Contributions to the 3D city modeling : 3D polyhedral building model reconstruction from aerial images and 3D facade modeling from terrestrial 3D point cloud and images

Karim Hammoudi

UNIVERSITÉ PARIS-EST

IGN INSTITUT GEOGRAPHIQUE NATIONAL

UNIVERSITE PARIS-EST - DOCTORAL SCHOOL N°532: MATHEMATICS, SCIENCES AND TECHNOLOGIES OF INFORMATION AND COMMUNICATION

INSTITUT GEOGRAPHIQUE NATIONAL - LABORATOIRE DE METHODES D'ANALYSES ET DE TRAITEMENT D'IMAGES POUR LA STEREO-RESTITUTION

Ph.D. THESIS BOOK

Thesis submitted for the degree of

Doctor of Philosophy in Signal and Image Processing from the *Université Paris-Est*

CONTRIBUTIONS TO THE 3D CITY MODELING

– **3D POLYHEDRAL BUILDING MODEL RECONSTRUCTION FROM AERIAL IMAGES**

– **3D FACADE MODELING FROM TERRESTRIAL 3D POINT CLOUD AND IMAGES**

By Karim Hammoudi

December 2011

Doctoral Committee:

Nicole Vincent	President & Examinator	Univ. Paris Descartes	France
Angel Sappa	Reviewer	Computer Vision Center	Spain
Véronique Cherfaoui	Reviewer	Univ. de Technologie de Compiègne	France
Beatriz Marcotegui	Examinator	MINES ParisTech / CMM	France
Nicolas Paparoditis	Ph.D. Director	Univ. Paris-Est / IGN	France
Fadi Dornaika	Ph.D. Supervisor	Univ. of Basq. Country / Ikerbasque	Spain
Bahman Soheilian	Ph.D. Supervisor	Univ. Paris-Est / IGN	France

---

INSTITUT GEOGRAPHIQUE NATIONAL
Laboratoire MATIS, 73 avenue de Paris, 94165 Saint-Mandé cedex, FRANCE

MATIS

Ph.D. THESIS BOOK

CONTRIBUTIONS TO THE 3D CITY MODELING

- **3D POLYHEDRAL BUILDING MODEL RECONSTRUCTION FROM AERIAL IMAGES**
- **3D FACADE MODELING FROM TERRESTRIAL 3D POINT CLOUD AND IMAGES**

To cite this thesis:

BibTEX entry

```
@PhdThesis{Hammoudi:PhD,
author = {Karim Hammoudi},
title = {Contributions to the {3D} city modeling.
{3D} polyhedral building model reconstruction from aerial images \&
{3D} facade modeling from terrestrial {3D} point cloud and images},
school = {Universite Paris-Est},
year = {2011}}
```

Classic entry

Hammoudi, Karim. Contributions to the 3D city modeling. 3D polyhedral building model reconstruction from aerial images & 3D facade modeling from terrestrial 3D point cloud and images, Ph.D. thesis, Université Paris-Est, 2011.

Thesis prepared at the Institut Géographique National (IGN) in France and carried out in collaboration with the University of Basque Country (UPV/EHU) and the Basque Foundation for Science (Ikerbasque) both located in Spain. Please report any errors, typos or misspellings (even minor ones) to karim.hammoudi@ign.fr

"N'allez pas où le chemin peut mener, allez la où il n'y a pas de chemin et laissez une trace."

"Do not go where the path may lead, go instead where there is no path and leave a trail."

Ralph Waldo Emerson (1803-1882)

"L'imagination est plus importante que la connaissance."

"Imagination is more important than knowledge."

Albert Einstein (1879-1955)

Contents

Avant-propos / Foreword

Cette thèse a été conduite dans un contexte qui allie des activités professionnelles variées. D'une part, des activités de recherche ont été menées, notamment, par des échanges scientifiques réguliers en interne à l'Institut Géographique National Français et en externe par une collaboration avec le département informatique et intelligence artificielle de l'Université du Pays Basque situé en Espagne. D'autre part, des activités d'enseignements en programmation informatique ont été réalisées en parallèle des activités de recherche au département mathématiques et informatique de l'Université Paris Descartes (poste au tiers-temps d'enseignement durant 3 ans).

Le présent manuscrit s'articule essentiellement autour de quatre parties organisées de la façon suivante. La Partie I présente le contexte de la thèse ainsi que la définition du problème. La Partie II comprend un chapitre qui porte sur la modélisation 3D de bâtiments à partir d'un jeu d'image aériennes multiscopiques (Chapitre 2). Plus précisément, l'objectif consiste à modéliser les toits des bâtiments de manière à obtenir des modèles 3D de formes polyédriques. L'approche proposée est appropriée pour la modélisation de bâtiments comprenant des toits aux topologies variées tels que des toits multi-pans et ce même en présence de superstructures. La Partie III comprend trois chapitres. Ces chapitres portent essentiellement sur la modélisation du bâti à partir de données terrestres. Respectivement, en Chapitre 3, le système d'acquisition terrestre, le protocole d'acquisition ainsi que les données acquises sont présentés. De plus, les challenges et les difficultés dans le traitement des données sont discutés afin de mettre en évidence la compléxité du problème pour atteindre efficacement une solution recherchée. En Chapitre 4, une approche est proposée en vue de la segmentation et de la classification des nuages de points de rue dans les données laser terrestres. Finalement, une approche est proposée en Chapitre 5 en vue de la modélisation et de le texturage sans occultations des façades de rues urbaines en combinant des données multi-sources; à savoir, des images optiques, du laser et un plan casdastre. Nous mentionnons aux lecteurs que les Parties II et III peuvent être lues indépendamment.

La Partie IV Chapitre 6 présente une conclusion générale en retraçant les recherches majeures conduites au cours de la thèse au niveau de l'aérien et du terrestre et en proposant des pistes sur des travaux de recherches qui pourraient être entrepris à l'avenir.

Bonne lecture !

This thesis has been carried out in a context that includes various professional activities. On the one hand, research activities have been achieved, notably by regular scientific discussions in internal at the French National Mapping Agency and by an external collaboration with the department of Computer Science and Artificial Intelligence from the University of Basque Country located in Spain. On the other hand, teaching activities in computer programming have been led in parallel of the research activities at the Department of Mathematics and computer science at the University of Paris Descartes (tierce-time position in teaching during three years).

The presented manuscript is essentially organized into four parts as follows. The Part I presents the context of the thesis as well as the problem statement. The Part II includes one chapter which deals with the 3D modeling of buildings from multiple aerial images (Chapter 2). More specifically, the goal is to model the roofs of buildings as a 3D polyhedral models. The proposed approach is suitable for modeling various roofs topologies such as multi-sided even in presence of superstructures. The Part III contains three Chapters. These Chapters are essentially focused to the 3D building modeling from terrestrial data. Respectively, in Chapter 3, the system of terrestrial acquisition, the acquisition protocol as well as the acquired data are presented. Moreover, the challenges and the difficulties that lie in the data processing are discussed in order to enhance the understanding of the problem complexity for efficiently reaching a searched solution. In Chapter 4, an approach is proposed to the segmentation and classification of street point clouds in urban terrestrial laser data. Finally, an approach is proposed Chapter 5 to the modeling and the occlusion-free texturing of urban street facades by combining multi-source data; namely images, laser and cadastral map. We mention to the readers that the Parts II and III can be read independently. Part IV Chapter 6 presents a general conclusion regarding the researches conducted at aerial and terrestrial level as well as future works that could be investigated.

Happy reading!

Remerciements / Acknowledgments

Je tiens en premier lieu, à remercier ma famille, mes grand-parents disparus et qui ont toujours cru en moi, ma grand-mère Mme Yvonne Dulouard, toujours à mon écoute et mon grand-père M. Raymond Dulouard, Chevalier dans l'Ordre National du Mérite Agricole, pour ses encouragements. Je pense également à mes grand-parents paternels, ma grand-mère Mme Aicha Hammoudi et mon grand-père M. Mohamed Hammoudi, artisan forgeron-serrurier.

Je remercie mes parents Viviane et Larbi, mon frère Sami, mon épouse Malika et mes beaux-parents Fatima et Azzouz pour m'avoir énormément soutenu et encouragé au cours de la thèse. Durant mon parcours, j'ai bénéficié de l'éducation de mes parents, dans l'attachement de ma mère au suivi de mes études et de mon père, Compagnon du Devoir et du Tour de France, Meilleurs Ouvrier de France et Chevalier dans l'Ordre National du Mérite. Je dédie ce travail à ma famille car il n'aurait pas eu lieu sans leur soutien.

Je souhaite exprimer ici ma sincère gratitude à Mme Nicole Vincent, professeur à l'Université Paris Descartes et responsable de la spécialité Images du Master mathématiques et informatique dont je suis issu. Je la remercie pour m'avoir permis d'intégrer dès le début de la thèse l'équipe des enseignants de l'Université Paris Descartes et de m'avoir confié durant 3 ans des enseignements en informatique dans le cycle Licence.

J'aimerais ensuite exprimer mes profonds remerciements et toute ma reconnaissance à M. Fadi Dornaika, professeur à l'Université du Pays Basque et à la Fondation Basque pour les sciences (Ikerbasque). Il m'a encadré de façon permanente dès le début du stage de Master jusqu'à l'aboutissement de la thèse. À ses côtés, j'ai évolué tant au niveau méthodologique que scientifique. J'ai apprécié cette passation du savoir et du savoir-faire et je le remercie pour sa disponibilité et sa patience dans la validation étape par étape des recherches par des publications. Nous avons échangé continuellement sur les idées, les pistes de recherche à investir, les méthodologies à mettre en œuvre. Je n'ai que trop apprécié ces moments de réflexions animés de précieux conseils, de raisonnements scientifiques rigoureux et de remarques pertinentes.

J'exprime mes profonds remerciements et toute ma reconnaissance à M. Nicolas Paparoditis, directeur du laboratoire MATIS de l'IGN et également directeur de la thèse. Durant mon Master, ce fut mon professeur en imagerie 3D et il m'a conduit de fait à m'intéresser particulièrement aux travaux menés à l'Institut Géographique National. Je le remercie pour la confiance et le soutien qu'il m'a accordés dans les recherches menées ainsi que pour son investissement dans les échanges qui ont contribué à orienter, positionner et faire progresser les recherches. Je le remercie pour avoir toujours su veiller à entretenir dans l'environnement de travail des relations humaines et chaleureuses avec chacun d'entre nous. Je le remercie de m'avoir impliqué dans les projets de recherches IGN très enrichissants et de m'avoir permis de présenter régulièrement l'avancement de mes travaux dans des manifestations internationales.

Je tiens également à remercier M. Didier Boldo, ancien directeur du laboratoire MATIS pour m'avoir accordé son appui lors de mes recherches d'un financement de thèse pour poursuivre à l'IGN les travaux entrepris lors de mon Master. Je remercie également M. Patrice Bueso, ancien

directeur du service de la recherche de l'IGN, pour m'avoir très bien accueilli dans l'équipe du MATIS.

Je remercie M. Bahman Soheilian, chargé de recherche au laboratoire MATIS pour m'avoir permis de bénéficier de son encadrement en cours de thèse. Cette étape délicate à entreprendre fut entre autres importante pour maintenir en interne une veille sur les besoins du laboratoire au niveau des recherches à mener. Je le remercie donc pour avoir su renforcer mon encadrement, pour les travaux de recherche que nous avons entrepris, pour les remarques pertinentes et pour le soutien technique qu'il m'a apporté.

Je remercie le chef du département informatique et intelligence artificielle de l'Université du Pays Basque en Espagne, M. F. Xabier Albizuri pour m'avoir reçu lors de mes séjours doctoraux.

J'exprime mes profonds remerciements à Mme Nicole Vincent, Mme Beatriz Marcotegui, Mme Véronique Cherfaoui et M. Angel Sappa pour avoir accepté de représenter mon jury de thèse. Je vous suis particulièrement reconnaissant pour les corrections et les remarques que vous m'avez transmises et qui ont contribué à améliorer la qualité du manuscrit.

Je tiens à remercier mon collègue, M. Jean-Pascal Burochin avec qui j'ai partagé ces trois années toujours dans la joie et la bonne humeur. Je te remercie pour ton enthousiasme et tes implications actives dans le travail qui ont conduit à des partages de connaissances et des échanges d'expériences enrichissants tant sur le plan humain que sur le plan scientifique.

Je tiens à remercier Mme Sylvie Cach, responsable administrative de l'école doctorale ED MSTIC, M. Robert Eymard, professeur et directeur de l'ED MSTIC et M. Bernard Lapeyre, directeur du département des études doctorales de l'Université Paris-Est pour ces rencontres très enrichissantes dans les réunions du conseil de l'école doctorale, les réunions de l'association des doctorants REDOC Paris-Est et les manifestations doctorales environnantes. Elles m'ont permis de me tenir informé et de bénéficier des formations, dispositifs, bourses mises en place pour la progression des doctorants. Je les remercie pour leur attention dans le suivi de mon avancement; et pour leurs encouragements en vue de réaliser la thèse dans les temps impartis.

Je remercie M. Nicolas Loménie, professeur d'analyse vidéo à l'Université Paris Descartes (durant mon master) pour ses encouragements dans ce qu'on peut effectivement appeler, comme il me l'a dit à mes débuts "une aventure très enrichissante".

Je tiens à remercier M. Matthew Rigby, physicien originaire d'Ottawa, venu à l'époque dans le cadre des échanges internationaux pour finaliser son Master à Paris à l'Université Pierre et Marie Curie et qui fut mon voisin en résidence universitaire lors de mes débuts en thèse. Je me souviens des longues soirées à échanger et il m'a énormément soutenu afin que je progresse dans la langue de Shakespeare. Aussi, je le remercie ainsi que sa famille pour leur accueil chaleureux lors de nos retrouvailles à une conférence en vision qui s'est déroulée à Ottawa.

Je tiens également à remercier M. Sami Metari, chercheur en vision au laboratoire LVSN de l'Université Laval au Canada et M. Abdelaziz Khiat, chercheur en vision au centre de recherche Nissan Motor de Kanagawa au Japon pour les discussions enrichissantes et les conseils avisés pour progresser dans la recherche.

Je remercie en particulier M. Mahieu Brédif et M. Olivier Tournaire, chercheur au MATIS pour leur aide technique lors de mon arrivée à l'IGN. Je remercie M. Lâmân Lelégard, ingénieur de recherche au MATIS pour ces discussions sur les recherches entreprises par le passé à l'IGN et pour sa contribution dans la partie délimitation de façades. Je remercie M. Bertrand Cannelle et M. Jean-Pierre Papelard, ingénieurs au MATIS pour m'avoir donné les moyens techniques et les informations afin d'exploiter les données du véhicules de cartographie de l'IGN. Je remercie profondément M. Bruno Vallet, chercheur au MATIS pour son aide dans la partie mosaïquage.

Je remercie aussi tous les autres membres du laboratoire MATIS qui par leur professionnalisme, ont rendu mon séjour au laboratoire agréable et constructif, Nicolas David, Isabelle Cléry, Fabien Coubard, Nicolas Champion, Alexandre Devaux, Marc Pierrot-Deseilligny, Laure Chandelier, François Boyero, Antonio Ferraz, Athanasios Georgantas, Patrick Julien, Clément

Mallet, Mahzad Kalantari, Melanie Sedze, Frédéric Bretar, Nesrine Chehata, Daniela Craciun, Erwann Houzay, Corina Iovan, Lionel Pénard. Enfin, mes remerciements ne sauraient être exhaustifs sans que je ne remercie tous mes amis et camarades de l'Université Paris-Est, de l'association REDOC Paris-Est, de l'Université Paris Descartes et de l'Université de Picardie Jules Verne.

Un grand merci à tous.

Abstract / Résumé (In French)

The aim of this work is to develop research on 3D building modeling. In particular, the research in aerial-based 3D building reconstruction is a topic very developed since 1990. However, it is necessary to pursue the research since the actual approaches for 3D massive building reconstruction (although efficient) still encounter problems in generalization, coherency, accuracy. Besides, the recent developments of street acquisition systems such as Mobile Mapping Systems open new perspectives for improvements in building modeling in the sense that the terrestrial data (very dense and accurate) can be exploited with more performance (in comparison to the aerial investigation) to enrich the building models at facade level (e.g., geometry, texturing).

Hence, aerial and terrestrial based building modeling approaches are individually proposed. At aerial level, we describe a direct and featureless approach for simple polyhedral building reconstruction from a set of calibrated aerial images. At terrestrial level, several approaches that essentially describe a 3D urban facade modeling pipeline are proposed, namely, the street point cloud segmentation and classification, the geometric modeling of urban facade and the occlusion-free facade texturing.

Keywords: *3D City Modeling, Building Roof Reconstruction, Street Modeling, Facade Modeling and Texturing, Calibrated Aerial Images, Multiscopy, Mobile Mapping System, Terrestrial Laser data, 3D Point Cloud, Geometry, Hypothesize-and-verify Strategies, Computer Vision, Photogrammetry, Remote Sensing, Geomatics*

L'objectif principal de ce travail est le développement de recherches en modélisation 3D du bâti. En particulier, la recherche en reconstruction 3D de bâtiment est un sujet très développé depuis les années 90. Malgré tout, il paraît nécessaire de poursuivre les recherches dans cet axe étant donnée que les approches actuelles consacrées à la reconstruction 3D de bâtiment (bien qu'efficaces) rencontrent encore des difficultés en terme de généralisation, de cohérence et de précision. Par ailleurs, les récents développements des systèmes d'acquisitions de rues tel que les systèmes de cartographie mobile ouvrent de nouvelles perspectives d'amélioration de la modélisation des bâtiments dans le sens ou les données terrestres (très précises et résolues) peuvent être exploitées avec davantage de cohérence (en comparaison à l'aérien) pour enrichir les modèles de bâtiments au niveau des façades (la géométrie, la texture).

Ainsi, des approches de modélisation aériennes et terrestres sont individuellement proposées. Au niveau aérien, nous décrivons une approche directe et

dépourvu d'extraction et d'assemblage de primitives géométriques en vue de la reconstruction 3D de modèles polyédriques simples de bâtiments à partir d'un jeu d'images aériennes calibrées. Au niveau terrestre, plusieurs approches qui décrivent essentiellement un pipeline pour la modélisation 3D des façades urbaines sont proposées; à savoir, la segmentation et classification de nuage de rues urbaines, la modélisation géométrique des façades urbaines et le texturage des façades urbaines comportant des occultations causées par d'autres objets du mobilier urbain.

Mots clés : *Modélisation 3D des villes, Reconstruction des toits de bâtiments, Modélisation des rues, Modélisation et texturage des façades, Images aériennes calibrées, Multiscopie, Système de cartographie mobile, Données laser terrestre, Lidar, Nuage de points 3D, Géométrie, Génération et vérification d'hypothèses, Vision par ordinateur, Photogrammétrie, Télédétection, Géomatique.*

List of Figures

List of Tables

Glossary of Acronyms and Notations

ANR	Agence National de la Recherche
CCA	Connected Component Analysis
CSG	Constructive Solid Geometry
DCM	Digital Cadastral Map
DE	Differential Evolution
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
GA	Genetic Algorithms
GIS	Geographic Information System
GPS	Global Positioning System
GSD	Ground Sample Distance
IBMR	Image-Based Modeling and Rendering
IGN	Institut Géographique National
IMU	Inertial Measurement Unit
iTOWNS	image-based Town On-line Web Navigation and Search Engine
LiDAR	Light Detection And Ranging
LSA	Least Squares Adjustment
MATIS	Méthodes d'Analyse et de Traitement d'Images pour la Stéréo-restitution
MBR	Minimum Bounding Rectangle
MLS	Mobile Laser Scanning
MMS	Mobile Mapping System
MSD	Mean Squared Deviation
NIR	Near-InfraRed
SA	Simulated Annealing
SHT	Standard Hough Transform
PPHT	Progressive Probabilistic Hough Transform
RANSAC	RANdom SAmple Consensus
RJMCMC	Reversible Jump Markov Chain Monte Carlo
ROI	Region Of Interest
SAD	Sum of Absolute Differences
SLAM	Simultaneous Localization And Modeling
SSD	Sum of Squared Differences
TIN	Triangulated Irregular Network
TLS	Terrestrial Laser Scanning
TRAPU	Tracé Automatique de Perspectives Urbaines

$\mathcal{A}$	Accumulation map
$\mathcal{B}$	Binary image
$\mathcal{C}$	Cluster of points or point cloud
$\mathcal{C}_i$	Camera i
$\mathcal{D}$	Dataset
λ_X	Linear coordinates of the point X
$\mathcal{I}$	Optical image
Π	Plane
$\mathcal{S}$	Score
$\mathcal{T}$	Threshold
$\mathbf{w}$	Vector of model parameters

Part I

Introduction

Chapter 1

Introduction

Abstract

This Chapter exposes in a global context the common interests of some fields towards the 3D terrestrial mapping as well their ways to tackle the topic under different points of view. Also, some of concrete applications in 3D city modeling in which these fields contribute are presented. In a second time, the thesis is positioned amongst the whole of research projects led at the IGN. Each described project should contribute from near and far at the enhancement of the building description towards cartographic applications. The project as well as their stakes are then described. Finally, the problem statement of the thesis, the major contribution as well as the manuscript organization are presented.

Contents

1.1 Global context

This section aims at providing to the reader a short overview on the 3D city modeling topic by describing certain specific aspects as well as concrete objectives of usage.

1.1.1 Towards 3D Mapping

1.1.1.1 Geography and Informatics: Geomatics

In 1960, strong links between the studies of the geo-localization (e.g. in geodesy, geography, cartography) and the automatic processing of information have been significantly observed and have progressively conducted to a novel field called geomatics. Henceforth, this field is officially renowned by international research organizations and consists of the management of geospatial data (e.g.; by using geographic information systems (GIS)) and to the use of surrounding technologies and fields related to their acquisition, their storage, their processing and their broadcast such as computer science, mapping,

remote sensing or photogrammetry. The geomatic field represents then an essential support as well in planning of urban environments as in development of the urban 3D modeling approaches. In particular, an overview of research in geomatics carried out at the French National Mapping Agency can be found in (De Blomac, 2011).

1.1.1.2 Computer Vision vs Photogrammetry

3D Modeling consists in a geometrical description of an object or a scene in a three-dimensional space. The generation of 3D models of scenes and environments is one of the major research topic for the computer vision and photogrammetry communities. However, the problem of modeling was tackled in these fields under different angles. In computer vision, one of the main aims consists in providing to the robots the three-dimensional map (e.g. depth map) of the surrounding scene. This map allows to the robots to locate the obstacles in order to carry out manoeuvres and displacements. The approaches are specifically developed to be usable in real-time. In cartography, the data are collected at large scales then post-processed. The 3D information is employed to produce Digital Cadastral Maps (DCM) or 3D city model. Research deals more with the accuracy and the detail of the modeling. Nowadays, the researchers issue from these disciplines strongly interact and confront their approaches of this problem.

1.1.1.3 A Photogrammetric Challenge

The 3D city modeling is a real challenge that depends on the data acquisition and its succeeding processing. Although it appears enigmatic at first glance, active research are carried out in photogrammetry to the extraction of 3D information from single or multiple 2D optical images. This extraction of information strongly depends on the quality of the acquired images. The employed images generally are acquired by digital cameras (e.g.; CCD) that record the light backscattered from the scene. These cameras include a matrix comparable to a set of photoelectric cells that produce an electric current according to the quantity of received light. This electric current is then converted into conventional image intensity. These digitalization stages (discretization, quantification) involves an important loss of information. Also, the collected information is not perfect due to the noises related to the sensor (e.g. lens distortions). Consequently, the 3D modeling is a reverse problem that consists in recovering the initial 3D scene by considering these factors. Although inexhaustive, we mention that an overview dealing with the photogrammetry history is presented in (Burtch, 2008). Also, we mention that specific techniques to the digital photogrammetry can be found in ((Kasser et al., 2002), renowned book to the IGN).

1.1.1.4 Evolution of the Mobile Mapping Systems

As previously mentioned, the 3D modeling is a large and emerging research field and the current study is focused to the modeling of urban objects. One can observe meaningful progression of the usage of various, parameterizable and adaptable sensors (panoramic camera, rotary or profile scanner) to collect data according with the viewed scenes, the size of the scanned objects and

the considered application. These sensors that integrate the mapping systems can be classify into two major categories, static (e.g., tripod) and mobile such as Mobile Mapping System (MMS). Mobile Mapping Systems can use robots, drones, helicopters or vehicle.

In our study, we are particularly interested in the MMS devices in reason of their ability for covering wide urban areas. The terrestrial MMS are relatively recent in comparison to the conventional aerial acquisition systems. The first operational land-based MMS was developed by the Center for Mapping at the Ohio State University (called GPSVan^{TM}) and that is then become a pioneer in this technology. In (Ellum et al., 2002), the authors describes in more detail the history of the Land-Based Mobile Mapping Systems. Nowadays, it is observed that these systems are more and more widespread since particularly suitable to the massive acquisition of images and laser street data in dense urban environments.

1.1.1.5 Complementarity of Image and Laser Data

The remote sensing is a science which consists in acquiring of remote information by the usage of a measuring instrument. Thus, digital camera and laser sensor are respectively employed in photogrammetry and lasergrammetry fields in order to collect urban data. Actually, the devices of acquisition tend then to combine digital camera with laser sensor. Indeed, this combination appears interesting as input data for 3D city modeling approaches in the sense that the laser data locally provides very accurate geometric measurements of the scene and are robust to the shadows whereas the optical images provide a radiometric information that is less accurate but very rich. The laser data can be used for estimating the object surfaces and the radiometric information can then be exploited for the accurate extraction of object boundaries at surface junctions in the sense that these junctions are often missed by the laser beam. Also, the radiometric information can be used for texturing generated laser-based 3D models. Since complementary, the scientific exchanges between the photogrammetry and the lasergrammetry communities are more and more strong and some studies investigate the generation of 3D model by merging these data (e.g.; (Abdelhafiz, 2009)).

1.1.1.6 3D Shape Recovering

Pattern recognition and image analysis are essential disciplines used to the generation of 3D city model. The pattern recognition is more related to the geometric description of the scene into 3D whereas the image analysis deals with the extraction of meaningful information from images. In conventional photogrammetry, the third dimension is calculated using pairs of stereoscopic images inspired by the binocular human visual system. Indeed, the principle of stereoscopy is a usual technique employed to produce a 2.5D depth map of the scene from calibrated images. This depth map provides a geometrical description of the relief and constitutes a supplemental type of data, often employed in input of 3D modeling approaches such as a 2.5D draft. In cartography, the depth maps that describe the urban environment are in general the Digital Surface Maps (DSMs) that include the ground and building surfaces as well as the Digital Terrain Maps (DTMs) that represent the ground surface alone.

The handle data to the modeling description usually are either raster (pixels or voxels) or vectorial data (segments, polygons).

1.1.1.7 Usage of 3D City Models

Cities are composed of numerous urban objects (e.g.; buildings, roofs, facades, roads, road signs). In our case, the urban objects of interest are the building roofs as well as the facades in general. The topic of 3D building reconstruction interests various institutional and industrial sectors:

- Autonomous navigation of vehicles in urban areas (car industry).
- Realism increasing of video games (video game industry).
- Terrestrial recognition for the guidance of missiles (army, defense).
- Enrichment in 3D model of flight simulators (aeronautic industry).
- Digitalization of architectural building (storage of the patrimony).
- Seismic simulations for territory planning (urban planning).
- Virtual navigation/visits through urban environments (Digital Globe, Google Earth).
- Optimization of antennas position for data transmission (telecommunications industry).

According to the needs, research are led in order to improve the quality of the 3D city model (e.g., accuracy, level of detail, runtime, rendering). The modeling pipelines are usually composed of several modules that will be discussed in the problem statement section.

1.2 Specific Context of the Thesis

In this section, we describe the context of the thesis, namely, the positioning of the thesis, the achieved collaborations as well as surrounding research works.

1.2.1 Positioning of the Thesis in the IGN Research

Since almost two decades, the 3D reconstruction topic is addressed at the MATIS laboratory of the *Institut Géographique National* (French National Mapping Agency). The MATIS laboratory aims at satisfying two kinds of needs in this field. On the one hand, the laboratory responds to a set of IGN, ministerial and regalian requests. On the other hand, the laboratory aims at leading research for the general interest in close cooperation with institutions proficient in the field. Various research on modeling topics have been led and conducted to become the heart of competence of the laboratory.

In the literature, the research in building modeling from terrestrial laser data are still very recent in comparison to the existing research developed from aerial data. In the past decade, numerous IGN researches dealt with the 3D building modeling from aerial data such as e.g.;(Jibrini et al, 2000, Fuchs, 2001, Taillandier et al, 2004, Lafarge et al, 2010, Bredif, 2010) and significant researches have then been achieved at the building roofs and superstructures reconstruction. At present, one of the aim of this thesis particularly addresses 3D building modeling from terrestrial street data.

This thesis has been carried out according to several surrounding IGN research projects. The projects, the stakes as well as the thesis positioning are

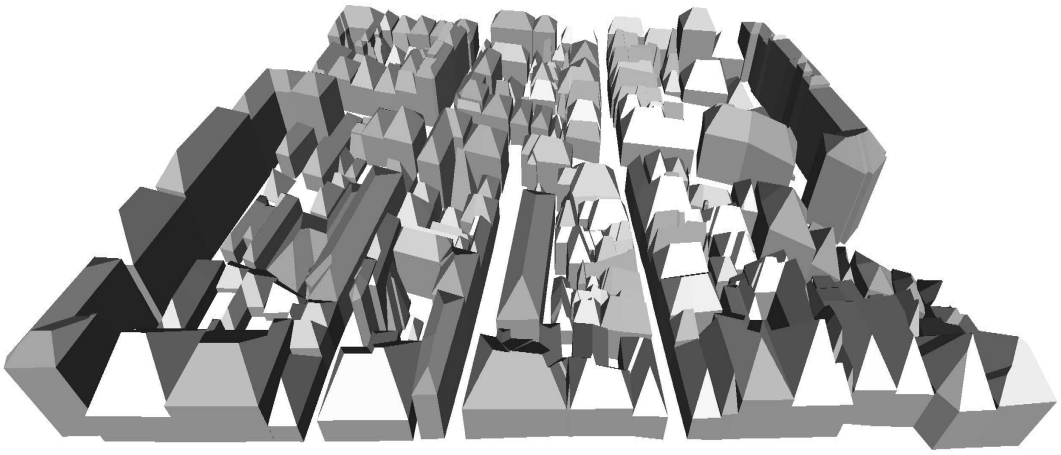


Figure 1.1: Illustration of 3D polyhedral building models of BATI-3D® prototype (Figure retrieved in (Bredif, 2010)).

described in more detail in the following subsections as well as in the section of problem statement. In particular, the project are presented by chronological order.

1.2.1.1 Aerial Pipeline for 3D Building Modeling

A prototype software (BATI-3D®) is developed at the IGN in order to generate 3D building model in urban environment. The generated models will be useful for the real-time navigation in a detailed model of the territory. The modeling pipeline is built upon high resolution aerial images and generated dense Digital Elevation Models of urban environments. The building models are automatically generated and include textures by suitable aerial ortho-images. Besides, the shape of the roofs are modeled in detail and Digital Terrain Model can also be generated. An illustration of a 3D city model (geometry) resulting of this pipeline is shown in Figure 1.1.

1.2.1.2 Street Mobile Mapping System

The IGN develops a top-notch Mobile Mapping System called Stereopolis[1] that aims at mapping into 3D the urban streets (see Stereopolis vehicule in Figure 1.2). First version of this MMS has been developed in 2006. Actually, this MMS is a vehicle equipped with laser sensors, panoramic and stereo digital camera in order to collect accurate street data. The laser data acquired from the street can reach a centimetric relative accuracy. These sensors are connected to a positioning technology combining GPS, odometer and IMU (Inertial Measurement Unit) data. The IMU is notably integrated in the submarines provides the vertical orientation, the north location and the speeding of a solid in the space. Since the Satellite reception of the GPS can be erroneous in dense urban environments, the IMU is used to rectify the ill-location of GPS points and to fill the hole in the signal. The whole of these devices are connected to on-board computers and units of data storage. More details

1. Stereopolis is the contraction of two greek words; *stereo* that refers to the stereoscopy and *polis* that means the city.

Figure 1.2: IGN street Mobile Mapping System.

about the georeferencing theory for the estimation of point coordinates measured from mobile acquisition platforms are presented in ((Ellum et al., 2002)). The data collected at street scale can cover large areas such as a district or a city. The acquired data bring massive and various information on streets that can enrich GIS databases and 3D city models. In particular, one can notice that such MMS-based data are employed at the IGN for the reconstruction of road marks (Arlicot et al., 2009) and road signs from stereo-images (Soheilian, 2008) or for the extraction of vertical urban posts from street laser data (Liberge et al., 2010). Besides, the collected data are also employed for the development of platforms dedicated to the street visualization such as the IGN Google-like Street Viewer presented below. We finally mention that a detailed description of the acquisition system and its experimentation is provided in Chapter 3 of Part III.

1.2.1.3 Street Viewer from Panoramic Images

The iTowns project is financed by the French National Research Agency and Cap Digital, i.e. cluster of research institutes). The goal is to merge an immersive urban viewer available online for net surfers (as Google Street View (Anguelov et al., 2010)) with a search engine multi-source (e.g., text or image queries, linked to GIS urban databases). This viewer exploits high resolution panoramic images of street provided by the IGN Mobile Mapping System previously presented. The users directly navigate across the image flow running without the use of pre-defined 3D city model. The challenges are fluid navigation and visualization at street level, the automatic import/export of urban street features (e.g.; exploitation of this features to semantically enrich the GIS database or external 3D city model) as well as the generation of learning databases with intelligent systems in data mining (e.g., intelligent navigation of machines). The objective is to create a national service addressed to the large public as well as to professionals. For the needs of this project, the whole of the 12th district of the city of Paris has been covered. This acquisition has

Figure 1.3: Google like Street-viewer developed to the IGN. Assembly of optical images that constitutes high resolution panoramic images useful for the street vizualisation and navigation at 360°.

required $12Hrs$ of driving to sweep $180km$. $4Tbytes$ of raw data have been collected. More than 25000 panoramic images have been composed from the 10 full HD camera. Conjointly, thousands of points have been acquired at street level and are also used to enrich the platform in functionalities. The Figure 1.3 illustrates the prototype platform developed to the immersive visualization and navigation of urban streets. More details on this platforms and its numerous functionalities are presented in (Devaux et al., 2010).

The stakes of this thesis according to these three major projects are specified in details in the following section.

1.3 Problem Statement

The development of novel approaches for 3D building reconstruction from aerial images is still of great interest. Some building modeling pipelines have been developed in the past but even if the rates of generalization are often high, it happens that approaches fail the reconstruction process in certain cases. This deficiency is in general related to the input data quality that can be noisy, incomplete or to the building shapes that can be complex. Consequently, the full building modeling pipelines still require some human operators in order to manually rectify the ill-reconstructed building models. As for some full pipelines, it is still the case for the BATI-3D® project.

The data used as input of this pipeline also includes a pseudo-cadastral map used to initially select the building footprints in the aerial images (focusing stage). More precisely, this map is a derivation of the conventional cadastral map into footprints of building block. The digital cadastral map is the output of a set of successive stages such as footprint measurements achieved by surveyors, paper map digitalization, map vectorization (2D segments of dwelling parcels) and map simplification (2D segments of building block). It appears

that the map simplification stage often causes over or under inclusions of segments and requires then a manual rectification stage. The resulting map of building footprint can then be very inaccurate and include building shapes with topological ambiguities. Since these maps are not very accurate, the intermediary processing included into the modeling pipeline could be affected as well as the accuracy of resulting 3D building models. Furthermore, the whole of the manual stages mentioned above are extremely expensive and active mapping researches aim at reducing these tasks. An ideal alternative to these problems will be the automatic and massive generation of more accurate building footprints from terrestrial laser data.

Besides, the Digital Elevation Models are usually employed as support data for the 3D building modeling. The DEM is usually generated by a multi-correlation process from aerial images. However, the DEM is known to be inaccurate at discontinuities (e.g., ground/roof junctions). In this case, terrestrial building footprints could also be useful to generate discontinuity maps in order to rectify the Digital Elevation Model.

In a point of view independent of the 3D modeling aspect, it should be also interesting to study the automatic generation of Digital Cadastral Map (DCM) from street MMS data in order to refine and to update the surfacic information of the residence sizes. This information could be of great interest to the French National State since the DCM is initially employed for the individual taxation of the citizens according to their dwelling surfaces.

In short, researches should be investigated for the generation of accurate facade planes from street laser data. Thus, 2D segments associated to the delimited facade planes could be used for improvements/rectifications of usual building footprints at the initialization of the focusing stage in aerial-based modeling pipelines. In addition, the improvements of the building footprint qualities can enrich the cartographic data and the related applications (e.g., DEM's rectification and national residential taxation).

Besides, very few works have addressed the hybrid aerial/terrestrial 3D modeling pipelines. The reasons are: (*i*) the difficulty of data accessibility, (*ii*) the street shape complexity (heterogeneous environments), and (*iii*) the difficulty for fusing data of different scales. Since the building shapes appear different at centimetric and decimetric resolutions, the data fusion implies representation and generalization problems. However, the merge of data is necessary in order to complete the existing aerial modeling pipelines by integrating street modeling and texturing processes. The IGN research aims at continuously improving the quality of the 3D city model in addition to the conventional 2D urban maps. Then, one major aim of this thesis consists of conduct researches towards the development of a pipeline dedicated to the 3D urban facade modeling that could be used as a framework complementary to the existing IGN aerial modeling pipeline (BATI-3D®) or as an independent framework for external applications (e.g., 3D enrichments in video games, studies for robotic navigation).

It is worth mentioning that the presented work addresses the street-scale facade modeling by fusing terrestrial image and laser data towards producing a FACADE-3D modeling pipeline complementary to the BATI-3D® modeling pipeline. Also, in parallel to the need of producing virtual 3D city model of high quality (e.g.; for urban visualization), it can be noticed the strong evolution of online image based street viewer that graphically simulated a three-

dimensional environment although any existing 3D urban model was used (e.g.; Itowns Project). Hence, the IGN Mobile Mapping System previously described should rapidly collect required terrestrial urban data to deserve this large wave of needs.

Also, we have previously discussed about the usefulness and limitations of the conventional cadastral map in aerial-based modeling approaches. It will be also interesting to evaluate the usefulness and limitations of this map in the processing of the terrestrial data acquired by the MMS. In particular, we think to experiment the segmentation of the laser data by exploiting the vectorized cadastral map (set of building segments) to automatically crop the facade points and to individualize coplanar dwelling facades. We stress the fact that this kind of map is available and widespread in major cities. As a long-term goal, we would like to propose an approach that will minimize or fully avoid the use of the conventional cadastral map in the data segmentation or in the modeling pipeline. Indeed, since these novel ground-based data are collected in tremendous amount and with very high quality, they should foster the online generation of such map and active research are conducted across the world in this sense.

Besides, the calibration and registration problems concerning the aerial and terrestrial data that have been employed are beyond the scope of the current study. Since we are interested in facade modeling from laser points, it has been decided to limit the geometric representation of the studied scene to an intermediary level of modeling (e.g. polygonal meshes, polyhedrons).

At present, the thesis problems have been stated. In the following section, the contributions have been listed.

1.4 Major Contributions

We have previously described the positioning and aims of the thesis in some of national research projects. Concretely, the major contributions of this thesis are:

1. A featureless approach to generate simple 3D polyhedral building model from aerial images.
2. A model-less approach to the segmentation and classification in raw terrestrial urban street point cloud.
3. An approach to the modeling and occlusion-free texturing of urban facades from street point clouds and optical images.

These contributions have been described in detail in the respective chapters II2, III4, III5. The organization of these chapters and their contents is presented in the following section.

1.5 Organization of the Thesis

The manuscript includes 5 Parts that are organized as follows: The Part I describes the objectives of thesis and the problem statement. The Parts II and III respectively deal with the aerial and terrestrial building modeling. The achieved research follows the order of description of the IGN research projects

previously mentioned. Each contribution is then confined in a specific corresponding chapter. The Part IV concludes on the led researches and proposes perspectives. The Part V corresponds to the appendix of the thesis.

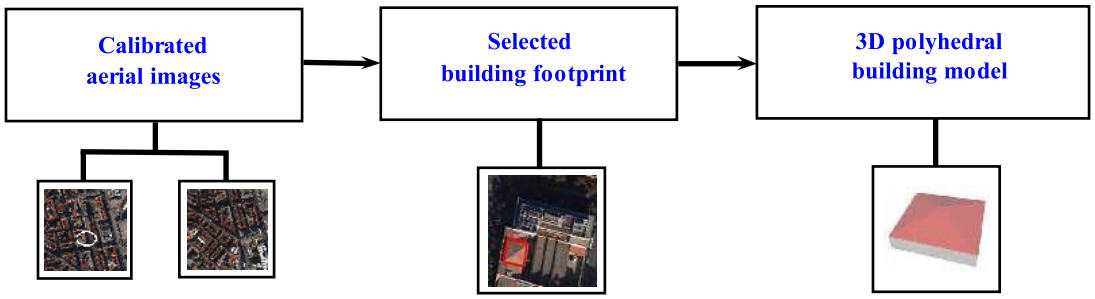


Figure 1.4: A featureless approach to generate 3D polyhedral building model from aerial images.

