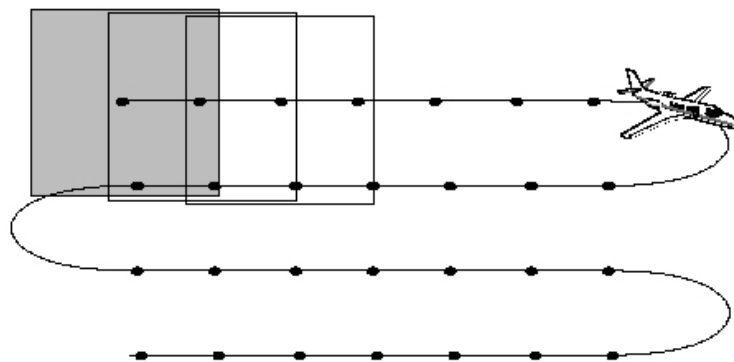


FIG. 3.8 – Schéma d'une prise de vue classique

Ces images numériques sont généralement composées de deux fichiers : un fichier image.tif et un fichier image.ori . Le fichier image.tif contient les données images codé en 3 canaux (RVB). Le fichier image.ori quand à lui contient les conditions de prises de vue et notamment :

- Le positionnement 3D de la caméra
- Et les caractéristiques de calibration de la caméra.



FIG. 3.9 – Exemple de photographie aérienne sur le centre ville d'Amiens

3.5 Les différents types des données

Ici je tiens à préciser le type de données utiliser dans l'article et plus généralement en géographie sont d'un format particulier c'est ce format que je vais vous présenter.

Les coordonnées géographiques.

Coordonnées angulaires définies par rapport à deux plans de référence : le premier étant le plan défini par l'axe de rotation de la terre et le méridien de référence, le second par le plan de l'équateur. Tout point du globe peut alors être connu par ses coordonnées géographiques : longitude et latitude.

Les coordonnées cartographiques.

Coordonnées cartésiennes définies par rapport à deux axes de référence positionnés sur une carte. Tout point du globe peut alors être connu par ses coordonnées cartographiques X,Y dans un système de projection et un système de coordonnées donné. Le système Lambert. Système de projection conique direct conforme défini en 1772 par Jean-Henri Lambert (1728- 1777). Utilisé en France métropolitaine avec l'ellipsoïde Clarke 1880, comme référence pour tous les grands travaux cartographiques.(Figure:3.10)[Lam60] Le cône est tangent ou sécant à un parallèle que l'on appelle le parallèle origine ou isomètre central (de latitude j) Les méri-

diens sont représentés par des droites concourantes au pôle. Les parallèles sont représentés par des cercles concentriques au pôle.

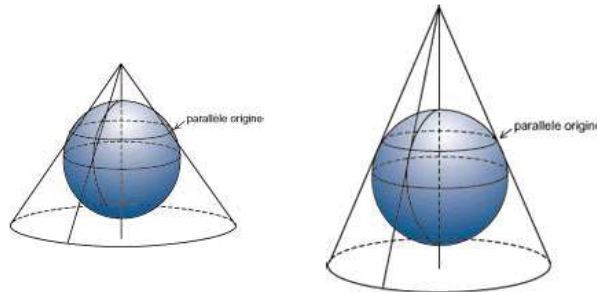


FIG. 3.10 – Système de projection conique

Les longueurs sont de plus en plus dilatées au fur et à mesure que l'on s'éloigne du parallèle origine. C'est pour cela que la France a été fractionnée en 4 zones (dont la Corse). Ces projections sont associées au système géodésique NTF en France et sont appelées Zone Lambert. (Figure:3.11) En fait, pour atténuer les déformations, on utilise en France des cônes sécants à la sphère en 2 parallèles qui sont à des latitudes équidistantes du parallèle origine. Il y a 4 cônes sécants à des latitudes différentes (parallèles origines) :

- Lambert I (NORD) ($j=55$ grades)
- Lambert II (CENTRE) ($j=52$ grades)
- Lambert III (SUD) ($j=49$ grades)
- Lambert IV (CORSE) ($j=46,85$ grades)