In Chapter 2 Part II, a featureless approach has been presented in order to generate 3D polyhedral building model from aerial images. Across studies on 3D building modeling, we have observed that the massive pipeline of 3D building modeling requires the intervention of operators and local modeling methods to rectify ill-reconstructed building model (incoherent shape, inaccurate estimation). Hence, we have decided to pursue researches on building modeling started during the Master internship. The extended research employs high resolution aerial images in order to reconstruct 3D polyhedral building model. The global diagram of the proposed approach is illustrated in Figure 1.4. This approach could be integrated into the BATI-3D® modeling pipeline or used independently in external research on aerial building modeling. This approach has the originality to be featureless contrarily to some modeling strategies in the sense that the conventional stages of feature extraction and assembly are avoided.

In Chapter 3 Part III, some architectural properties of the urban streets of Paris as well as similarities with big capitals and cities across the world have been highlighted. The analysis aims essentially at extracting universal geometric characteristic that could be used by the terrestrial modeling approaches. Furthermore, IGN mobile mapping system conception, optical and laser sensors as well as positioning instruments have been described. This description is followed by the analysis of the collected data and their managements. Amongst other things, it appears that some acquisition effects can affect the processing of the collected data. The delivered explanations justify the fact that the data must be manipulated with much precautions.

In Chapter 4 Part III, a model-less approach is proposed to the segmentation and classification of terrestrial urban street point cloud. The research deals with point cloud of wide street portions. In this study, we have undertook the pursuit of research from a data subset including full facades. Indeed, we have localized in the data initially collected samples of streets that include parallel alleys. In this way, the acquisition from the major roads allows to cover the facade top since the vehicle is sufficiently distant to the targeted facade. Besides, the raw street point cloud contains some street structures (posts, vegetations, cars, facades) and the detection of urban objects by labelling the set of 3D points is necessary. In our case, the interest lies in the extraction of the facade wall points, dominant facade planes as well as points of facade microstructures

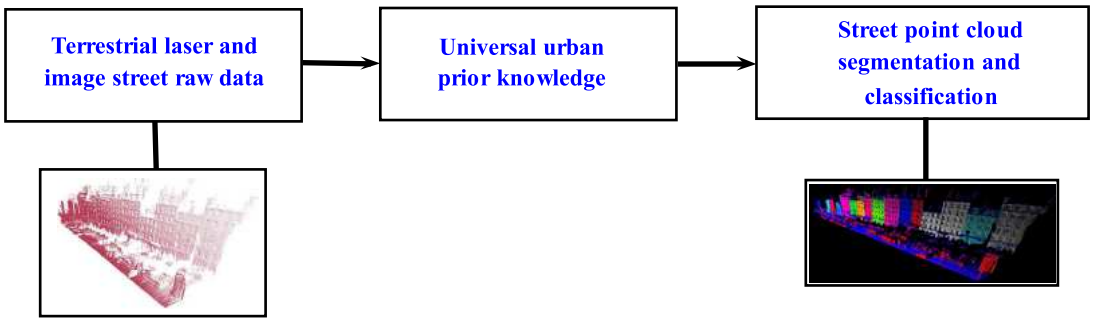


Figure 1.5: A model-less approach to the segmentation and classification in terrestrial urban street point cloud.

such as windows. The specificity of this approach lies in its use of a low level of a priori knowledge related to the scene, on the one hand, since relatively universal geometric properties on building topology have been used and on the other hand, since any advanced and pre-defined 3D street models has not been used. In addition, this approach takes advantage of speediness in runtime own to the used algorithm and accuraccy own to the laser data. The cadastral map is minimally used in order to individualize coplanar dwelling facades. It is shown that the presented approach reaches a satisfying compromise between efficiency and generalization. The global diagram of the proposed approach is illustrated in Figure 1.5.

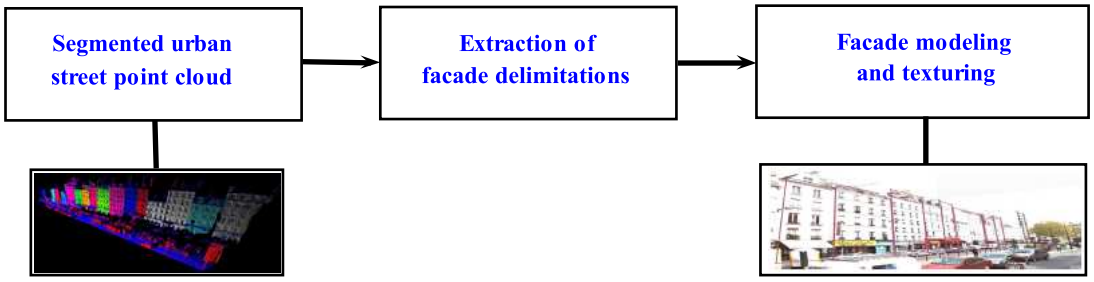


Figure 1.6: An approach to the modeling and occlusion-free texturing of urban facades from street point clouds and optical images.

Chapter 5 Part III is focused to the modeling and occlusion-free texturing of urban facades from street point clouds and optical images. Indeed, the interest of this chapter is twofold: *i*) a study on altimetric variability of the facade clusters computed in Chapter 4 lies to classify the facade models into quadrilateral or detailed polygons and thus to parametrize the wire-frame of street facades according to the facade LODs (Level Of Details). Then, the dominant facade plane (map-based or laser-based) can be horizontally and vertically delimited. In cases of rectangular facade, several strategies based on geometric heuristics are amongst others proposed. In case of detailed facades, a key idea consists of 3D facade top tracing by combining the estimated facade plane and the image-based skyline detection (limit between the building and the sky). Left and right extremities of each facade are delimited by using the cadastral map. *ii*) the points of occluding objects are detected and employed to generate images with masks that hides the regions reached. For each estimated facade model (wire-frame model), a set of facade images that optimally visualizes facade portions is attached. These images rectified according to the associated estimated

facade plane are rawly mosaicked in order to produce occlusion-free textures in case of high occlusions. The proposed approach employs a set of data that is composed of laser points acquired by a sensor specifically oriented to cover the facade top (facade bottom missing) and a set optical images. The image and laser data acquired by the MMS have then be merged in this Chapter. The global diagram of the proposed approach is illustrated in Figure 1.6.

Then, the approaches presented in the Chapters 4 and 5 of Part III can be considered together as constituting essential modules (i.e., components) of a terrestrial prototype pipeline for 3D urban street facades modeling. Besides, we stress the fact that the choice of the topics investigated in the Part III (terrestrial facade modeling) have been regulated by the arrival of the acquired MMS datasets that have naturally changed from one period to the next during the thesis.

Finally, the Chapter 6 Part IV concludes this study and discusses investigations that could be led in the future.

Transition

Undoubtedly, the aerial data are of great interest for the coverage of building at wide scale. More particularly, active researches are led in photogrammetry and computer vision communities in order to automatically generate building models from aerial images. In spite of great scientific advances in these fields, it appears that the research on 3D building modeling from aerial data requires to be pursue since the actual approaches for 3D massive building reconstruction (although efficient) still encounter problems in term of generalization, coherency, accuracy. Recently, a strong effort of exchanges between the communities bring research tracks that can be exploited towards solutions.

Part II

Aerial Building Modeling

Chapter 2

Extracting 3D Polyhedral Building Models from Aerial Images

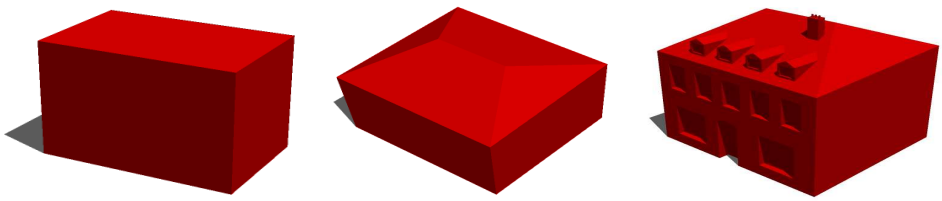
Abstract

In this Chapter, a featureless approach has been presented in order to generate 3D polyhedral building model from aerial images (Dataset $\mathcal{D}_1$). Indeed, we have observed across studies on 3D building modeling that the massive pipeline of 3D building modeling requires the intervention of operators and local modeling methods to rectify ill-reconstructed building model. The proposed approach employs high resolution aerial images in order to reconstruct 3D polyhedral building model. This approach has the originality to be featureless contrarily to some modeling strategies in the sense that the conventional stages of feature extraction and assembly are avoided. Experiments show the potential of such approaches towards the 3D building modeling.

Contents

2.1 Introduction and Motivation

In the two past decades, the cartographic field has evolved significantly; mainly in order to provide a digital and 3D geometric description of urban environments in addition to 2D conventional paper urban maps. More precisely, some active work in the photogrammetric, remote sensing and computer vision communities is focused on the 3D building modeling approaches since the buildings constitute urban objects of great interest for the 3D city modeling. The 3D building modeling approaches are more and more developed due to the increasing needs of institutional and industrial applications in the civil and military contexts. The visualization of urban environments (e.g., virtual tourism), the urban planning, the site recognition (military applications) or the conservation of architectural work (cultural heritage) are some of the many applications requiring 3D building modeling approaches. For these reasons, several approaches are proposed across the literature and provide more or less accurate, detailed and adapted 3D building models according to the targeted applications. Globally, the proposed approaches tend to produce 3D building models with a quality closer to the physical reality. The prior knowledge of the urban areas under study (e.g., cities topology, environment densities, shape complexity, existing surveys, urban GIS databases) and the remotely sensed rawdata collected are very rich sources of information that can be used to develop sophisticated building modeling approaches. The 3D building reconstruction is a complex task due to the diversity of building shapes (e.g., architectural and contemporary buildings). The building facades usually have some microstructures (e.g., windows, doors) and the building roofs present some superstructures (e.g., chimneys, attic windows). The representations of 3D building models can thus be divided into three main categories (see Figure 2.1).



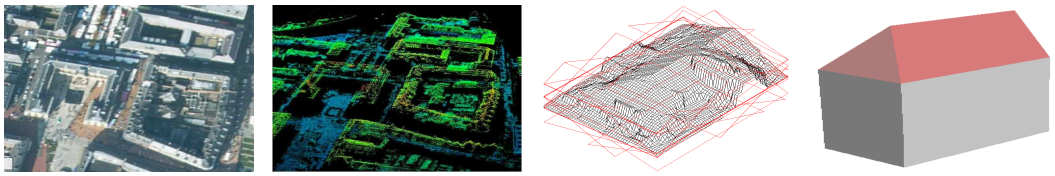
(a) Prismatic model. (b) Simple polyhedral model. (c) Complex polyhedral model.

Figure 2.1: Examples of generic model representations. Three illustrations of the same building with different level of details (from low to high).

The complexity of 3D building models can be planimetric (complex polygonal ground footprint) as well as altimetric (e.g. heights variation). Aerial data are very useful for the coverage of large areas such as cities. In the literature, several aerial or satellite data-based approaches are proposed to extract 3D prismatic and polyhedral building models. The data usually employed as input to these approaches are either optical aerial or satellite images, aerial or satellite Digital Surface Model (DSM) or aerial 3D point clouds such as aerial LIDAR data (Light Detection And Ranging data). Some data samples usually employed are shown in Figure 2.2.

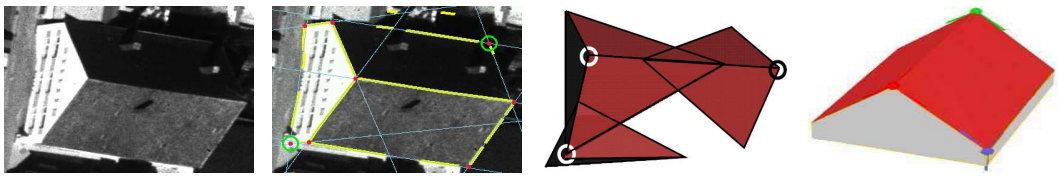
Figure 2.2 (Top) illustrates the building modeling using Digital Surface Models. Figure 2.2 (Middle) illustrates the building modeling using recon-

3D building modeling using DSMs



(a) High resolution aerial image (overview). (b) Generated 3D point cloud (e.g., DSM). (c) Extracted 3D roof planes from DSM. (d) 3D building model (simulated example).

3D building modeling using extracted 2D and 3D features



(e) Targeted building (aerial image). (f) Extracted 2D corners. (g) Extracted 3D roof corners. (h) 3D building model.

3D building modeling using our proposed featureless approach

(i) Targeted building (aerial image). (j) 3D building model.

Figure 2.2: The upper part of this Figure illustrates an example of 3D building modeling process using a DSM. The middle part of this Figure shows image-based feature extraction and assembly. The lower part shows our proposed direct and featureless image-based approach. Figures 2.2(b) and 2.2(c) are retrieved from a confidential document of Paparoditis and (Jibrini et al, 2000), respectively. Figures 2.2(e), 2.2(f), 2.2(g), and 2.2(h) are retrieved from (Fischer et al, 1998).

structed geometrical features (e.g., 2D vertices and lines). Figure 2.2 (Bottom) illustrates our proposed featureless approach.

The flowchart of the two first strategies (image-based building modeling) is illustrated in Figure 2.3. In the first strategy, a DSM is generated or directly employed as input (e.g., (Lafarge et al, 2010). The very dense aerial DSM can be comparable as an aerial 3D point cloud such as shown in Figure 2.2(b). The succeeding stages consist of the use of the DSM as reference for the extraction of high level geometric features (e.g., 3D segments or 3D planes). The extracted features are finally assembled into a polyhedral building model using optimization methods. However, these successive estimation stages inevitably introduce some inaccuracies that propagate from one stage to the next, which

can affect the final 3D model. If these inaccuracies are large enough, then, one can note, that the obtained shape can be erroneous (e.g., see Figures 2.4(a) and 2.4(b)). In the second strategy, geometrical features are extracted from aerial images (e.g., 2D segments, junctions, corners, lines) and then converted into 3D features. The final polyhedral model is then estimated using these 3D features (e.g., Figure 2.2(g)). As in the first strategy, the extraction and matching stages inevitably affect the accuracy of the final 3D model. (Taillandier et al, 2004) and (Fischer et al, 1998) are well-know references in the literature which respectively illustrate the two strategies described above.

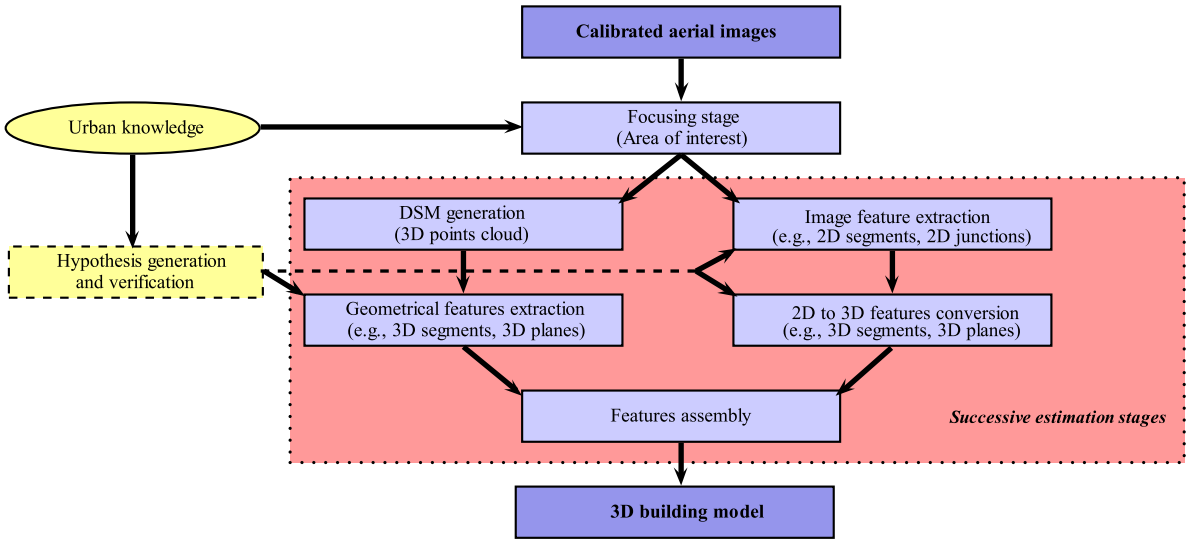
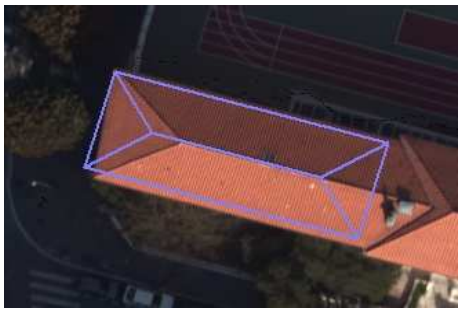
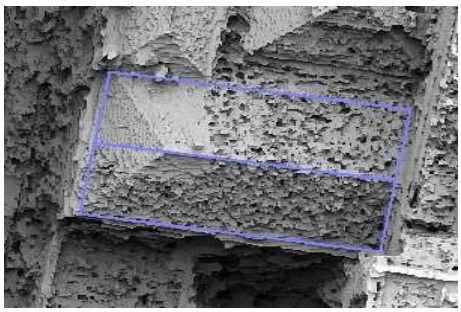


Figure 2.3: Flowchart diagram currently adopted by some image-based building modeling approaches. The diagram presents two paths conducting to 3D polyhedral building models. These two paths are illustrated by the first two rows of Figure 2.2.



(a) Four modeled facets versus three for the real building.



(b) Two modeled facets versus three for the real building.

Figure 2.4: Some erroneous reconstructed buildings resulting from a known feature-based framework for massive building reconstruction (BATI-3D® prototype software—a large scale building modeling pipeline developed at the French National Geographical Agency). The estimated 3D models are projected onto the image or DSM.

The 3D building reconstruction of a full urban environment requires automatic or semi-automatic methods. The massive reconstruction approaches usually employ a feature extraction stage. However, this stage is very sensi-

tive since it can induce some missed-detections, false alarms, false negatives or false positives. To control these effects, the 3D building modeling approaches employ computer vision strategies. These strategies are regrouped into two paradigms. More precisely, the first paradigm is a bottoms-up scheme and consists in the assembly of geometric features with a low-level of prior knowledge of the sought model. The second paradigm, called top-down, exploits a library of models and searches the model that best fits with the input data (images, DSMs).

As previously mentioned, several approaches for 3D reconstruction of polyhedral building models currently employ as input Digital Surface Models (see Figure 2.2(b)). The classical DSMs are usually generated from calibrated aerial images by a multi-correlation based optimization process such as the graph cut optimization. The DSMs (derived data) are generally maps comprising only one value of altitude z for each ground location (x, y). These 2.5D maps can be considered as 3D point clouds. However, the obtained 3D surface does not accurately model the physical surface especially at height discontinuities such as at roof and superstructure boundaries due to the correlation criterion used. Hence, the DSMs provide an approximated geometrical description of building surfaces and can be noisy. Other modeling approaches employ multi-source data, for example optical images combined with LIDAR data (e.g., (Brenner, 2005)). Although less dense, LIDAR data can be employed in place of DSMs (e.g., (Verma et al, 2006, Dorninger et al, 2008, Sohn et al, 2008, Sampath et al, 2010, Elberink et al, 2009)).

2.1.1 Chapter Contribution

In this chapter, we propose a direct and featureless approach for the extraction of 3D simple polyhedral building models from aerial images (Figure 2.2(a)). The novelty of our approach consists in using a genetic optimizer which bypasses all the intermediary estimation/extraction stages previously mentioned. First results of our approach were presented at the MVA and ACIVS conferences, respectively in (Dornaika et al, 2009) and (Dornaika et al, 2010). This work presents a substantially extended version which describes in more detail the models as well as the core of the proposed methodology and processes.

We are interested in modeling residential buildings having simple polyhedral shapes and whose ground footprints are represented by quadrilaterals. We note that in most cases, these quadrilaterals are rectangles. However, this requirement is not a limitation to our approach. Indeed, any complex shape can be considered as a union of simple models with rectangular footprints.

The input data are calibrated aerial images. Hence, our research deals with the intermediary degree of generic modeling such as described in Figure 2.1(b). In our case, the proposed approach can be considered as a top-down scheme (model driven) in the sense that a library of parametric building models is employed. However, our top-down approach is not conventional in the sense that the 3D model estimation is direct and only uses image intensities. Moreover, the exhaustive search for the best model is avoided. The proposed approach employs aerial images as illustrated in Figure 2.2(a). The building footprint (focus area) is selected by an operator in one aerial image. The building footprint corresponds to the building gutter limits in images. The

building footprint could also be retrieved from a cadastral map (existing 2D map of building footprints) (Jibrini et al, 2000, Suveg et al, 2004, Durupt et al, 2006, Chen et al, 2008, Suveg et al, 2001). In this case, the 2D footprint is expressed in a georeferenced world coordinate system.

In this study, we are essentially focusing on the approaches producing polyhedral building models (as shown in Figure 2.1(b)) from a single source of data, namely the high resolution aerial images. For this reason, several image-based modeling approaches are detailed in the following Section 2.2 (related work).

The rest of the work is organized as follows. Section 2.2 describes various existing image-based approaches for 3D polyhedral building modeling. Section 2.3 presents the global strategy of the proposed approach. Section 2.4 describes the optimization process of the approach. Section 2.5 gives several intermediary results and evaluations of major steps.

2.2 Related Work

Many interesting building modeling approaches have been addressed in the literature for the reconstruction of 3D polyhedral building models (e.g., (Rau et al, 2002, Lafarge et al, 2006, Lafarge et al, 2008, Madhavan et al, 2006, Vestri, 2006, Hongjian et al, 2006, Habib et al, 2010, Haala et al, 2010, Wang et al., 2007, Poullis et al, 2009, Kada et al, 2009, Khoshelham, 2005)). The intention here is to briefly describe some aerial-based approaches that are discerned by their optimization process, their global methodology or their efficiency.

In (Jibrini et al, 2000), Jibrini *et al.* propose a 3D polyhedral building modeling approach from a very high resolution aerial stereo-pair using a cadastral map. A cadastral map is a 2D ground map (detailed register) showing the parcel delimitations of each building. This standard 2D map is often used by governments for the annual taxation of their residents according to the size of their homes. Their proposed method is generic in the sense that it can be used to estimate the polyhedral shape of buildings without pre-existing knowledge about the real shape. Firstly, the corresponding volume of interest is set as an extrusion of the 2D footprint into 3D. This volume is then discretized and a correlation score is calculated for each voxel using a stereoscopic principle and a block matching method. Hypothesis of 3D planes are then detected using the Hough Transform (HT) weighted by the correlation score of each voxel. Several arrangements associated with these 3D planes inside the delimited volume are calculated. The research of admissible shapes will be equivalent to the research of maximal clicks in a compatibility graph. The last step selects the best admissible model by optimizing a term related to the data (compatibility between the model and the images) and a term of regularization related to the model complexity.

Taillandier *et al.* (Taillandier et al, 2004) present another generic approach that can be considered as an extension of the approach described in (Jibrini et al, 2000). The reconstruction is directly achieved using a Digital Elevation Map generated from multi-view images. A building is modeled by a polyhedral shape, without overhangs. The building boundaries are modeled by vertical walls. This proposed method is generic and allows the modeling of almost all building categories. For each building, an operator manually selects a focus area as well as a ground altitude. 3D planar features (horizontal, vertical and oriented planes) and 3D segments are then automatically extracted in this

area. A 3D graph of arrangements is generated by the intersection of all the planes. After a graph simplification step, the search of admissible 3D models is proved to be similar to the search of maximal clicks. The model is finally selected using a Bayesian modeling method. In another work (Durupt et al, 2006), Durupt and Taillandier have proposed operational approaches useful to adapt the generic algorithm to more realistic data. These approaches are mainly focused on the calculation of the arrangement of 3D planes.

In (Lafarge et al, 2010), Lafarge *et al.* propose an approach for the 3D building reconstruction in dense urban environments using high resolution satellite images. The approach employs a DSM and a set of parametric models. A marked point process is employed to automatically extract rectangular building footprints from the DSM. The best model parameters with a rectangular footprint are searched using pre-existing knowledge of classical models and their interactions. The data term minimizes the error between the models and the DSM. The model parameters associated with a block of buildings are obtained by searching the maximum a posteriori. This maximum is obtained using a RJMCMC method (Recursive Jump Monte Carlo Markov Chain) and a SA method (Simulated Annealing).

Fisher *et al.* present in (Fischer et al, 1998) a model-based approach to the 3D building model extraction from aerial images. This proposed approach allows the reconstruction of various types of polyhedral buildings. The building parts are classified according to their roof types. The approach employs the extraction of low-level image features, the matching of these features according to building part models and the aggregation of the model into complete building models.

In (Jaynes et al, 2003), Jaynes *et al.* present a model-based approach to the automatic detection and reconstruction of buildings using aerial imagery. Optical aerial images are first segmented in order to detect the buildings. The corresponding DEM is employed to reconstruct the buildings. Each segmented DEM region is associated with a class of building roof shape either peaked, flat or curved. The segmented regions are extruded and fitted to the DEM by an optimization process. The segmented DEM region allows the decomposition of the building area into sub-area according to its roof shape. The final building model is obtained by the union of roof part models independently estimated. This strategy allows the reconstruction of a wide variety of polyhedral building models.

Zebedin *et al.* propose in (Zebedin et al, 2008) an approach for the automatic building reconstruction from aerial images. An approach is proposed to meet the need for realistic and accurate building models for virtual applications. Line features that characterize the height discontinuities are detected and combined with dense depth data providing the roof surface by using a global optimization process based on Graph Cuts technique. The proposed algorithm generates elegant building models.

In (Tseng et al, 2002), Tseng *et al.* propose a promising 3D building reconstruction approach that uses Genetic Algorithms (GA) for model-image fitting. The buildings are reconstructed piece by piece and each CSG feature (Constructive Solid Geometry) is fitted according to the edge pixels of aerial images. CSG boolean set operators are employed in order to combine building parts into a single building. The theory of the GA method for model image fitting has been analyzed and demonstrated in several examples.

Paper	Process	Input data	Strategy
(Jibrini et al, 2000)	Automatic	Urban map/Aerial Images	Bottom-up
(Taillandier et al, 2004)	Automatic	Aerial Images	Bottom-up
(Fischer et al, 1998)	Automatic	Aerial Images	Hybrid
(Lafarge et al, 2010)	Automatic	DSM	Top-down
(Jaynes et al, 2003)	Automatic	DEM/Aerial Images	Hybrid
(Zebedin et al, 2008)	Automatic	Aerial Images	Bottom-up
(Tseng et al, 2002)	Interactive	Aerial Images	Top-down

Table 2.1: Some feature-based approaches developed for 3D polyhedral building modeling from aerial images.

Table 2.1 briefly presents some feature-based building modeling approaches available in the literature. The regrouped approaches demonstrate the high diversity of employed techniques in 3D building modeling from aerial images. The presented approaches propose building modeling advances at various levels of generalization, geometry, accuracy, and realism. The priority characteristic is guided by the targeted application. In our case, the main goal is (i) to improve the accuracy of 3D polyhedral building models using images, and (ii) to rectify the erroneous estimated shape of building model issues from certain feature-based approaches (as shown in Figure 2.4).

2.3 Problem Statement and Model Parametrization

In this section, we present our formulation of the problem and the adopted parametrization. In the previous section, we described several approaches that have been addressed in the literature. Here we state the characteristics of our approach.

Since aerial images are employed, the proposed approach only deals with roof models due to the angle of view. Indeed, an aerial image allows the visualization of two facades at best, since the building generally has a rectangular footprint. Nevertheless, the building facades can actually be determined using the prior knowledge of the ground-height of the area under study (from urban database) and by the assumption that the dominant facade planes are vertical. In this work, we restrict our study to simple polyhedral models (several roof varieties). Some are illustrated in Figure 2.5. The shown models present either horizontal and/or vertical symmetry assumptions and the inner and outer vertices respectively have the same height. These parametric building models with roofs having two, three, or four facets can also be described by a more generic building model (see Figure 2.6). In this model, any simple polyhedral model can be obtained by varying the 3D location of the inner vertices (*i.e.*, a deformable model) and by setting the height of all external vertices. Furthermore, the multi-facet model (Figure 2.6) and the one facet model (Figure 2.5(b)) can describe all typical situations: asymmetric shapes, sloping roofs or ground (*i.e.*, every vertex can have a different height). Hence, the proposed generic model describes more various building models than the model set shown in Figure 2.5. We mention that this library represents an extension (models, naming) of the library shown in (Lafarge et al, 2006).

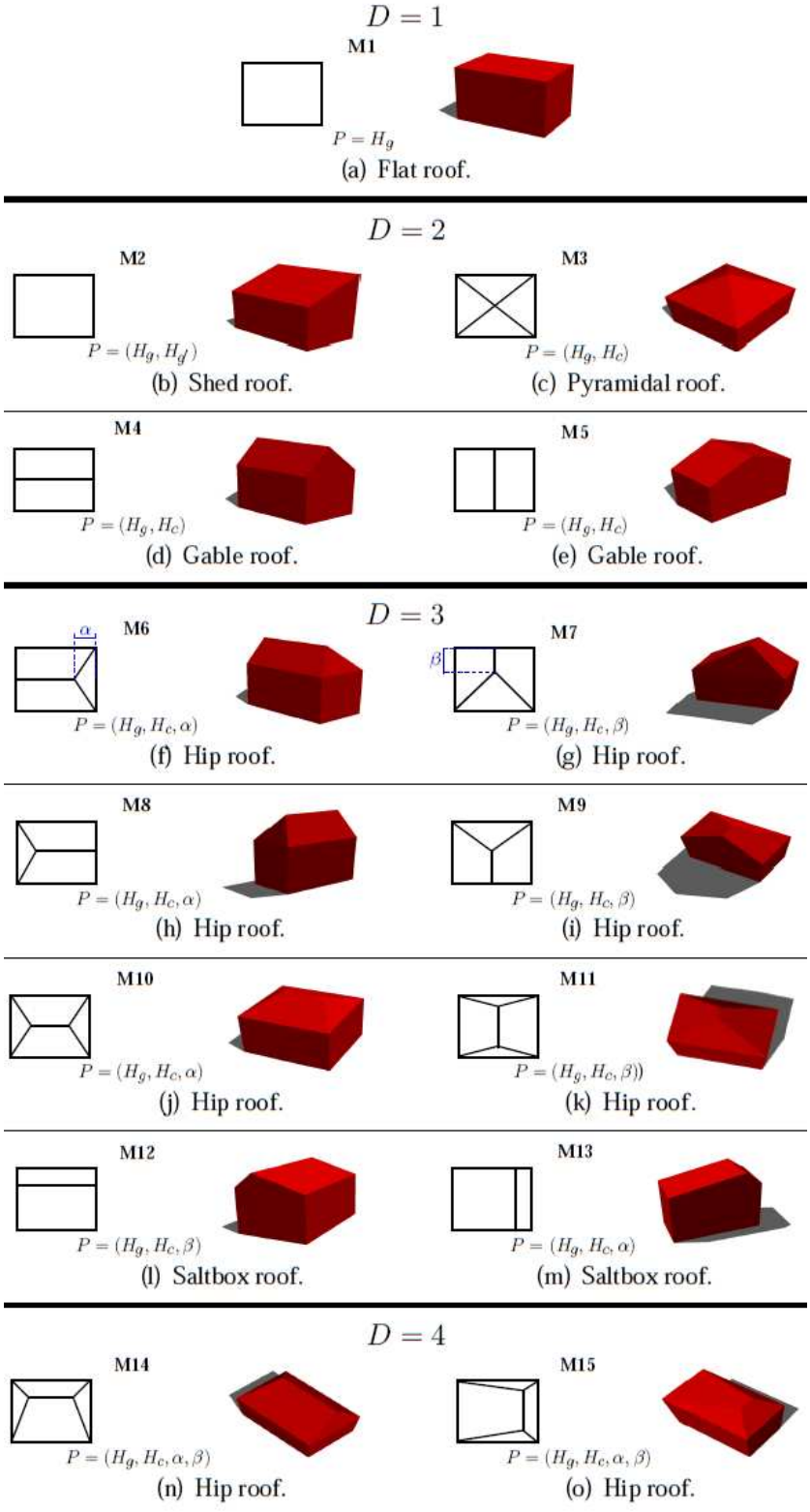


Figure 2.5: Samples of parametric building models M that can be reconstructed by our proposed featureless approach. Standard polyhedral shapes and their corresponding ground footprints are shown. D corresponds to the number of parameters. P denotes the model parameters. H_g and H_c correspond to the gutter height and the central line height. α and β respectively represent the horizontal and vertical recess as illustrated in blue.

Since a complex building can be described as an aggregation of simple polyhedral building models, our approach can also deal with complex buildings once a partitioning of the building into simple building-parts is done.

In this case, vertices are estimated for each simple model. The vertices having adjacent models are replaced by the barycenter of these points in order to reconstruct the final model.

The adopted multi-facet roof model comprises six vertices A, B, C, D, M, N (see Figure 2.6). In theory, the estimation of the roof model is equivalent to the estimation of the three-dimensional coordinates (X,Y,Z) of each vertex. As previously mentioned, the rectangular building footprint is manually selected in one image by an operator (interactive method). This footprint is considered as the footprint of reference for the succeeding processes. Moreover, the calibration of the aerial images is known (intrinsic and extrinsic parameters of the cameras).

Consequently, the perspective 3D lines of sight passing by the vertices of the building footprint in the image are known. The 3D vertices that we seek to determine (A, B, C, D) are 3D points located along these perspective lines with unknown heights. In other words, by varying the height value, the corresponding 3D point slides along the perspective line. In this condition, the outer vertices we are searching for each have one degree of freedom. Hence, our polyhedral model can be simplified to ten parameters instead of eighteen: four parameters for the heights of the outer vertices and six parameters for the three-dimensional coordinates of the inner vertices. These ten parameters are encapsulated into one single vector $\mathbf{w}$:

$$\mathbf{w} = (X_M, Y_M, Z_M, X_N, Y_N, Z_N, Z_A, Z_B, Z_C, Z_D)^T \tag{2.1}$$

Moreover, since the images are calibrated the 3D coordinates of the inner vertices M and N can be replaced by the triplets (U_M, V_M, Z_M) and (U_N, V_N, Z_N), respectively. (U,V) represent the image coordinates in the reference image.

Furthermore, it is easy to show that our polyhedral model can be fully described by the 3D coordinates of the inner vertices and of two outer vertices that are diagonally opposite (coplanarity constraint). Indeed, the building can be parameterized by eight parameters: four parameters for the image location of the inner vertices M and N but also four parameters for the height of the vertices A, M, N, and C. The remaining vertices are determined by intersecting the corresponding lines of sight with the estimated support planes. Indeed, we assume that B belongs to the estimated plane (ANM) and D belongs to the estimated plane (CMN) since the roof shape is supposed to be composed of planar facets. For these reasons, Equation (2.1) can be simplified to:

$$\mathbf{w} = (U_M, V_M, U_N, V_N, Z_A, Z_M, Z_N, Z_C)^T \tag{2.2}$$

where (U_M,V_M) and (U_N,V_N) are the image coordinates of the vertices M and N, respectively.

In other words, our method has the obvious advantage that the coplanarity constraints are implicitly enforced in the model parametrization. By contrast, the feature-based approach requires fitting the planes to DSM or 3D points.

Recall that the $3D$ coordinates are expressed in a local coordinate system

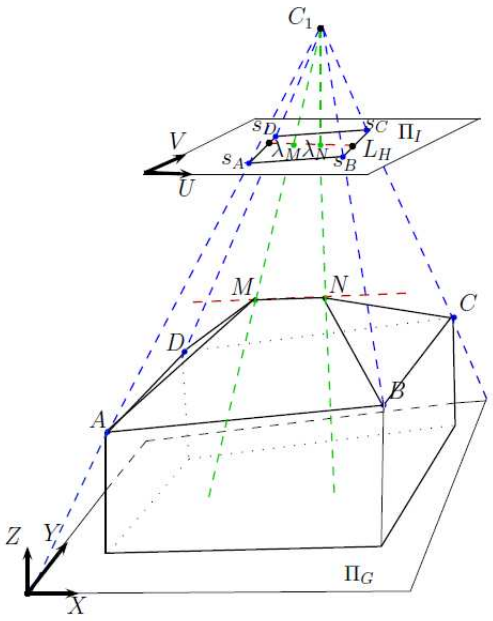


Figure 2.6: The adopted generic 3D polyhedral model. The multi-facet model (*i.e.*, deformable model) is parameterized by Equation (2.3). s_A, s_B, s_C, s_D correspond to the footprint vertices selected in the image plane Π_I (master image). λ_M and λ_N correspond to the linear coordinates of the inner vertices M and N (unknown) along the detected Hough line L_H. C_1 corresponds to the center of projection of camera 1. Blue and green lines are outer and inner lines of sight (perspective lines), respectively. Π_G represents the ground plane in a geo-referenced world coordinate system.

whose Z-axis coincides with the ground normal (the aerial images are geo-referenced). In practice, although the location of inner vertices is not known, the $2D$ line (the projection of a ridge segment) going through them can be easily extracted from the image by using a conventional edge detector (e.g., Canny edge detector) followed by a Hough transform. Once the equation of this line is known, the parametrization of the building model (Equation 2.2) can be further simplified to:

$$\mathbf{w} = (\lambda_M, \lambda_N, Z_A, Z_M, Z_N, Z_C)^T \tag{2.3}$$

where λ_M and λ_N parameterize the location of the inner vertices along the 2D segment obtained by intersecting the 2D line with the building footprint.

Thus, finding the model boils down to finding this vector $\mathbf{w}$. Henceforth, we have defined the parametrization of the adopted generic building model. The succeeding section aims at describing a global methodology in order to determine the kind of model (one-facet or multi-facet) that corresponds to the reality as well as to compute the corresponding numeric parameters. To this effect, some computer vision mechanisms and strategies are described for 3D building shape recovery.

2.4 Proposed Approach

In this study, we present a novel modeling approach which is direct and image-based. The challenge consists in the reconstruction of 3D polyhedral building shapes using directly photometric information of aerial images. In computer vision, direct approaches have been essentially proposed for the image registration in order to generate mosaic images. Featureless image registration techniques strive to compute the global motion of the brightness pattern (e.g., affine or homographic transforms) without using matched features (e.g., (Romero et al, 2007)). We were inspired by this kind of approach and we propose a direct method for 3D building reconstruction. The flowchart diagram of the proposed approach is depicted in Figure 2.7.

2.4.1 Multiscopic Context and 3D to 2D Projection

As previously mentioned, our approach employs calibrated aerial images. The building under study is observed by n different points of view ($n \geq 2$), in other words, in a multiscopic context. The visible area common to all the associated images is called overlapping area or overlapping volume. This area potentially characterizes the reconstructible area into 3D. We mention that geometric principles for 3D scene reconstruction from multiple views are described in detail in (Hartley et al., 2004). Besides, if the camera's calibration is known and if the images are properly georeferenced, as assumed in our case, then an hypothetic physical 3D point $M(X,Y,Z)$ (expressed in the world referential) that belongs to the overlapping volume can be projected in each image acquired by the camera $\mathcal{C}_i$ onto a corresponding image point $p_i(u_i,v_i)$ where $1 \leq i \leq n$. These image points called homologous points can be calculated using a 3×4 projective camera matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{K} \cdot [\mathbf{R}|\mathbf{T}] \tag{2.4}$$

where $\mathbf{K}$ corresponds to the matrix of intrinsic parameters related to the camera $\mathcal{C}_i$, $\mathbf{R}$ and $\mathbf{T}$ correspond to the extrinsic parameters which denote the coordinate system transformations from 3D world coordinates to 3D camera coordinates.

2.4.2 Measuring Model-to-Data Consistency

In computer vision, the homography principle is employed in image registration, auto-calibration of cameras, motion estimation and also for stereoscopy and 3D scene reconstruction. Mathematically, the homography is a projective collineation that describes an image-to-image transformation that can be used either in the case of a pure 3D camera rotation, or a planar scene (see Figure 2.8). The homography matrix can be estimated by different techniques. An overview of these techniques are described in (Agarwal et al, 2005). We are particularly interested by the homography principle since it can be used to transfer a roof facet of an image to another image if the 3D support plane of the facet is known.

The homography matrix is the transfer matrix that allows the transfer of the point $p_1(x_1,y_1)$ of the reference image (image 1) to its homologous point $p_2(x_2,y_2)$. The equation that links each pair of homologous points can be defined as:

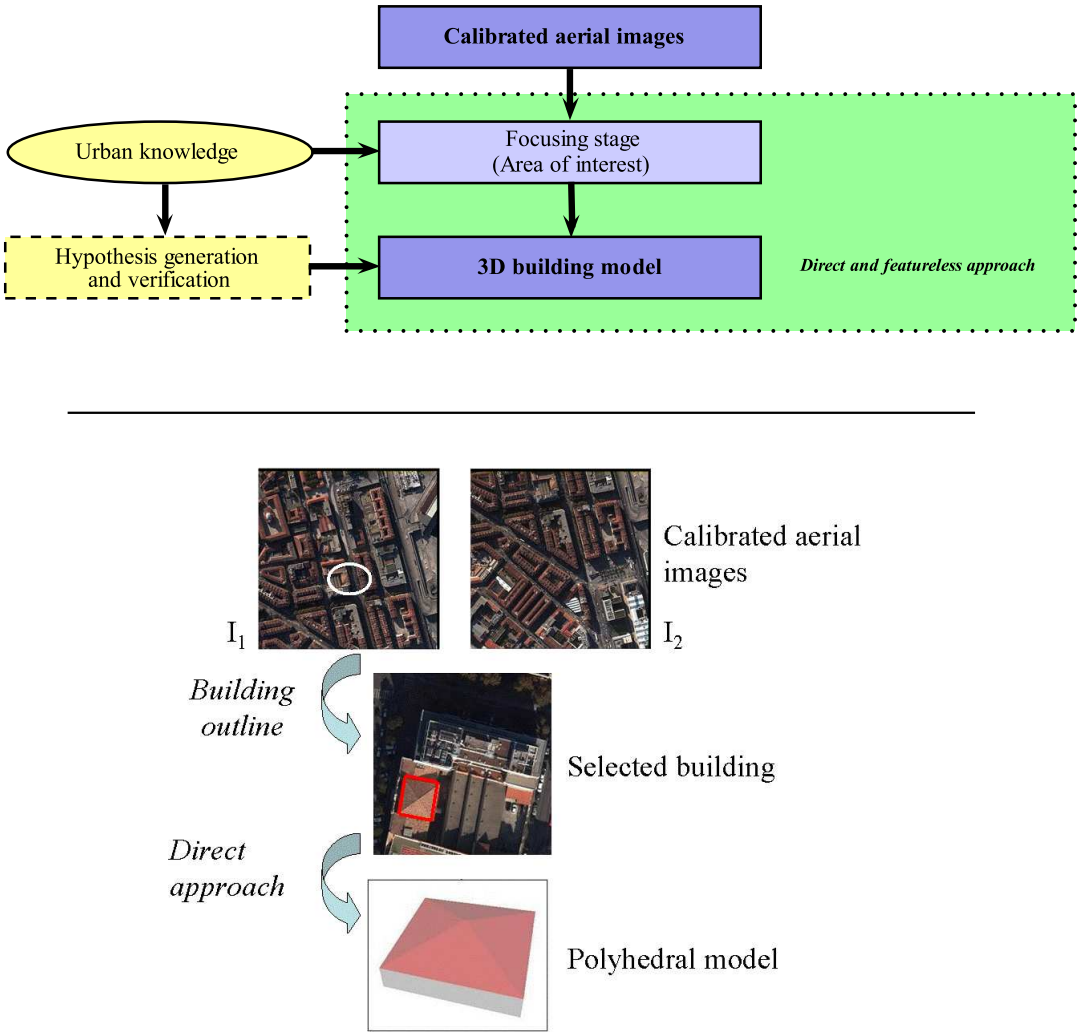


Figure 2.7: Flowchart diagram of the proposed approach (top) and illustrations of the main steps (bottom).

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \cong \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \Leftrightarrow \mathbf{X}_2 \cong \mathbf{H}\mathbf{X}_1 \tag{2.5}$$

where the last row is fixed to $H_{31} = H_{32} = 0$ and $H_{33} = 1$; and where $\cong$ denotes the equality to a given scale factor. Equation (2.5) provides:

$$\begin{cases} x_2 = \dfrac{H_{11}x_1 + H_{12}y_1 + H_{13}}{H_{31}x_1 + H_{32}y_1 + H_{33}} \\ y_2 = \dfrac{H_{21}x_1 + H_{22}y_1 + H_{23}}{H_{31}x_1 + H_{32}y_1 + H_{33}} \end{cases} \tag{2.6}$$

The matrix $\mathbf{H}$ has eight degrees of freedom. For this reason, it is possible to judiciously select four points in another image in order to solve the system. We can note that this technique is generally employed with key points detectors. The coefficients of the matrix $\mathbf{H}$ depend on intrinsic and extrinsic parameters of the cameras as well as on the parameters of the plane:

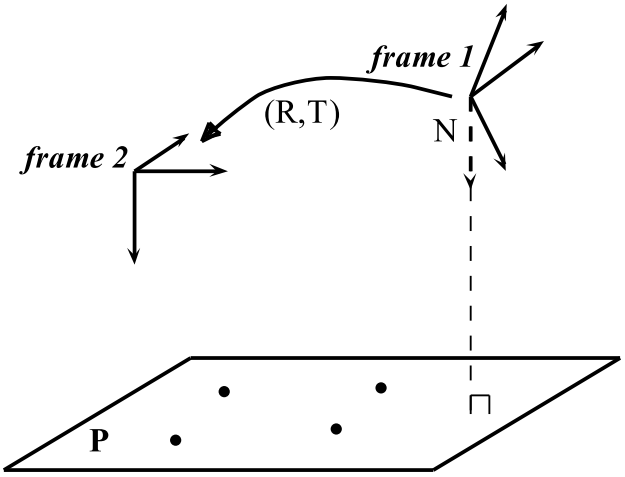


Figure 2.8: Homography induced by a plane.

$$\mathbf{H} \cong \mathbf{K}_2 \cdot [\mathbf{R} + \frac{\mathbf{T}}{d}\mathbf{N}^t] \cdot \mathbf{K}_1^{-1} \tag{2.7}$$

where the matrix $\mathbf{K}_1$ and $\mathbf{K}_2$ respectively are the intrinsic matrix of the two cameras, $\mathbf{R}$ represents the rotation, $\mathbf{T}$ represents the translation vector ($\mathbf{R}$ and $\mathbf{T}$ represent the motion between the two cameras), $\mathbf{N}$ and d represent the parameters of the plane in the camera 1.

In our case, the intrinsic and extrinsic parameters of the cameras are known (calibrated cameras). The parameters of the plane need to be determined for each facet that compose the model. If the planes' parameters are known, the homography matrix will directly transfer, facet by facet, sets of master pixels to their homologous pixels.

2.4.2.1 Measuring Facets-to-Data Consistency

In this subsection, a measure has been defined in order to value the accuracy of hypothetical facets according to the data. As we recall, in the multi-facet case, the facets are rigidly joined as shown in Figure 2.6. Firstly, we stress the importance of having a rigorous matching between the homologous points. The homologous points are 2D pixels representing the same 3D physical point in the scene. For this reason, the pixel intensities of homologous points have very close numeric values.

More precisely, our basic idea relies on the following fact: if the shape and the geometric parameters of the building (encoded by the vector $\mathbf{w}$) correspond to the real building shape and geometry, then the pixel-to-pixel mapping (induced by homographies) between the master image I_m (the one containing the selected 2D footprint) and any other aerial image (in which the building is visible) will be correct for the entire building footprint. In other words, the dissimilarity associated with the two sets of pixels should correspond to a minimum.

Recall that $\mathbf{w}$ is defining all support planes of all the building's facets and thus the corresponding pixel p' of any pixel p is estimated by a simple image transfer through homographies (3×3 matrices) based on these planes. Therefore, the associated global dissimilarity measure reaches a minimum. For an

arbitrary model instance $\mathbf{w}$, the global dissimilarity is given by the following score:

$$e = \sum_{j=1}^{n-1} \sum_{p \in S} \rho(|I_m(p) - I_j(p')|) \tag{2.8}$$

where n is the number of aerial images in which the whole building roof is visible (in practice, n is between 2 and 5), S is the footprint of the building in the master image I_m, p' is the pixel in the image $I_j \neq I_m$ that corresponds to the pixel $p \in I_m$, and $\rho(x)$ is a robust error function.

The choice of the error function $\rho(x)$ will determine the nature of the global error (2.8) which can be the Sum of Squared Differences (SSD) ($\rho(x) = \frac{1}{2}\, x^2$), the Sum of Absolute Differences (SAD) ($\rho(x) = x$), or the saturated Sum of Absolute Differences. In general, the function $\rho(x)$ could be any M-estimator (Chen et al., 2003). In our experiments, we used the SAD score since it is relatively robust and its computation is fast.

We seek the polyhedral model $\mathbf{w}^\star = (\lambda_M^\star, \lambda_N^\star, Z_A^\star, Z_M^\star, Z_N^\star, Z_C^\star)^T$ that minimizes the above dissimilarity measure over the building footprint:

$$\mathbf{w}^\star \;=\; arg\min_{\mathbf{w}} e \tag{2.9}$$

We can also measure the fitness of the 3D model by measuring the gradient norms along the projected 3D segments of the generated 3D models. In general, at facet discontinuities the image gradient is high. Thus, for a good fit, the projection of the 3D segments will coincide with pixels having a high gradient norm in all images. Therefore, we want to maximize the sum of gradient norms along these segments over all images. Recall that we have at most nine segments for our simple 3D polyhedral model. Thus, the gradient score is given by:

$$g = \frac{1}{n} \sum_{j}^{n} g_j \tag{2.10}$$

where g_j is the gradient score for image I_j. It is given by the average of the gradient norm over all pixels coinciding with the projected 3D model segments.

Since we want the dissimilarity measure (2.8) and the gradient score (2.10) to help us determine the best 3D polyhedral model, we must combine them in some way. One obvious way is to minimize the ratio:

$$\mathbf{w}^\star \;=\; arg\min_{\mathbf{w}} \frac{e}{g} \tag{2.11}$$

It is worth noting that during the optimization of (2.11) there is neither feature extraction nor matching among the images. Furthermore, the use of the image gradient norms in (2.11) is not equivalent to a feature-based method.

The image-to-image transfer can be carried out pixel-to-pixel by combining 3D point construction (line of sight intersected with plane) and 3D-to-2D projection; or more directly facet-to-facet by using homographic transfer (Equation 2.5).

In order to minimize (2.11) over $\mathbf{w}$, we will use an evolutionary optimizer

that will be described in Subsection 2.4.3.

2.4.3 Computing the Polyhedral Building Model

In this subsection, we briefly describe the mechanisms and the goals of optimization processes. Moreover, we select an optimizer adapted to the considered modeling problem.

2.4.3.1 Computing the Prismatic Building Model

In our case, our approach begins by approximating any building model by one horizontal facet, *i.e.*, adopting a prismatic model. An urban database is employed in order to know the minimum and maximum ground altitude Z_{ground_min}, Z_{ground_max} of the area under study as well as the minimum and maximum heights amongst all the included buildings H_{min}, H_{max} (e.g. $H_{min} = 5m$ and $H_{max} = 50m$). The ground altitudes are provided according to the sea altitude. Consequently, we know with certainty that the building altitude z is in the interval $\mathcal{I} = [Z_{ground\_min} + H_{min}; Z_{ground\_max} + H_{max}]$. If we sweep this interval for z values with a step $\triangle z$, then we obtain a set of candidates from which the best prismatic model is selected—the one that minimizes the objective function (2.11).

2.4.3.2 Computing the 3D Model using the Differential Evolution Algorithm

The Differential Evolution algorithm belongs to the family of Genetic Algorithms and to the evolutionary strategies. The genetic algorithm modifies the structure of individuals using the mutation and the crossover. The evolutionary strategies achieve the auto-adaptation by geometric manipulation of individuals. These ideas have been formulated by a simple and powerful operation of vectors mutation proposed in 1997 by Price and Storn ((Storn et al., 1997)). Since then the Differential Evolution has become an essential method for a large quantity of real problems and benchmarks. The algorithm of the Differential Evolution can be described such as shown in Alg.(1).

The DE algorithm is employed in order to compute the 3D model and integrates the minimization process guided by the dissimilarity measure previously defined. This algorithm achieves generations of solutions—populations. The population of the first generation is randomly chosen around a rough solution. The rough solution will thus define a given distribution for the model parameters. The rough solution is simply given by a zero-order approximation model (the prismatic model) which is also obtained by minimizing the dissimilarity score over one unknown (the average height of the roof).

In our case, the use of the DE algorithm is described in Figure 2.9 which illustrates the main steps performed in one single iteration for one facet. (1) The prismatic model is estimated, (2) a population (set of facets) is generated from the prismatic model, (3) the best model is determined, (4) a new solution is generated by using crossover, mutation and evaluation steps.

We use the Differential Evolution optimizer since it has four interesting properties: (i) it does not need an accurate initialization, (ii) it can integrate geometric constraints according to the context (adaptability properties), for example, constraints can be imposed in order to ensure that the polyhedral

Algorithm 1 The *Differential Evolution.*

1: **Parameters:**
N=The size of the population
D=dimension=number of unknown parameters
F=constant of tuning, 0<F<1.2
CR=constant of crossing, 0<=CR<=1
$x[i][j]$=table of dimension $D \times N$ that includes N individuals (vectors)
where X_i include the parameters
$trial[j]$=T=trial vector
$best[j]$=**B**=best vector
$f()$=objective function
a, b, c, d, i, j, k=indexes of type integer

2: **Initialization:**
Randomly generation of all the individuals (vectors) X_i
Evaluation and storage of the score for each vector X_i
Determination of the best vector (sample) amongst the set of X_i and storage in **B**

3: **Main loop:**
for (i=0 ; i<N ; i++)
//Randomly selecting 4 individuals (vectors).
a ←rnd1()*N;
b ←rnd1()*N;
c ←rnd1()*N;
d ←rnd1()*N;
j ←rnd1()*D;
//Generating a novel individual by mutation and crossing.
for (k=1; k<=D; k++) {
if(rnd1()<CR || k==D) {
D_{abcd}←$x[a][j]$-$x[b][j]$+$x[c][j]$-$x[d][j]$;
$trial[j]$←$best[j]$+F*D_{abcd};
}
else {
$trial[j]$←$x[i][j]$;
} j←(j+1)%D;
}
//Selecting the descendent population.
Evaluate $f(T)$
if ($f(T)$ <= $f(X_i)$) **then** T **replace** X_i
if ($f(T)$ <= $f(B)$) **then** T **also replace** B
}
End of the main loop

roof model are an assembly of facets with slopes inferior to 60° (standard information coming from urban databases concerning the area under study), (iii) it does not need the computation of partial derivatives of the cost function, and (iv) theoretically it can provide the global optimum. Hence, this algorithm is easy to implement and to integrate into the applications. In our case, the experiments show that only a few iterations lead to convincing results.

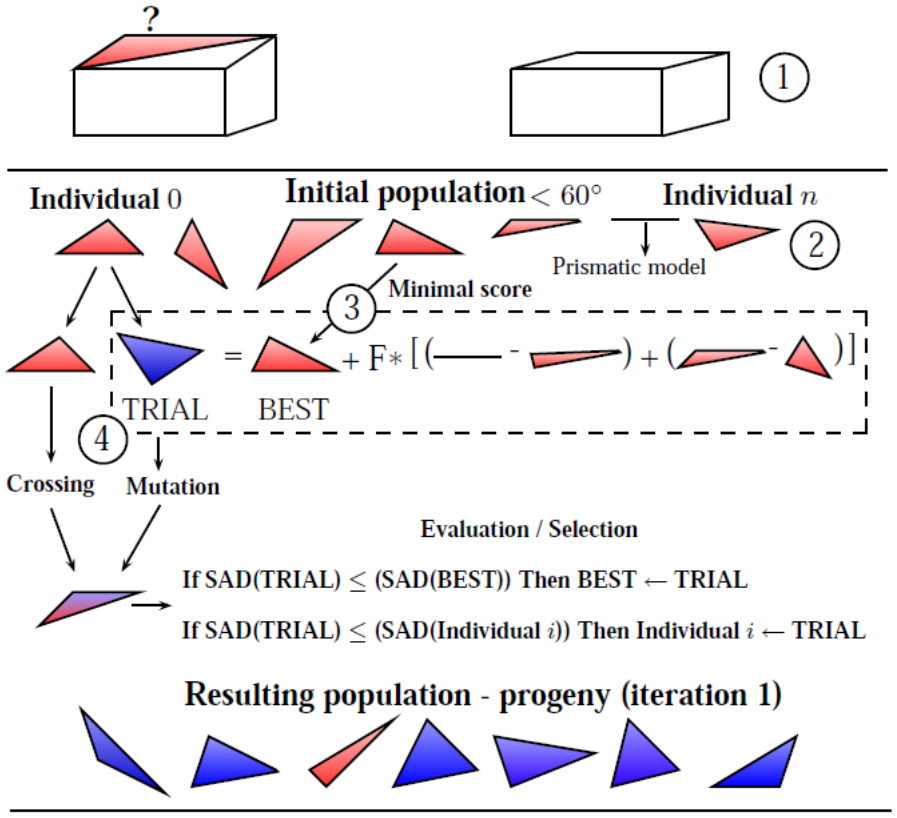


Figure 2.9: Illustration of one iteration of the DE algorithm.

2.4.3.3 Selecting One or Multi-Facet Building Modeling

This subsection deals with the one-facet or multi-facet selection. Several automatic strategies can be employed:

- The first strategy consists in reconstructing the two models independently (one-facet model (sloped roof) and multi-facet model). Each calculated model provides a SAD score. The 3D model finally obtained will be the solution providing the minimum score among the models shown in Figures 2.10(b), 2.10(c), 2.10(d) in the sense that this score characterizes a better correspondence between the 3D model and the image data set.
- The second strategy (adopted) exploits the putative estimation of 4 facet normals. The building footprint (rectangular) is divided into 4 triangular facets as a pyramidal model (see Figure 2.10(a)). The estimation of the pyramidal model is driven by the evolutionary algorithm as described above. A geometrical constraint is imposed; namely, the central 3D point is located along the line of sight that passes across the central pixel of the footprint in the master image. In this first step, a pyramidal model is computed. In a second step, a score $\mathcal{S}_{\angle}$ is set the highest deviation between the 4 facet normals and the vertical direction. If this computed score is less than a predefined tolerance threshold for normals verticalness denoted $\mathcal{T}_{\perp}$ (low angular deviation empirically fixed) then the retained model is the prismatic model initially calculated (e.g., Figure 2.10(b)). If this score exceeds this tolerance threshold (normals are non vertical), a score $\mathcal{S}'_{\angle}$ is set to the highest deviation between the facet normals. If this calculated score is less than a predefined tolerance threshold for normals

parallelism denoted $\mathcal{T}_{\|}$ then the one-facet estimation (e.g., Figure 2.10(c)) is carried out (estimation of a triangular facet using the rectangular footprint and estimation of the fourth vertex by intersecting the associated line of sight with the estimated plane). Otherwise, it means that the normal vectors are neither vertical, nor parallel (e.g., Figure 2.10(d)). In this condition, the multi-facet modeling is carried out.

It is worth noting that erroneous feature-based solutions as illustrated in Figures 2.4(a) and 2.4(b) (DEM-based) or coming from existing, less accurate, modeling pipeline could be used as initial solutions for the prismatic model estimation. Alternatively, they can also be used more directly as initial solutions for the Differential Evolution algorithm (multi-facet). In summary, the proposed approach proceeds in two parts. First, the algorithm decides if the building contains one or more facets. This decision is carried out by analyzing the 3D normals associated with four virtual triangles forming a partition of the whole building footprint. Second, once the model is selected, its associated parameters are then estimated by minimizing the defined dissimilarity score.

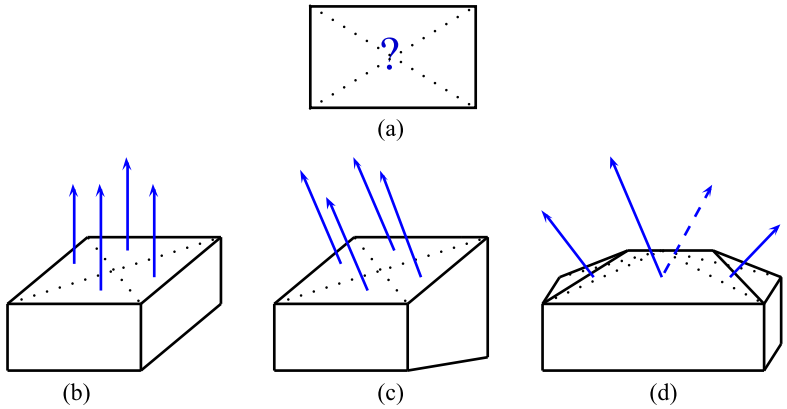


Figure 2.10: (a) Building footprint initially selected. No prior knowledge of the model shape is known. (b) Estimated prismatic model (algorithm initialization). (c) Estimated one-facet model (sloped roof). (d) Estimated multi-facet model (hip roof).

2.5 Experimental Results and Performance Study

In this section, we present the dataset employed as input of the proposed approach as well as the evaluations and the results obtained by our reconstruction method. We carry out several evaluations in order to analyse the convergence, the robustness and the accuracy of our image-based approach. These evaluations demonstrate the high potential of our modeling approach.

2.5.1 Input Dataset

The considered input dataset contains multiscopic gray-scale aerial images (see sample Figure 2.12). Each acquired image is described by a set of data specifying the georeferencing, the intrinsic parameters of the camera (e.g., focal, distortion coefficient, principal point) as well as the extrinsic parameters of camera location and orientation (rotation matrix, coordinates of point of

(a) Aircraft model *Beechcraft Super King Air 200T* belonging to the IGN aerial fleet.

(b) Officer that controls aerial acquisition system from an on-board station (back- ground) connected to the central camera (foreground).

(c) Similar scene observed from the opposite point of view. The central camera is installed in the floor hatch.

Figure 2.11: An IGN aircraft with its conventional configuration of devices employed for the aerial photography. Figure 2.11(a) is retrieved from the website spottingaviation. Figures 2.11(b) and 2.11(c) are retrieved from the IGN website and (Duffaut et al., 2008), respectively.

view). These images have been acquired from an airplane, equipped with three cameras, coming from the aerial office of the city of Toulouse. A central camera was oriented vertically to the Nadir point. Two other cameras were mounted, one to the front of the plane and the other to the back of the plane, with a front and back oblique view. The photography shows in Figure 2.11 illustrates a conventional scenario of aerial acquisition achieved from an IGN aircraft. This kind of aircraft flies with an altitude that can reach in practice $8800m$ (maximal altitude), a flight range of $8h$ and an approximative cruising speed at $440km/h$. The flight axis generally are parallel and foster the stereoscopic vision by respecting a longitudinal and lateral overlapping of 60% and 20%, respectively. Here, the resolution of the presented digital images is around 10 centimeters. There are two key parameters related to acquisition, namely; $\mathcal{B}$ that corresponds to the distance between two cameras (two positions) and $\mathcal{H}$ that corresponds to the altitude of the flight. Generally, these two param-

(a) View point 1 (b) View point 2

Figure 2.12: A pair of aerial images extracted of the multiscopic dataset. Each image covers a common area of the city of Marseille acquired from different points of view (partial overlapping). The size of the images is $\mathcal{N}_c \times \mathcal{N}_r = 4158 \times 4160$ where $\mathcal{N}_c$ and $\mathcal{N}_r$ correspond to the number of columns and rows, respectively.

eters are provided by the ratio $\frac{\mathcal{B}}{\mathcal{H}}$. A high ratio allows for a more accurate reconstruction. A low ratio allows a more reliable matching but reduces the accuracy. For the considered dataset, $\frac{\mathcal{B}}{\mathcal{H}} = \frac{230}{1280} \approx 0.18$. The acquired images are well-spaced and demonstrate a sufficient level of overlap. Hence, the input dataset is appropriate to carry out the 3D reconstruction.

Comments:

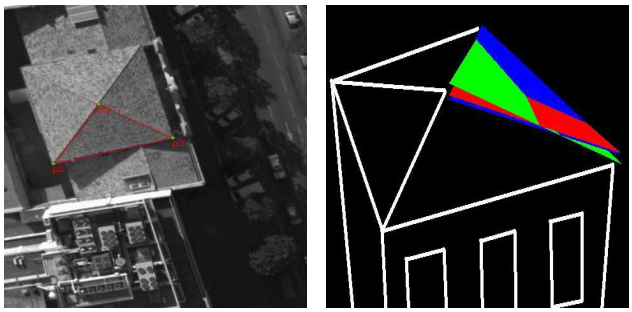
- *The image resolution is sufficient in order to reach the intermediary level of modeling (Figure 2.1(b)).*
- *In order to reduce self-occlusion of roofs and perspective effects, we use images captured by the vertical camera only. Moreover, for a given set of Nadir images in which the building under study is visible, the master image is the one having its center the closest to the 2D building footprint. These choices for the camera and for the master image maintain a certain matching robustness.*
- *As previously mentioned, no Digital Elevation Model has been used in our approach.*

2.5.2 Reconstructed 3D Models and Convergence Study

In this subsection, we measure the quality of the reconstruction obtained using the DE algorithm. We carry out a 3D modeling of one generic 3D facet running the DE algorithm with 30 iterations. The considered facet contains 5424 pixels. The number of individuals that compose the population is fixed to 30.

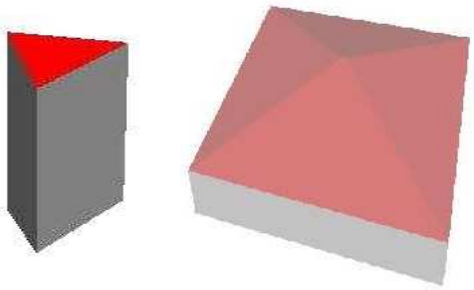
Figure 2.13(b) shows the estimated 3D facet corresponding to three iterations of the DE algorithm. The facets green, blue and red respectively correspond to the solution associated with iterations 1, 11 and 26, respectively. We can observe the facet evolved in the three-dimensional space. Each emerging facet represents a more accurate solution than the solution estimated in the preceding iteration. The intermediary solutions tend to correlate with the 3D

ground truth. The image facet initially selected corresponds to a roof portion of a pyramidal building (Figure 2.13(a)). In Figure 2.13(b), the coarse structure drawn in white is only shown to accentuate the perspective effect for a better comprehension. The final 3D models (3D quarter and pyramidal model) resulting from the estimation are respectively shown in Figures 2.13(c) and 2.13(d).



(a) the selected facet in the master image (top view).

(b) three estimated 3D facets resulting from consecutive iterations (perspective view).



(c) the estimated one-facet 3D model (a quarter).

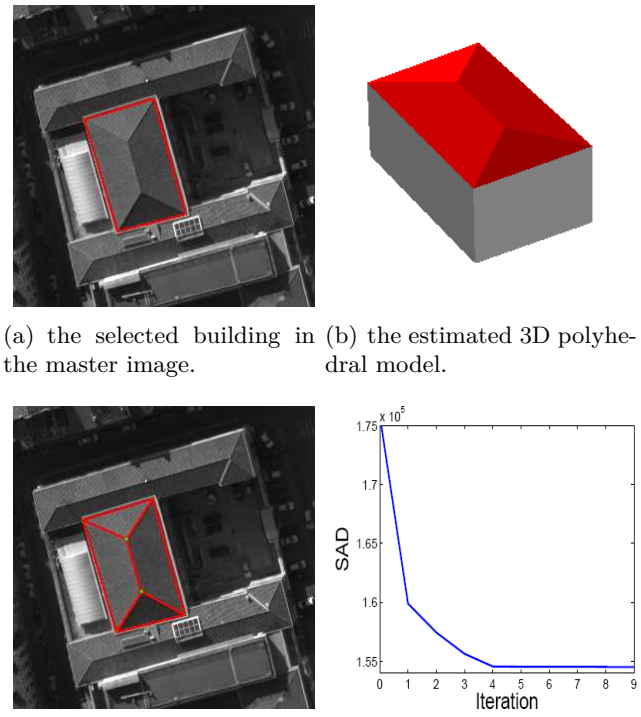
(d) the estimated multi-facet pyramidal (full).

Figure 2.13: (a) illustrates the facet in the master image. (b) illustrates the successively estimated 3D facets during the evolution of DE algorithm. (c) and (d) illustrate the final estimated 3D model.

Figure 2.14 shows the reconstructed 3D model (Figure 2.14(b)) associated with the building footprint selected in the master image (Figure 2.14(a)). The estimated 3D model projected into 2D images (e.g., 2.14(c)) enables the qualitative verification of the geometric coherence of the reconstructed 3D model. Figure 2.14(d) illustrates the global SAD score normalized by the number of pixels corresponding to the best individual obtained by each iteration. The graph evolution of the SAD shows the progressive evolution of the reconstructed 3D facets towards the ground truth configuration.

Additional evaluations and results of reconstructed 3D building models and convergence are illustrated in Figures 2.15 and 2.16. In particular, Figure 2.16 shows in detail the evolution of a 3D polyhedral building model using the DE algorithm. The model convergence is analyzed in the aerial images (column 1 and 2) and in the three dimensional space (column 3) at different iterations of the algorithm (iterations 1, 3 and 5). We observe that the registration of the full building footprint (boundaries and inner line segments) between the master image (fixed footprint) and the other images (moving footprints) from the multiscopic dataset allows the correct 3D structure of the associated

polyhedral model to be inferred. The registration process is guided by the DE algorithm. The estimated 3D models show a coherent converging shape from one to the next iteration. This continuity demonstrates the convergence reliability of the process.



(a) the selected building in the master image.

(b) the estimated 3D polyhedral model.

(c) the projection of the estimated model onto the image.

(d) the evolution of the best SAD as a function of the iteration number.

Figure 2.14: Estimated 3D polyhedral building model and related convergence.

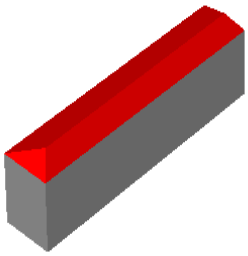
2.5.3 Accuracy Evaluation

Figure 2.16 shows the results of reconstructing a building composed of four facets. The images shown in the middle column represent the projection of the estimated 3D model. At the end of the optimization process, we observe that the model points projected onto the other image coincide from one image to the other. This demonstrates that the estimated three-dimensional model is accurate. For a more quantitative evaluation, we selected one facet of this building and compared the 3D model with that obtained by a DEM-based modeling approach.

Table 2.2 provides a comparison associated with one 3D facet using on the one side, a DEM-based approach and on the other side, our proposed direct approach. The DEM-based approach from the known modeling pipeline previously mentioned has been used as a reference in our evaluation. Notice that the presented 3D facet solution (3D plane equation) has been estimated using a robust estimator by considering the complete set of 3D points included in the facet footprint.

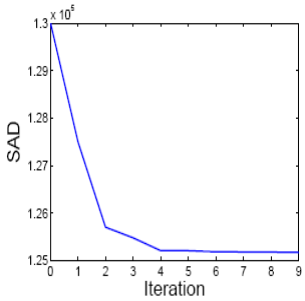
The direct image-based and featureless approach provides a satisfying three-dimensional modeling since the results are very close to the ground truth data. As can be seen in Table 2.2, the results indicate an average deviation of around one decimeter.

(a) the selected building in the master image.



(b) the estimated 3D polyhedral model.

(c) the projection of the estimated model onto the image.



(d) the evolution of the best SAD as a function of the iteration number.

Figure 2.15: Estimated 3D polyhedral building model and related convergence.

Accuracy evaluation will also be studied in the following subsections that deal with the robustness evaluation of the proposed approach in complex cases.

	DEM-based approach	Featureless proposed approach
$p1$	$(117.59, 396.80, 26.95)$	$(117.66, 396.75, 27.21)$
$p2$	$(123.96, 387.79, 23.70)$	$(124.05, 387.74, 23.80)$
$p3$	$(108.36, 390.32, 24.51)$	$(108.33, 390.26, 23.85)$
Barycenter	$(116.70, 391.62, 24.97)$	$(116.66, 391.67, 25.06)$

Table 2.2: Comparison of 3D modeling results obtained in the first case from a DEM-based approach and in the second case from our direct image-based approach.

2.5.4 Performance in the Presence of Image Noise

In order to get a quantitative evaluation of the 3D accuracy of the proposed approach, we adopted a simple and cheap scheme. For the sake of simplicity, we limited the study to a triangular facet that is viewed in two aerial images. In this scheme, we employ semi-synthetic aerial images (see Figure 2.17). Starting from a 3D facet model associated with a master image we synthesize the intensities of this facet in the second image by simply warping its intensities in the master image to the second image. The 3D selected solution will be considered as the ground truth. Each pixel belonging to the footprint in the second image is synthesized by its match from the master image using a bilinear interpolation and the corresponding ground-truth homography. In this case, the associated SAD value will be close to zero. We then add image noise to the transferred intensities. The proposed reconstruction approach is then invoked in order to compute the 3D model of this facet. The deviation between the ground-truth

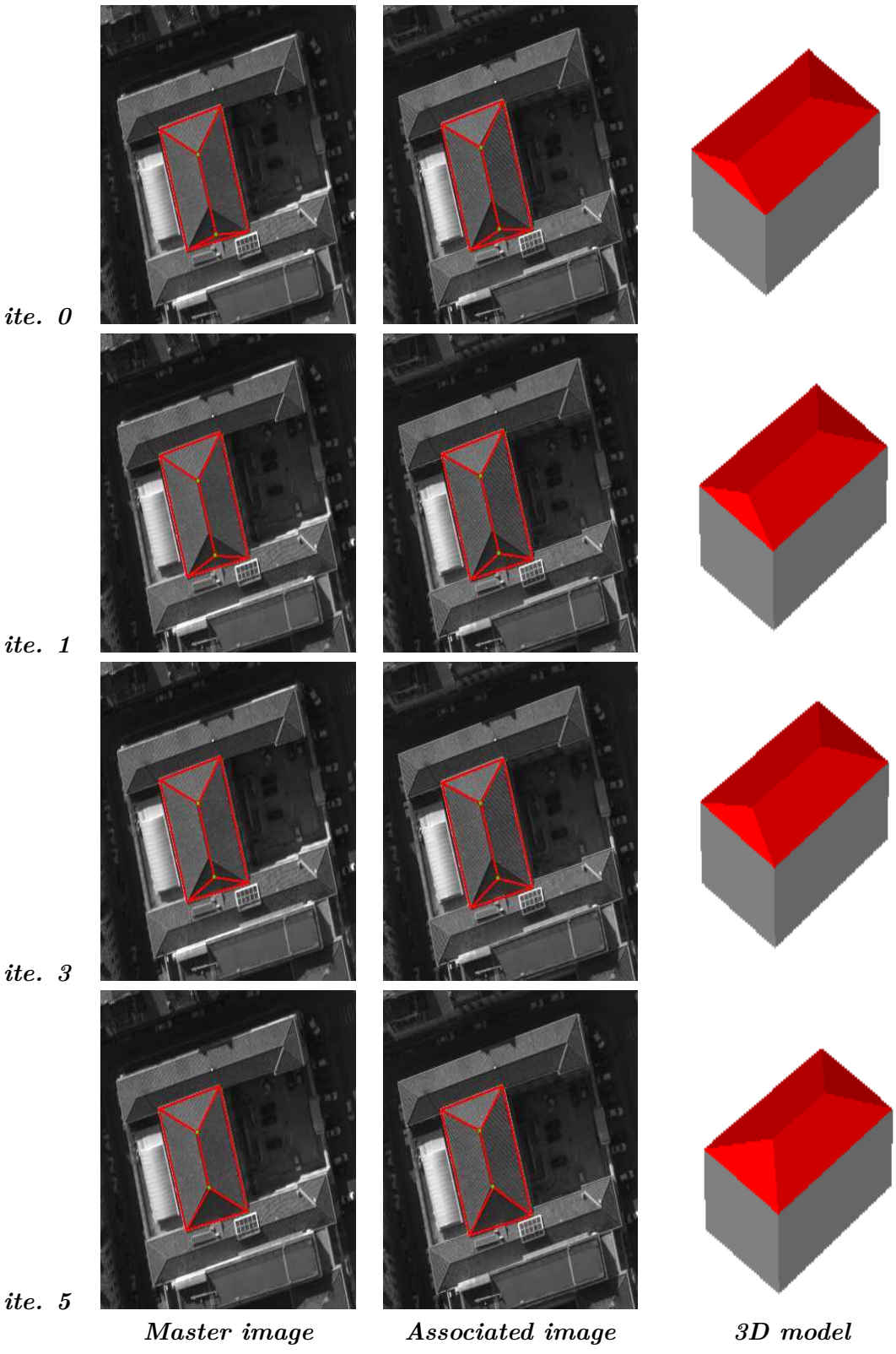
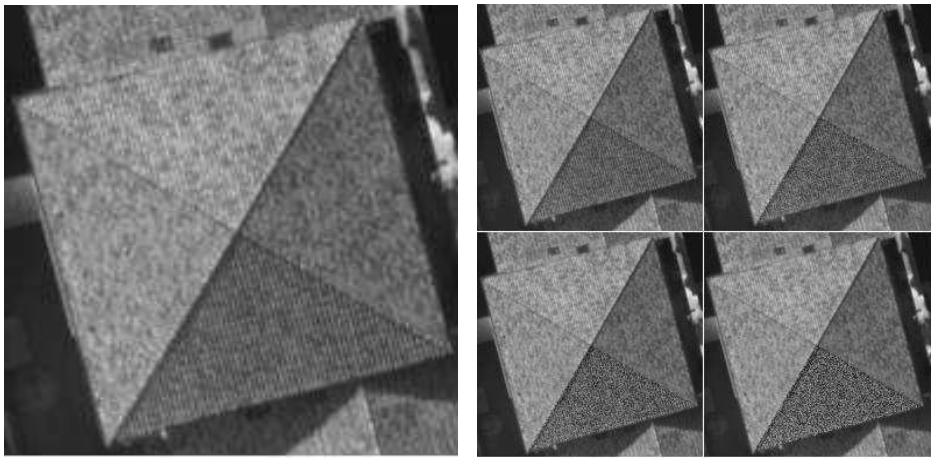


Figure 2.16: The best solution at several iterations of the Differential Evolution algorithm. The evolution of the 3D model and the footprint in the associated image is shown. The proposed algorithm converges to an optimal final solution in a few iterations.

3D model and the estimated 3D model is calculated as a function of the noise magnitude.



(a) Targeted pyramidal building model. (b) Noisy facets (bottom facet).

Figure 2.17: Adding noise to a facet for robustness evaluation. The intensities of the gray-scale pixels belong to the interval [0,255]. The magnitude $\mathcal{M}$ of the uniform noise progressively increases according to the respective intervals $\mathcal{M}_1 = [-4,4]$, $\mathcal{M}_2 = [-8,8]$, $\mathcal{M}_3 = [-16,16]$ and $\mathcal{M}_4 = [-32,32]$. These four levels of noise are shown in (b) (the corresponding random noise affected the bottom facet).

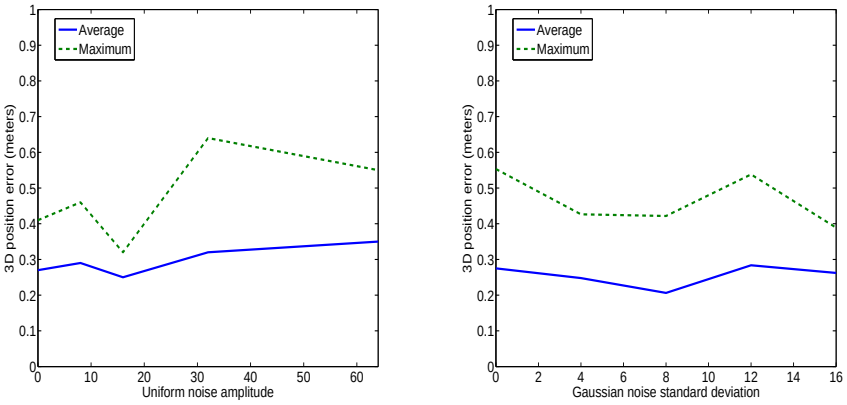
We have used two kinds of image noise: uniform and Gaussian. The three first tests correspond to a uniform noise. The three succeeding tests correspond to a Gaussian noise (see Figure 2.20).

Figures 2.18, 2.19 and 2.20 show the errors of 3D reconstruction of one facet as a function of the level of noise, respectively uniform (to the left) and Gaussian (to the right). These errors are obtained by calculating an average over ten trials for each noise level, *i.e.*, ten reconstruction solutions. The 3D errors are expressed in meters. The x-axis shows the four levels of noise magnitude. The square deviation of the Gaussian noise is equal to 32 when the level of magnitude is equal to 4.

In this way, the added noise that progressively increases simulates images of buildings with different levels of quality and tests them for a 3D reconstruction. Thus, the first level of noise simulates slight defaults in the acquisition. In this way, we can test the robustness of our reconstruction method according to the quality of the acquired images.

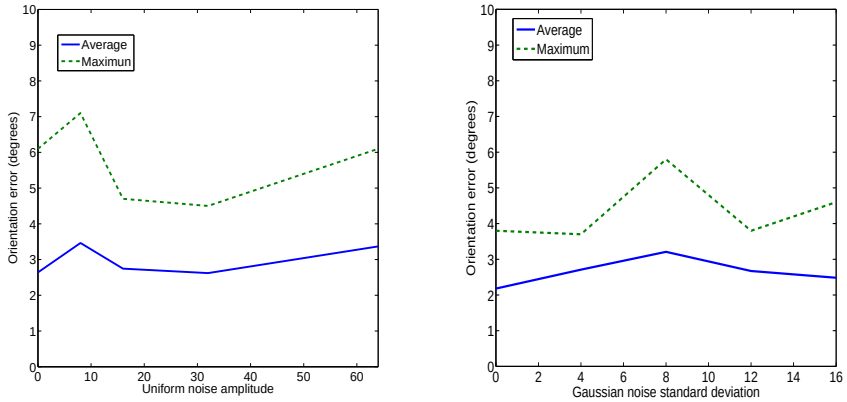
We observe that the noise added to the image of footprints does not severely affect the accuracy of the 3D reconstruction. Depending on the type of noise, the average errors associated to the vertices can reach $33cm$, the average error associated with the sloping angle can reach $3.5°$ and the average error associated with the vertices altitude can reach $31cm$. Moreover, we can observe that the maximum errors have an inaccuracy multiplied by two to three with a maximum sloping angle of $7.2°$, a maximum location deviation of $62cm$ and a maximum altitude deviation of $53cm$. The values concerning the location and the sloping errors seem to oscillate while the altitude error seems to increase more with the noise. Nevertheless, despite the presence of high noise magnitude, the location deviations remain inferior to one meter and the sloping angle deviation is inferior to $10°$. These values prove that the quality of

the acquired images and their resolution are sufficient to allow accurate 3D building reconstruction using the proposed method.