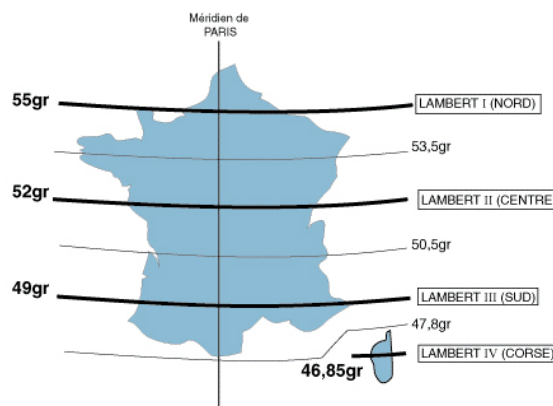


FIG. 3.11 – Delimitation des quatre zones Lambert

Remarques

- Pour des besoins de représentation globale et d’homogénéité des coordonnées en France, le Lambert II étendu a été créé : il s’agit de la zone Lambert II dont les limites kilométriques ont été élargie de façon à couvrir tout le territoire.
- Le Lambert 93 : c’est une nouvelle projection qui a été définie pour le nouveau système géodésique RGF93. C’est aujourd’hui la projection rendue légale en France depuis l’arrêté du 26 Décembre 2000.

Ici, l’ensemble des données dont on dispose sont les images aériennes, les MNE , les MNT ,qui peuvent se retrouver dans trois types très différents de coordonnées.

- Coordonnée de type MAP (POINTS CARTO)
Coordonnées cartographiques dans le système de projection de Lambert.
- Coordonnée de type GROUND (POINTS TERRAIN)
Coordonnées dans un repère euclidien local.
- Coordonnée de type PHOTO (POINTS IMAGES)
Coordonnées images classiques.

3.6 La radiométrie sur multicanaux

Les radiométries dans les images nous fournit une image à N canaux de radiométrie au même format que le MNE. L’information de ces canaux est redondante, mais elle est plus importante pour l’image N canaux que pour une seule image, et ainsi cette image possède certaines caractéristiques qui guideront plus efficacement la segmentation. Par analogie l’apport de cette image pourrait être comparé à l’apport de la couleur par rapport aux niveaux de gris. Même si les 3 canaux sont redondants l’information qu’ils apportent permet de distinguer des pixels de même luminance mais de chrominance différente, alors que l’image en niveaux de gris ne l’aurait pas permis. Les différents canaux permettent de discriminer des régions sur l’orientation du plan par rapport à la prise de vue. En effet selon le type de matériaux, la lumière n’est pas réfléchi de la même manière. La modélisation de la fonction de réflectance bidirectionnelle (BRDF) nous donne des éléments de réponse sur ce problème. Les modèles de réflexion pour des matériaux comme l’ardoise ne sont pas isotropes, mais spéculaires. Le maximum d’intensité réfléchi sera dans la direction du symétrique du rayon incident par rapport à la normale.

Ceci génère donc des différences entre les différentes prises de vue. La position du soleil pouvant être considérée comme constante sur toutes les prises de vue, la direction spéculaire variera selon le plan sur lequel la lumière se réfléchit. Ainsi dans le cas de le schéma 3.12 la caméra 1 recevra une intensité beaucoup plus importante du plan A que les autres caméras de ce même plan, de même pour la caméra 3 et le plan B. Dans l’image multicanaux que nous construisons la luminance du canal 1 (resp 3) sera beaucoup plus importante que celle des autres canaux pour les points