(a) Uniform noise (b) Gaussian noise

Figure 2.18: Error on the vertex 3D positions.



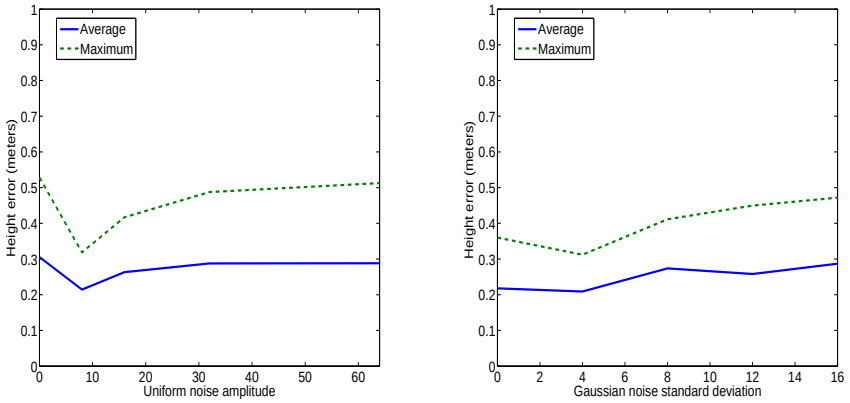
(a) Uniform noise (b) Gaussian noise

Figure 2.19: Error on the sloping angle.

2.5.5 Performance according to the Image Resolution

In this subsection, we propose to study the performance of the approach when the image resolution is reduced. A simple experiment was conducted. A triangular facet was selected. This facet contains 5432 pixels. We generated a sub-sampled facet by dropping every other column in the original image. Thus, we simulated a facet image with a reduced resolution.

Table 2.3 illustrates the 3D reconstruction of the tested triangular facet with the original resolution (first column) and with the reduced resolution (second column). The deviation between these two solutions is very small. The estimated 3D coordinates of the facet vertices globally varied for a few centimeters. In extreme cases, this deviation was around $20cm$. In spite of a decrease in the image resolution, the accuracy of the estimated 3D models does not considerably downgrade.



(a) Uniform noise (b) Gaussian noise

Figure 2.20: Error on the facet height.

	Original resolution images	Sub-sampled images
$p1$	$(117.65, 396.75, 27.14)$	$(117.65, 396.77, 26.87)$
$p2$	$(124.05, 387.75, 23.70)$	$(124.05, 387.75, 23.76)$
$p3$	$(108.35, 390.24, 24.44)$	$(108.34, 390.25, 24.27)$

Table 2.3: Comparison of 3D modeling results in the cases original resolution and sub-sample images.

2.5.6 Performance in the Presence of Superstructures

As previously mentioned, we aim to reconstruct planar roofs from aerial images. An important question comes to mind: what is the effect of superstructures on our reconstruction method? Indeed, a large majority of buildings incorporate superstructures. Consequently, the superstructures may generate unwanted noise since their 3D structures are not included in the dominant plane associated with the facet.

In this section, we present a method which increases the robustness of the facet reconstruction having several superstructures. The aim is to prevent the superstructures from distorting the estimation of the planar roofs. The idea consists in (i) detecting the pixels of the superstructures, and (ii) in using the footprint removed from these pixels. Assuming that the 3D plane calculated by the Differential Evolution algorithm is relatively accurate, we can thus classify the associated pixels into two categories: the pixels belonging to the dominant plane and the outlier pixels (pixels that do not belong to the plane). The proposed method proceeds in two passes:

- In the first pass, the DE algorithm is used with the totality of the reference footprint.
- In the second pass, the DE algorithm is used with only those pixels considered as belonging to the dominant plane.

Once the plane has been estimated in the first pass, several techniques can be used in order to carry out a coarse classification of the pixels. We observe that the pixels that do not belong to the theoretical plane of the facet will have a significant residual (absolute difference between the gray levels in different images) since the transfer pixel-to-pixel will not be correct. The idea is then to detect the pixels having a significant residual. We present then two techniques

based on the threshold of individual residuals:

– The first technique selects the outlier pixels by determining a threshold for the residuals. The empirical threshold $\mathcal{T}_{emp}$ is defined as follows:

$$\mathcal{T}_{emp} = \mu + k \cdot \sigma \tag{2.12}$$

where k is the coefficient of weight and has been manually determined. μ is the average of the individual residuals and σ is the associated square deviation.

For k=0.2, we observe that the major part of the superstructures included in the facet (chimney and trap of roof) are detected.

– The second technique uses another formula for the threshold. This threshold noted $\mathcal{T}_{gen}$ is defined by:

$$\mathcal{T}_{gen} = 2.5 \cdot Me \tag{2.13}$$

where Me corresponds to the median value of the residuals associated with the whole facet footprint.

We observe in Figure 2.22 that the major parts of the numerous superstructures belonging to the facet are detected. The detected white pixels shown in Figures 2.22(b) and 2.21(b) are ignored in the calculation of the final solution. All the pixels having higher residual values than $\mathcal{T}_{gen}$ are removed. The adopted threshold $\mathcal{T}_{gen}$ provides satisfying results for massive and generic filtering of the facet superstructures.

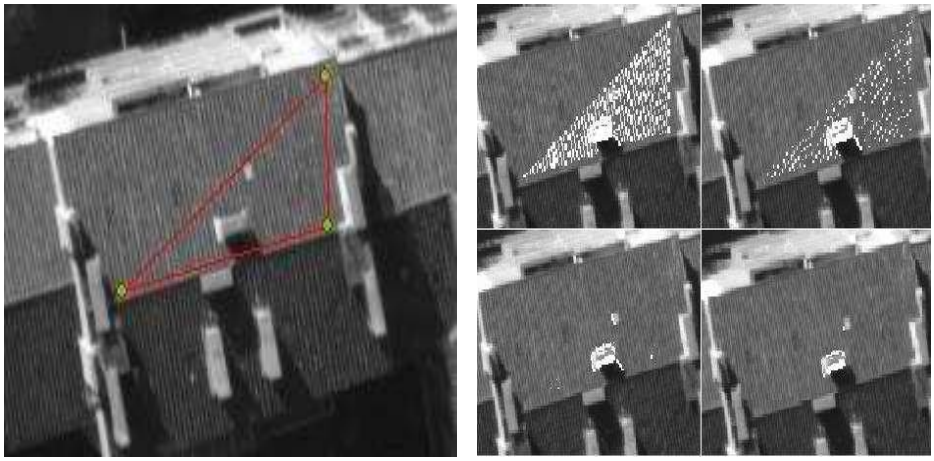
We have carried out a comparison of modeling methods by using the solution provided by the DEM as a reference solution. We have applied the reconstruction methods on one facet including superstructures namely several chimneys (see Figure 2.22(a)). The solution provided by the DEM has been obtained by estimating a 3D facet from all the associated elevation points via a robust estimator. Two measures of dissimilarity have been applied in our DE algorithm, in the first case, the SAD measure and in the second case, the SSD measure. Table 2.4 summarizes the 3D reconstruction of the facet using five strategies. We only report the height of the estimated vertices.

We observe the reconstructions to the altitude of the vertices since it is the parameter that varies the most (see Table 2.4). We have used the first technique in order to filter the superstructures. The coefficient k has been empirically tuned to 0.2. Although a part of the pixels belonging to the roof plane is filtered out, this part is low in comparison to the number of the considered pixels (several thousands) in the facet estimation.

	DEM	SAD	SAD	SSD	SSD
Including		superstructures	filtering	superstructures	filtering
$z1$	$41.96m$	$42.92m$	$42.22m$	$43.61m$	$42.75m$
$z2$	$41.36m$	$41.10m$	$40.98m$	$40.84m$	$40.87m$
$z3$	$39.78m$	$39.62m$	$40.22m$	$38.88m$	$40.10m$
Average deviation in z	$0.0m$	$0.46m$	$0.36m$	$1.02m$	$0.53m$

Table 2.4: Comparing the modeling results obtained with the SAD and SSD scores using facets including superstructures with and without the filtering process.

We observe that the most accurate method seems to employ the SAD measure. Moreover, we find that the presence of superstructures can affect the 3D reconstructed model. A filtering stage is thus necessary in order to increase



(a) Targeted facet part with superstructures.

(b) The superstructure detection with different values for k.

Figure 2.21: Filtering out the superstructures. (b) Tuning the k coefficient for the determination of the residual threshold $\mathcal{T}_{emp}$ (k=0, k=0.1, k=0.2, k=0.3, respectively,). The removed pixels are shown in white.

the accuracy of the solution. To this step, we envisage testing several methods integrating by different ways the superstructure filtering with the aim of model improvement. We stress on the fact that the DEM-based reference solution does not correspond to the ground truth.

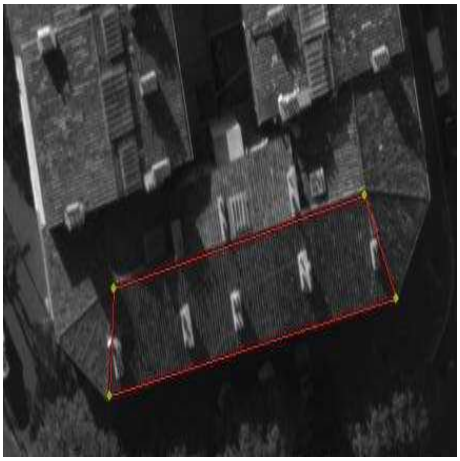
2.5.7 Performance in the Presence of Significant Shadows

As previously mentioned, the correct image registration of the building footprint leads to a correct 3D building model. Figure 2.23 illustrates the 3D modeling of some buildings using aerial images containing significant areas of shadow. In the Figure, we visualize the 2D projection of the obtained 3D model in the non-reference image. As can be seen, for a variety of different buildings, the registration process is robust even in the presence of significant shadows.

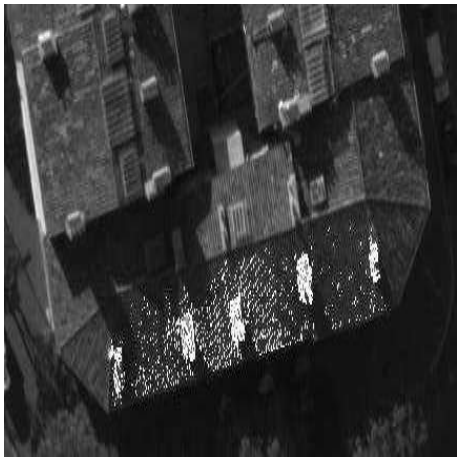
2.6 Conclusions

2.6.1 Contribution

In this work, we provided an overview of some problems and solutions dealing with 3D building modeling. We proposed a new methodology for 3D building reconstruction based on a featureless process. To the best of our knowledge, this method has never been exploited in the 3D building modeling problem. Unlike existing methods, the pixel-to-pixel matching process is avoided. However, it is a by-product of the proposed method in the sense that once the 3D shape of the building is known, the image-to-image transfer is known from the associated homographies. The optimization associated with the proposed method has been carried out using the Differential Evolution algorithm. The method has been validated using real and simulated images. The proposed approach was compared with DEM based modeling approaches. It is beyond the scope of the current work to compare the proposed approach with all existing feature-based approaches. Indeed, it is well known that featureless approaches



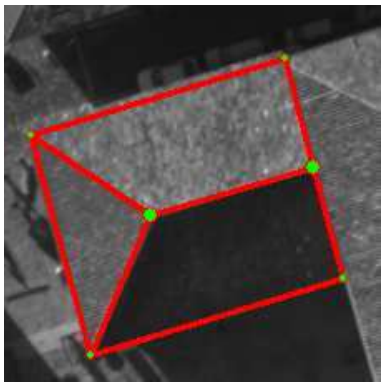
(a) Targeted facet part with superstructures.



(b) Detected superstructures.

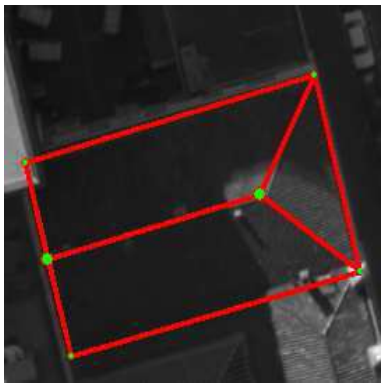
Figure 2.22: Automatic detection and filtering of the superstructures. The threshold $\mathcal{T}_{gen}$ is proportional to the median of all residuals. The removed pixels are shown in white.

(a) Gable roof.



(b) Hip roof.

(c) Gable roof.



(d) Hip roof.

Figure 2.23: Correct building modeling in the presence of significant shadows. The master images are not shown.

outperform feature-based approaches regarding the accuracy of the estimated geometric transforms used for image registration. The proposed method provides a satisfying polyhedral building reconstruction from gray-scale calibrated aerial images. The proposed top-down approach is also able to rectify erroneous

reconstructed polyhedral building models in existing 3D city models whenever the corresponding aerial images are available. Furthermore, we argue that the proposed modeling method provides a novel tool that can be used in existing large-scale urban modeling pipelines as a main or complementary tool.

2.6.2 Future Work

Future work will be concentrated on the following directions:

- Testing other dissimilarity measures. New dissimilarity measures could be used and evaluated.
- Improving the reconstruction of roofs having superstructures. One possible solution is the integration of the outlier pixel filtering in the Differential Evolution algorithm. Indeed, each individual provides a normalized SAD score which considers only the pixels belonging to the roof plane. The pixels belonging to the superstructures will not be considered in the score calculation. The superstructure filtering will generate a different distribution in the progeny and provide a more accurate solution. The second scenario envisions running the proposed two passes several times.
- As we have previously mentioned, the registration process is carried out between a reference footprint (fixed boundaries) selected by an operator in the master image and the intensities of the aerial images of the multiscopic data set in which the building is visible. Consequently, the initial boundary line segments of the reference footprint could be slightly shifted. In a future work, we intend to use a method that is able to deform a rough 2D footprint into a precise 2D footprint. This rectification process should precede the image based 3D reconstruction. We stress the fact that the use of cadastral maps can release the requirement of having an accurate 2D footprint.
- In our work, we assume that the building roof has conventional and simple shapes. Research could be done in the future in order to extend the direct approach to the case of generic buildings and roofs with atypical shapes.

Transition

The recent developments of street acquisition systems such as Mobile Mapping Systems open new perspectives for improvements in urban building modeling in the sense that the terrestrial data (very dense and accurate) can be exploited with more performance (in comparison to the aerial investigations) to enrich the building models at facade level (e.g., geometry, texturing). For this reason, studies should be undertaken at terrestrial level in order to analyse the urban buildings such as seen from the street as well as to design and to develop complementary and/or appropriate 3D facade modeling pipelines that could bring solutions to the actual needs (e.g.; hybrid aerial-/terrestrial-based building modeling, urban facade visualization).

Part III

Terrestrial Facade Modeling

Chapter 3

Properties of Terrestrial Urban Laser Data

Abstract

This Chapter analyses amongst others the architectural complexity of the street facades in order to extract urban knowledge and characteristics that could be exploited by street facades modeling approaches. In addition, the IGN Mobile Mapping System (MMS) as well as the acquisition protocol, the collected data and their management are described in detail. Finally, the acquired laser data are meticulously analysed and the challenges and difficulties are described and illustrated. This Chapter provides then a starting point and key information useful to the understanding of the different modules that compose the proposed street facade modeling pipeline.

Contents

3.1 Urban Street Facades Modeling

This section essentially aims at understanding some properties related to the terrestrial laser data acquired at street level in dense urban environment.

3.1.1 A Priori Knowledge on Street Facades

The city of Paris is one of the world cities the most rich in term of architecture.

Figure 3.1: A part of a panoramic image of a Parisian street under study. This street belongs to the 12^{th} district. Some of frequent urban objects can be observed between the acquisition point and the facades.

Figure 3.2: Sample of street facade images that illustrates the accumulation of diverse architectural shapes and material of constructions.

Indeed, the urban environment of the city of Paris is composed of streets with architectures that notably merge facade styles such as Haussmannian and contemporary. In reason of the diversity of facade styles, the complexity of architectural shapes, the huge amount of urban objects, the street activity (see Figure 3.1), the modeling of building facades and street facades from street raw data can be considered as a hard problem.

Besides, the study of the terrestrial modeling in dense urban environment such as the city of Paris (area under study) presents the high interest that the developed approach should be more easily generalizable to process city with less complex architecture. Notably, it will be helpful to process some of the American contemporary cities that are often composed of street with highly regular, linear shapes and orthogonal networks (characteristics observed in urban morphogenesis).

Figure 3.1 illustrates a typical street (panoramic image) acquired in the 12^{th} district of the city of Paris. From left to right, this Figure shows quantity of frequent urban objects located in front of the facades (3D objects); namely a bus shelter, a parkmeter, a small urban post, barricades of construction, a streetlight, a green street garbage, barricades of park, an advertising panel, a mast of flag, a street seats. We also observe pedestrians, vegetation, parked vehicles. Numerous frequent urban objets are not represented here (e.g., road signs, phone booths). Figure 3.1 exhibits thus the presence of diverse urban objects in the scene.

Also, numerous street facades have the architecture that change from one floor to the next and from one facade to the next. The ground floors of the facades often include shops devantures of stores, coffee terraces, cinemas, theatres, hotels, scaffoldings. The upper floors vertically and horizontally alternates range of wall and facade structures (e.g.; disjointed windows, balconies, bay-windows, protrusions). Also, these floors sometimes contains small decorative structures. Nevertheless, these upper floors globally have the aspect of a grid and correspond to the devantures of dwellings. The whole of these facade characteristics are illustrated in Figure 3.1.

Besides, building facades with different styles of architecture are usually joined together for hundreds of meters in a row and form in some cases a compact street facade. Beyond to the facade style diversity, the use of numerous construction materials and the diversity of assembly techniques conducts to architectural shapes and surfaces with different level of roughness such as granular surfaces (rock, brick, coating, joint) as it can be seen in Figure 3.2 or smooth surfaces (glasses of windows and shop windows). For this reason, the dataset of 3D points acquired at the facade level generally includes a variable thickness.

One of the most current and elemental facade represention is employed in the cadastral map in which each facade is generally represented by one single 2D segment line. However, the direct segmentation of joined and aligned facades (e.g., coplanar dwelling facades) into a set of facade segments such as those included in a conventional cadastral map is a complex task using street laser data.

Visually, individual facades can be distinguished by the high discontinuities of the facade top points located between two dwelling facades or by the periodicities of the windows. We also observe that aligned facades are sometimes separated by vertical gutters and downspouts (see Figure 3.2). However, the

direct use of this knowledge to the complete facade individualization remains complex in the sense that it depends on the efficiency of several previous stages; notably, street cloud segmentation, facade points extraction, windows points and facade top points detection and analysis.

Additionally, the street point cloud segmentation is an essential and delicate stage since the set of street points contains a tremendous amount of points of urban objects located out of the facades (i.e., outlier points) that will be necessary to detect and to remove towards facade processing.

To help in this stage, we stress the fact that many big cities throughout the world are composed of buildings having some common geometric characteristics. The walls of building facades are vertically and linearly constructed. The maximum height of facades and the minimum and maximum width of facade portions of the area under study can be known from existing urban databases. The minimum width of streets can also be known. It can be used to characterize holes in facade clusters, e.g. crossroads or occlusions. Then, these parameters can be employed to generate low-level hypothesis to the delimitation of facade regions (e.g., 3D bounding box).

In this way, although the terrestrial data usually contain several small street objects such as pedestrian, cars, posts, street lights, barriers; the geometric characteristics of street facades such as universal characteristics (low-level prior knowledge) or elaborated ground truths (e.g., cadastral maps, urban databases, libraries of parametric models as in (Chevrier et al., 2010)) can thus be used to roughly reduce the search space required to determine the facade regions as well as for the choice of relevant parameters to their accurate extraction. These parameters particularly are key information to drive the modeling approaches that employ strategies of hypothesis generation and verification.

3.2 Dataset Acquired by the Mobile Mapping System

Nowadays, the Mobile Mapping Systems are actively developed in order to massively collect more and more targeted urban data at street level towards the generation of future GIS databases. Notably, an introduction to the Mobile Mapping Systems has been achieved in (Petrie et al., 2010). Similarly to these described systems, this section aims at describing the street Mobile Mapping System that is developed at the IGN (see Figure 3.3), the acquisition protocol as well as the datasets that have been employed in the current study.

3.2.1 The IGN Street Mobile Mapping System

In this description, we mention that we are positioned as user of the data collected by this Mobile Mapping System. The works that consist in the mounting of hardware and the sensor calibration has been achieved by IGN technicians.

As illustrated in Figure 3.3, the vehicle surface has been painted in matte black in order to limit the effect of reflectance that could affect the acquired data.

The Mobile Mapping System is equipped with a mast that is composed of many heterogeneous embedded sensors (see Figures 3.4 and 3.5(a)). More precisely, the mast head is composed of eight cameras that are oriented in cardinal and ordinal directions in order to cover the scene at 360°. These cameras mainly are oriented in a plane fronto-parallel to the road plane.

Figure 3.3: IGN Street Mobile Mapping System.

Figure 3.4: View of a set of devices located at the roof of the vehicle.

(a) To the roof, instruments of measurements have been mounted on a mast.

(b) At the back, a unit of secure data storage and a power system have been fixed.

(c) At the passenger side, a station control with an onboard computer has been installed.

Figure 3.5: Focus on the vehicle conception.

(a) 2D range laser sensor (two cylinders) sloped to the facade top directions.

(b) Optical cameras organized at 360°to cover the surrounding scene (omnidirectional device).

(c) Optical cameras organized in stereoscopic pair.

Figure 3.6: Focus on embedded sensors used for data acquisition.

(a) Global Positioning System (GPS).

(b) Inertial Measurement Unit (IMU).

(c) Odometer.

Figure 3.7: Focus on the instruments used for data positioning.

Table 3.1: Focus on the vehicle planning and the multi-source measurement devices of the IGN Mobile Mapping System. We mention that only the sensors used in the thesis are describes here.

Two additional cameras are laterally located with an angular sloping at around 45°in direction of the facade top (see Figure 3.6(b)). In addition to these cameras, the vehicle is also equipped with two cameras located in the front and in the back of the vehicle and oriented in direction of the road. These camera pairs constitute two stereoscopic acquisition systems (see Figure 3.6(c)).

In short, the vehicle is equipped with Full HD digital cameras AVT Pike 210C (made by the Allied Vision Tec company)[1] that acquire images with a high radiometric resolution, namely 16-bit RGB channels. In geometric resolution, the Ground Sample Distance of these images can reach up to $\approx 5cm\ GSD$ (i.e., each pixel potentially represents a ground area of 0.05 square metre). The cadence of acquisition is one capture each 4.5m or each 10 seconds (grouped captures). The size of the acquired images can then reach $\mathcal{N}_r \times \mathcal{N}_c = 1920 \times 1080$ to wit around $20\overline{M}$ of pixels ($\mathcal{N}_r$ and $\mathcal{N}_c$ correspond to the number of rows and columns, respectively). Each image of the image dataset has undergone similar optical corrections (e.g., flat-field and white balance corrections). This stage has been post-processed by IGN engineers. These sensors will essentially intervene in the stage of facade models texturing (Chapter 5). For this reason, the dataset of optical images will be then presented in more detail in the final Chapter of this Part.

Moreover, the mast of the vehicle is also equipped by a high precision 2D laser sensor LMS-Q120i made by the RIEGL company[2] (see Figure 3.6(a)). The laser sensor is positioned at the mast of the vehicle and oriented to each street side in direction of the facades. Its beam plane is perpendicular to the vehicle trajectory. The system allows us to carry out up to 10000 measurements per second and the beam vertically sweeps with an opening of 80°($-20°$ to 60°with respect to the horizontal). This vertical sweeping provides then a frame of points (i.e., range or profile). In this study, the angular resolution was configured to 201 points by frame. The ground based laser range transmits laser pulses using a laser beam wavelength in the Near-InfraRed (NIR). The angular precision of the beam is equal to the one hundredth degree. More specifically, the accuracy of laser-based measurements is approximately $3cm$ at $150m$. The raw measurements provided by the laser sensor are points that are parametrized by a distance δ and an angle ϕ.

For each measure achieved by the laser sensor, the laser sensor also provided a reflectance information. This information corresponds to the energy of retro-diffusion back-projected by the surface targeted by the beam. The reflectance information is provided by a value that is normalized between 0 and 1.

Since the angular sampling is fixed for each acquired frame of points, a set of acquired frames provides an image of intensity which is represented by the temporal frames in x-axis and by 201 rows of reflectance intensities in y-axis (see Figure 3.12). Besides, the coordinates of the 3D points are expressed in the laser sensor coordinate system and also in a common coordinate system, namely the ground reference (absolute) Northern, Eastern and Altitude in Lambert 93 (French standard reference system). The accuracy of a 3D point depends on the laser beam accuracy (e.g., size of the laser footprint), the calibration accuracy of the sensor (intrinsic), the georeferencing system accuracy, the system rigidity and the device synchronization. Besides, the whole of the sensors have been

1. Link to the Allied Vision Tec company: *http://www.alliedvisiontec.com/*, Kodak optic
2. Link to the RIEGL company: *http://www.riegl.com/*

mounted on a shock-absorbing bracket in order to reduce the data inaccuracy due to vehicle vibrations and eventual chocks. In our case, we assume that the accuracy interval of points acquired in a street portion in one single pass do not exceed few decimetres.

The combining of a Global Positioning System (GPS) made by the Trimble company [3], an Inertial Measurement Unit (IMU) POS LV 220 made by the Applanix company [4] and an odometer allows the direct georeferencing of the data (respectively, Figures 3.7(a), 3.7(b), 3.7(c)). In the literature, some similar processes are described in detail (devices, sensors calibration) such as in (Ellum et al., 2002).

The whole of the embedded sensors and devices are connected to a reloaded power system and the data are collected in a unit of storage of Terra-bytes capacity. This hardware has been fixed at the vehicle back as it can be seen in Figure 3.5(b).

3.2.2 Acquisition Protocol, Collected Data and their Management

The laser data are acquired under realistic conditions in dense urban environments (e.g., street activity, uncontrolled environment, various topologies of buildings).

The acquisition of the mobile systems is generally separated in two modes: *Stop-and-Go* or *On-Drive*.

- The *Stop-and-Go* mode consists of acquiring data at multiple fixed location. The vehicle stops during the scanning operation (static platform orientation and location). The acquisition time is high enough and the acquired data usually are often very dense.

- The *On-Drive* mode consists to acquire data even when the vehicle is in motion. The scan rate, the coverage and the accuracy vary with the speed and the distance to the objects.

In our case, the acquisition has been carried out in On-Drive mode (i.e., in continuous). This mode has been initially chosen to simplify the condition of acquisition. Moreover, the laser data are also collected when the vehicle stops. Indeed, the activation of acquisition during the displacements is not directly managed by the current system (e.g., odometer) and the human intervention will be very fastidious to achieve this task. Hence, the acquisition has then been completely operated on-the-fly. The control station can be appreciated in Figure 3.5(c).

In the current study, the acquisition is focused on areas of the 12^{th} district of the city of Paris. The acquisition protocol used in our study has not been specifically designed to address the facade modeling topic in the sense that the strategy of acquisition has been to massively collect urban data (multi-purpose dataset) for the needs of several research projects based at street level (e.g., road sign detection and reconstruction). Also, we stress the fact that only the major axis (main streets) of the city of Paris have been acquired. The information related to the quantity of collected data has been presented in

3. Link to the Trimble company: *http://www.trimble.com/*
4. Link to the Applanix company: *http://www.applanix.com/*

Collected information	Stored quantity	Allocated memory size
Points of view (grouped images)	45 310	n/a
Images	543 720	2 250*Go*
Laser points	882 849 787	6*Go*
Post-processed images	n/a	6 800*Go*
Post-processed 3D points	n/a	70*Go*
Post-processed panoramic images	n/a	550*Go*

Table 3.2: Amount of the data collected towards IGN research projects for terrestrial urban applications. The set of laser and image data have been conjointly acquired.

Table 3.2.

Figure 3.8: On-drive acquisition of street raw data by using the IGN Mobile Mapping System in dense urban environment (longitudinal view). Example of high disparity in facade size, alignment and visibility from one street side to the other. Typical sensor configuration employed in our case.

In our case, the initial sensor configuration was not specifically adapted to the facade modeling. Indeed, the first dataset that we have used do not contain points at the facade top in reason of the laser sensor orientation and the associated angular resolution that cover both of street roads (including street objects) and facade at half-height. Also, only one of the two shown laser sensor has been exploited. The laser sensor configuration has slightly changed during the thesis essentially since the narrow field of a single laser sensor is not sufficient to cover the facade from the ground to the top. A

typical sensor configuration in condition of urban facade acquisition can be visualized in Figure 3.8. Besides, a stake of the thesis lies in the research of solution to process the tremendous amount of data towards producing models of urban street facades.

Moreover, the datasets are collected in a specific data structure related to the 2D laser scanner (frame of 3D points). Then, the laser sensor data can be processed in two ways; either as a container with data organized linearly in temporal sequences of 201 points (data frames) or like a cloud of 3D points. A set of information is labelled to each 3D point such as the date of acquisition and raw measurements (i.e., meta-data). Since the use of 2D laser sensors is relatively widespread, we have decided to exploit the sensor topology to our advantage. The scanner is oriented perpendicularly to the vehicle trajectory and facing the facades. The third dimension is then induced by the vehicle displacement.

3.2.2.1 The Input Dataset

As mentioned earlier, the acquisition has been achieved in realistic conditions of dense urban environments. In reason of the huge amount of 3D points, our first investigations have dealt with the processing of subset of 3D points corresponding to building block then facades of wide urban street. Figure 3.9 illustrates the acquired 3D point cloud corresponding to three dominant facade sides of a building block. The acquired point cloud is composed of facade points, occluding vegetation and set of points belonging to surrounding facade. Numerous external points overflows of the facades of interest. In Figures 3.10, 3.11 and 3.12, we respectively observe the acquired central facade of the city hall of the 12^{th} district in front view and Nadir view as well as associated image of laser intensities.

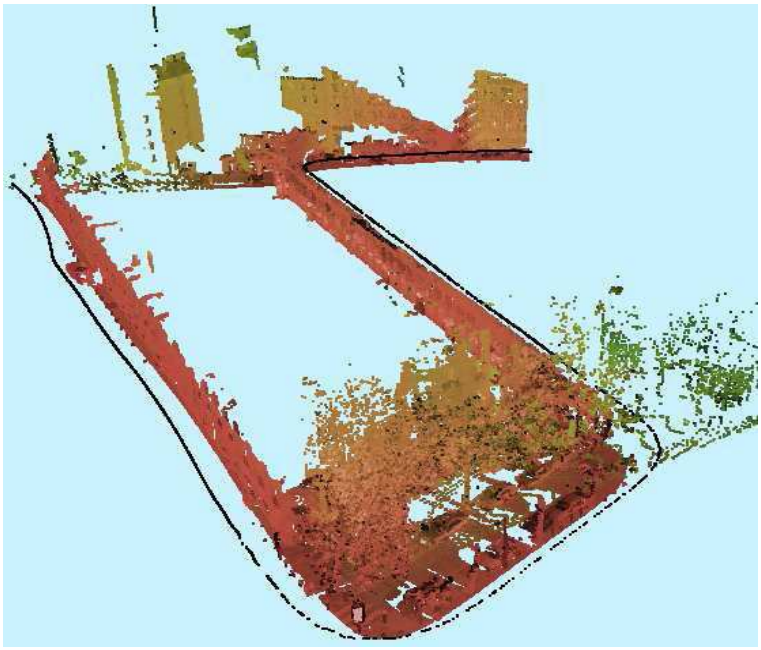


Figure 3.9: 3D point cloud of very high density corresponding to a portion of a building block (bird eyes view). These data represent the building facade acquisition (up to 10000 measurements per second). The black dotted line represents the trajectory of the laser sensor.

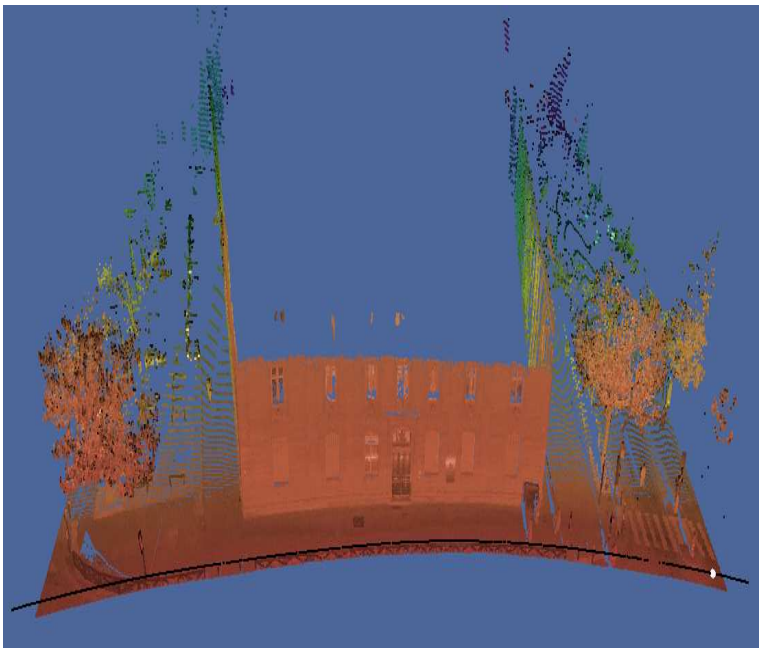


Figure 3.10: Illustration of an acquired 3D point cloud corresponding to the facade of a Parisian city hall in case of non-occlusions. The beam vertically sweeps the scene with an opening of 80°(−20° to 60°with respect to the horizontal). The narrow field of the sensor covers street and facade portions. The facade top is missing here.

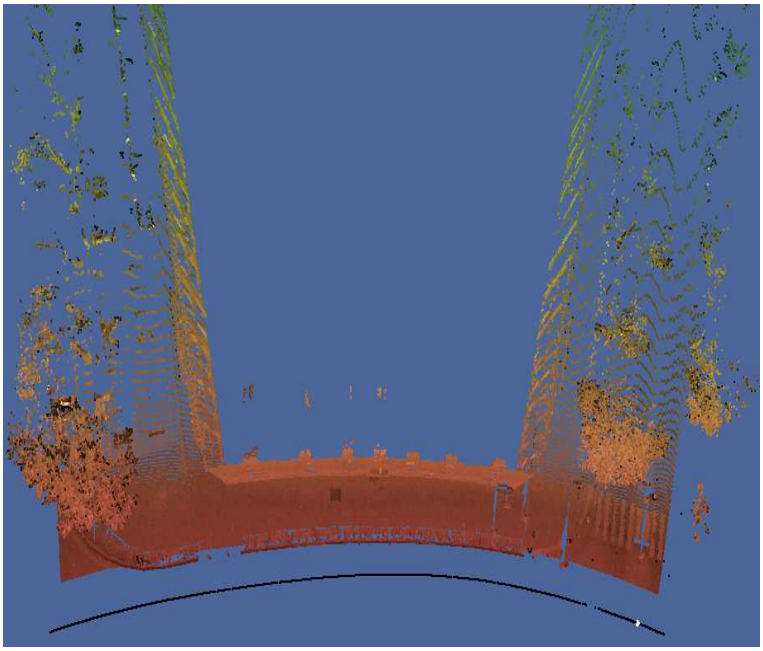


Figure 3.11: Three facades are visualized from the Nadir point of view. The Laser sensor has only swept the frontal facade. The silhouette of the building footprint is perceptible. Extrusions of facade microstructures such as windows are visible to the central facade in reason of the acquisition points of view (black line).

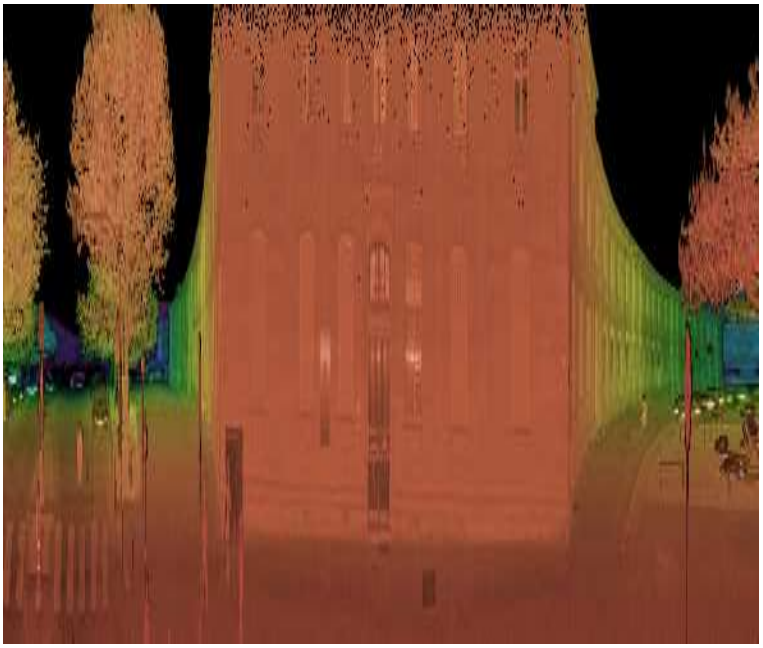


Figure 3.12: Image of returned laser intensities related to the Figure 3.11. The size of the image is $\mathcal{N}_c \times \mathcal{N}_r = \mathcal{N}_c \times 201$ where $\mathcal{N}_c$ and $\mathcal{N}_r$ respectively correspond to the number of columns (i.e., number of acquired frame related to the acquisition time) and rows (parametrized frame resolution).

3.3 Exploiting Laser Data: Challenges and Difficulties

As previously mentioned, the data are acquired under realistic conditions in dense urban environments. Moreover, the datasets are collected in On-Drive mode and stored in a container specific to the laser scanner (frames of 3D points). Here we will describe the main observed effects associated with the acquired laser data. These effects have been organized from minor to major as follows.

3.3.1 Laser Reflectance

The laser reflectance could cause confusions in the 3D data interpretation. Certain points don't model a physical surface. This effect appears on a retro-reflector surface. Observations sometimes show an aureole of points around road signs. These dispatched points represent erroneous data as it can be seen in Figure 3.13. Moreover, certain points model a different surface other than the surface of interest. Sometimes, the beam of the laser either rebounds off of the outside of the window or it passes through the window and models the inside of the dwelling. In our case, this effect can be neglected. However, it appears that the effect can be strongly restricting in some cases. Indeed, this effect seems changed according to the wave length of the used laser beam. These scattered points represent erroneous information for the facade modeling. In addition to this, other less frequent effects could arise due to poorly reflective surfaces.

3.3.2 Variability of Point Density

The density of the acquired 3D points vary according to the speed of the vehicle. In general, the density of points turns around 1 point to each 5 to 10 centimetres in horizontal and vertical directions when the vehicle is traveling at around 30 kilometers per hour. Moreover, the density is also depending on the spatial orientation of surfaces and their distance for each acquired frame (heterogeneous density). For example, in the ground plane, the points acquired close to the vehicle have a high density whereas the far-off points have a more low density (see the ground in Figure 3.13 or Figure 8 in (Yoo et al., 2009)).

3.3.3 Data Redundancy

The redundancy of data is due to many factors. The acquisition is running continuously and independently of the vehicle motion. Thus, even when the vehicle is stopped the laser sensor is still providing sets of 3D points. Then, redundant frames of points are collected (i.e., redundant 3D profiles). This effect is not easily perceptible in a visualization platform for 3D point cloud but clearly observed in the associated image of intensity (see Figure 3.14). Moreover, due to sensor characteristics (orientation and linear scanning), we could sometimes have up to three acquisitions of the same facade part caused by the graining of the laser beam in the turns. The redundancy of data (points, frames, parts of the facade) presents an inconvenience for the feature extraction techniques based on vote schemes or random trials.

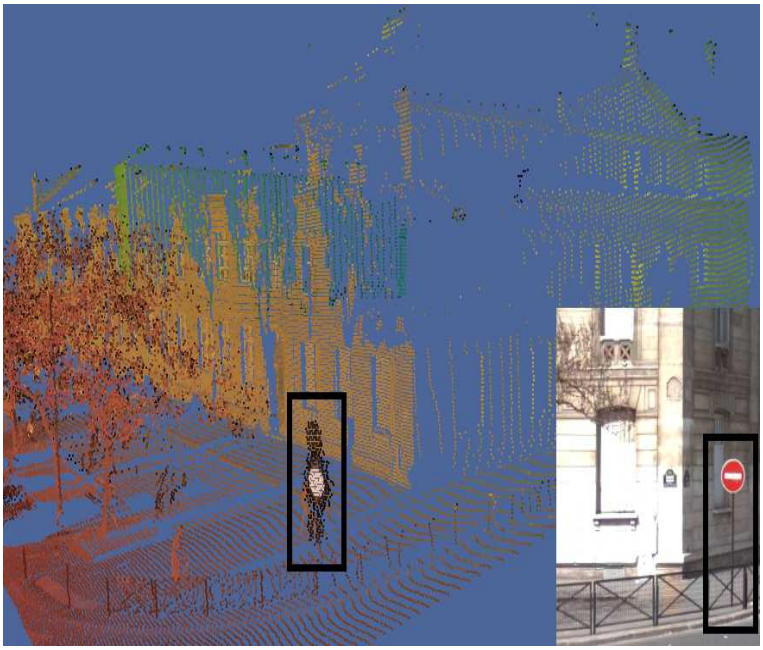


Figure 3.13: 3D points of occluding objects such as barriers and a road sign are included between the vehicle and the facade. The acquired set of points includes 3D points belonging to the three-dimensional physical reality and also non-existing 3D points included by error as it can be observed around the circular panel (aureole of points).

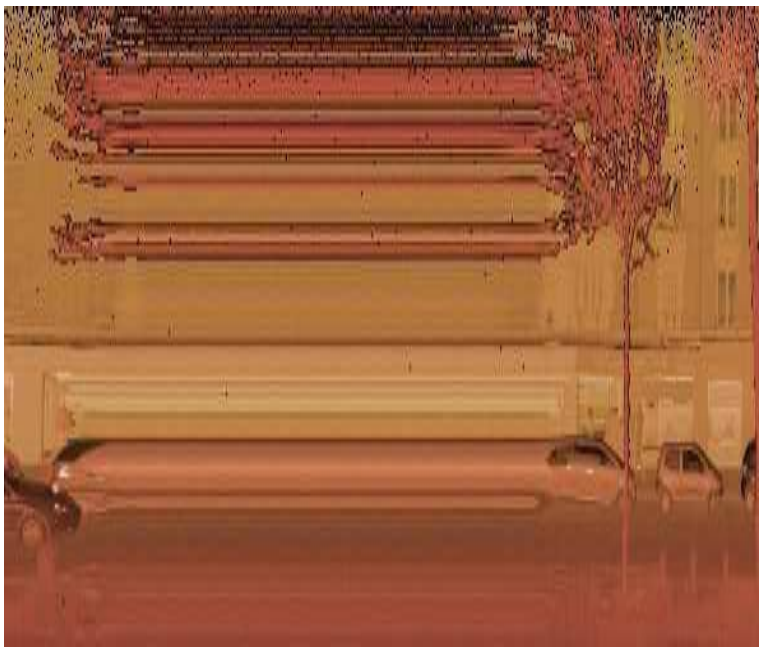


Figure 3.14: Returned intensities of the 2D scans. The size of the image is $\mathcal{N}_c \times \mathcal{N}_r = \mathcal{N}_c \times 201$ where $\mathcal{N}_c$ and $\mathcal{N}_r$ respectively correspond to the number of columns (i.e., number of acquired frame related to the acquisition time) and rows (parametrized frame resolution). The redundancy effect appears when the vehicle temporarily stops. The vehicle and the branches seem stretched.

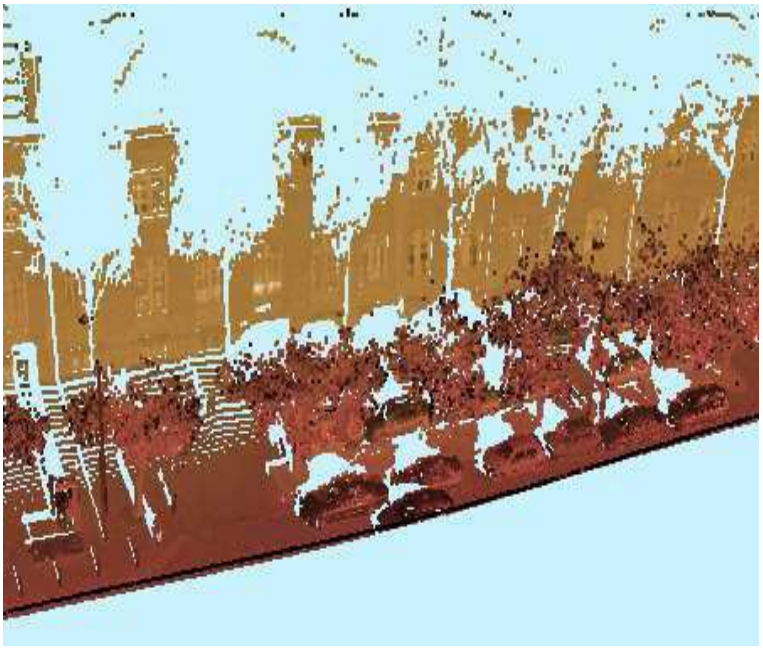
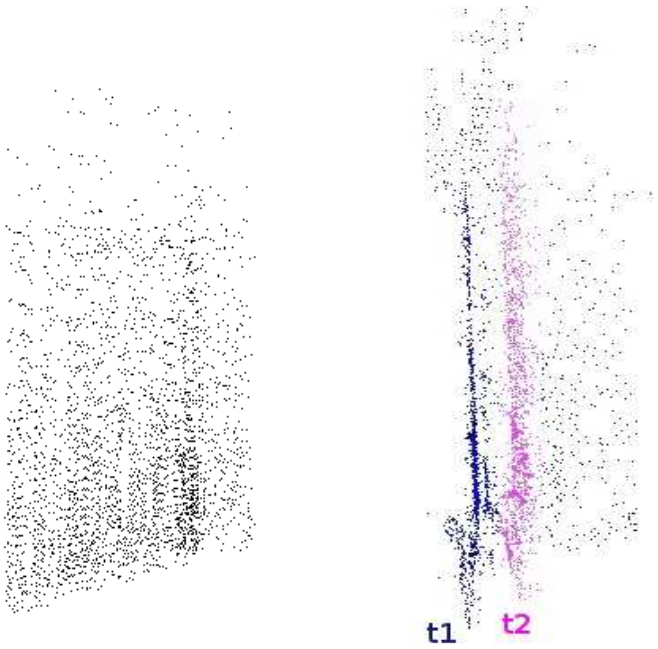


Figure 3.15: A street in the city of Paris (perspective view). The vehicle trajectory is plotted in black. The beam vertically sweeps the scene with an opening of 80°(−20° to 60°with respect to the horizontal). Yellow and red colors respectively denote facade and ground surfaces. The building facade is strongly occluded at the facade top in reason of the foliage of trees and at the facade bottom in reason of parked vehicles.



(a) (b)

Figure 3.16: Effect of 3D misalignments in case of multiple acquisitions of a same facade. (a) Perpective view of a 3D crop of one facade point cloud coming from two facade sweeping with an acquisition interval of a couple of hours. (b) Profil view of the same cloud reveals two overlapped facades.

3.3.4 Static and Mobile Occlusions

Occlusions cause a problem for the complete acquisition of building facades in urban environments. The occlusions could be caused by two categories of objects, static or dynamic created by man-made and natural objects. The raw cloud may suffer from missing data due to the presence of pedestrians (minor effects), street lights, mobile and parked vehicles, vegetation (major effect) and many others objects (see Figure 3.15). Mobile objects cause a thickening of the acquired cloud when the vehicle stops and travels. Certain acquired points model an ephemeral surface and could be considered as erroneous points. Besides, we alas observe that the vegetation is strongly present in urban environment and will often strongly affects the processing to generate detailed facade models in the sense that facades can be half-hidden in case of laser sensor oriented perpendicularly to the vehicle trajectory. Although beyond to our acquisition configuration, we mention that the adding to the MMS of a complementary inclined sensor will be useful to reduce this effect. Also, autumnal acquisitions could be envisaged in order to reduce the effect causes by the foliage of trees.

3.3.5 Visibility and Coverage of Urban Objects

The visibility and coverage of urban objects of interest (e.g.; facades) is highly depending of the employed laser sensor (e.g.; rotative sensor or range laser sensor) and its configuration and location on the MMS vehicle. In our case, we employ a range laser sensor that is oriented perpendicularly to the vehicle trajectory in direction of the facade. In this configuration, the structures such as walls parallel to the laser beam planes are not acquired. Moreover, the occluding objects are only acquired in surfaces that face the vehicle. The back and lateral sides of occluding objects are lowly or not reached by the laser beam; i.e., partial acquisition of occluding objects. Furthermore, if several occluding objects are located between the vehicle and the facades, then the acquired occluding objects can hide others occluding objects in front of the facade. These missing parts are not visible in the laser data in reason of the unidirectional acquisition. Otherwise, we should bear in mind that all these missing parts could be visible in the optical images in reason of the image field of view. Besides, the visibility and coverage of the objects is also limited by the discretisation own to the laser data. This discretisation often affects the contours of the objects such as the facade top boundaries. In addition, it happens in some cases that the range laser sensor can not cover the totality of the facades from the bottom to the top in reason of its limited opening (e.g.; see Figure 3.15). Also, we notice that these effects explain in part the distortions of the image of laser intensities that differ of the scene observed into 3D or into a conventional optical images.

3.3.6 3D Data Misalignment

Across our analysis, we have observed that all sides of one building block can difficultly be acquired with a single passage of the MMS due to the directions of circulation that changes from one street to the next. The merging of several clouds of 3D points will be necessary. However, the filling of the missing gaps remains at a delicate stage. Indeed, two acquisitions of the same facade at two

different moments could present a lag up to a few decimeters in planimetry (see Figure 3.16) and up to meters in altimetry according to the location provided by the positioning systems. Although the topic of registering laser data is beyond the scope of this study, we mention that active researches are conducted in the literature in order to registered terrestrial laser scanning with a robust global map (existing ground truth) such as 2D floor plan (Wang et al., 2010) or DSM (Ridene et al., 2009). For the sake of efficiency of the experimentations undertaken on block of buildings (requiring processing of the totality of sides), we will exploit building facades acquired in a single passage.

3.4 Conclusion

In (Cahalane et al., 2010), the authors presents investigations into the effect of vehicle speed on laser scan lines acquired from Mobile Mapping Systems. Our study as well as analysis results of MMS-based surrounding research works have then been regrouped and provide an overview of this topic. This description allows us to acknowledge the major problems associated with the raw laser data. In this context, the data must be manipulated with much precaution. The 3D data should undergo several preprocessing steps before becoming exploitable. Thus we need a process robust to some outliers and noisy data.

Although still expensive, such Mobile Mapping Systems are booming. It appears that these systems, although sophisticated, employs some similar devices and progressively become standardized Furthermore, we mention that an interesting methodology is proposed in (Yoo et al., 2009) to compare point-cloud data quality from various MMS designs by modifying spatial configuration of laser imaging system. The methodology presented can be used to design and validate new concepts of mobile mapping systems according to the needs in data quality.

In the next chapters, several topics are investigated from the data collected by the MMS.

Transition

The IGN Mobile Mapping System has then been used in order to acquire multi-source dataset at street level. The collected data deals with dense urban streets and include then a huge amount of urban objects that belong to the scene (e.g.; ground, urban posts, cars). Here, the objectives consist of segmenting and classifying the point cloud of the scene into object categories. Then, these stages will foster the initial processing for the generation of models with a high accuracy according to the targeted objects. More particularly, the main interest in our case lies in the localization and the extraction of the facade points (i.e.; Region Of Interests) towards providing suitable features for the succeeding stages that consist of geometric facade modeling. For this reason, the investigations presented in the following chapter deals with the segmentation and the classification of the street laser data.

Chapter 4

Segmenting and Classifying Terrestrial Urban Street Point Cloud

Abstract

This chapter presents techniques for segmenting and classifying urban street facades in raw laser data. It is intended to provide clusters of points classified into georeferenced urban objects. These results will be useful as primary inputs for high-level approaches devoted to the fine geometric modeling of urban scene. Notably, we adapt the Progressive Probabilistic Hough Transform (PPHT) in order to simultaneously retrieve Regions of interest such as the dominant planar facades clusters (e.g.; walls) and the clusters of facade microstructures (e.g. windows, balconies). This approach uses 3D point cloud alone without use of predefined and elaborated 3D model. A cadastral map is minimally used (last stage) to individualize the detected microstructures by dwelling. Experiments show that the proposed method is able to segment street facades and their windows in massive raw street laser data.

Contents

4.1 Introduction and motivation

In parallel to the aerial-based investigations, active researches are also led for the ground-based object recognition (e.g.; urban street object) from terrestrial data acquired at street level by emerging mobile measurement systems. In both of aerial and terrestrial researches, the object recognition approaches are exigent in term of detail, accuracy, and target very various and specific objects.

More precisely, the current work is specifically devoted to the urban object extraction and classification in street raw laser point clouds (Figure 4.1(c)) acquired by the IGN terrestrial Mobile Mapping Systems (MMS) equipped by range laser sensor. Indeed, the approaches that provide georeferenced and identified urban object clusters are of great interest as primary input to the high level geometric features generation and modeling stages (Hernández et al., 2009) (e.g., in 3D city modeling pipelines). High interest lies also to the autonomous robots navigation in the cities, e.g. by localizing clusters of pedestrian obstacles (e.g.; posts) or urban indicators (e.g.; road signs).

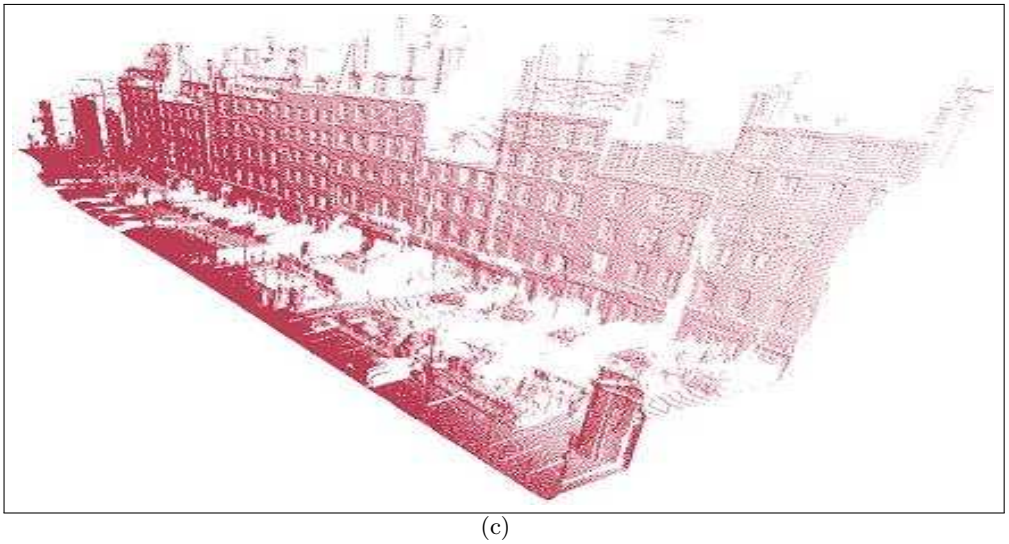
In our case, the interest essentially deals with the extraction of facade clusters of points delimited by dwelling. More precisely, an objective will be the extraction of facade clusters by separating the facade points of the dominant walls and the facade points of the dominant openings such as facade microstructures (e.g.; windows). Then, these clusters of points could be used in more high level processing for the modeling of the facades; namely, the facade plane estimation and delimitation as well as the modeling of openings.

More atypical, a research topic in virtual navigation aims at detecting and blurring windows in street images to respect the private life of citizens. This visibility problem could be treated if the windows points are detected. Indeed, since georeferenced and calibrated images are acquired by the MMS in parallel to the georeferenced laser data, the 3D points can be projected into the associated optical images. Thus, detecting the windows in the raw 3D point cloud will be roughly equivalent to windows segmentation in the optical images. Hence, a blurring mask can be applied to each projected point in the image.

Also, a research topic in sustainable development aims at quantifying the urban heat losses. Then, the automatic detection of facade wall points and windows points could respectively provide information on thermal insulation by dwelling at wall and windows levels by similarly projecting the points in associated acquired thermographic images. The detection of facade dominant walls and microstructures in urban street point cloud will be then of great interest in many applications.

(a)

(b)

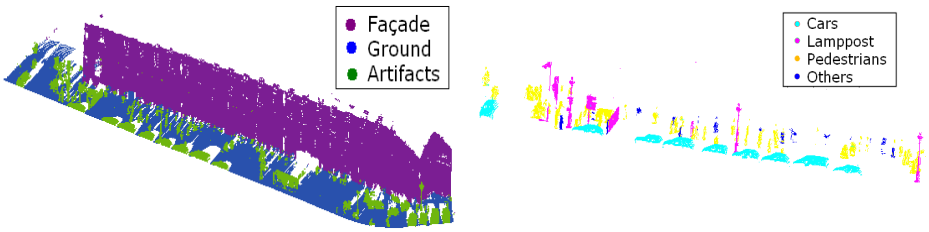


(c)

Figure 4.1: (a) Street scene under study (profile facade views). The area of interest delimited in black. (b) Street side of interest (frontal facade view). (c) Acquired georeferenced raw street point cloud (downtown Paris) corresponding to the scene visualized in Figures 4.1(a) and 4.1(b) (perspective view).

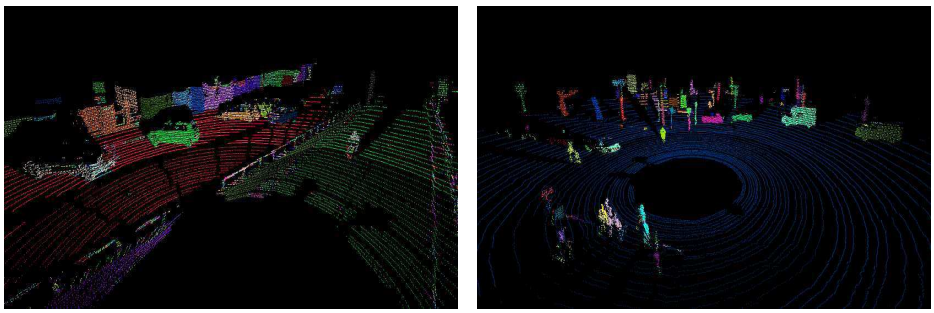
4.2 Related Work

Segmented and classified urban street point cloud acquired from range laser sensor.



(a) Segmented street point cloud. (b) Classified street artifacts.

Segmented urban street point cloud acquired from rotative laser sensor.



(c) Segmented street point cloud. (d) Segmented street point cloud.

Figure 4.2: Segmentation and classification of urban street from terrestrial laser data acquired by MMS equipped with range (upper row) and rotative (lower row) laser sensors. Respectively, Figures are retrieved from (Hernández et al., 2009), (Moosmann et al., 2009).

Foremost, we mention that the aerial-based approach are beyond to the scope of the current study since the terrestrial laser data are known to overpass aerial data in data resolution, accuracy and visibility at urban facade level. Besides, some actual approaches (e.g., (Hernández et al., 2009, Moosmann et al., 2009)) for large scale recognition in urban street point cloud acquired from mobile acquisition are foremost very developed to the street microstructures classification (e.g. lampposts, pedestrians, cars) such as illustrated in Figure 4.2.

Moreover, some works as (e.g., (Ning et al., 2009)) aim at classifying building points acquired from static device (e.g.; tripod sensor). These approaches are valid to finely process specific buildings (patrimonial or individual) but restricted in our case since experimented at small scale from acquisition sensors with narrow field (e.g.; tripod sensors). The applicability of the static device approaches in the context of MMS and massive 3D point cloud processing (i.e., wide scale processing) is not guaranteed and can be unsuccessful due to the street complexity (e.g., street activities, uncontrolled environment, partial occlusions, facade misalignments).

Furthermore, the direct segmentation of urban point cloud from pre-existing

3D model is still hard to generalize at large scale in reason of the model availability, the model quality and the data fusing problems. Besides, we observe that the approach dedicated to the massive segmentation of point clouds are essentially focused to the extraction of street objects. In our case, we search to reach an additional degree of segmentation at facade level in order to separate facade points and points of facade microstructures.

The difficulty of the actual segmentation approach lies amongst others in the splitting of the urban point cloud into clouds of urban objects by respecting a good compromise between several parameters such as runtime, accuracy, level of detail, degree of generalisation. It appears that the approaches undoubtedly privilege certain characteristics to the expense of others. Thus, we can observe approaches that provide a fine segmentation of the 3D point cloud are computationally expensive and are lowly generalizable.

In our case, we search to process the segmentation of urban 3D point cloud at street scale. Also, we would like to be able to reach a satisfying level of detail in the segmentation essentially at facade level. The segmentation of urban point cloud is the essential stage for generating models of urban objects. Indeed, this stage brings key low-level features (sets of points of identified objects) that can be used for the generation of geometrical features (e.g.; 3D segments, planes) and their matching for the generation of 3D models.

4.3 Proposed Approaches

We present two approaches to the segmentation and classification of a raw street point cloud (see Figure 4.1(c)) that initially employ common stages. In Stage 1, the street raw point cloud is denoised and the area is focused by a set of 3D data pre-processing. In Stage 2, an accumulation map is generated by disretizing the area covered by the street point cloud. In Stage 3, a map-based segmentation is carried out by a global filtering and the point cloud is separate into vertical and surface clusters. The remain of the stages is own to each proposed approach and will be detailed in a second time.

4.3.1 3D Data Pre-processing

In this section, we describe three stages which consists of the 3D cloud points pre-processing and denoising, respectively.

4.3.1.1 Partial Filtering of Redundant Points

As we have mentioned earlier, the laser sensor sweeps the building facade even when the vehicle is stopped. Consequently, the acquired raw data may contain many redundant frames due to this continuous acquisition. For this reason, we have computed the point-to-point distances between two consecutive frames [1]. The redundant frames are detected if too close from one to the others and thus removed from the dataset by automatic thresholding. Moreover, this temporal effect is observed in the images of the laser beam intensity and appears as multiple repetitions of similar frames of intensity (i.e., redundant signal). Therefore detecting redundant frames can also be based on differences between

1. A "frame" denotes the sequence of 3D points acquired by given vertical scan.

the returned intensities of two consecutive frames. This step solves only the problem of data redundancy related to the sensor immobility.

4.3.1.2 Region Of Interest (ROI)

The sensor characteristics can be used to crop the 3D points belonging to the facade walls. In our experimentation, an horizontal band has been defined between two horizontal planes. The lower plane passes through the sensor center. The upper plane is shifted by a certain distance that is related to the height of the buildings under study. In this way, we specifically targets a set of 3D points that is essentially composed of facade points and that do not include points of ground.

Although the selected lower plane passes through the sensor center in our case (sensor altitude known), we precise that the height of this plane (horizontal) could also be tuned according to the ground altitude since this one can be simply deduced by measuring manually the laser sensor height. Thus, the visibility interval of the facade band can be more important. Finally, the Region Of Interest (i.e., volume of interest) represented by a 3D bounding box is determined by the georeferenced trajectory of the vehicle and the horizontal band defined above. The 3D points not included in this volume will then be removed from the dataset. This processing has sometimes been used in our study according to the experimentations.

4.3.1.3 Exploiting Alignment of Facade 3D Points

After the preceding filtering steps, the frames have undergone a 3D cropping. The data structure initially controlled by frames is now composed of sequences of 3D points. We can also exploit the fact that in this representation facade points are locally aligned. We can seek facade points which are principally organized linearly.

Thus, the dataset in this sequence is parsed by triplets. The central point of each triplet is kept in the dataset if the triplet is aligned, otherwise it will be removed from the dataset. Therefore, the collinear points of the building facade are kept. Besides, we observe that the 3D points belonging to other linear structures are also kept (e.g., points of tree trunks). We stress the fact that the retained 3D points are collinear but not necessarily vertical. Although very dependant of the structure of the laser data, this processing has the advantage to provide a structured representation of the scene by operating a single parsing of the triplet of points from the dataset.

4.3.2 Accumulation Map Generation (Euclidean)

The verticalness of the street facade walls and other vertical urban objects is exploited. An accumulation 2D map has been generated by projecting the 3D point cloud vertically into an horizontal regular grid, i.e. image of accumulation noted $\mathcal{I}$. The grid step $\Delta_{(u,v)}(5 \times 5cm)$ is chosen finely since the laser data resolution reaches up to hundreds of points per square meter in the horizontal and vertical planes. For each 3D point (X, Y, Z) of the street point cloud, the associated 2D planimetric coordinates noted (u, v) produce a vote in a cell of the map. The score of a cell is noted $\mathcal{S}_{(u,v)}$. Since the relation between 3D

points and 2D cells is kept, an image data $\mathcal{I}$ complementary to the 3D point cloud is generated.

4.3.3 Map-based Segmentation into Vertical and Surface Clusters

A global threshold is applied to the cells of the generated map. The 3D points that vote in cells with a score $\mathcal{S}_{(u,v)} > 1$ are labeled as potential vertical objects. This set of 3D points characterizes in majority the facades and street road microstructures (streetlights, posts, road signs, tree trunks) since their cells have a high vote. If visible, the roof superstructures are also contained into the obtained cluster (e.g. chimneys, antennas). Otherwise, the remaining 3D points (cells with a score $\mathcal{S} = 1$) are labeled as potential non-vertical flat surfaces. The associated cluster constitutes in majority ground and roof plane portions.

From this stage, we propose two strategies for the extraction of facade point clouds in urban street point clouds and each strategy extends a specific approach, map-based and model-less respectively. The first strategy consists in the direct use of the cadastral map (map-based approach) and its description is confined in the section 4.3.4. The second strategy consists in reducing the use of the cadastral map by limiting its usage as providing the rough 2D location of the two endpoints of a facade whose plane is derived from the raw cloud. The approach corresponding to this strategy is essentially model-less and is described in detail from the section 4.3.5.

4.3.4 Approach 1: Direct Urban Point Cloud Segmentation from the Cadastral Map

In our case, we assume that the building facades have dominant vertical planes in urban environment. Here, an existing cadastral based-map is directly employed for providing the location of the facades in planimetry (a set of 2D segment lines). Indeed, the cadastral map is available for many cities and large capitals throughout the world. The planimetric junction (i.e., boundary) between two joined facades having the same dominant support plane is difficult to delineate across a single 3D point cloud of street. Consequently, the cadastral map is useful for delimiting the facades at planimetric level.

Besides, the laser data and the cadastral map are georeferenced in the same coordinate system. However, the cadastral map is issue from the digitalization and the vectorization of paper map modeling approximately the building footprints. These processes are often semi-automatic and some footprints are missing. Hence, the laser data and the cadastral map are more or less misaligned or incomplete. Nevertheless, we assume that the misalignment between laser data and the cadastral map is relatively small and that the great majority of the facades are represented.

For this reason, a ground neighborhood is considered around all the line segments provided by the cadastral map as input. The 3D points having their planimetric coordinates (x, y) included inside the neighborhood are retained; providing thus a point cloud of piecewise planar facades (hypothetic facade clusters). Finally, each hypothetic facade cluster is retained if the number of included 3D points is above a given threshold. The threshold is empirically determined according to the laser data resolution and the length of the line segment. The planimetric neighborhood is selected sufficiently large in order

to guarantee that both of the facade points and their closed microstructures (e.g. windows).

Although this approach is simple, it relies on the assumption of the availability of a 2D cadastral map that is well aligned with the point cloud. Therefore, in practice, this approach cannot be always adopted.

4.3.5 Approach 2: Model-less Urban Point Cloud Segmentation

At this stage, we stress the fact that the input of our approach is given by only a raw 3D point cloud and some prior knowledge about the urban objects. The next stage segments the cluster of vertical objects in three categories by using the adapted PPHT algorithm (Matas et al., 1998); namely, clusters of street road microstructures (posts, street lights, road signs, vegetation as tree trunks), dominant facade walls and facade microstructures (windows, doors, balconies).

4.3.5.1 PPHT-based Segmentation for Getting Clusters of Facade Walls and their Microstructures

The processing deals on the preceding cluster of 3D points that is labeled as vertical objects.

Here, our interest lies in the extraction of linear features such as planar clusters of facade points. The whole of acquired data set deals with the acquisition of the major linear streets. Then the data observation shows that the urban facades are relatively linear at a coarse scale. We recall that the 2D straight lines are equivalent to 3D planes since the facades are assumed to be vertical.

In our field, some techniques allow the detection and extraction of 2D straight lines among a dataset of 2D points. In particular, pioneering pattern recognition techniques such as the Standard Hough Transform or the RANSAC algorithm are well-known for their ability in fitting linearity from a set of 2D points included in binary images. These methods that process raw dataset of points can be categorized as low-level and bottom-up methods.

Besides, in mobile robotics, a high interest lies in the study of algorithms for the extraction of lines from a set of 2D points. Indeed, these algorithms initially are very employed to the localization of robots in indoor environment by the dynamic map generation or the path planning. In the city modeling field, the RANSAC algorithm and variants are foremost very often used to the estimation of planes from 3D points such as in roof plane estimation from aerial data (Bretar et al., 2005, Tarsha-Kurdi et al., 2008) or in facade plane estimation from terrestrial data (Boulaassal et al., 2007). In this last case, although promising on specific urban portions such as individual facades, the generalization of such terrestrial RANSAC-based approaches for the extraction of planar facade cluster at street scale could be studied.

The principle of the RANSAC algorithm is briefly recalled here. The RANSAC algorithm (Fischler et al., 1981) is an iterative algorithm that consists to the estimation of the parameters of a mathematical model from a considered dataset including outliers. The RANSAC acronym corresponds to RANDdom Sample Consensus. This algorithm (non-deterministic) employs the paradigm of hypothesis generation and verification. A random subset of

the original dataset is iteratively selected and verifies a criterion of model-to-data consistency (e.g.; the number of inlier points). The detection probability is progressively increased to each iteration since an unsatisfying model can not be re-pulled on more time. Moreover, the better model is kept from one iteration to the next. Then, the optimization process search to fit the formulated model hypothesis (a set of parameters) to the data. This algorithm is often applied to estimate straight line parameters from sets of points.

Besides, the Standard Hough Transform (SHT) (Hough et al., 1962) is actively applied in image processing and patter recognition for the line detection in images. More precisely, the employed data as input often are binary images or images of contours. The principle of the Hough Transform is briefly recalled here. The linearity of the set of points is analysed in the (ρ,θ) space accumulation that is generated by verifying the constraint Eq.(4.1).

$$\rho = u \cdot \cos\theta + v \cdot \sin\theta \tag{4.1}$$

where ρ is the length of the perpendicular from the line to the origin and θ is the orientation associated to the normal vector expressed in the Cartesian accumulation map. Each point (ρ,θ) is unique if $\theta \in [0, 2\pi]$ and $\rho \geq 0$.

Then, each voting point produces a set of votes in the accumulator space along a sinusoid. The accumulation of votes (incrementation) involves accumulation of points (sinusoid intersections) in case of detected point collinearity. For this reason, the parameter of dominant lines corresponds to the local maxima in the Hough space accumulator.

Although the high linearity of street, the direct use of the SHT algorithm is complex in the sense that the neighborhood facades of the street often have walls with close but different orientations. These geometric characteristics can be observed in the Hough space accumulator (in high resolution) as several close peaks. The segmentation of these peaks into dominant facade walls is a complex task.

To achieve this task, we propose to use a non classic Hough Transform based on a probabilistic and randomized process. More precisely, we use an approach (subfamily of HT) based on the PPHT algorithm (Matas et al., 1998) to detect the planar clusters of points corresponding dominant facade walls. This algorithm propose a compromise between the SHT and RANSAC techniques in the sense that the over-detection own to the conventional Hough transform is relatively overcomes and the extraction of linearities remains progressive as for a recursive RANSAC.

The PPHT vectorization algorithm has been applied in part for its low computational cost and its partial data exploration. This algorithm is a variant of the conventionnal algorithm of the Hough Transform (HT). The PPHT algorithm Alg.(2) such as employed with binary image data is reminded below from (Matas et al., 1998):

The proposed approach employs a sequential algorithm and a random sampling scheme. In our case, the points are randomly selected one by one. Moreover, the probability to detect a potential line rapidly enhances in the sense that each selected point in the input dataset is progressively memorized in the Hough space. Thus, just a subset of points is employed to detect the lines. This voting strategy alleviates thus processing by reducing the number of voting points.

Algorithm 2 The *Progressive Probabilistic Hough Transform.*

1: **Check** the input image, if it is empty then finish.
2: **Update** the accumulator with a single pixel randomly selected from the input image.
3: **Remove** pixel from input image.
4: **Check** if the highest peak in the accumulator that was modified by the new pixel is higher than threshold l. If not then goto 1.
5: **Look** along a corridor specified by the peak in the accumulator, and **find** the longest segment of pixels either continuous or exhibiting a gap not exceeding a given threshold.
6: **Remove** the pixels in the segment from input image.
7: **Unvote** from the accumulator all the pixels from the line that have previously voted.
8: **If** the line segment is longer than the minimum length
9: **Then add** it into the output list.
10: **Goto 1**.

The 3D points are randomly selected one by one. At each selection of a 3D points, a sinusoid is progressively traced in the Hough space until the score of a cell exceeds the preset threshold (preferably a low value). In this condition, it means that a certain number of aligned 3D points having vote in this cell is reached and that these points are implicitly in the same plane. Considering the planarity of facade walls in general, there is a high probability that other 3D points are also aligned along the same plane. Moreover, the main streets are often relatively straight in urban environments. The overpass of the threshold of peak behaves like a starter, which launches the core of entire process.

As we have previously mentioned, the dominant facade walls have various surfaces. Thus, the dominant facade walls are quasi-planar structures. Nevertheless, the set of the 3D points associated with the dominant facade walls commonly has a roughness defined by a threshold of thickness. According to the street facade linearity, several planar clusters of facades can be extracted from one detected line.

The points positioned inside the neighborhood, i.e. inlier points are divided into several segments respecting properties of connectivity between the parsed points. The connectivity is defined according to the minimum width of streets from the a priori knowledge. For each extracted cluster, if the density of points belonging to the neighborhood is sufficient then the cluster is labeled as a facade part. The criterion for stopping the process is met when a preset number of random sampling no longer produces clusters or when the full dataset is extracted (other existing statistical criteria can be used).

Besides, we present the major steps of the PPHT algorithm applied in our case to the laser data to extract clusters of facade wall in street (output 1) are summarized as follows:

Repeat the following steps until the stopping criteria is fulfilled, namely stop when a given iteration number does not produce outputs:
1.By using (4.1) randomly selected points (u, v) vote in Hough space (ρ,θ).
2. If a predefined accumulation score (e.g. up to 500 votes here) is reached in any cell in Hough space then:
(a) Fragment in 3D space the line of parameters (ρ,θ) into facade segments (non-empty segments) by using $\mathcal{I}_{(u,v)}$ density.
(b) Extract the wall points from the facade segments (extremely close, neighboring $\mathcal{N}_1 = \pm 5cm$).
(c) Remove the wall points from the dataset (vertical point cloud).
(d) Reset to zero the whole Hough space accumulator.

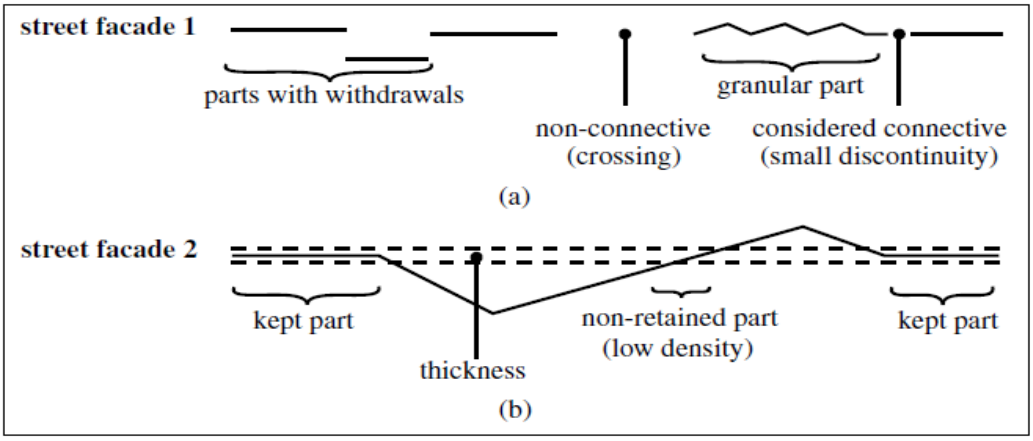


Figure 4.3: Comprehension scheme illustrating general characteristics of street facades in (a) and a step of segment keeping (one pass of the algorithm) in (b).

The earlier mentioned step that consists of fragmenting in 3D space the line of parameters (ρ,θ) into facade segments (non-empty segments) by using $\mathcal{I}_{(u,v)}$ density is described by the comprehension scheme illustrated in Figure 4.3.

Additional steps to conjointly reveal the facade microstructures category (output 2):
3. Then extract facade features as follows:
(e) Re-use the computed facade segments (step 2(a)) to extract points of facade microstructures in neighboring $\mathcal{N}_2 = \pm 1m$.
(f) Remove in addition to (step 2(c)) the points of facade microstructures from the dataset.
(g) Stopping criteria reached or return to the step 1..

Comments:

- *The facade microstructures are assumed to be in recess to the dominant facade planes. The facade microstructures are originally and fastly detected since their detection step is integrated into the PPHT algorithm.*

- *The probability of peak detection in the Hough accumulator is increased since the clusters of facade microstructures are progressively removed from the dataset.*
- *The locations of the vertical structures that are detected by using the generated accumulation map could be used as key points (interest points) for a registration in planimetry of the laser data acquired by the vehicle.*

4.3.5.2 Classification Refinements by Fusing of PPHT Outputs with High Density Segmented Points

The whole point cloud and the cells of the map $\mathcal{I}_{(u,v)}$ initially generated in Stage 1 are reconsidered. A global threshold is tuned to high values in order to reveal major points of street walls (vertical bands empirically appear). Within these bands, the removed facade points indicate potential windows regions. Besides, facade and windows clusters issue from the automatic Stage 3 are sometimes over/under segmented in reasons of facade missing parts, then the accumulation in vertical bands can rectify in part these effects by adjusting wall and windows point segmentation as follows: *i*) Extract 3D points with very high accumulation in $\mathcal{I}_{(u,v)}$ (essentially walls and streetlights). *ii*) Keep the subset of wall points included in neighboring $\mathcal{N}_2$ of facade segments computed in Stage 3 (retained wall points Stage 3 are only included in a small neighboring $\mathcal{N}_1$). *iii*) For each wall point (step *i*), compare its label in the Stage 3 outputs (walls and windows clusters). If the point is labelled as windows, then re-label as walls. If it is already labelled as walls, then do not achieve modification. If it is not included, then add the wall point to the wall cluster. According to the needs, this optional stage that passes the approach into semi-automatic can bring more accuracy to the expense of the runtime.

4.3.5.3 Cadastral Map-based Segmentation for Facade Individualization

The cadastral map is a set of 2D segments that delineates the parcels of buildings. A vectorized cadastral map is combined to the terrestrial laser data. This map is used for the clusters individualization of street facade walls and facade microstructures by dwelling. For each segment, the facade and windows points that are respectively located inside the orthogonal band and a defined neighborhood are kept. Then, 3D dominant facade structures can be estimated from the wall points. It could be also envisaged the use of the retained segments for getting the endpoints that delimit the facades. Furthermore, windows geometry can be studied to recognize facade variety and semantic information can be labelled to the facade model (e.g.; address database).

4.4 Experimental Results and Performance Study

In this section, we present the dataset employed as input as well as the evaluations and the results obtained by the proposed approaches for the segmentation and classification of urban street point clouds. Also, we perform several experiments in order to evaluate the robustness and the accuracy of our laser-based approach. These evaluations demonstrate the high potential of our approach.

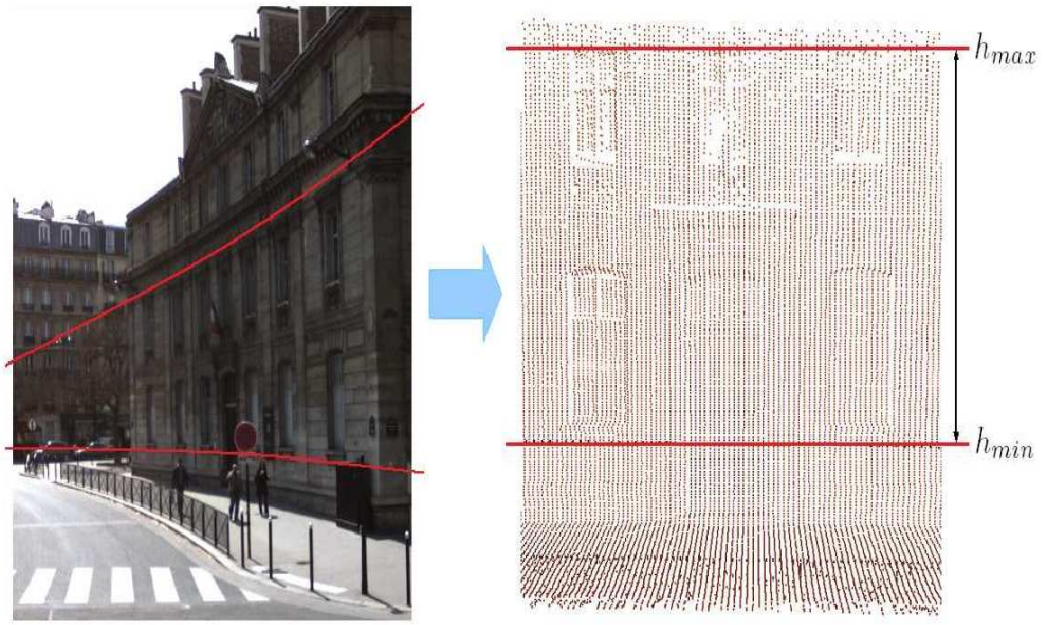


Figure 4.4: Acquisition of the 3D point cloud using the 2D laser sensor. The left image shows the delimitations of the point cloud induced in planimetry by the vehicle trajectory (i.e., vehicle displacements). The right image (frames) shows a selected band (without occlusions) that delimits the point cloud in altimetry. The whole of these delimitations directly form a specific Region Of Interest (ROI) for the facade point extraction.