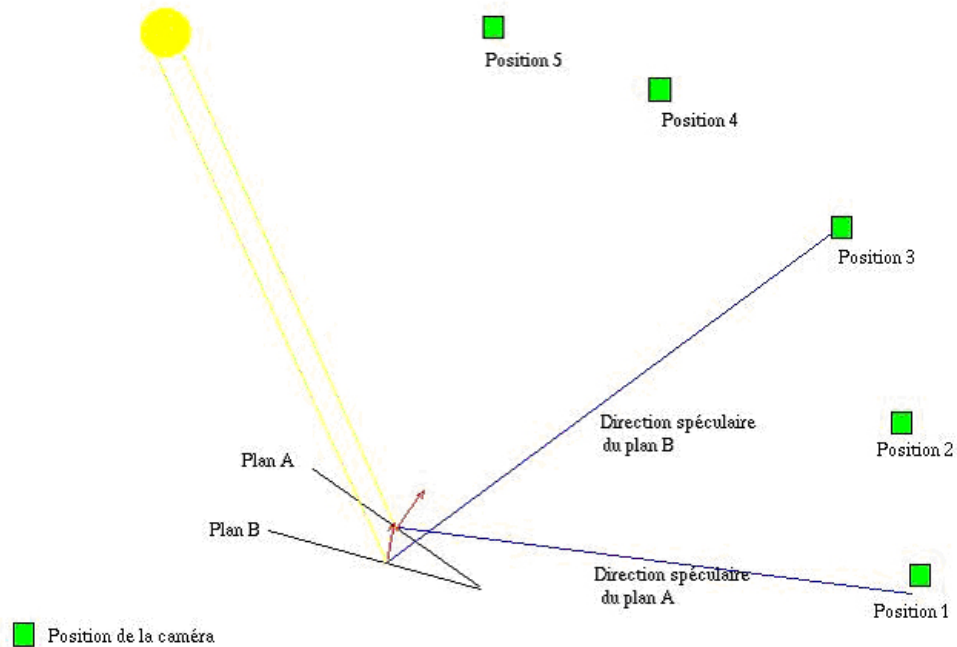


FIG. 3.12 – Exemple de réflexion

du plan A (resp B). Ceci nous permettra donc dans certains cas de discriminer en utilisant uniquement la radiométrie, des plans différents même si les matériaux sont identiques et que le contour est peu marqué. En poussant l'approche un peu plus loin, la différence de BRDF entre différents matériaux peut aussi permettre de réaliser de telles discriminations.

3.7 La détection de contours

3.7.1 Introduction

La détection de contour est une étape préliminaire à de nombreuses applications de l'analyse d'images. Les contours constituent en effet des indices riches, au même titre que les points d'intérêts, pour toute interprétation ultérieure de l'image. Les contours dans une image proviennent des :

- discontinuités de la fonction de réflectance (texture, ombre),
- discontinuités de profondeur (bords de l'objet),

et sont caractérisés par des discontinuités de la fonction d'intensité dans les images. Le principe de la détection de contours repose donc sur l'étude des dérivées de la fonction d'intensité dans l'image : les extréma locaux du gradient de la fonction d'intensité et les passages par zéro du laplacien. La difficulté réside dans la présence de bruit dans les images.

3.7.2 Quelques définitions

Le filtrage linéaire d'une image. Filtrer une image consiste à convoluer sa fonction d'intensité $I(x, y)$ avec une fonction $h(x, y)$ appelée réponse impulsionnelle du filtre.

$$I'(x, y) = h(x, y) * I(x, y)$$

$$I'(x, y) = \int_{-\infty}^{+\infty} \int_{+\infty}^{-\infty} h(u, v) I(x - u, y - v) dudv$$

$$I'(x, y) = \int_{-\infty}^{+\infty} \int_{+\infty}^{-\infty} h(x - u, y - v) I(u, v) dudv$$

Dans le cas discret

$$I'(x, y) = \sum_{u=-H/2}^{+H/2} \sum_{v=-H/2}^{+H/2} h(u, v) I(x - u, y - v)$$

Le gradient d'une image est le vecteur

$$\nabla I(x, y) = \left(\frac{\partial I(x, y)}{\partial x} \frac{\partial I(x, y)}{\partial y} \right)$$

caractérisé par son module et sa direction

$$m = \sqrt{\left(\frac{\partial I(x, y)}{\partial x} \right)^2 + \left(\frac{\partial I(x, y)}{\partial y} \right)^2}$$

$$\phi = \arctan \left(\frac{\partial I(x, y) / \partial y}{\partial I(x, y) / \partial x} \right)$$

La détection de contour Deux approches sont possibles dans la détection de contours :

1. Approche gradient : détermination des extréma locaux dans la direction du gradient.
2. Approche laplacien : détermination des passages par zéro du laplacien.

Les différentes approches existantes se classent ensuite suivant la manière d'estimer les dérivées de la fonction d'intensité :

1. Différences finies.
2. Filtrage optimal.
3. Modélisation de la fonction d'intensité.

Ici on s'intéressera seulement à l'approche gradient et par filtrage optimal. En utilisant les filtres de lissage et de dérivation de Canny Deriche. (voir figure 3.13.a, 3.13.b, 3.13.c).

3.7.3 La méthode de Canny Deriche

Critères de Canny (1983,[Can86]) On se place dans le cas monodimensionnel. On suppose que la détection est effectuée en convoluant le signal par un filtre de réponse impulsionnel h , les contours étant caractérisés par les extrêmes de la sortie du filtre. Les contours envisagés ici sont des contours de types marche et le bruit est supposé blanc (de moyenne nulle).

Les critères d'optimalités sont :

- détection : le contour doit être détecté, il faut minimiser les fausses réponses,
- localisation : le contour doit être localisé avec précision, il s'agit de minimiser la distance entre les points détectés et le vrai contour.
- réponse unique : il s'agit de minimiser le nombre de réponse pour un seul contour.

La minimisation de ces critères dans le contexte du modèle proposé débouche sur une équation différentielle dont la solution est de la forme :

$$h(x) = a_1 e^{\alpha x} \cos(\omega x) + a_2 e^{\alpha x} \sin(\omega x) + a_3 e^{-\alpha x} \cos(\omega x) + a_4 e^{-\alpha x} \sin(\omega x)$$

Le filtre de Deriche Deriche([Der90],[Der93]) a proposé un filtre de lissage dont la dérivée est la solution exacte de l'équation de Canny étendue à supports infinis. Le filtre de lissage correspondant est

$$h(x) = k(\alpha |x| + 1)e^{-\alpha|x|}$$

avec

$$k = \frac{(1 - e^{-\alpha})^2}{(1 + 2\alpha e^{-\alpha} - e^{-2\alpha})}$$

$$h'(x) = -k' x e^{-\alpha|x|}$$

$$k' = \frac{(1 - e^{-\alpha})^2}{e^{-\alpha}}$$

3.7.4 Le seuillage par Hystérésis.

Le filtre de Canny Deriche permet de calculer le gradient d'une image mais ne donnent pas les points de contours. Un traitement ultérieur est donc nécessaire pour pouvoir passer de l'image des dérivées à l'image aux contours. Les points de contour dans une image sont caractérisés par des extrémas locaux du gradient. Une première approche consiste donc à :

1. Calculer la norme du gradient en tous point de l'image,
2. Sélectionner les pixels à l'aide d'un seuil a priori pour la norme du gradient.

Ne permet pas de différencier efficacement les points de contour du bruit.

L'approche classiquement utilisée se décompose en deux étapes qui sont :

1. Extraction des extrémas locaux du gradient dans la direction du gradient. Cela revient à déterminer, pour un pixel p donné, les valeurs du gradient sur la droite passant p et de direction celle de son gradient. On vérifie ensuite que le gradient en p est bien localement maximal sur cette droite. (voir figure 3.13.d)
2. Seuillage par hystérésis des extrémas. Cette étape repose sur une hypothèse de connexité. Le principe est d'utiliser deux seuils pour la norme du gradient : sb et sh et de sélectionner les pixels pour lesquels :
 - la norme du gradient est supérieure à sb ,
 - le pixel donné est connecté, par un chemin constitué de pixels dont la norme du gradient est supérieure à sb , à un pixel pour lequel la norme du gradient est supérieure à sh . (voir figure 3.14 et 3.15).