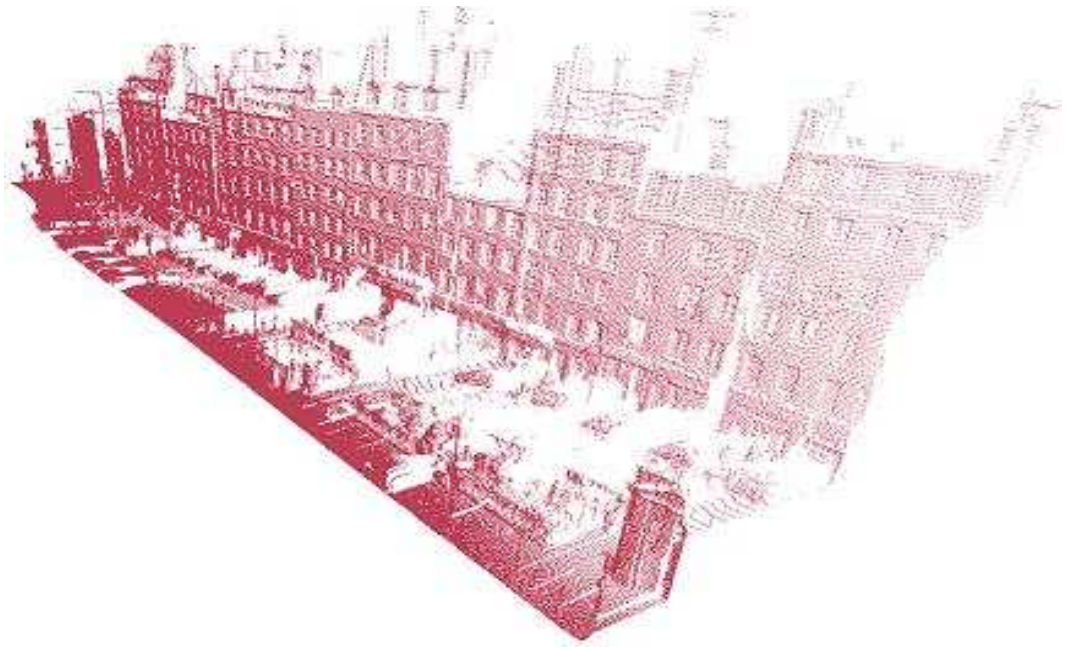


Figure 4.5: Acquisition of raw street point cloud (downtown Paris) corresponding to a complete scene under study. The stake lies in the segmentation and classification at street scale of a dense urban point cloud into clusters of urban objects. The observed scene includes several occlusions at facade level.

As mentioned earlier, the acquisition has been achieved in realistic condition of dense urban environments. In reason of the huge amount of 3D points, our initial investigations have dealt with the processing of subset of 3D facade points corresponding to building facade sides. A sample of non-occluded facade under study is illustrated in Figure 4.4. The Region Of Interest (ROI) has been selected according to the pre-processing presented in section 4.3.1.2. This pre-processing appears efficient for the extraction of facade points in case of non-occlusion. Then, the facade plane associated to this set of points could be estimated. However, the extracted set of facade points is partial and the whole of facade points will be then necessary for the facade delimitation and modeling.

Besides, the acquisition area has been restricted in this study to the major urban street of the 12^{th} district of the city of Paris. Although a huge amount of points has been collected at facade level, it appears that any building block (the whole of sides) has been acquired in one single sweeping in reason of constraints due to the road circulation sense (i.e., road signalization) that changes from one street to the next and also since some building sides are located in areas inaccessible to the vehicle.

A few sets of building blocks could be obtained by grouping several independent street acquisition–acquisition achieved in different time intervals. However, the reliability of this processing is not guaranteed in reason of geometric registration problems that can arise from one acquired dataset to the next (e.g.; (Ridene et al., 2009)). Since this problem has previously been assumed beyond to the scope of the current study, we have not adopted this solution. For this reason, our study have dealt with experimentations and processing of urban street portions such as shown in Figure 4.5.

4.4.1 Experimental setup

We remind that the acquisition is achieved *On-Drive* in realistic condition of circulation (e.g.; street activity, uncontrolled environment). The acquisition system allows us to carry out 10000 measurements per second and the beam vertically sweeps with an opening of $80°(-20°$ to $60°$with respect to the horizontal). The angular accuracy of the beam is equal to $0.01°$. More specifically, the nominal accuracy of laser-based measurements is approximately $3cm$ at $150m$. In this study, the angular sampling was configured to 201 points by frame. The vehicle is also equipped by a Global Positioning System (GPS), a Inertial Measurement Unit (IMU) and an odometer that are combined and related to the sensor in order to provide the direct data georeferencing.

4.4.2 Performance of Major Segmentation Stages

Before analysing the segmentation stages in details, we illustrate final results of segmentation produced by using the two proposed approaches, namely the direct segmentation by using cadastral maps and the model-less segmentation, see Figures 4.6 and 4.7 respectively. Differences between these results are explained in more details in next sections.

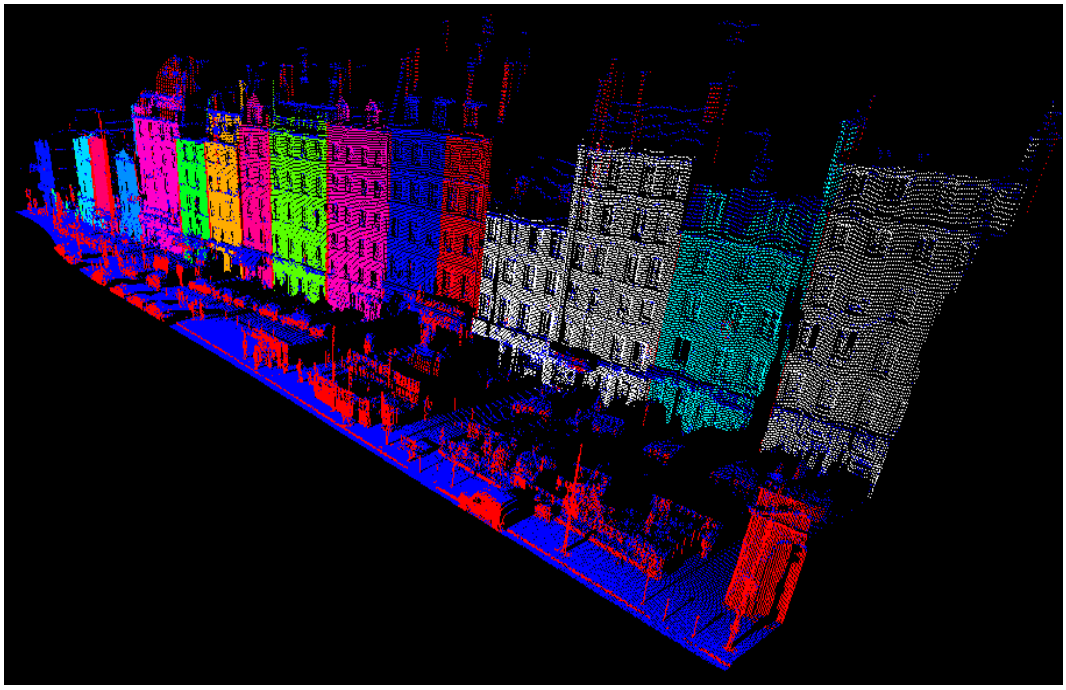


Figure 4.6: Approach 1: Resulting segmentation of the street point cloud by using the accumulation map and the cadastral map.

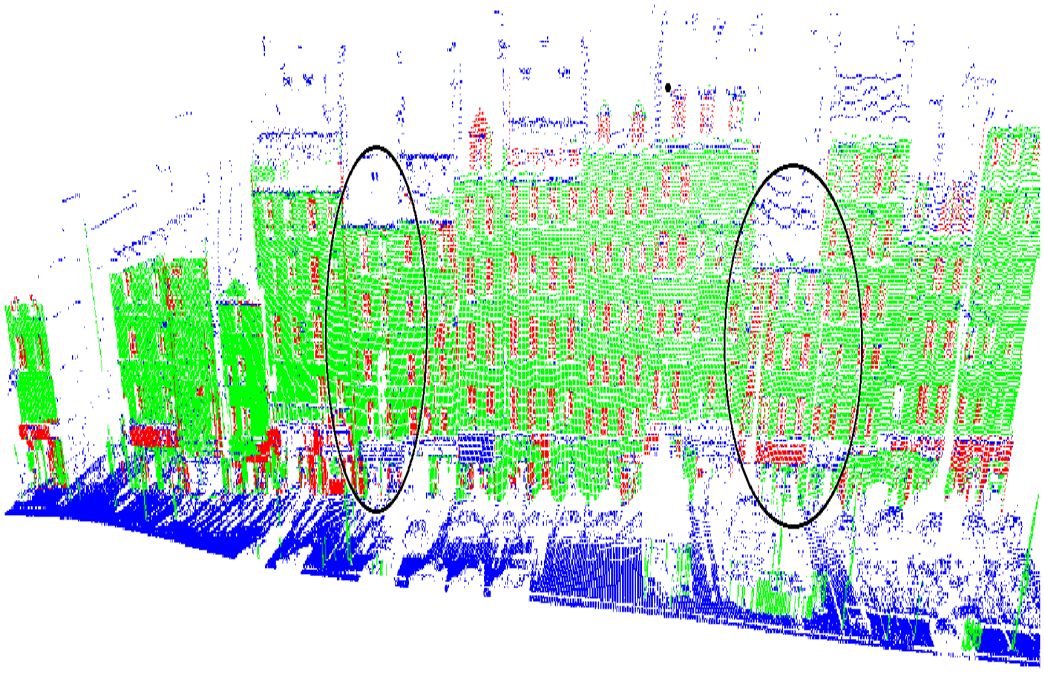


Figure 4.7: Approach 2: Resulting model-less segmented street point cloud by using the accumulation map, PPHT algorithm and map-based refinements. The oval areas show sensitive regions well classified inspite the presence of small occlusions in front of facades.

4.4.2.1 Segmentation Results of a Basis Common to the Two Approaches

In Figure 4.8, we visualize the original raw 3D point cloud acquired by the MMS as well as a set of intermediary segmentation stages common to the two proposed methodologies.

More precisely, the original raw 3D point cloud that is observed in Figure 4.8(a) deals with a typical Parisian street. This street includes then a huge amount of urban structures. Here, the field of the sensor covers the facade from the street ground to the roofs in the sense that the vehicle was sufficiently far of the facades. For this reason, the facades are covered in their totality.

Most of all, an accumulation map that recovers the corresponding area of this raw 3D point cloud has been generated by mapping the 3D points onto a 2D discretized plane. This accumulation map is produced by the votes of the totality of 3D points. The generated accumulation 2D map can be observed from a Nadir point of view in Figure 4.8(b).

Then, a heat map representation is shown. The surface points appear in blue. On the other hand, the dominant vertical structures (high accumulation) appear in black on the map ("hyper-points") and they correspond essentially to the facades and small vertical urban objects.

We observe in Figure 4.8(b) that the set of facades that seems coplanar in Figure 4.8(a) is in reality composed of a set of planes (see segment lines) that are more or less in withdrawal from the one to the others. Besides, we observe at the top part of Figure 4.8(b) that planes perpendicular to the targeted facade planes can sometimes appear.

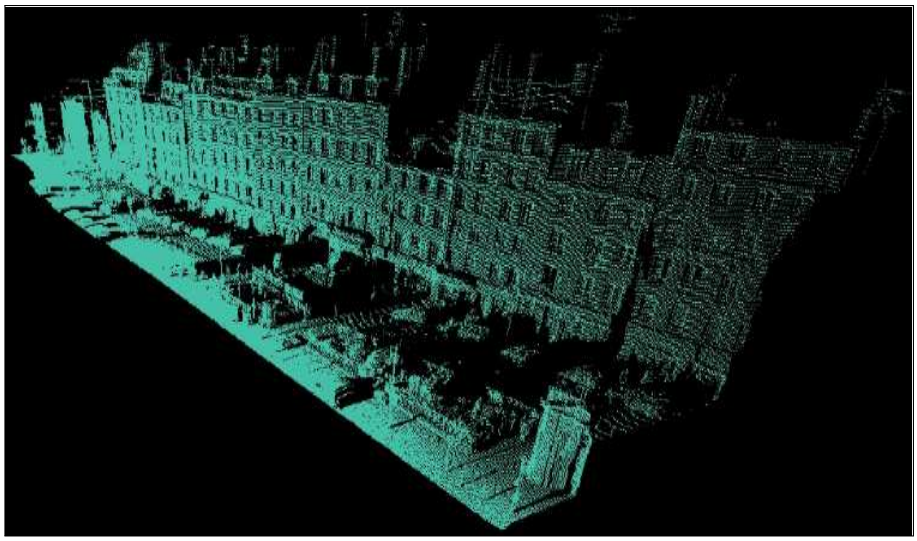
This phenomenon is due to the fact that the facades can have various heights (e.g., see Figure 4.8(a)). In the case of rectilinear building, this phenomenon can then appear when the trajectory of the vehicle is not perfectly parallel to the targeted facade planes (see curved line at the bottom of Figure 4.8(b)).

The small vertical object can correspond to urban posts, street lights, tree trunks, roof antennas, chimneys. More atypically, we remark that parked vehicles are detected by accumulation of points at the level of the vehicle doors.

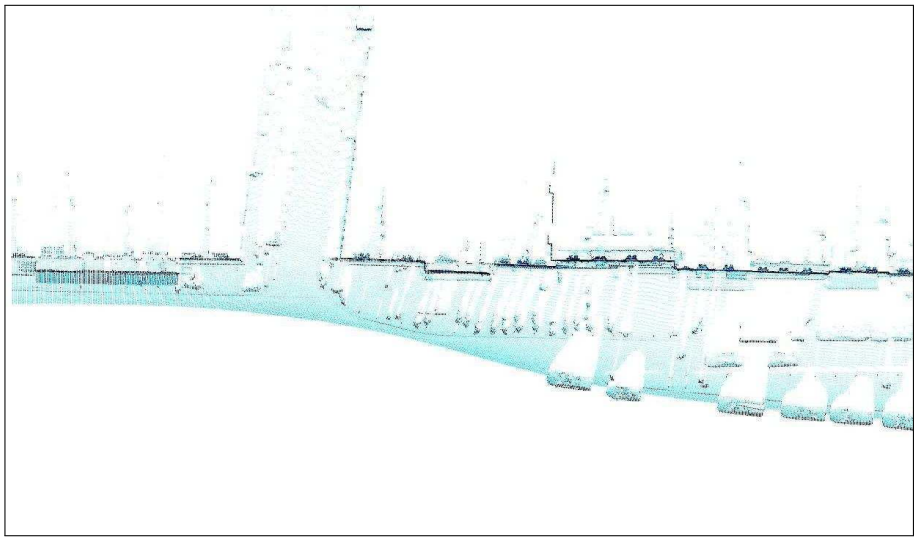
These objects are more visible in the corresponding 3D point cloud shown in Figure 4.8(c). A global threshold has been applied to the cell score values of the accumulation map. We observe that the map is sufficiently discretized in the sense that the street structures are clearly separated and identified in the scene.

Consequently, it demonstrates here that a good compromise has been found between the density and the repartition of points that are provided by the laser acquisition (1 point each $5cm$ to $10cm$ in planimetry and altimetry) and the 2D accumulation map discretization (square cells of $5cm \times 5cm$).

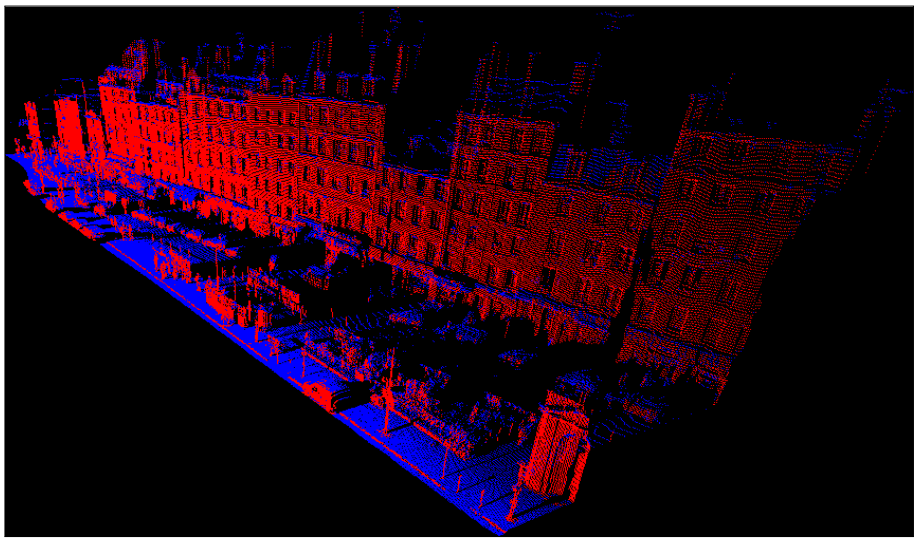
These stages offer an interesting flexibility in order to handle the urban street point cloud segmentation in the sense that only low-level constraints have been employed (spatial distribution of points, object verticality).



(a) Original raw 3D point cloud of a Parisian street including several urban structures.



(b) Generated accumulation map by voting of the totality of 3D points (Nadir view).



(c) Applying a global threshold to the cell score values of the accumulation map.

Figure 4.8: Intermediary segmentation stages common to the proposed approaches 1 and 2.

4.4.2.2 Results of Approach 1 for Segmenting the Street Point Cloud (Map-based)

The section presents results of street point cloud segmentation by using the generated accumulation map and the cadastral map. A global threshold of the map has been carried out in order to separate the vertical street structures and the surface non-vertical structure (see Figure 4.9(a)).

Besides, the national standard cadastral map of the area under study has been employed. The used map is currently a reference used for many applications in urban planning and in computational building modeling.

This map provides a set of 2D building segments that corresponds to the building parcels. The included building parcels are relatively synthetic in the sense that the whole of the residential facades are represented by single facade segments even in case of detailed facades (e.g.; facade with fine overhangs of walls).

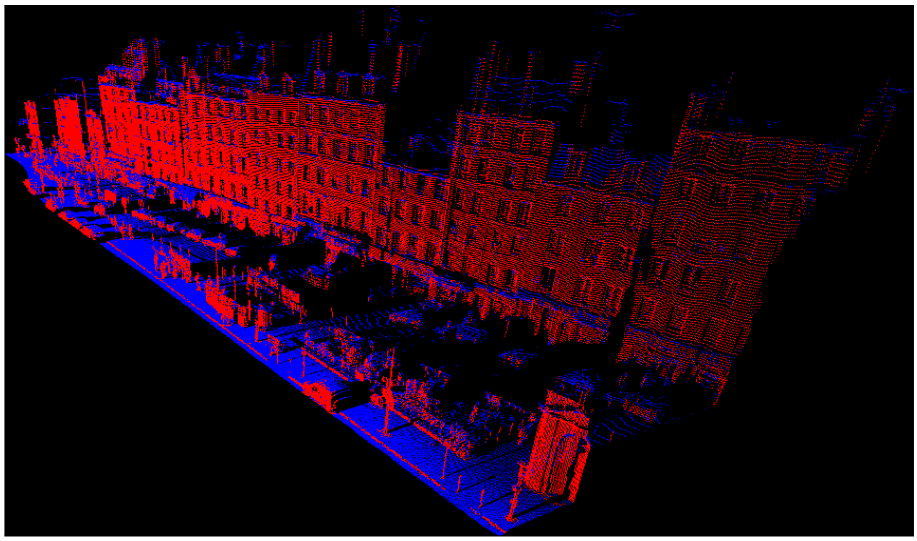
Generally, it is observed that this kind of map is heterogeneous in accuracy (depending on the surveyor proficiency) but often respects a bounded inaccuracy interval of some decimetres according to the reality.

Then, a ground neighborhood is considered around all the line segments provided by the cadastral map as input (see Figure 4.9(b)). The 3D points of the vertical street structures that have their planimetric coordinates (x, y) included inside the neighborhood are retained; providing thus a the facade point cloud for each dwelling.

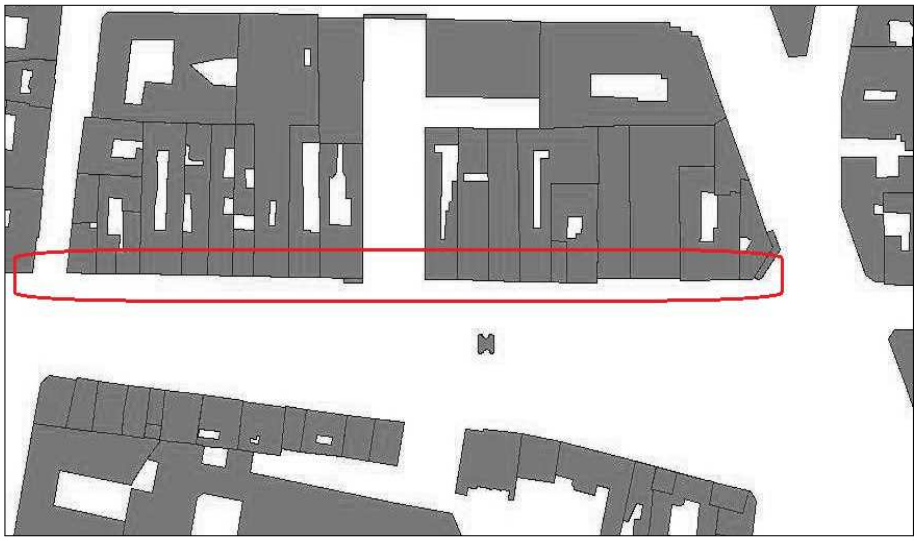
Consequently, the facade segments that belong to the cadastral map (retained facade supports delineated in Figure 4.9(b)) are employed in order to crop the facade points. The planimetric neighborhood around the facade segments is selected sufficiently large in order to guarantee that both of the facade points and their closed microstructures (e.g., windows) are extracted.

Figure 4.9(c) illustrates the raw street point cloud that is essentially segmented into points of ground, street superstructures (street light, urban posts, street artifacts) and individualized facades. As it can be seen, this approach efficiently provides a segmentation of the urban scene and a satisfying classification.

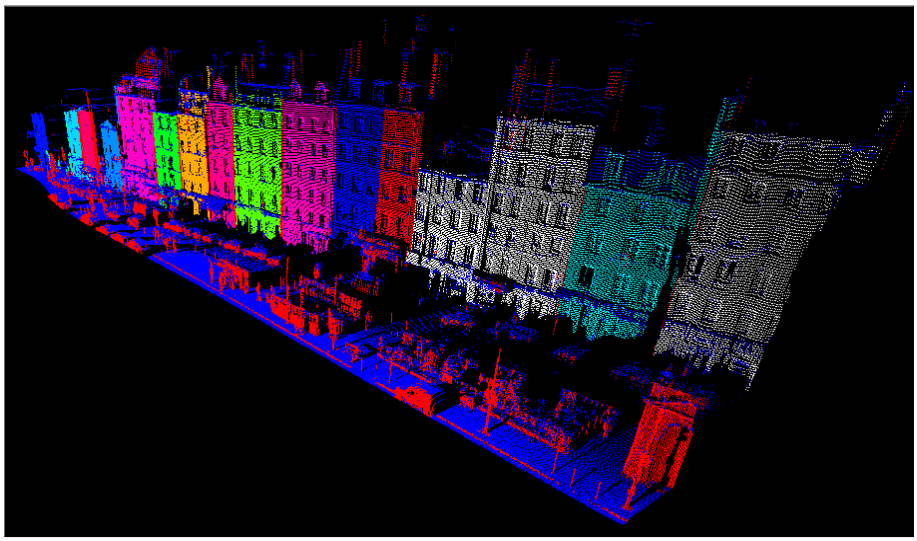
Besides, since the planimetric neighborhood around the facade segments is initially selected sufficiently large, the facade points includes points of walls and points of openings. Consequently, further processing will be required for enhancing the level of details in the segmentation and classification stages. The feasibility in the exploitation of cadastral map segments for the cropping of individualized facade clusters at wide scale has been demonstrated.



(a) Segmented vertical objects are represented in red.



(b) Set of segments belonging to the georeferenced standard French cadastral map.



(c) Resulting segmented street point cloud.

Figure 4.9: Intermediary segmentation stages complementary to the approach 1.

4.4.2.3 Results of Approach 2 for Segmenting the Street Point Cloud (Model-less)

In Figure 4.10, we analyse the characteristics of the straight urban streets that are commonly called portions of linear in the mapping field. The Figure 4.10(a) illustrates a frontal perspective view of a portion of urban laser data associated to our area under study. At first sights, the facades seem at this scale relatively aligned the one to the others and give the illusion that their associated planes are coplanar.

However, the accumulation map of this street shows that in reality, the facade are often slightly recessed (withdrawals) the one to the others (see Figure 4.10(b)). For this reason, the use of a conventional Hough transform in the detection of the facade planes will be difficult to apply in the sense that the detection and the separation of the local maxima in the Hough space will be complex for neighbouring facades that have different but very close locations and orientations. Consequently, we have operated to a Hough variants that is the Progressive Probabilistic Hough Transform.

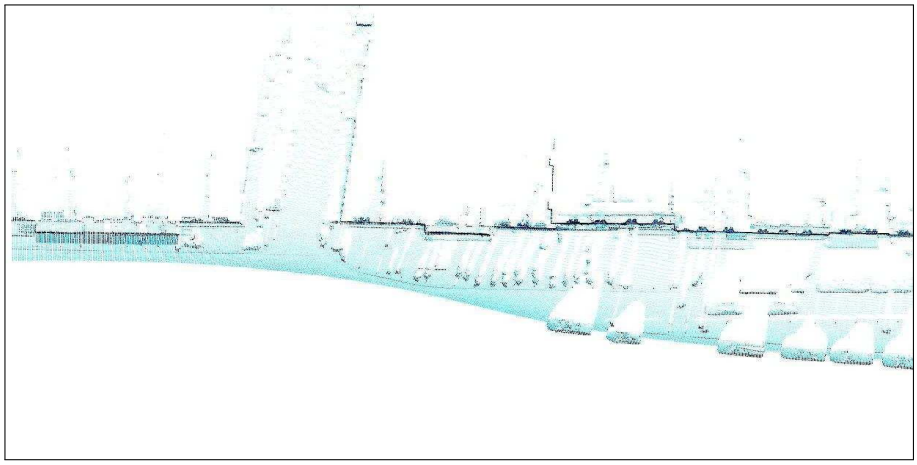
As it can be observed in Figure 4.10(c), the PPHT algorithm provides satisfying results for the extraction of coplanar point ROI at facade level. By using this technique, the clusters of coplanar points are progressively extracted of the data and the accumulation of close peaks in the Hough space are avoided. At this stage, no cadastral map has been employed. Besides, the bounding box associated to the extracted facade point ROI provides a pseudo 3D facade meshes that appear promising towards further investigations for the full model-less urban facade mapping.

Besides, we remark that the pseudo 3D meshes of facades are sometimes slightly overlapped. This effect is caused by the facades that are more or less thick according to their roughness such as previously mentioned. Nonetheless, the retained solution consists of keeping the facade wall points of each ROI and then regroups them in order to constitute a global point cloud of facade walls such as shown in Figure 4.11(b).

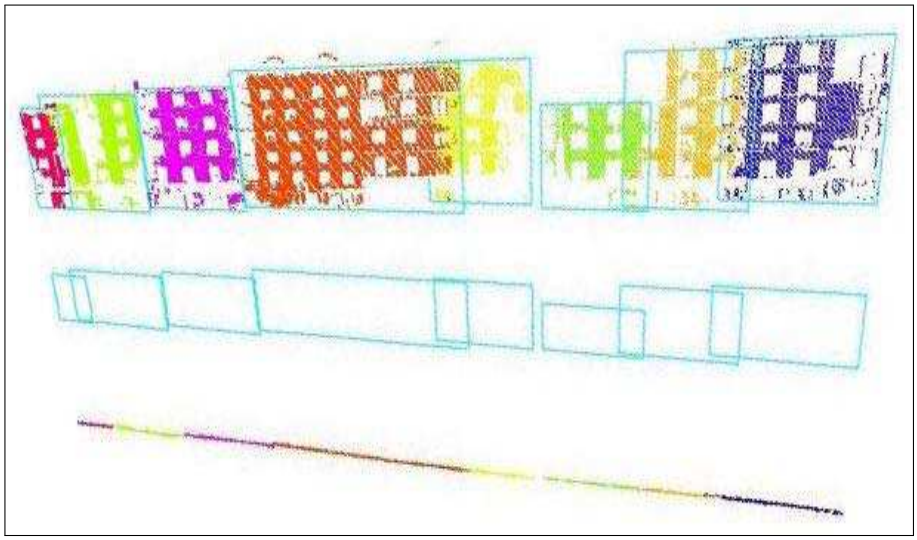
More precisely, the proposed approach employs the generation of the accumulation map and a global threshold in order to extract the vertical cluster of points (see Figure 4.11(a)). Then, the use of the PPHT algorithm can be twofold: *i*) On the one hand, this algorithm can be used for the street point cloud segmentation by extracting the planar cluster of points of the facades along the street linear features such as walls (Figure 4.11(b)). *ii*) the residuals that corresponds to the complementary of the PPHT extraction can be exploited to reveal the street microstructures. In other words, the upshot is that the 3D points that remain after the extraction of the facade wall points are the windows points (close of the facade walls) as well as the points of other vertical structures located in the scene (see Figure 4.11(c)).

For this reason, the PPHT algorithm has been completed here with additional stages that consists of extracting the facade wall points that are located along the linearities as well as extracting for each linearities the set of microstructures that is included in the close neighbouring of the facades (see Step 3 of the section 4.3.5.1).

(a) Portion of urban laser data of the associated area under study (Frontal perspective view).

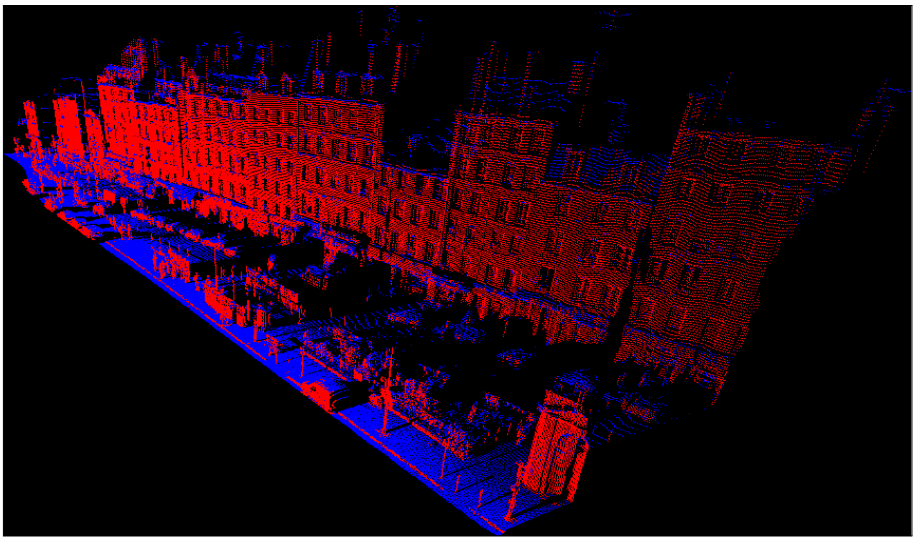


(b) Generated accumulation map by voting of the totality of 3D points (Nadir view).

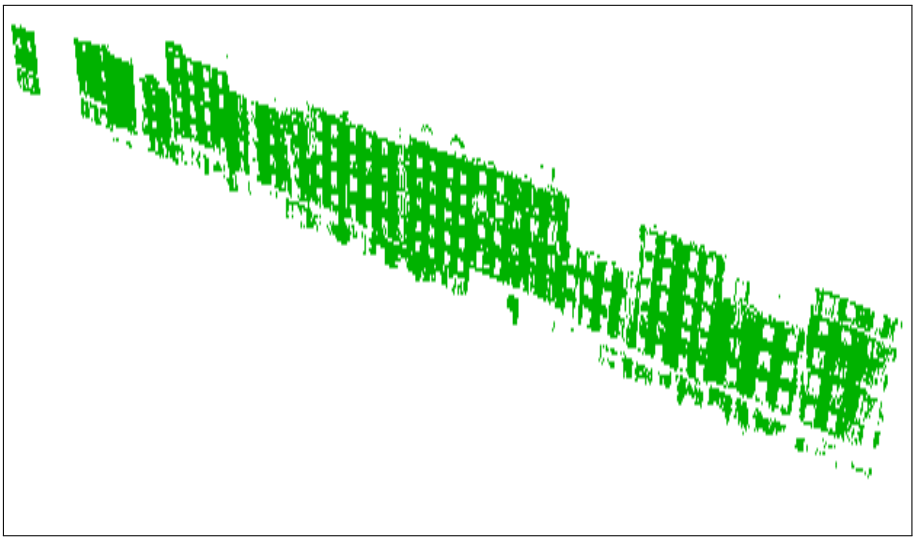


(c) PPHT-based extracted regions that include sets of coplanar points, i.e. Region Of Interests at facade level. (perspective view). From the top to the bottom, we observe: i) regions with 3D points, ii) regions alone (3D meshes) and iii) footprints of respective regions of facade points.

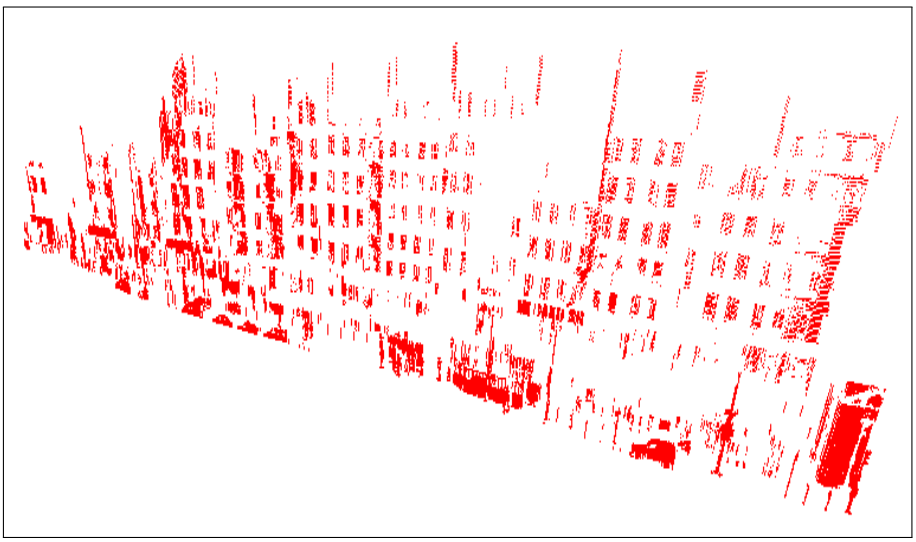
Figure 4.10: Results of facade point ROI extraction in the approach 2 by exploiting the verticality of the urban facades as well as the linearity by pieces of the dominant facade walls.



(a) Segmented vertical objects are represented in red.



(b) Extraction of facade point cloud by using the PPHT algorithm. This cloud is composed of planar clusters, i.e. piecewise planar.



(c) Complementary of the extraction after the use of the PPHT algorithm.

Figure 4.11: Intermediary segmentation stages complementary to the approach 2.

Hence, the level of detail in the segmentation and classification of points can be enhanced. The PPHT algorithm is employed for roughly extracting linear features (coplanar facade wall points) as well as for extracting small microstructures (linear or non-linear) that essentially corresponds to points of openings such as windows or balconies to name only a few.

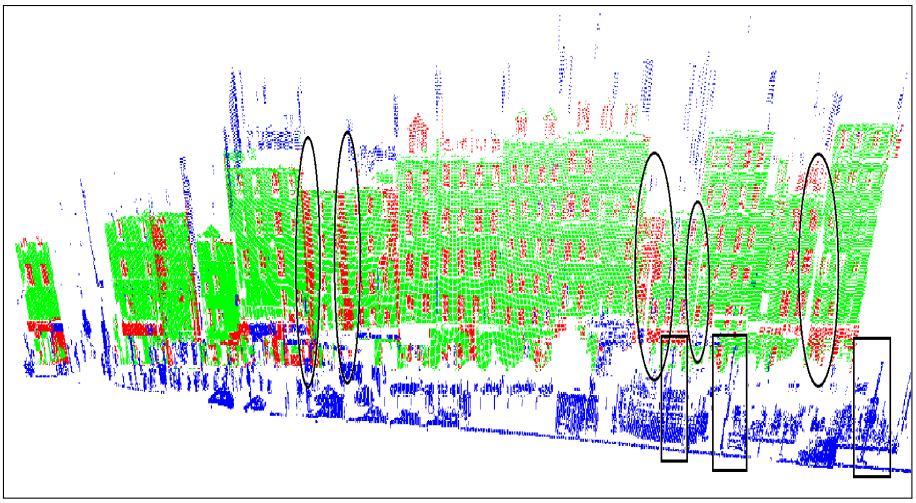
Here, the set of points of the openings are also regrouped into a single point cloud that constitutes the point cloud of facade microostructures. The remaining vertical points can be considered as a point cloud of street structures. It includes for example points of urban posts, street lights or vehicle parked. This segmentation results can be observed in Figure 4.12(a).

Moreover, the surface points have also initially been extracted by thresholding the accumulation map. This set of points essentially corresponds to the points of ground but can also includes points of building roof portions according to the sensor visibility. This set of surface points is illustrated in Figure 4.12(b). At this stage, we emphasis that the street point cloud is segmented and classified into four dominant categories without the use of a predefined and elaborated 3D urban models.

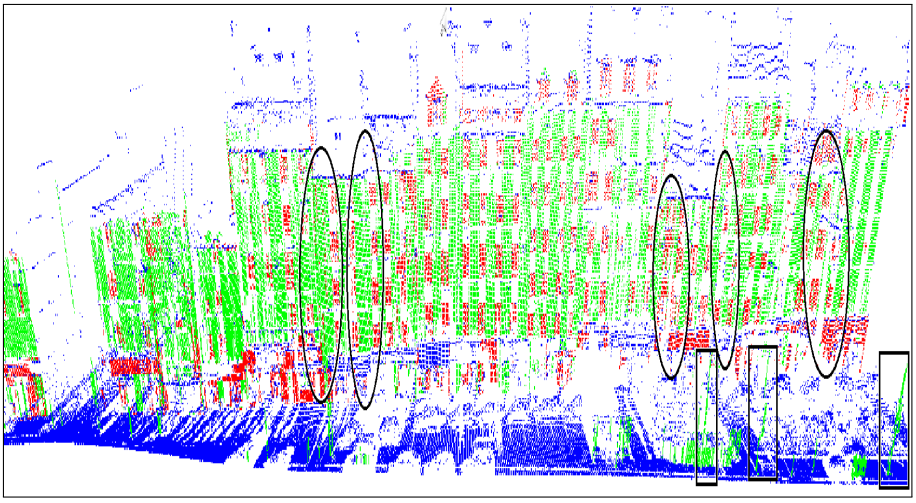
Also, further segmentation could be carried out by using the retained facade clusters for the separation of the points in front of the facades of the points in back of the facades. In this way, the surface points will be divided into two more categories, namely points of roof surfaces and points of ground surfaces. Furthermore, the street vertical structures could be also separated into two categories, namely points of vertical ground-based structures (e.g., street lights, cars) and points of vertical roof structures (e.g., antennas, chimneys).

Besides, the extraction of coplanar points of facade walls can sometimes be slightly incomplete (see oval areas in Figure 4.12(a)) in reason of missing parts due to facade occlusions (see rectangular areas) or in reason of biased extraction in sensitive facade parts (facade discontinuities, withdrawals, surfaces non planar). Nevertheless, the extraction of the majority of facade wall points should be sufficient in facade modeling for the final facade plane estimation in the sense that only three points by clusters will be necessary for parametrizing the corresponding facade planes. However, it is worth mentioning that the exhaustiveness in the extraction of coplanar facade walls will be interesting for the delimitation of the facade planes at the level of the boundaries of openings (e.g., windows, doors).

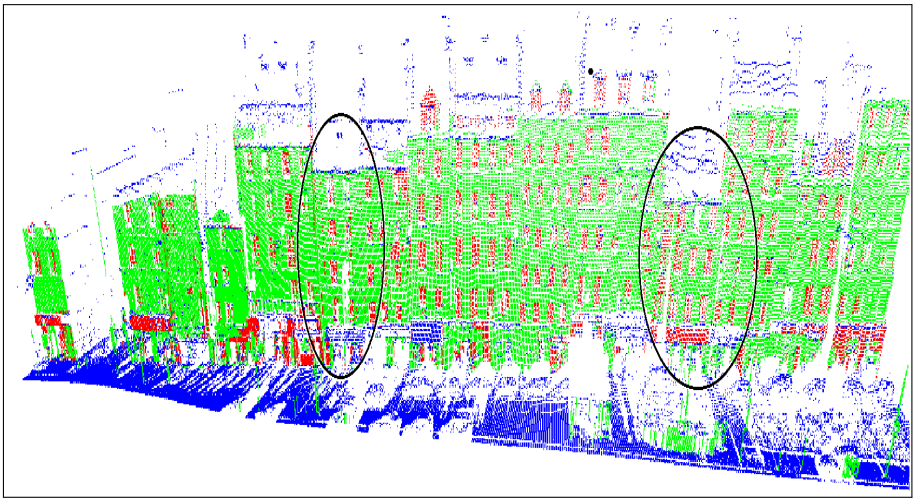
Moreover, the accumulation map initially generated can be reemployed in order to recover facade points that have been missed by the PPHT extraction process as well as for slightly rectifying the detection of points at the level of sensitive facade parts. The threshold of this accumulation map can be visually tuned in order to retain major facade wall points. More precisely, Figure 4.12(a) shows the joint extraction of walls (in green) and windows (in red) provided by using the PPHT algorithm. Minority of points are ill-labeled essentially due to facade missing parts (black ovals) caused by street occlusions (black rectangles). These effects are in part solved by exploiting the accumulation in vertical bands proposed (see Figure 4.12(b), $\mathcal{S} > 20$). Finally, the Figure 4.12(c) illustrates the merging of the ouputs of the PPHT algorithm and the accumulation in vertical bands. In this way, the segmentation between the points of facades and the points of facade microstructures can be refined.