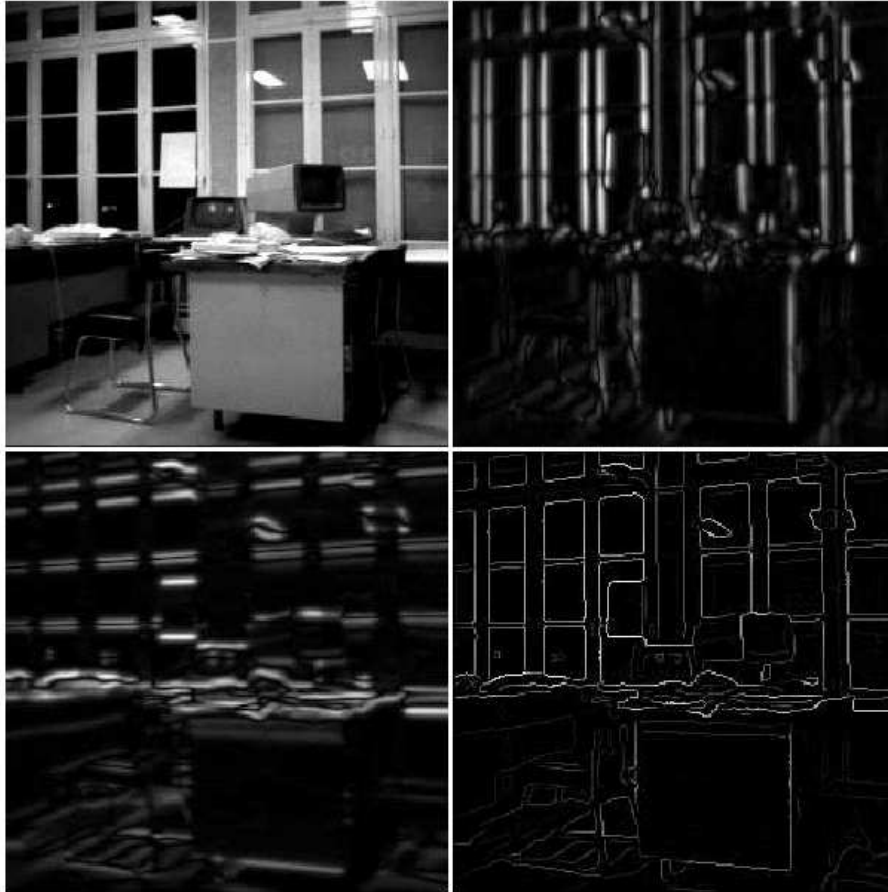


FIG. 3.13 – Image Originale (a), Gradient en x et y (b) et (c), et extréma locaux du gradient (d)

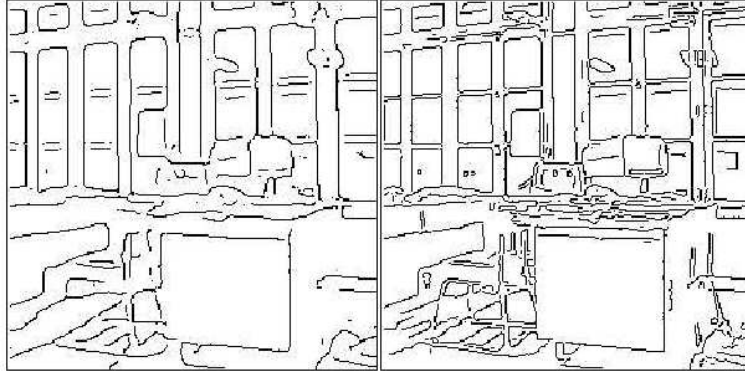


FIG. 3.14 – Seuillage par hystérésis des extrêmes locaux (avec alpha 0.5 et 1.5)



FIG. 3.15 – Image Originale + Contours détecté et par la méthode de Canny-Deriche

Chapitre 4

Méthode développée dans l'article

4.1 Classification

Dans l'approche de Haala et Brenner, on utilise l'information multispectrale fournie par une image aérienne couleur-infrarouge (CIR) et on la combine avec l'information géométrique d'un module de balayage à laser (MNE). De ces données on tire la classification selon les 5 classes suivantes :

- Rues
- Buissons
- Arbres
- Bâtiments
- Ombres & Non Classé

Les images CIR utilisées ont été prises à une échelle de 1 :5000. Pour la numérisation, les images sont obtenues par un balayage à une résolution de 60 micromètres, ayant pour résultat trois images numériques dans les bandes spectrales près de l'infrarouge, du rouge et du vert avec une empreinte de 30 centimètres par pixels.

L'idée fondamentale de l'algorithme proposé est d'employer simultanément les données géométriques du MNE et l'information radiométrique en appliquant une classification basée sur le pixel, par lequel le MNE est employé comme canal additionnel. En effet, il serait facile de faire la distinction (sur le MNE) entre les rues et les bâtiments ou entre les buissons et les arbres. Mais d'autre part, il sera difficile de faire la différence entre les arbres et les bâtiments et entre rue et buisson puisqu'ils sont au même niveau de terrain. Pour cette raison, la taille au-dessus du terrain comme source additionnelle d'information améliore la discrimination et évite les classifications fausses. Mais si on utilise seulement le MNE il sera difficile de faire la différence entre les informations qui sont au même niveau c'est là que l'on fait intervenir les données images qui nous permettent de faire la distinction entre ces

objets.

Mais en premier lieu il faut procéder à plusieurs prétraitements aussi bien sur les images que sur le MNE. Sur l'image on effectue une orthorectification (cf 3.1), et sur le MNE on dérive la surface du terrain par morphologie mathématique en utilisant l'approche décrite par Weidner et Forstner (1995 [UW95]) afin d'obtenir le MNT.

Ensuite, on crée un modèle hybride en soustrayant le MNT au modèle résultant de la fusion du MNS et du modèle obtenu par les images. C'est à partir de ce modèle que l'on segmente l'image en fonction des 5 classes choisies. La classification permet de détecter des groupes de Pixels. Pour cela on utilise un algorithme de technique d'analyse de données. On peut par exemple utiliser une ACP¹ ou les auteurs utilisent un algorithme du type de celui de Richards (1993 [Ric93])

Avec cette approche, le nombre optimal de clusters spectraux est automatiquement déterminé en appliquant itérativement des opérations de découpe et de fusion tout en exécutant les étapes suivantes :

1. Un certain nombre de paramètres doivent être initialisés par l'opérateur. Ces paramètres sont le nombre de clusters désirés, le nombre minimum d'itérations, le nombre minimum de Pixels dans un cluster, l'écart-type minimum pour lancer le dédoublement de cluster, et la distance maximale dans l'espace entre les centres des clusters pour lancer la fusion des clusters.
2. Un pixel est assigné à un des clusters prédéfinis par un critère de distance minimum.
3. On élimine tous les clusters contenant moins de membres que le nombre prédéfini de Pixel minimal.
4. Les nouveaux centres des clusters sont calculés à partir des Pixel assignés dans l'étape 2.
5. Les clusters agrégés sont dédoublés, si l'écart-type entre les clusters est plus grand que le seuil indiqué. Des clusters voisins sont fusionnés, si la distance est plus petite que le paramètre prédéfini et si le nombre de pixels maximums des clusters n'a pas été atteint.
6. L'algorithme est terminé, si le nombre maximum des itérations est atteint, sinon il est itéré à partir de l'étape 2.

Après que le critère d'arrêt soit atteint, les centres des clusters détectés sont employés pour classifier entièrement l'image sur le critère minimum de distance. Ensuite on affecte chaque cluster à l'une des 5 classes thématiques de base. Pour cela sont utilisés les techniques de base de la photogrammétrie (cf. 3). Les résultats de l'algorithme(cf (Figure:4.1)) dépendent, naturellement, fortement de la qualité

¹ACP : Analyse en Composante Principale.

des données d'entrée. En effet, il faut que le bruit soit minimal car il peut engendrer de grosse erreur.

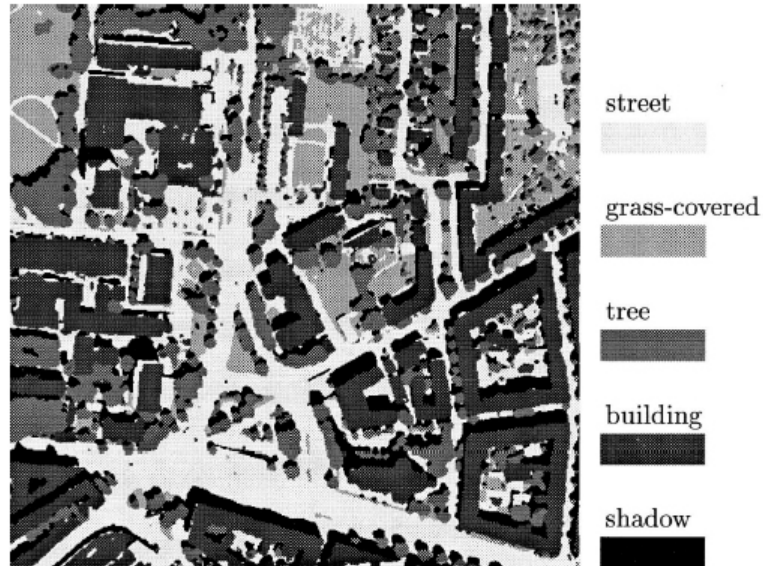


FIG. 4.1 – Result of ISODATA classification using CIR orthoimage and normalized DSM.

Cette méthode a de nombreux avantages comme de faire la distinction entre arbres et bâtiments. Néanmoins, les auteurs n'utilisent pas toutes les données fournies par l'acquisition laser Toposys©. En effet, le laser Toposys© fournit des données multi-écho c'est-à-dire que le faisceau laser présente la caractéristique de pouvoir se couper en plusieurs parties, enregistrées sous forme d'échos multiples. Il arrive en effet fréquemment que le rayon se divise en deux parties : une partie est réfléchiée par un objet en sursol (végétation, bâti) et l'autre atteint le sol pour s'y réfléchir ((Figure:4.2)). La plupart des systèmes n'enregistrent que deux échos (comme pour le cas de Toposys©), certains peuvent cependant en enregistrer jusqu'à 5. En faisant la différence des cartes altimétriques du premier et du dernier écho, on voit apparaître distinctement la végétation ainsi qu'une partie des bords de bâtiment ((Figure:4.3)). C'est cette propriété qui aurait pu être intéressante à utiliser. De plus, en France et partout où le cadastre est disponible, il peut être intéressant d'utiliser le cadastre (cf 3.3) car il peut nous donner les bases pour la segmentation, et notamment pour la classe bâtiment.