(a) Extraction of the facade and microstructures point clouds by using the PPHT algorithm.

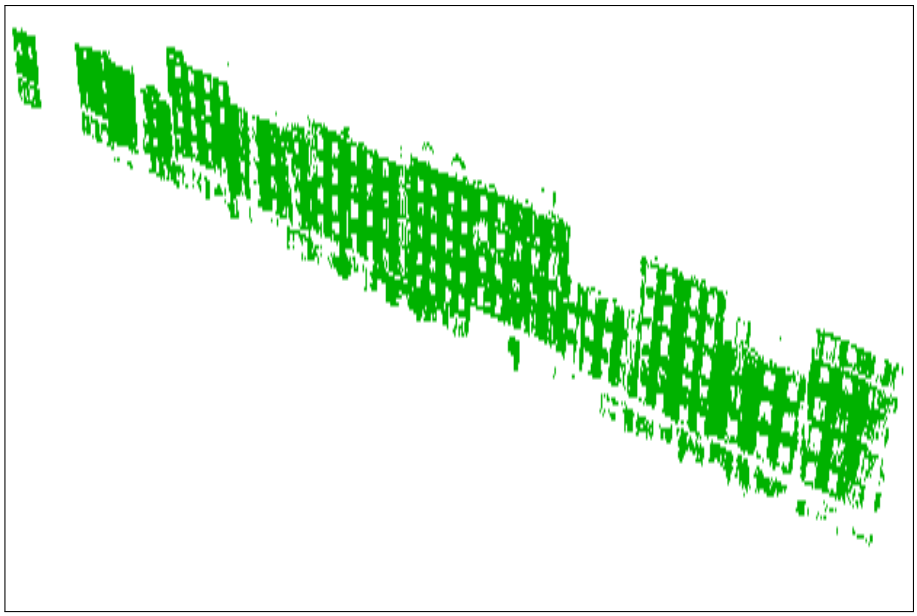


(b) facade walls and windows by exploiting the generated accumulation map (tuned empirically).

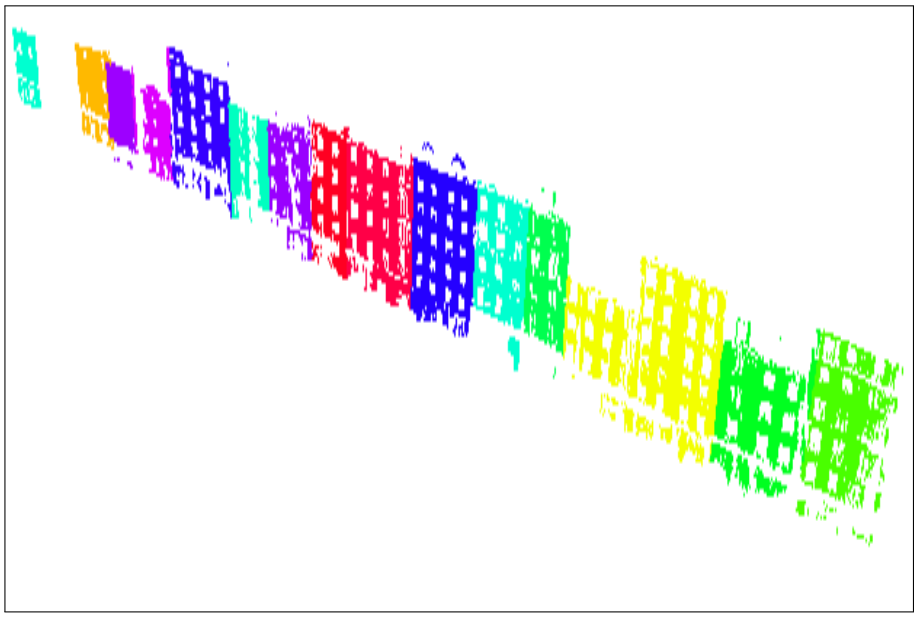


(c) Combining the PPHT and wall points adjustments.

Figure 4.12: Refining the segmentation of point cloud associated to the facades and their microstructures. Stages complementary to the approach 2.



(a) Extraction of facade point cloud by using the PPHT algorithm. This cloud is composed of planar clusters, i.e. piecewise planar.



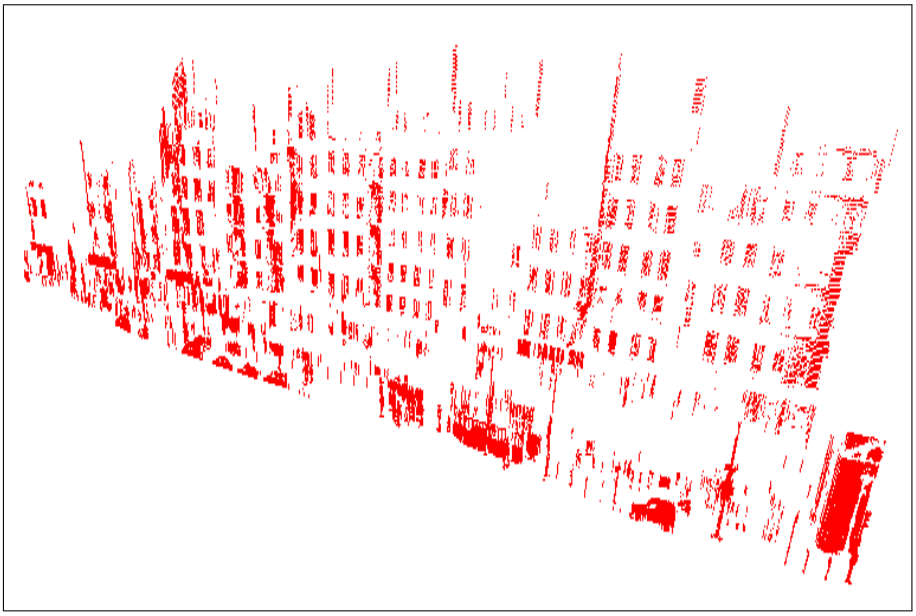
(b) Semantization of the facade. The cadastral map is only employed to split the facade point cloud by dwellings

Figure 4.13: Final facade wall segmentation and semantization stages of approach 2.

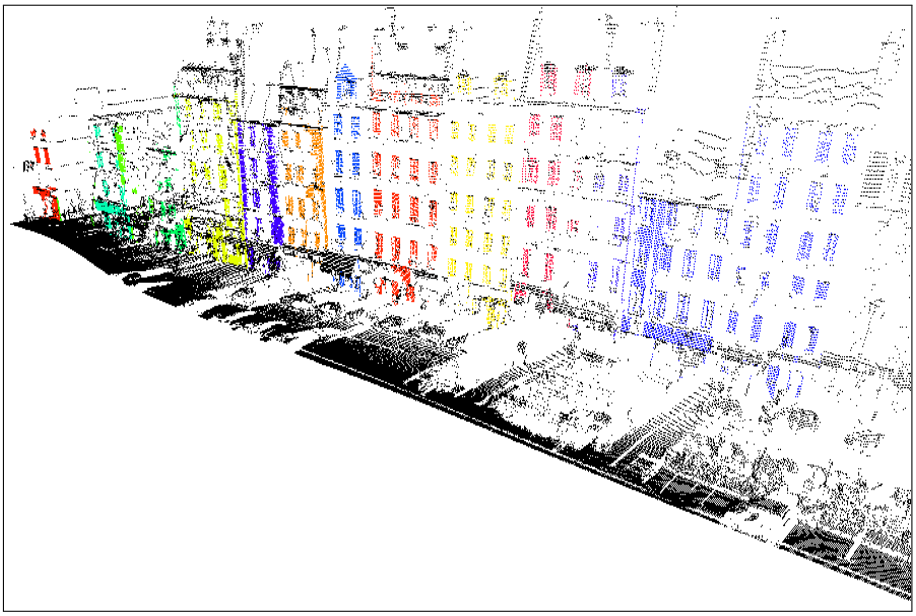
At this stage, it is worth mentioning that a satisfying level of detail in the segmentation and the classification of the street point cloud has been reached by producing four dominant categories of specific urban features without the usage of predefined and elaborated 3D urban models . The usage of the accumulation map and the PPHT algorithms form an interesting combination for the segmentation and the set of point extraction of urban features.

In brief, the proposed approach is principally based on the consecutive spatial analysis of the street point distribution (e.g., accumulation of points) in the Euclidean space as well as in the Hough Space. Figure 4.10, Figure 4.11 and Figure 4.12 illustrate results of segmentation stages by using the model-less

approach (approach 2).



(a) Complementary of the extraction after the use of the PPHT algorithm.



(b) Semantization of the facade superstructures. The cadastral map is only employed to split the point cloud of facade superstructures by dwellings.

Figure 4.14: Final facade windows segmentation and semantization stages of approach 2.

The last stages of the approach 2 are shown in Figures 4.13 and 4.14. The use of the cadastral map is reduced in the sense that the associated 2D segments are only employed to the facade individualization stage.

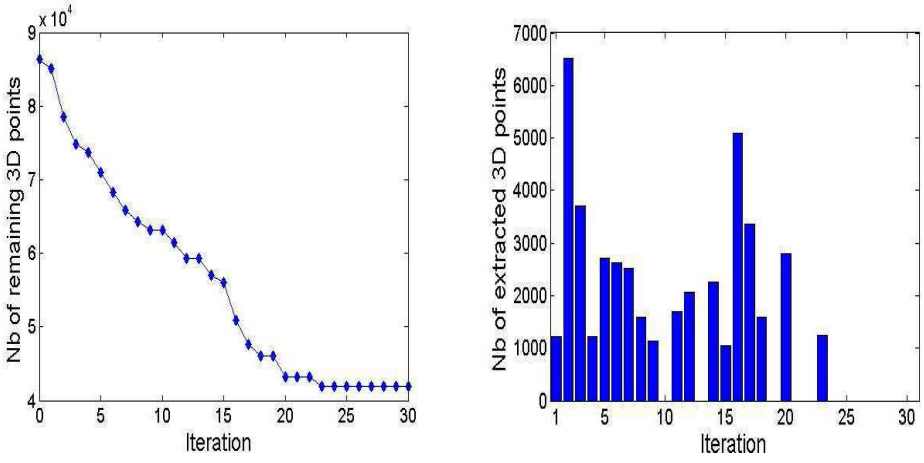
The facade wall points and the facade points of microstructures that are individualized by using segment lines of the cadastral map are respectively shown in Figures 4.13(b) and 4.14(b). In Figure 4.14(b), we have included the set of surface points. Globally, we observe that the individualization stage is carried out with success.

Finally, it is worth pointing out the fact that the standard cadastral map are in general connected to a set of GIS databases. Consequently, the level of semantization of the facade clusters can be enhanced by labelling the cluster with various information inferred by these databases.

The approach 2 allows the consideration of facades corresponding to recent building constructions or destructions. In this way, the generated street facades model can be more frequently updated. Moreover, the approach 2 allows the enhancement of the level of detail in the facade segmentation in comparison to the approach 1 in the sense that the points of facade walls and superstructures (e.g.; openings, windows) are separated.

Furthermore, since the dominant facade clusters are extracted in the approaches 1 and 2, the back and front points can be detected by estimating facade support planes. Then, the surface points can be segmented into points of street roads and points of roof planes. Identically, the points of vertical structures can be segmented into points of street superstructures (e.g. urban posts) and points of roof superstructures (e.g. chimneys). Thus, the level of details of the urban street segmentation could enhanced.

4.4.3 Overall Analysis and Accuracy Assessment



(a) Graph of convergence. (b) Histogram of clusters density.

Figure 4.15: Behavior of the adopted PPHT algorithm in the facade wall extraction.

Our approaches have been tested on raw laser data that correspond to portions of urban streets of the 12^{th} district of Paris. The approach speed allows to extract facade point clouds in 20 to 30 seconds for 30 iterations; near real-time in the sense that the time for the data acquisition is of the same order of the processing time.

Comment:

– *It is worth mentioning that the fast extraction of street geometric features is of great interest for urban mapping. Indeed, urban mobile robots are more and more developed from low cost technologies and can suffer of trajectory derivations for example in reason of lack of GPS signals. Consequently, the fast generation of street maps should allow the registration of trajectory by the*

matching between such dynamic street maps with global and standardized street maps generated from top-notch MMS technologies or coming from existing urban GIS.

An iteration of the algorithm is achieved when the threshold associated to the peak is reached. Figure 4.15(b) shows the number of points extracted at each iteration. The extraction process stops when the size of the dataset is stabilized (see Figures 4.15(a) and 4.15(b)). The approach 2 is promising for high-level investigations and approaches that aims at mapping urban street online.

Figure 4.14(b) depicts the successful individualization by dwelling of facade microstructures from the map segments (orthogonal neighboring of $2m$). Here, we analyze the presented windows extraction by comparisons with segmentation results manually obtained in the area under study. The accuracy is expressed in completeness and correctness (recall and precision respectively) such as shown in Eq.4.2 and Eq.4.3.

$$completeness = \frac{number\_of\_detected\_windows}{number\_of\_observed\_windows} \tag{4.2}$$

Error assessments are performed in the synoptic Table 4.1. The windows detection algorithm achieved here a completeness of 90.54%. It corresponds to the number of detected windows amongst the whole of the windows observed in the scene. Non-detected windows are essentially due to missing parts or windows quasi-coplanar to the facades.

$$correctness = \frac{number\_of\_complete\_windows}{number\_of\_detected\_windows} \tag{4.3}$$

The measured value of classification correctness (defined complete or partial) that deals with the detected windows set equals 85.82%. It corresponds to the number of windows observed as complete amongst the set of windows previously detected.

	Windows detection (%)
Completeness	90.54
Correctness	85.82

Table 4.1: Statistical measures for the extraction of windows regarding about fifteen buildings and several tens of windows in a typical street under study.

Although only universal hypotheses on cities geometry are used (verticalness, planarity, facade proportion), the segmented windows points Figure 4.14(b) can be used by applications mentioned above. A satisfying compromise is reached between efficiency and generalization.

4.5 Conclusions

4.5.1 Contribution

We have proposed two approaches for the segmentation and the classification of urban street point cloud. The first approach is directly based on the use of the cadastral map. This approach has the advantage to exploit available

standard building maps for the extracting facade points. It is worth pointing out the fact that the production of these maps are very expensive in reason of the human interventions and this map is still present in the aerial-based production of building models. Since the development of urban construction arises a boom in numerous cities across the world, the perennity of this kind of map is in phases of questioning at long term since their quality depends of our ability in maintaining them updated. Nonetheless, the proposed approach provides a good compromise for overpassing the stages between actual systems and future full map-less systems.

The second approach intends to segment and to classify the urban street point cloud by reducing the use of the cadastral map in the major segmentation stages such as the facade point extraction. The approach deals with the research of a long-term solution that could be applied by reducing the human interventions and that could be more easily maintained and updated. Besides, the proposed approach also bring refinements by revealing facades structures and their microstructures at street scale in taking advantage of the PPHT properties. The proposed approach is model-less in the sense that no existing predefined and advanced urban 3D model has been used. In the last stages, a map is finally used to individualize the street facades and their microstructures useful for the residential facade recognition (separating dwellings) and to bring a superior dimension to the classified clusters of points; namely the feature semantization.

The common basis of the two proposed approaches has been underlined since we think that it could inspire others work extensions.

Experimentations have been carried out on real and dense urban raw data acquired to the heart of the city of Paris. The efficiency of the proposed approaches has been demonstrated for the processing of wide urban street portions and the shown results prove a potential of great interest for numerous applications in city modeling.

Transition

Here, the investigations specifically deals with the facades. In first, the aerial-based and terrestrial-based investigations for facade modeling are presented and discussed. Besides, the terrestrial laser that have been acquired at street level data have been segmented and classified. By this way, the ROI of facade clusters are extracted and localized. Then, in second, these features will be exploited in order to geometrically represent the facade model as well as to detect the object out of the facades. This information will then be employed by conjointly exploiting the terrestrial images in order to produce occlusion-free textures that could be mapped onto produced terrestrial-based facade models or mapped onto facades of external aerial-based building models.

Chapter 5

Facade Modeling and Occlusion-free Texturing from Image and Laser Data

Abstract

This chapter is focused to the modeling and occlusion-free texturing of urban facades from street point clouds and optical images. Firstly, a study on altimetric variability is conducted in order to classify the facade into rectangular and detailed categories from the facade clusters estimated in the precedent Chapter. A parametrization of the wire-frame model of street facades is then given. Secondly, a methodology is proposed in order to delimit the dominant facade plane (map-based or laser-based) according to the level of details of facades. Thirdly, the facade occluding objects are detected into 3D and exploited in order to generate images that include masks at the location of the occluding objects. Finally, these images are employed to produce occlusion-free textures.

Contents

5.1 Introduction and Motivation

Nowadays, city modeling has become an important research topic for architectural lasergrammetry, photogrammetry and computer vision communities. There is an increasing need for 3D building descriptions in urban areas in several fields of application such as city planning and virtual tourism. In this section, we remind the context as well as the important elements that motivate us for enriching the quality of the building representation in the urban environments.

As presented in the Part II, some research in the cartographic field deal with city modeling using aerial data. Effectively, aerial data are very useful for the coverage of large areas such as cities. A huge amount of aerial-based approaches have been developed and provide in majority 3D city models with a satisfying reconstruction quality at roof level.

Besides, we can observed that the facade models that compose the aerial-based 3D building models correspond in majority of cases to quadrilateral planes produced either by the extrusion into 3D of a cadastral map or by the extrusion into 3D of building footprints extracted from the aerial data (set of 2D segments). Consequently, the current accuracy of aerial-based delimited facade boundaries is highly depending on the aerial data characteristics (angle view, scale of view, data sampling, data covering) and on the quality of the cadastral maps or the extracted building footprints. Above all, we will study the compatibility and the divergence in the building representation between the aerial-based building modeling and the ground-based facade modeling for determining possibilities in model fusing.

5.1.1 Compatibility and Divergence between Aerial-based Building Modeling and Ground-based Facade Modeling

As it has been described in Part II, many research activities on city modeling have focused on the automatic generation of 3D building models from aerial data. Most aerial-based pipelines have been developed to recover the 3D shape of roof surfaces employ footprints of building blocks or footprints of individual buildings as input to the focusing stage.

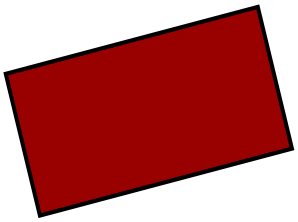
Besides, it is worth point out the fact that first operational investigations in the generation of aerial-based building model have been carried out around the 70's. Otherwise, the experience in the wide-scale generation of facade model from terrestrial data acquired from mobile mapping systems is inferior in comparison to the aerial investigations certainly in the sense that the first MMS system has been operational around 2002 and the terrestrial MMS-based experimentations for the facade modeling are still very recent.

At first glance, it appears that the terrestrial data can be employed complementary to the aerial data in the generation of building model (e.g.; facade plane estimation and delimitation). However, the analysis of buildings of urban street and the aerial and terrestrial data analysis show that the merging of information is a complex task in reason of the differences of data scale, data visibility or data accuracy. Undoubtedly, it will cause questionings in

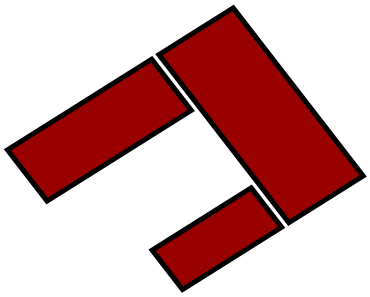
the selection of a common building model representation and in the degree of generalization of the developed approaches.

For the sake of understanding, we are interested here in the current representation of the buildings such as seen from the sky as well as by the limitation of aerial-based approaches in the building model representation. More precisely, we can analyse in particular the building footprints in the sense that the associated segment lines of the building facades can be extracted from aerial data as well as from terrestrial data.

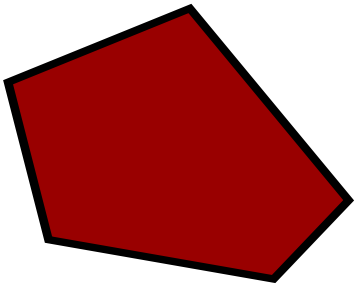
In aerial data, the extraction of building footprint is a complex task in reason of the diversity of the building shapes and their arrangements. The representations of building footprint can essentially be divided into three main categories (see Figure 5.1).



(a) Individual building with rectangular footprint.



(b) Building with rectilinear footprint; i.e., a set of rectanglular building footprint (frequent).



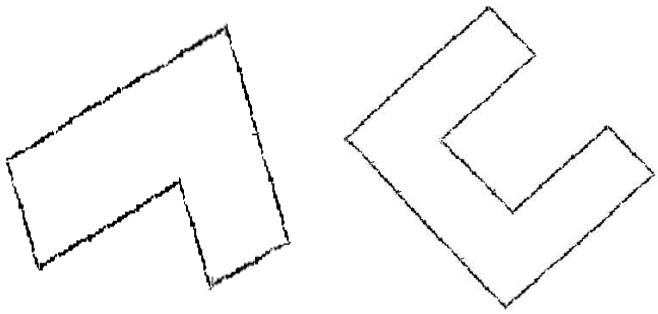
(c) Building block with polygonal footprint.

Figure 5.1: Examples of three current types of building arrangement that can be encounter in an urban environment.

The representation of the building footprints can be varied, namely, simple quadrilateral (e.g.; Figure 5.1(a)), assembly of multiple rectangles (e.g.; Figures 5.1(b) or 5.2), closed polygons (e.g.; Figure 5.1(c)). The Figure 5.2 illustrates a family of frequent rectilinear polygonal shapes $\mathcal{F} = \{L\text{-shape}, T\text{-shape}, U\text{-shape}, F\text{-shape}\}$ (category 2). Such shapes are common building configuration that are currently studied in urban planning (e.g.; (Gomes et al., 2003)). From the sky, the building shape representations are very rich and the aerial-based approaches allow the modeling of the whole building shapes. From the ground, the visualized facades can not easily inform on the global building topology since only parts of the building facades are observed.

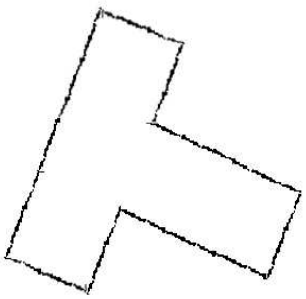
5.1.2 Related Work

Consequently, the aerial data are very useful for the coverage of large areas such as cities and several aerial-based approaches are proposed for the extrac-

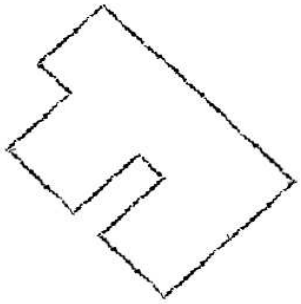


(a) L-shaped building footprint

(b) U-shaped building footprint



(c) T-shaped building footprint



(d) F-shaped building footprint

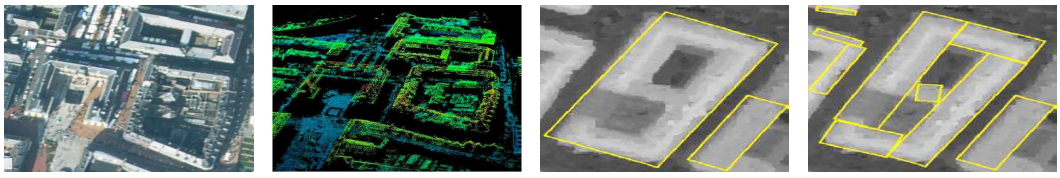
Figure 5.2: Illustrations a family of building footprints with rectilinear shapes $\mathcal{F} = \{L\text{-shape}, T\text{-shape}, U\text{-shape}, F\text{-shape}\}$. This family represents a subset of frequent rectilinear building shapes (second category) that are usually studied in aerial-based building modeling.

tion of building footprints. Moreover, the open areas located inside building blocks such as hidden garden can be detected. More precisely, the data essentially employed as input to these approaches are either optical aerial images and derived Digital Surface Model (e.g., (Tournaire et al., 2010)) or aerial LiDAR 3D point clouds (e.g., (Wang et al., 2006)) such as shown in Figure 5.3.

Besides, footprints of building block are also generated from surveying measurements of building facades achieved by surveyors equipped of tachymeters. As for the cadastral map generation, key points such as building angles can be measured. These tasks are then very fastidious in order to process large scale areas such as building blocks. Moreover, the interventions of several operators are very expensive. Other approaches aim at digitizing and vectorizing existing paper cadastral maps. These maps are composed of set of 2D segments that delineates the buildings by dwelling. The vectorization stage is then succeeded by a map simplification stage into polygonal footprint of building blocks. These processing often are automatic and it happens that the resulting building maps can be incomplete. Additional stages are then necessary to refine the resulting map such as the map completing.

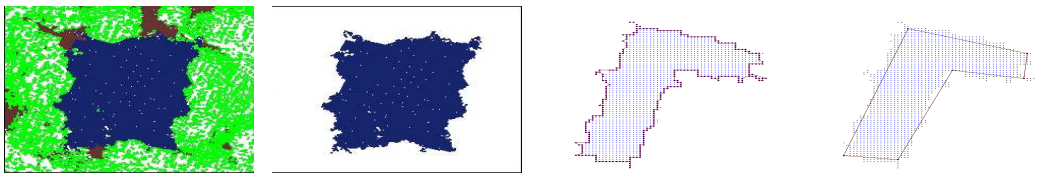
Nonetheless, the building footprints have demonstrated their usefulness in the wide scale modeling pipelines. Also, approaches are developed in order to enhance the building footprint geometry. In (Vallet et al., 2009), an approach is proposed for subdivided a cadastral map into into smaller regions proper for fostering subsequent 3D reconstruction. The authors mention that this app-

Rectangular footprint of building block generated from aerial or spatial DSM



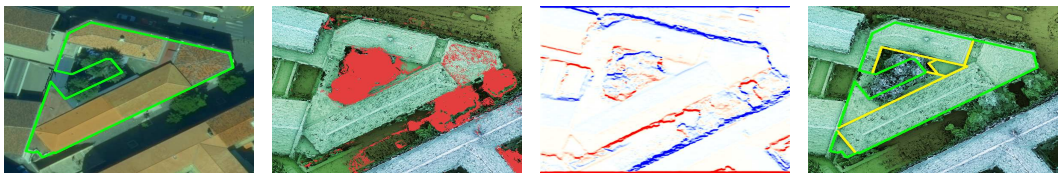
(a) High resolution aerial image (overview). (b) Generated 3D point cloud (e.g., DSM). (c) Generated building footprint (tuning 1). (d) Generated building footprint (tuning 2).

Polygonal footprint of building block generated from aerial LiDAR data



(e) Laser points of a rectangular building (with close vegetation). (f) Building points (boundary noise). (g) Detected boundary points. (h) Approximated building boundaries.

Based-map polygonal footprint of building block rectified from aerial DSM



(i) Orthophotography and existing footprint. (j) Shaded DEM and vegetation mask. (k) Edges detection using gradient mask. (l) Improvement of footprint geometry.

Figure 5.3: The upper part of this Figure illustrates an example of building footprint generation using a aerial or satellital DSM. The middle part of this Figure shows an example of building footprint generated using aerial LiDAR data. The lower part shows an example of footprint improvement by using footprints of map and aerial DSM. Figures are retrieved from (Tournaire et al., 2010), (Wang et al., 2006) and (Vallet et al., 2009), respectively.

proach could be a useful preprocessing step to any 3D reconstruction method based on the cadastral map or any other vectorial footprint of the building to reconstruct.

Consequently, building footprints can be extracted either in an automatic way using the aerial data, either in a manual way requiring many surveyors to make measurements in the terrain. However, the building footprint databases sometimes do not exist (e.g., in less developed countries) or can be very difficult to obtain (e.g., in areas with difficult access such as prohibited overflights areas). Moreover, the full aerial approaches are known to be of insufficient geometrical quality to satisfy the requirements of urban cartographic applica-

tions at street level (e.g., accurate facade modeling). Indeed, the resolution in aerial data is very limited (often decimetric) in comparison with the resolution of terrestrial data that can reached few centimetres. Furthermore, as it can be observed in (Wang et al., 2006), the aerial laser points are very irregular at the building junction. Also, the DSM are known to be very noisy at the building junction. Then, the accurate building footprint extraction using aerial data is a hard task.

The buildings are often partially visible in aerial data in reason of the vegetation that causes missing building portion. This effect is illustrated in Figures 5.3(e) and 5.3(j). Imprecise and/or incomplete focusing will affect the modeling process in the sense that the final 3D building model will lack accuracy and details. Further to limitations of aerial data (e.g., sampling, lack of visibility), the use of existing building maps for footprint estimation cannot tackle all problems. The resulting building footprints are issue of multiple intermediary stages namely digitization, vectorization, simplification, rectification. All these successive stages inevitably affect the resulting building footprint in accuracy.

Also, both of DSM and aerial laser points are currently combined with the building footprint to directly compute 3D building model by footprint extrusion (e.g.; (Ledoux et al., 2009)). Indeed, one of the simplest and most efficient technique consists of assigning for each building footprint a single mean height value. This technique that has the advantage to rapidly produce 3D city models (set of building model) is frequently used in the urban applications.

For these reasons, it is fundamental to develop the research at street level to improve the building footprint geometry in planimetry. From terrestrial data, it should be then possible to extract accurate facade model and facade 2D supports for rectifying the building boundaries in aerial-based building models.

In the literature, many interesting approaches deal with the exploitation of vectorial maps of building footprints such as existing cadastral maps or building footprints generated from aerial (or satellital). The intention here is to briefly describe some of these approaches (e.g., (Wang et al., 2006, Tournaire et al., 2010, Vallet et al., 2009)).

In (Wang et al., 2006), it is pointed out the fact that the automatic processing can be complex in reason of building edges that are often noisy. A Bayesian technique is presented to extract building footprint from a pre-classified aerial LiDAR point cloud. The algorithm first computes an approximated boundaries of building footprint by using an application of the shortest path algorithm. Then, the most probable building footprint is determined by maximizing the posterior probability using linear optimization and Simulated Annealing techniques (SA techniques). The algorithm has been applied to more than 300 buildings and it is observed that accurate building footprints are obtained in comparison to a ground truth. The authors highlight the flexibility of this automatic algorithm that can be applied to other man-made shapes (roads or telecommunication lines) by carrying out minor modifications.

In (Tournaire et al., 2010), an automatic approach has been proposed in order to extract building footprint from aerial and satellital Digital Surface Model(DSM) with resolution from metric up to decimetric. As described in (Ortner et al., 2007), marked punctual processes of rectangles are employed. The approach targets buildings with footprints that can be decomposed into a set of rectangles. More precisely, neither the number of rectangles that

composed the building footprint, nor the rectangles parameters are previously known. The objective consists in the determining of the optimal object configuration by minimizing an energy function. The optimization process is achieved by combining an RJMCMC sample (Reversible Jump Markov Chain Monte Carlo) with a Simulated Annealing process in order to determine the optimum of this non-convex function. The computation time is relatively reduced in comparison with the preceding approaches, on the one hand, the energy formulation is simplified, on the other hand, achieved geometric optimizations lead to fast computations.

In (Huang et al., 2008), the authors proposed an approach to the generation of building boundary from aerial LiDAR points. The convex hull based boundary tracing algorithms often are not suitable to the generation of building boundaries that can be composed in some cases of many concave parts. Moreover, the authors emphasize the difficulty in the stability of boundary tracing algorithms that often require a hull adjustment guided by a tolerance threshold (i.e.; regularization term). Then, the paper proposes a side ratio constraint for boundary tracing algorithm from discrete points. The proposed approach is able to trace building boundaries of concave polygons including holes.

Paper	Process	Input data	Strategy
(Wang et al., 2006)	Automatic	Aerial LiDAR data	Bottom-up
(Ortner et al., 2007)	Automatic	Aerial DSM	Bottom-up
(Tournaire et al., 2010)	Automatic	Aerial or Satellital DSM	Bottom-up
(Vallet et al., 2009)	Automatic	Urban map/Aerial DSM	Hybrid
(Huang et al., 2008)	Automatic	Aerial LiDAR data	Bottom-up

Table 5.1: Some aerial-based, satellite-based and map-based approaches developed for generating building footprint.

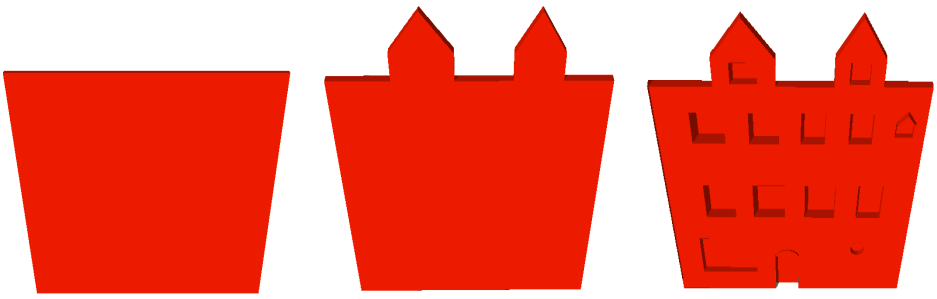
Table 5.1 briefly presents some aerial-based, satellite-based and map-based approaches available in the literature for the building footprint generation or the improvement of cadastral map splitting. The regrouped approaches show an overview of techniques that are known to be efficient to process large scale areas. The presented approaches propose building footprint shapes under various geometric representation. We observe that major limitations of the aerial-based approach is due to the quality of aerial dataset (e.g.; data sampling, facade visibility).

5.2 Terrestrial-based Facade Modeling

Besides, we have remarked that the generation of complete building footprint from urban laser data acquired from the ground is a topic few addressed in the literature. At long term, it appears that the fusing between aerial-based building model and terrestrial-based facade model will be unavoidable. Above all, our interest deals here with the wide scale urban facade modeling from terrestrial data.

Henceforth, the 3D city models need geometric, graphic and semantic enrichments at street level. In this study, one of the aim consists to estimate the 3D planes and the boundary delimitations of street facades from terrestrial multi-source data in order to bring more accuracy and better localization at

facade level that can meet the requirements of current cartographic applications.



(a) Rectangular model. (b) Detailed polygonal model. (c) Polyhedral model.

Figure 5.4: Examples of generic model representations (Birds eye view). Three illustrations of the same facade with different level of details (from low to high).

The representations of 3D facade models can be divided into three main categories (see Figure 5.4). In the current work, we investigate research in order to propose solution for reaching the intermediary Level Of Detail (LOD) such as presented in Figure 5.4(b).

5.2.1 Problem Statement and Main Objectives

For sake of clarity, we recall below some requirements for concrete cartographic applications as well as the potential of use of street laser and image data acquired by the MMS for bringing solutions.

1. *Improving Building Geometry:*

   (a) *Generating Digital Cadastral Maps:* The research on mapping tends towards the automatic generation of Digital Cadastral Map (DCM). The DCM is a crucial data since employed for the national taxation of each citizen according to the size of his dwelling. The generation of this map in a manual way requires many surveyors to make measurements in the terrain and stays very expensive. The fast mapping (carried out by MMS in On-Drive mode) can be used in order to update the urban map since recently constructed or destructed buildings can be detected using the terrestrial laser data.

   (b) *Correcting the DEM:* Many urban modeling approaches employ as input Digital Elevation Models (DEMs) generated from aerial images. However, the DEM can be very inaccurate at facade limits and undoubtedly require geometric correction. The extraction of fine and accurate facade limits will be useful to establish DEM corrections at facade locations.

   (c) *3D Facade Modeling:* The fine mapping of building facade from aerial-based framework is a complex task. Research works on city mapping tend to develop approaches for facade mapping at large scale from terrestrial acquisition. The fine facade modeling at street

level requires the extraction of facade outlines by dwellings (i.e. generation of a wire-frame model of street facades).

- *Generating Aerial/Terrestrial Compatible Facade Models:* One of the field's long term goal consists of combining in a coherent way aerial and corresponding terrestrial data in order to accurately model all facets of full buildings (roofs and facades). Since the facades that compose the aerial-based building models are in the great majority rectangular, we are particularly interested here in the wide scale rectangular representation of facade models from street data such as represented in Figure 5.4(a). In this way, a common degree of representation in both of data outputs will foster the fusing of models into complete and accurate building models.

- *Generating Facade Models adapted for the Walk-through Street Visualization:* Independently of the aerial/terrestrial fusing, the generation of detailed facade models specific for the ground-based 3D visualization and navigation in the urban street will be interesting (e.g., 3D street view). This kind of visualization platform is employed at the quotidian to virtually visit a place and enhancing our ability of orientation. For this reason, we investigate researches for the facade modeling at an intermediary level of details (see Figure 5.4(b)) to develop the aspect of street exploration from ground-based point of view.

2. *Improving Building Graphic Rendering and Semantic:*

   (a) *Segmenting Street Images:* In the framework of our work, we possess a Google like web-based Street View. This street viewer employs data collected by the MMS and provides to the general public a tool for city visualization/navigation at street level through a flow of georeferenced panoramic images. The projection of the 3D georeferenced delimitations of each facade onto the corresponding georeferenced images allows the segmentation of street images into facades and the remain of the scene. Then, the delineated facade images (facade areas) are of great interest for research on urban scene semantization (e.g. facade keywords extraction or facade retrieval in image databases). Also, the facade images can be employed as input for texturing at facade level the generated wireframe models (our approach) as well as external 3D city models.

   (b) *Occlusion-free Facade Texturing:* As we have observed in Chapter 3, the facades often are strongly occluded by urban street occluding objects. In this case, the frontal facade images acquired by the MMS are maladapted to the direct facade model texturation. The laser data conjointly acquired could be combined to multi-view facade images in order to produce facade textures without occlusions, i.e. occlusion-free textures.

5.2.1.1 Chapter organization

In this scope, we aim at improving the geometric modeling of buildings (Point 1c) as well as the building graphic rendering and semantic (Points 2a and 2b). These set of Points cover principal modules for the generation of rich and realistic models of street facades. The textured street facade models can be strongly employed for deserving popular and ludic cartographic applications such as the visualization and navigation in dense urban environments.

More particularly, this Chapter brings research pieces in order to establish a prototype pipeline for generating virtual 3D model of urban street facades by fusing cadastral map with terrestrial images and laser data. A Mobile Mapping System (MMS) has been employed to massively gather georeferenced images of street scene in addition to the laser raw data previously processed. In the precedent Chapter, the extraction of facades point cloud is achieved by exploiting the cadastral map and street laser data. Some solutions have been proposed in order to carry out the street point cloud segmentation with a particular interest for the clusters of points extraction at facade level. The current Chapter presents then an extension that addresses the facade plane estimation, the geometrical delimitation of the outer facade boundaries with a high degree of realism and the generation of facade textures by exploiting image and laser data.

5.2.2 Related Work

Recent progress in technologies have allowed the construction of sensors more and more elaborated. These instruments are employed to the development of mobile mapping devices for rapid and massive acquisition of 3D cartographic terrestrial data with very high accuracy in dense urban environments.

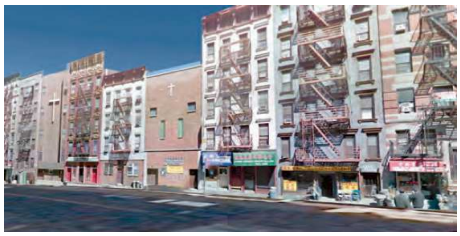
Henceforth, the Mobile Mapping System allows then an easy coverage of large scale areas such as districts and cities. The feasibility and performance of this kind of system has been largely demonstrated (Haala et al., 2008) and the data acquired by terrestrial laser scanning system are foremost employed for applications like the digitizing of historic district for the cultural heritage (e.g., digitized Venetian buildings (Studnicka et al, 2011), digitized buildings of the Istanbul peninsula (Baz et al, 2008)) or the terrestrial modeling of urban scene for the street visualization (e.g., Google Street View (Anguelov et al., 2010)). In cases of digitized buildings, we stress the fact that the building representation often consists in a colourized dense point cloud.

In (Anguelov et al., 2010), Google Street View data are used to create photorealistic 3D models for Google Earth. The authors highlight the fact that the terrestrial multi-source data are adapted to produce pleasant walk-through street environment. Street View's laser scans are employed to the generation of 3D facade models. Then, the objects in facade foregrounds are filtered and suitable facade textures are generated from multiple images by aligning, blending, and mosaicking.

In (Xiao et al., 2009), the authors propose an automatic approach to generate street-side 3D photo-realistic models from overlapping images acquired at street level. A multi-view semantic segmentation is operated at pixel level on each image and categorized in urban object (e.g., building, ground, car, vegetation). In addition, the major horizontal and vertical line structures of the

3D rectangular facade models textured from various image dataset

(a) Aerial-based textured street facade model.



(b) Terrestrial-based textured street facade model.

Terrestrial-based model of street facade models with various LOD

(c) Generated street facade models with quadrilateral facade representation (as LOD shown in Fig. 5.4(a)).

(d) Generated street facade models with polygonal facade representation (as LOD shown in Fig. 5.4(b)).