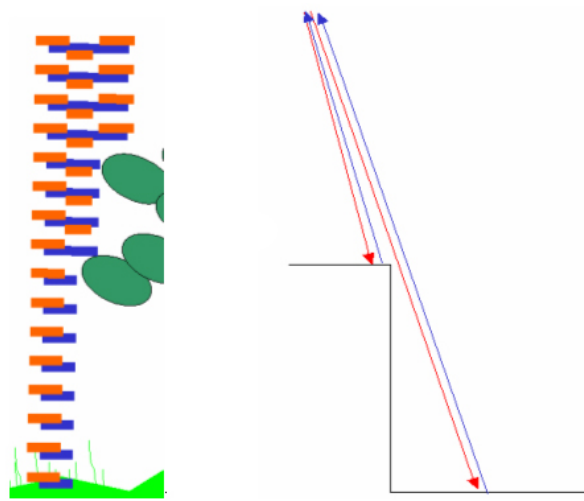


FIG. 4.2 – Mise en évidence de la séparabilité du faisceau : à gauche le faisceau se scinde en trois parties sur de la végétation ; à droite il se coupe en deux sur le bord d'un bâtiment

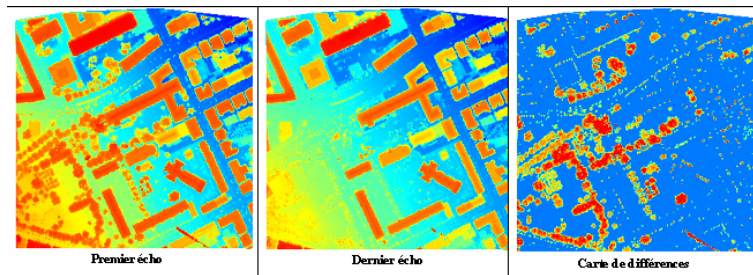


FIG. 4.3 – Visualisation d'une carte altimétrique premier écho, dernier écho et d'une carte de différences entre les deux

4.2 Reconstruction 3D

Cette partie peut se subdiviser en 2 parties distinctes car on a effectué séparément la reconstruction 3D des arbres et des bâtiments.

4.2.1 La Reconstruction 3D des arbres

Pour effectuer la reconstruction des arbres, on utilise l'image segmentée obtenue après la première étape. Et plus particulièrement, la classe arbre que l'on va commencer par squelettiser². Cette squelettisation va nous permettre de trouver où placer les troncs. En effet, on placera les troncs d'arbres sur les noeuds du squelette (cf. Point blanc sur la (Figure:4.4)). Les paramètres de taille de l'arbre seront donnés par le MNE.



FIG. 4.4 – Positionnements des arbres dans la classe arbre.

²Squelettiser : méthode des morphomathématiques qui permet de réduire un objet à sa forme la plus compacte tout en conservant son homotopie

Cette méthode est très simple et efficace, mais elle permet seulement de reconstruire un seul type d'arbre (modèle simple). Ce qui peut être un handicap lorsque l'on désire un rendu photoréaliste.

4.2.2 La Reconstruction 3D des bâtiments

La reconstruction des bâtiments est plus complexe. Comme pour les arbres, on utilise une classe obtenue dans la première étape de traitement (la classe bâtiment).

Dans un premier temps on va subdiviser la classe bâtiment en rectangles primitifs(cf (Figure:4.5)).



FIG. 4.5 – Décomposition de la classe bâtiments en rectangles primitifs.

Ces rectangles forment la base de la reconstruction, il ne reste plus qu'à déterminer la forme du toit et sa hauteur pour cela on utilise une évaluation au moindre carré qui réduit au minimum les distances entre la surface du MNE et les types de toit de bases. Les façades sont constituées de plans sur lesquels on viendra mapper des images terrestres pour obtenir le résultat de la méthode (Figure:4.6)

Pour ma part, je trouve que cette méthode est très simpliste. En effet, bien qu'il n'existe que peu de type de toit, on trouve tout de même beaucoup de variantes. Donc, cet algorithme ne peut effectuer une reconstruction que très grossière de certains bâtiments. Je pense qu'il pourrait être intéressant à faire un découpage plus poussé des bâtiments. Pour ma part, comme les auteurs, j'utiliserai la classe

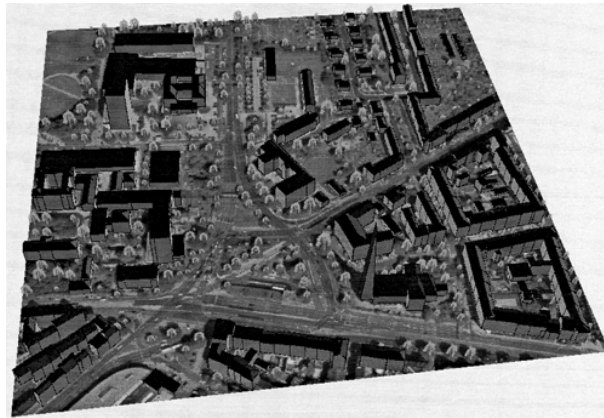


FIG. 4.6 – Résultat de la méthode.

bâtiment. Ensuite je commencerai par effectuer la détection de contour de Deriche (cf. 3.7) ce qui me permettra trouver les contours de chaque pan de toit, ainsi que les fetses de toit. Puis, pour simplifier les futurs traitements je supprimerais tous les segments d'une taille inférieure à 50cm par exemple car ils ne sont pas significatifs d'un pan de toit. Après cette étape, je chercherai à extraire chaque bâtiment, afin de pouvoir appliquer sur chaque bâtiment des contraintes géométriques. En effet, les murs d'une maison sont généralement perpendiculaires ou parallèles. L'application de ces contraintes nous permet d'obtenir des forment très proche de la réalité. Puis je découperais chaque bâtiment en un ensemble de pan de toit et, pour chaque pan, je ferais une estimation au moindre carré du plan à associer aux pans de toit (donc de dessiner un polygone 3d). Ce qui permettrait d'obtenir une reconstruction plus fine, pouvant modéliser tout type de toit.

Chapitre 5

Conclusion

Tout d'abord, je commencerai par conclure sur la méthode. En premier lieu, la méthode de classification, je pense qu'elle est bonne, bien qu'elle puisse être améliorée par l'utilisation du cadastre. Par contre, je trouve la méthode de reconstruction très simple, en effet comme je vous l'ai présentée rapidement on peut augmenter la finesse de la reconstruction en utilisant un algorithme plus élaboré. Néanmoins, je pense que l'approche des auteurs est très intéressante car elle utilise les données laser et les images ce qui permet d'obtenir de meilleurs résultats.

Sur l'article lui-même, mon avis est qu'il reste très général. En effet, il n'entre pas dans les détails de l'implémentation. Il m'a donc été difficile de présenter avec précision la méthode élaborée par les auteurs.

Bibliographie

- [Bol02] D. Boldo. *Fabrication conjointe de Modèles Numériques de Surface et d'Orthoimages pour la visualisation perspective de scènes urbaines*. PhD thesis, Université René Descartes-Paris V, 2002. 11
- [Can86] J. Canny. A computational approach to edges detection. In *IEEE trans. On pattern analysis and machine intelligence PAMI-8*, pages no. 6,679–698, 1986. 21
- [Der90] R. Deriche. Fast algorithms for low level vision. In *IEEE PAMI 12*, pages no. 1,78–87, 1990. 21
- [Der93] R. Deriche. Techniques d'Extraction des contours. INRIA Sophia-Antipolis, 1993. 21
- [Fau93] O. Faugeras. Three-dimensional computer vision. In *MIT Press*, Cambridge, 1993. 6
- [HB99] N. Haala and C. Brenner. Extraction of buildings and trees in urban environments. In *ISPRS Journal of Photogrammetry and Remote Sensing Vol. 54*, pages 130–137, 1999. 7
- [Lam60] J.H. Lambert. *Photometria sive de mensura et gradibus luminis*. In *colorum et umbrae*, 1760. 16
- [Maa99] HG. Maas. Fast determination of parametric house models from dense airborne laserscanner data. In *ISPRS Workshop on Mobile Mapping Technology*, 1999. 7
- [Ric93] J.A Richards. Remote sensing and digital image analysis. In *Springer-Verlag*, Berlin, 1993. 26
- [UW95] W. Forstner U. Weidner. Towards automatic building extraction from high resolution digital elevation models. In *ISPRS J. 50 (4)*, page 38–49, 1995. 26
- [VS01] G. Vosselman and I. Suveg. Map based building reconstruction from laser data and images. In Basel : Birkhaeuser Verlag, editor, *Automatic Extraction of Man-Made objects from aerial and space images*, pages 231–239, Ascona, 2001. 7