Figure 5.5: The upper part of this Figure illustrates the benefit of terrestrial images for the texturing of street facade models. The lower part depicts results of street facade modeling approaches. Figures are retrieved from (Anguelov et al., 2010), (Xiao et al., 2009) and (Kang et al., 2007), respectively.

scene are exploited to separate the buildings into independent blocks. Finally, for each block, the authors propose an inverse patch-based orthographic composition and structure analysis method for facade modeling with pushed and pulled rectangular shapes. These methods regularize the noisy and missing reconstructed 3D data. The approach is validated on a large amount of data dealing with a typical city example.

In (Kang et al., 2007), the authors present a 3D reconstruction approach for rapid 3D visualization from street level, which is based on a combination of vehicle-based image sequence and 2D vector map. The approach consists of two major steps: the rapid reconstruction of facade along the street based on sideshooting vehicle-based image sequence and associated 2D map. This approach is quite appropriate for fast reconstruction and visualization towards disaster management (e.g. flood or earthquake).

Figure 5.5 illustrates some modeling results coming from the approaches described above. The facade textures that are exhibited are relatively clean in the sense that they not contain major occluding objects (e.g., vegetation, street objects). In Figures 5.5(a) and 5.5(b), 3D rectangular facade models have been rendered with textures coming from aerial images and terrestrial images, respectively. Undoubtedly, the quality of the terrestrial facade textures exceeds that of the aerial facade textures (in case of non-occluding facades). Figures

5.5(c) and 5.5(d) depict facade models with LOD which respectively trigger a high interest in our case.

Particularly, we can observe in some cases that the illustrated facade LOD (quadrilateral or detailed polygons) generally are adapted to the walk-through street navigation in virtual environment. Also, the lacks of details in the geometry of the fine objects is often compensated by the use of photo-realistic facade textures as well as the inclusion of artificial structures as sky or mist.

5.3 Proposed approach

The proposed approach aims at automatically extracting a model of street facades by combining 3D point clouds of streets, the corresponding cadastral map and digital facade images. Furthermore, these multi-source terrestrial data are fused for the generation of occlusion-free textures.

5.3.1 Parametrization of Street Facade Models vs. Facade LOD

5.3.1.1 Estimation of the Dominant Facade Planes

In the precedent Chapter, we have amongst others presented an approach combining laser data and a cadastral-based map for the automatic extraction of dominant facade planar clusters in very dense and real urban environment.

In particular, the planar clusters have been extracted using the Progressive Probabilistic Hough Transform (PPHT) that has several advantages. First, the PPHT algorithm contributes in the global urban street point cloud segmentation. Second, it can generate an updated urban map by incorporating or removing some building facades. Third, it provides a fine estimation of planar facade clusters (there is always some misalignment between the georeferenced laser data and the cadastral map). Since the extracted facade wall points are quasi-planar, the facade planes can be directly parametrized by selecting 3 points for each extracted facade wall clusters. Furthermore, the facade planes can also be estimated by using a Least Squares Adjustment technique (LSA).

Consequently, we recall that the virtual support planes provided by the cadastral map are not directly employed. Just the extremities of the 2D segments are used and projected for delimitation. Although the conventional cadastral map is actually required, the works naturally tend to select the georeferenced data acquired by the MMS as reference since we hope that it will be possible to generate online the future cadastral maps.

5.3.1.2 Extraction of Facade Top Points from the Accumulation Map

In the precedent Chapter, the street point cloud has been segmented into urban objects. The facade point clouds have been extracted and individualized by using the cadastral map. The cluster of points associated to each dwelling facade are identified. Besides, the accumulation map $\mathcal{S}(u,v)$ has been previously generated and can be re-employed. Hence, the ground space has already been finely discretized into 2D cells (u,v). For each identified cluster, the associated 2D cells can be retained. Then, for each retained cell (u,v), the whole of 3D points included inside the associated voxel is analysed and the included 3D point having the maximum altitude is kept. The set of top points is then formed.

However, it is not guaranteed that the top points of the clusters are facade top points. Indeed, it often happens that the opening of the laser sensor is too low for covering buildings with high heights.

Besides, we remind that the acquired 3D points can be considered either as a global set of points or as a structured set of points by exploiting the topology of points specific to the employed sensor (frames [1] of points). Consequently, the set of points should be previously labelled as follows.

The 3D points have initially been acquired by sequence of 201 points (frame of points). For each frame of points, the distance of the top point (one extremity of the frame) is measured. If its distance is "infinite" (the sensor indicate a non-returned point) then all the facade points contained in the frame are labelled as complete top points (or only the extremity). Otherwise, the facade points contained in the frame are labelled as topless points. This information labelled to the points is independent of the location of the point in the space but only provide semantic information on the fullness of the acquired facade top according to the reality.

Then, if the majority of cluster top points extracted for the accumulation map are labelled as complete top points, then the facade top analyse and delimitation such as proposed in the next sections will be invoked (assumed here). The contrary case will be discussed later.

5.3.1.3 Analysing the Altimetric Variability of Facade Top

In this section, we have used a statistical method in order to automatically classify the facade according to their level of details; namely, the rectangular facade model and the detailed polygonal facade model.

More precisely, the method employs the Mean Squared Deviation (MSD) that is known to be a robust statistical measure of the variability. Hence, for each set of retained facade top points, the coordinates are sorted by altitudes and the corresponding median value is determined (selected as central tendency).

For each facade j, this measure is then calculated from the heights of the top points in order to evaluate the facade top discontinuities by using Eq.(5.1):

$$MSD_{\boldsymbol{j}} = \frac{1}{n} \cdot \sum_{i=1}^{n} (Z_{Xi} - Z_{X\,median})^2 \tag{5.1}$$

Therefore, the Level Of Detail (LOD) corresponding to a facade j can be automatically identified by applying Eq.(5.2) defined as follows:

$$\mathbf{w}_j = \begin{cases} \mathbf{w}_{quadrilateral}, & if(MSD_{\boldsymbol{j}} < \mathcal{T}_{\boldsymbol{discontinuities}}) \\ \mathbf{w}_{top\_detailed}, & otherwise \end{cases} \tag{5.2}$$

The MSD measure provides thus a score quantifying the altimetric variability. The nomenclature of the facade polygon is deduced. A threshold $T_{discontinuities}$ can be empirically adjusted towards a full street facade labelling. This information could then be employed to adapt the facade top modeling according to the identified LOD.

Besides, we mention that a very similar method is proposed in (Dash et al., 2004) for the fast extraction of building from aerial laser data by analysing the

1. A "frame" denotes the sequence of 3D points acquired by given vertical scan.

contour uniformity of urban object (although the data scale and the object of interests are different).

5.3.1.4 Generating a Compact Wire-Frame 3D Model of Urban Street Facades

Consequently, the optimal parameters of the street facade model noted Φ will be obtained by aggregating the facade model $\mathbf{w}_j$ individually estimated by Eq.(5.3) defined as follows:

$$\Phi^{\star} = \bigcup_{j=1}^{n} \mathbf{w}_j \tag{5.3}$$

This model parametrization is highly compact in the sense that it provides a very compressive representation of the urban facades at street scale.

Two strategies can be employed in the choice of the threshold $T_{discontinuities}$ that can be seen as a regularisation term. This value can be manually tuned. A very low value will lead to adopt the quadrilateral parametrization in case of facade top with slight discontinuities and the detailed parametrization in the remaining case.

A very high value will conduct to adopt the quadrilateral parametrization in majority of cases and a detailed parametrization for facade top with strong discontinuities (e.g.; atypical facade top architecture). In our case, the first strategy that behaves like a slight smoothing of targeted facades has been adopted.

5.3.2 Generating Aerial/Terrestrial Compatible Facade Models

In this section, robust strategies are proposed in order to model the building ground and top limit by an horizontal line for the facade cases associated to the quadrilateral parametrization.

5.3.2.1 Facade Bottom Delimitation by using the Terrestrial Laser Data

A global strategy consists in selecting the altitude of the points having the minimum height in the filtered point cloud. This solution is not very robust if the street is sloped. Then, we recommend the use of more local strategies described as follows.

- *Strategy* 1: This strategy consists of estimating the height from the altitude of the points having the minimum height for each individual facade cluster. In case of an observed discontinuity for a given facade, we reset its minimal height to the interpolation of the two neighboring facades. Indeed, occluding objects like vehicle parked often cause strong occlusions at facade bottom.

- *Strategy* 2: The altitude of the bottom limit can also be computed subtracting the vehicle height to the laser sensor altitude for each cluster and adding the height of a standard border pavement. In this way, the altitude of the facade bottom delimitation is estimated with a total independence of the importance of the occlusions located in front of the facade (facade cluster). Also, this strategy also allows to overpass the lack of laser data at facade bottom since the sensor opening is limited. In our case, this phenomena happens when the sensor is specifically configured to acquire the facade top. We remind that our experimentations were limited to the use of a single 2D laser sensor.

These two strategies are robust in the sense that they provide solutions to estimate the facade bottom delimitation even in presence of strong occlusions. They behave at the facade bottom as filling gap solutions.

5.3.2.2 Facade Top Delimitation by using the Terrestrial Laser Data

- *Strategy* 1: For each facade, the top points extracted in section 5.3.1.2 for the analysis of the facade top variability can be re-employed. More directly, the limit of facade top consists of fitting the facade top by the horizontal line corresponding to the maximum height of the facade top points. In this case, roof superstructures include in the continuity of the facade wall will be considered.

- *Strategy* 2: More directly, the median height computed in section 5.3.1.2 can also be re-employed. In this case, the facade top delimitation essentially characterizes the roof/facade junctions even if superstructures are included in the continuity of the facade wall.

These two strategies can be selected according to the needs in term of modeling and texturing mode.
Finally, the extracted facade planes are delimited by the assembly of four straight segments. In this way, the approach rapidly provides a rectangular wire-frame model of street facades. This elementary representation gives a first facade modeling solution that can satisfy with success the objective such as reported in Point (1c) of Section 5.1.

We stress the fact that this wire-frame model has been estimated only from the laser data. Besides, it is worth pointing out the fact that our investigations

were essentially constrained to the exploitation of the laser data for fitting the urban facade by accurate 3D rectangular models.

Obviously, the use of the images that are acquired conjointly to the laser data will be interesting for formulating hypothesis for the 3D segmentation of the rectangular facade models and for the inclusion of estimated and delimited facade microstructures.

5.3.3 Limitations of Laser Data in Facade Top Delimitation

Besides, we observe in the literature and after some experimentations that the exploitation of the laser data alone (model-less condition) remain complex towards reaching an intermediary level of detail such as the facade representation illustrated in Figure 5.4(b).

In (Boulaassal et al., 2011), the authors present a technique that aims at extracting edge points from a very dense planar cluster of points. The technique consists in two steps: *i*) The triangulation of the planar cluster by using a Delaunay triangulation. *ii*) It appears that the boundary points of the facades and those of the inner objects correspond to the extremities of the long sides. For this reason, a global threshold is applied in order to retain the extremities for the segment lines with a sufficient length. These points could be used for the delineation of the facades or the facade top.

We have also envisaged the extraction of key points such as corners of facade superstructures by using techniques based on accumulation of lines from consecutive pair of top points. For example, the accumulation of lines in the image space is currently employed in our field to the determining of vanishing points (e.g.; (Cantoni et al., 2001)). This technique could be applied to the facade top points by parsing the set of facade top points by pair of points from one facade extremity to the next. Indeed, the facade top points can be put in order by sorting the points according to one axis (abscissa or ordinate) since the facades are assumed quasi-linear. For each pair of points, the associated lines can be traced into 3D and projected in the facade plane discretized in voting space. The accumulation of lines will reveal geometric key points of the facades that could be linked.

However, these methods could be very approximative in the sense that the real facade top is often missed by the laser beam. Moreover, the resolution of the acquired laser data should be very high to ensure an estimation of the key points with a high accuracy. In this way, the direct linking of the facade top points often provides an inaccurate and partially real delimitation. Also, this representation is not sufficiently elegant to reach the requirements of visualization platforms in term of rendering in reason of the discretization effect own to the laser data even if this effect can vary according to the laser resolution.

For this reason, we have proposed to reduce the dicretization effects and to enhance the facade top geometry by exploiting the set of optical images acquired conjointly to the laser data in case of facade labelled as detailed polygon.

Furthermore, a strong limitation that we have observed in the laser dataset acquired is the lack of points at the facade top for the facade top delimitation. This limitation is due *i*) on the one hand by our range laser sensor that can not cover the totality of the facades from the bottom to the top in reason of its opening. We recall thus that the majority of the laser data processed in our

studies have been acquired by a laser sensor configured to cover both of ground and facade portions in order to also satisfy the needs of external research at the ground level (initial choice). We stress also the fact that the Laser sensor employed here is relatively standard in the sense that a huge amount of similar sensor equips the MMS developed in the field. *ii*) On the other hand, we have otherwise observed that the specific configuration of the range laser sensor to guarantee the covering at the facade top is delicate and also fails in several cases in the acquisition of facade top points. Indeed, the urban streets often are very close and the distance that separates the vehicle and the facades is then small. In these cases, the delimitation of the facade top can not be carried out by the lack of facade top points in the dataset. However, this strong limitation is overcome by exploiting the optical facade images conjointly acquired since the field of view of the camera obliquely oriented covers the facade top in all the observed acquisition cases.

Hence, if the majority of cluster top points previously extracted from the accumulation map are labelled as topless points, the facade top delimitation of the associated facade should directly be carried out from the optical images conjointly acquired (detailed polygonal facade). The procedure will be explained in the next section.

5.3.4 Investigations for Generating Facade Models adapted for the Walk-through Street Visualization

The proposed idea is schematized in Figure 5.6 and consists in the 3D realistic tracing of the facade top boundaries (relief in altimetry) by using the set of optical images and 3D facade planes (existing or estimated). The expected advantages will be as follows:

1. Increasing the resolution of top limits of detailed facades (polygonal wireframe model) since it will be set to the optical image resolution.
2. Microstructures close to the facade plane can be included in the top limit of the facades.

5.3.4.1 Computing Realistic Top Limits by exploiting the Terrestrial Images

More precisely, the images are employed in order to detect the demarcation line that separates the facade and the sky, i.e. skyline. Then the detected image skyline is converted into 3D by using existing or estimated facade planes. We recall that the set of multi-source data and geometric features are calibrated, georeferenced and expressed in a common referential system.

Indeed, we stress the fact that the facade planes can be either the facade planes estimated from the laser data or the facade planes obtained by extrusion of a cadastral map.

More precisely, the computation of the 3D detailed facade top boundaries requires *i*) assigning for each facade object a set of images that visualise this facade (Section 5.3.4.2), *ii*) detecting in the facade images the skylines (Section 5.3.4.3), and *iii*) converting the 2D skyline into 3D facade top boundaries (Section 5.3.4.4).

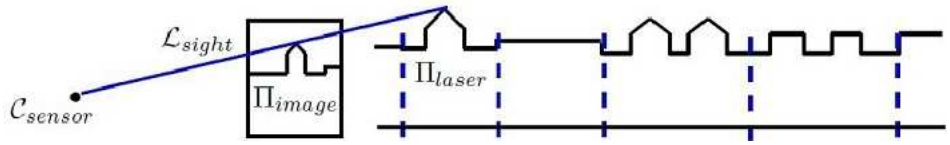


Figure 5.6: Comprehension scheme illustrating the methodology for 3D realistic mapping of urban street facades (3D skyline-based modeling). The line of sight $\mathcal{L}_{sight}$ intersects the associated laser plane Π_{laser} that is delimited to the sides by cadastral maps extremities (separators with dotted lines).

5.3.4.2 Matching Cadastral 2D Segments and Facade Images

The cadastral map provides a set of 2D segments. Each segment characterizes a virtual support plane of a facade (set of segments coming from estimated laser planes can also be used). On the other hand, the MMS provides a panoramic view (set of optical images) every 4 meters along the vehicle trajectory. Thus, it is necessary to match every cadastral 2D segment to one or more acquired images that fully or partially overlap with the facade of that 2D segment. Therefore, this is a matching problem that can be one-to-one or one-to-many depending on the size of the facade and the field of view of the camera.

This step aims to associate for each cadastral segment the corresponding image(s). The employed images are those acquired by the lateral camera oriented perpendicularly to the trajectory. Since the street facades are generally parallel to the street road, the dominant facade plane are more or less fronto-parallel to the image plane of the lateral camera. The one-to-one matching is as follows.

- For each cadastral segment $\mathcal{S}$, to define a neighborhood around the segment barycenter $\bar{X}_{\mathcal{S}}$.
- Keep the images having their projection centers inside this neighborhood. It provides a subset of images $\mathcal{D}_1$.
- Keep the images having an image plane orientation close of the facade plane orientation to $\pm 45°$. It provides a subset of image $\mathcal{D}_2$ such as $\mathcal{D}_2 \subseteq \mathcal{D}_1$.
- For each retained image in $\mathcal{D}_2$, project its optical center C orthogonally to the image plane onto the line $\mathcal{L}_S$ associated with the segment $\mathcal{S}$. The projected point is noted $v_\perp$, $v_\perp \in \mathcal{L}_S$.
- Keep the image providing the minimum value δ. δ represents the Euclidean distance between the projected point $v_\perp$ and the barycenter of the segment $\bar{X}_{\mathcal{S}}$ as follows:

$$\delta = \|v_\perp - \bar{X}_{\mathcal{S}}\| \tag{5.4}$$

In our case, this segment-to-image matching will be useful for the 3D geometric facade delimitation as well as for the facade plane texturing.

5.3.4.3 Detecting the Skyline in the Acquired Facade Images

This stage consists in detecting the skyline for each facade image associated to a segment. Some complex facade top images have been acquired by the

MMS. These images correspond to street facades of the city of Paris (12^{th} district). There is an overlap between the images even if the acquisition rate is not high. In this way, missing the modeling of some parts of facades is excluded.

The robustness of the approach depends on the skyline detection step. In our case, the acquisition are achieved using a MMS. In this way, the images could be acquired at large scale under perfect time condition (blue sky) in order to work with high quality street images. Nevertheless, complex facade images have been acquired in cloudy condition in order to assess the performance of the skyline detection step. In our case, we have used the image skyline detection that has been described in (Liao et al., 2007) namely:

- For each facade image associated to a segment, the image corresponding to the blue channel has been retained. A dynamic threshold $\mathcal{T}_{bin}$ is determined according to the histogram of the blue channel as follows.
- The histogram $\mathcal{H}_{blue}$ of blue intensity vs. frequency is calculated and divided in two parts (0 to 128 and 129 to 255).
- Intensity peaks the most high in each $\mathcal{H}_{blue}$ part, respectively P_1 and P_2, are retained.
- The interval of intensity $\mathcal{I}_{blue} = [P_1, P_2]$ is swept and the intensity value with the minimum number of pixel is chosen as threshold $\mathcal{T}_{bin}$.
- The threshold $\mathcal{T}_{bin}$ is employed for the image binarization process.
- Globally, the sky and the facade respectively appear in white and black. Each image column is swept from up to down and the first black pixel is retained.

In this way, an image facade skyline has been detected. Moreover, since the skyline can be very detailed, the skyline geometry can be simplified by applying the well-known Douglas and Peucker algorithm (Douglas et al., 1973). Our interest deals then on the transformation of this skyline into 3D.

5.3.4.4 Converting the 2D Skyline into 3D Facade Top Limit

If the 2D facade skyline detection is correct then the 3D shape of the facade top can be accurately and finely reconstructed. Indeed, the multi-source data are georeferenced and expressed in a common referential. Moreover, the camera calibration is known (intrinsic and extrinsic parameters). Consequently, the 3D facade polygons can be calculated as follows:

- For each pixel of the detected skyline, the associated line of sight vector can be calculated since the pixel location and the camera center $\mathcal{C}_{sensor}$ are known. Usually, the pixel line of sight is used with pixel of detected edges or corners for 3D recovering as in (Zlatanova et al., 2002).
- The associated perspective line of sight $\mathcal{L}_{sight}$ is then intersected with the virtual plane given directly by the urban map Π_{map} or by the estimated laser plane Π_{laser}.
- If the resulting 3D point is included inside the separators provided by the cadastral 2D segment, the 3D point is retained. In this way, a 3D facade skyline (3D facade top silhouette) can be computed (see Figure 5.6).
- As previously mentioned, the limit of the facade bottom can be deduced from the altitude value of the vehicle.
- For each facade, the 3D skyline and bottom limits are chained into 3D facade polygons.

This methodology fills the requirements in term of rendering that we have initially described in Points (1c) and (2) of Section 5.1.

5.3.5 Generating Occlusion-free Facade Textures

In this section, the aim consists in generating specific textures for the extracted wireframe facade models.

5.3.5.1 Selection of the Texturing Mode

Indeed, a tremendous amount of images acquired at facade level in urban environment are occluded in reasons of urban objects located between the sensor of acquisition and the facade walls. We propose to exploit the associated laser data in order to detect these occlusions and to produce occlusion free textures according the following stages:

1. *Detection of Facade Occluding Objects:* Previously, we have proposed a methodology to estimate and delimit the facade by using the terrestrial laser data. The facade planes issue of the cadastral map can also be used. An orthogonal band is defined in front of the facade planes that is borned by the vehicle trajectory. The associated volume potentially contain 3D points of facade occluding objects as well as points of ground. The points of ground can be removed by using the accumulation map that has been generated in segmentation Chapter 4.
2. *Texturing the Facade Model:* Each facade model (wire-frame model) is textured according to the quantity of detected occluding points as defined Eq.(5.5):

$$\textbf{Texturing Mode} = \begin{cases} \mathcal{M}_{\textbf{Direct}}, & if(\mathcal{Q}_{occlusion} < \mathcal{T}_{occlusion}) \\ \mathcal{M}_{\textbf{Mask-based}}, & otherwise \end{cases} \tag{5.5}$$

where $\mathcal{M}$ denotes the mosaicking that characterizes the texturing mode) and $\mathcal{Q}_{\boldsymbol{occlusion}}$ corresponds to the quantity of points contained in the facade neighbourhood and labelled as occluding points.

More precisely, if the quantity of facade occluding points do not exceed a predefined threshold $\mathcal{T}_{\boldsymbol{occlusion}}$ (negligible quantity) then it means that the facade is either occluded by very small objects (e.g.; urban posts) or not occluded. In this case, the wire-frame model (georeferenced) generated in Section 5.3.2 can be projected in georeferenced panoramic images (employed for the IGN Street Viewer). Then, the segmented panoramic images (i.e., segmented facade images) can be directly employed for the facade model texturing (texturing mode $\mathcal{M}_{\textbf{Direct}}$). Otherwise, the texture should be reconstructed by removing the occluding objects visible in the set of images associated to the facade. This texturing requires the generation of images with masks that hide the facade occluding objects and is finalized by a mosaicking that employs these masks (texturing mode $\mathcal{M}_{\textbf{Mask-based}}$).

5.3.5.2 Generating Images with Masks Hiding the Occluding Objects

The facade occlusions are caused by the laser beam that is intersected by intermediary urban occluding objects such as the vegetation or streetlights. For this reason, some facade parts are missing and can be qualified as laser shadows. Amongst others, this effect is strongly visible in case of occluding trees since they clearly cause ecliptic holes in the facade point cloud.

Since the facade planes have previously been estimated, the 3D points belonging to the facade occluding objects have been extracted for each facade by retaining all the 3D points located in a neighbouring defined by the facade plane and the vehicle trajectory. Also, a set of facade images has been matched to the facade plane according to a visibility criterion. Then, the retained 3D points that occlude the facades are projected onto the respective set of facade images that have been matched to the facade planes. We recall that the acquired images are calibrated and the whole of image and laser data are georeferenced.

In addition to this stage, the 3D points are used to generate binary images (masks) with a size identical to the original acquired images in order to generate associated binary images. Then, these binary images undergo fundamental morphological operations as follows:

1. *Morphological Dilation:* A morphological dilation is applied onto each projected point in the binary images with a large circular Kernel (radius of $50 pixels$) in order to easily cover the occluding objects in the facade images (background in white and occluding object masks in black). Indeed, it is worth pointing out the fact that only sets of points (discretized representation) of the occluding objects are employed. Also, the choice of a large mask allows the reducing the lack of object information (the presence of small holes) in areas located inside the masks.

   The dilation of the binary image $\mathcal{B}$ by the structuring element (i.e., kernel) K can be defined by:

$$\mathcal{B} \oplus K = \{x \in K \mid K_x \cap \mathcal{B} \neq \varnothing\} \tag{5.6}$$

   The dilation of $\mathcal{B}$ by K can then be understood as the set of points x

covered by K when the center of K (noted K_x) moves inside objects of $\mathcal{B}$. The dilation is an operation that grows the size of binary objects. As previously mentioned, K is a large circular Kernel (radius of 50 *pixels*) that operates onto the totality of the binary image. Here, the binary objects only are scattered single points.

By applying this dilation, the scattered points become connected the one to the others and the resulting masks should cover the occluding facade objects in majority.

2. *3D Morphological Dilation:* For each detected occluding point, the point is enhanced into a cube by considering a neighboring. The aim of this stage consists in intensifying the density of points in order to cover more the occluding objects. This technique can be interesting to produce masks (image masks) that cover more the occluding objects. However, this technique is very expensive in computational time in the sense that the number of points that is dilated is multiplied by 9.
3. *Morphological Erosion:* A morphological erosion is then applied to the dilated images with a small circular Kernel (radius of 20 *pixels*) in order to reduce the amplified size of the masks at the contours.

   The eroding of the binary image $\mathcal{B}$ by the structuring element (i.e., kernel) K can be defined by:

$$\mathcal{B} \ominus K = \{x \in K_x \mid K \subseteq \mathcal{B}\} \tag{5.7}$$

   The erosion of $\mathcal{B}$ by K can then be understood as the locus of points reached by the center x of K when K moves inside $\mathcal{B}$. The erosion is an operation that diminishes the size of the binary objects. Here, K is a small disk that operates onto the totality of the previous dilated binary image. The dilation was large enough to well cover the facade occluding objects. Otherwise, the erosion is less strong in order to produce masks with shapes that fits more the initial boundaries of the occluding objects.
4. *Smoothing Mask Edges:* Finally, a smoothing is applied to the contours of the masks in order to reduce the visual effect (non-elegant) caused by the accumulation of circular Kernel patterns. In this way, the rendering of a mosaicking with facade images by subtracting the masks, i.e. reconstructed textures will be not too affect at the multiple recovered junctions.

Then, the images with masks are applied to the original images. In other words, ortho-image of facade images (with masks) are generated. Finally, a mosaicking is achieved by overlapping the facade images subtracted of the masks of occluding objects. As previously mentioned, this mosaicking is depending on the presence of laser occluding objects detected in front to the estimated facade plane.

5.4 Experimental Results and Performance Study

5.4.1 Generated Wire-Frame 3D model

Figure 5.7 shows the extraction of wire-frame model from a street cloud. At the top, the 3D wire-frame of street facades as well as the segmented facade

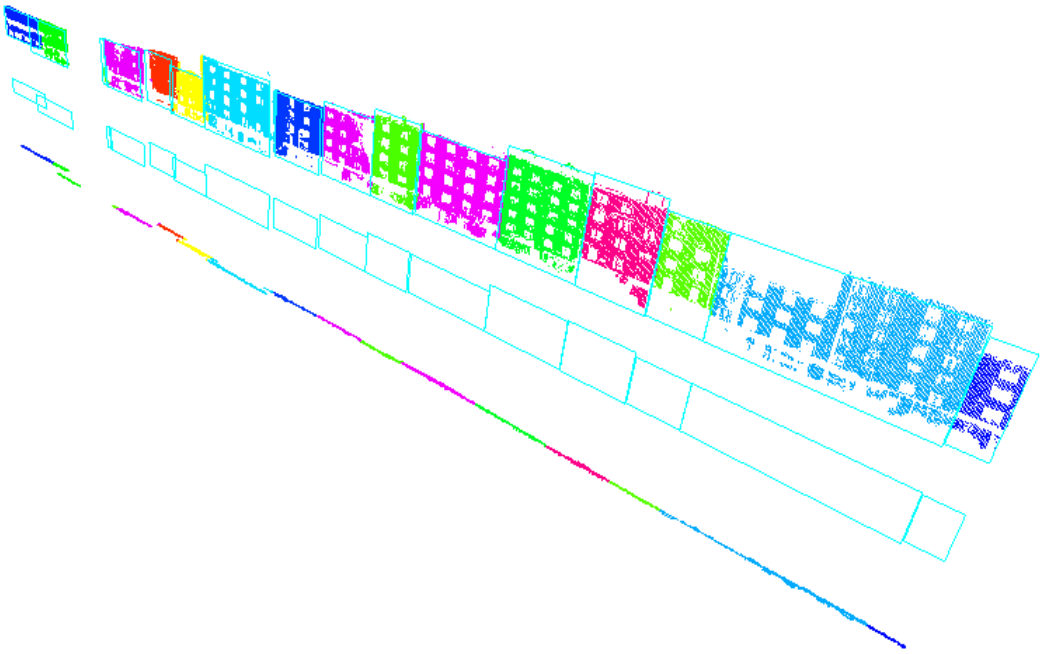


Figure 5.7: Result of a wide-scale wire-frame model generated for 15 urban facades (perspective view). (top) 3D wire-frame model and individualized cluster of street facades. The wire-frame model is delimited at the facade top with the maximum height of each cluster as described in section 5.3.1.4. (middle) The wire-frame is represented alone. (bottom) The planimetric footprints of each cluster are exhibited.

point clouds are shown together. The presented wireframe of street facades has been generated by using the geometric heuristics described in Section 5.3.2. At the middle, we visualize the wire-frame model of street facades alone. Finally, the footprint of the facade clusters are shown at the bottom.

5.4.2 Accuracy Evaluation

In this section, we have achieved qualitative evaluations of the wire-frame models of street facades by visually comparing the generated delimitations with two national standardized ground truths *i*) the corresponding image geometry of the IGN Street Viewer (see Figure 5.8) *ii*) the corresponding delimitations of the IGN aerial-based modeling pipeline (see Figure 5.9), respectively.

Thus, the 3D wire-frame model of street facades (approach 2) has been projected in the panoramic images of the IGN Street Viewer (see Figure 5.8). We recall that the delimitations of the generated model as well as the images acquired by the vehicle are georeferenced in a common coordinate system. As it can be observed, the delimitations at the bottom level pass across the vehicle since they characterize the junction between the facade bottom and the street road. The lateral facade delimitations correctly separate the facade of dwelling. These delimitations have been derived from the segments contained in the cadastral map. The estimated top delimitations correspond to the maximum height of each facade cluster. This delimitation depends on the location of the roof superstructures that can be either in the continuity of the facade or in recessed. It depends also on the narrow field of the laser sensor that can

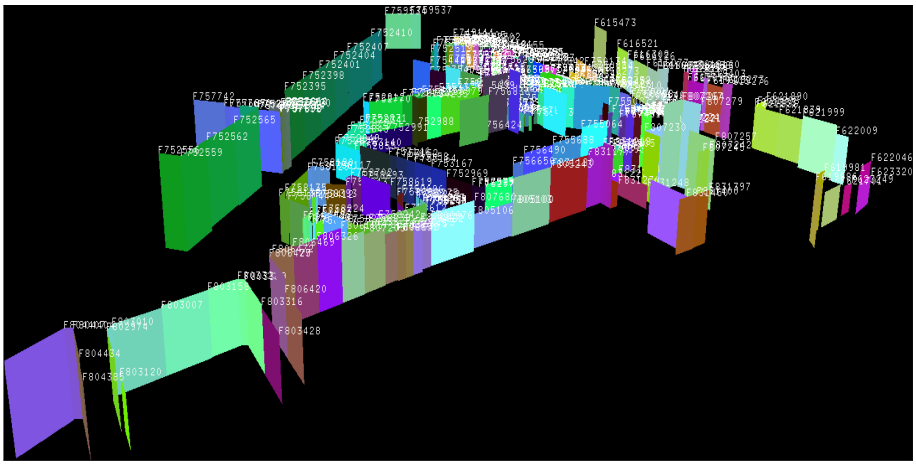
Figure 5.8: Estimated wire-frame model projected onto a panoramic image of the IGN Street Viewer.

covers or not the roof superstructures according to the laser sensor opening and to the distance of the vehicle to the facades. For these reasons, the top delimitations of the facades generally represent either the gutter limits or the top boundaries of roof superstructures (e.g. top boundaries of attic windows). Globally, the obtained 3D wire-frame models correctly segment the panoramic street image. In this way, the segmented images can be use to the texturation of the wire-frame model of street facades or more directly used in the Street Viewer as presented in Figure 5.8.

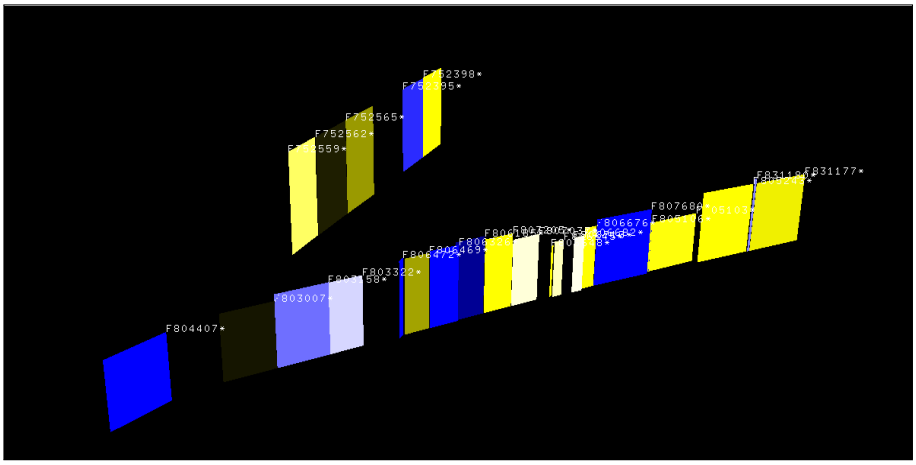
In Figure 5.9(a), a set of wire-frame models (facade models) has been generated by using the IGN aerial-based modeling pipeline. This set of wire-frame models corresponds to a part of the 12^{th} district of Paris and has been selected as a ground truth. Figure 5.9(b) illustrates a set of 25 facade models estimated from the terrestrial data (our approach). Figure 5.9(c) shows the two sets of facade models represented in a common referential. We observe that the two set of facade planes do not perfectly coincide. The overlapping of these two set of facade planes exhibits facade plane deviations.

	Terrestrial laser data
Deviation in planimetry	in (x, y)
Maximum deviation	$0.913m$
Minimum deviation	$0.014m$
Average deviation	$0.229m$

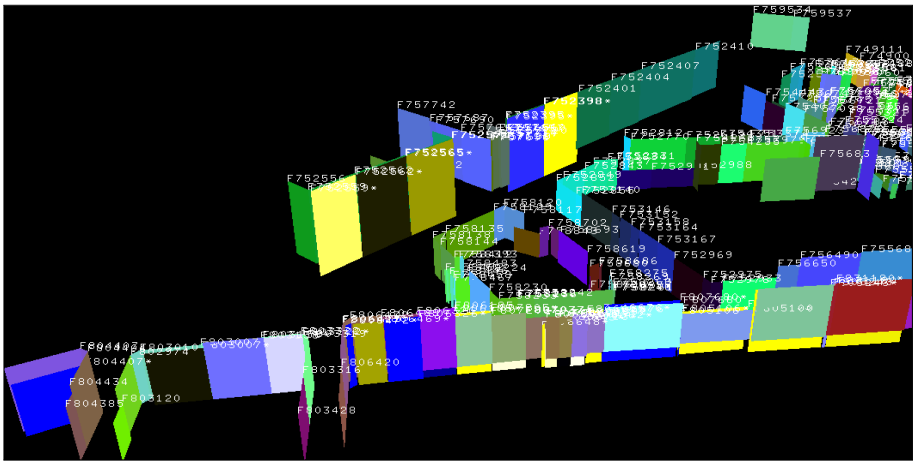
Table 5.2: Comparison between modeling results over 25 facades estimated from terrestrial laser data (best dataset) and those associated produce by the IGN aerial-based modeling pipeline.



(a) Set of facade models derived from the cadastral map. (Each facade model is labelled with an ID noted FNUMBER)



(b) Set of 25 facade models estimated from terrestrial laser data. (Each estimated facade model is labelled with an ID noted FNUMBER*)



(c) Facade models provided by the aerial modeling pipeline BATI-3D® and facade models estimated with the terrestrial laser data that are represented together in a common coordinate system.

Figure 5.9: Comparison between 25 facade models estimated from laser data and associated facade models provided by cadastral map (facade models produced by BATI-3D® prototype, i.e. IGN building modeling pipeline).

The deviations in planimetry [2] are quantified in Table 5.2. We mention the fact that the deviation in altimetry (i.e., in altitude) has not been estimated here since our laser dataset was incomplete at facade bottom and/or at facade top in the available ground truth area due to the acquisition specifications. However, we expect that the relative height of facades are estimated with high accuracy due to the accurate laser data.

In planimetry, the maximum deviation can reached $0.913m$. This deviation can occur in reason of certain cadastral-based facade segment that can be inaccurate. The average and the minimum deviation respectively are of $0.229m$ and $0.014m$. It denote then that the estimated facade planes are relatively accurate. It is worth pointing out the fact that the cadastral-based segment that are employed as ground truth for the comparison does not perfectly correspond to the reality in the sense that a facade with several withdrawals will be represented by a single support segment. Also, this ground truth is relatively biased since based on the use of a noisy cadastral map and aerial-based noisy data. Furthermore, the accuracy of these delimitations should strongly varied according to the quality and accuracy of the Global Positioning System.

More particularly, the deviations measured between 25 facade planes estimated from the terrestrial laser data and the associated facades planes coming from the IGN aerial-based modeling pipeline are also depending on some factors such as the intrinsic and extrinsic calibration accuracies of the multi-source sensors, the synchronisation between the positioning systems and the acquisition sensors (vehicle georeferencing), the cadastral-based map deformation and projection. In particular, a phenomena of deviations between two facade point clouds acquired at different times of a single facade has been illustrated in a previous Chapter.

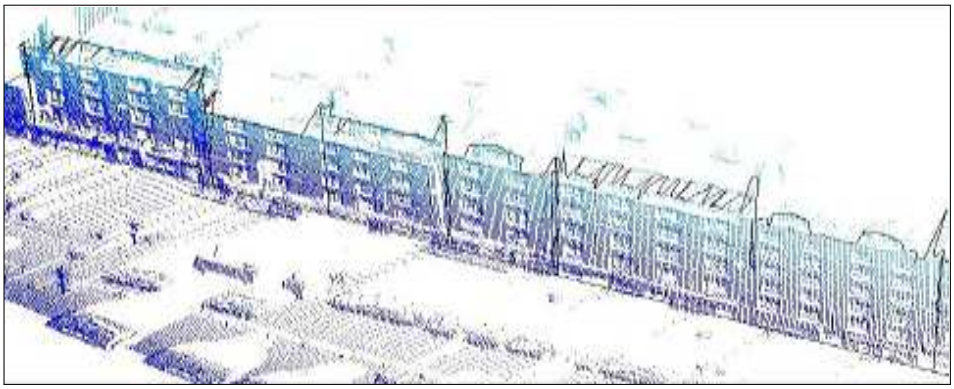
Nonetheless, the wire-frame model generated from the terrestrial data has been compared with a portion of a national standard ground truth. The global topology of the facades appears coherent and the deviation measured in this area are included in a submetric bounded interval. Also, we argue that the generated wire-frame model derived from the terrestrial laser data should be more accurate (ground-based data) since the details between the facade planes are more perceptible (from dense laser data) and since the acquisition is carried out in one single pass.

5.4.3 Performance in Enhancing Facade LOD from Laser Data

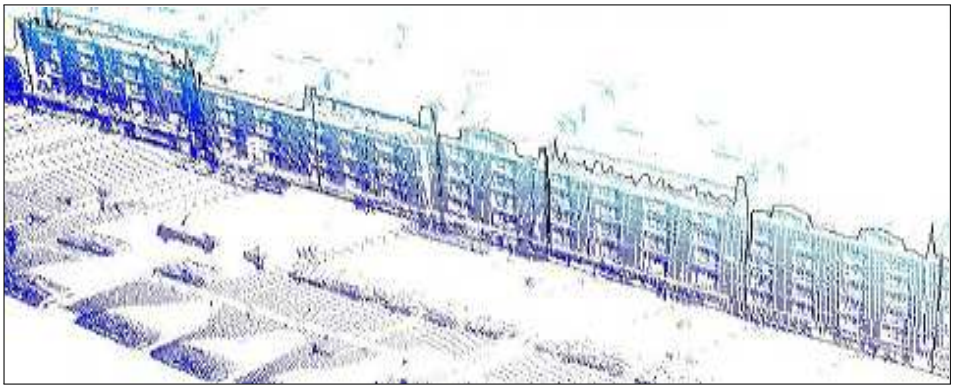
Figure 5.10 illustrates the generation of a street facade polygonal mesh by directly using the individualized cluster of facade points. As it can be observed, the delimitations of the generated model from laser data (without use of existing parametric facade model) are maladapted since the estimated shapes are too approximative and since the rendering is not sufficiently elegant and appealing for the urban visualisation. Nevertheless, these delimitations could be employed for the street facade semantization the associated regions by analysing the facade contents (e.g.; textual extraction). At this stage, the optical facade images can then be used in order to overpass the mentioned limitations. The proposed idea consists of the detection of the pixels associated to the image skyline as realistic facade top delimitations.

2. The *deviations in planimetry* correspond here to the deviations of distance in the (x, y) plane between the extremities of the facade models provided by our ground truth and those of the generated facade models.

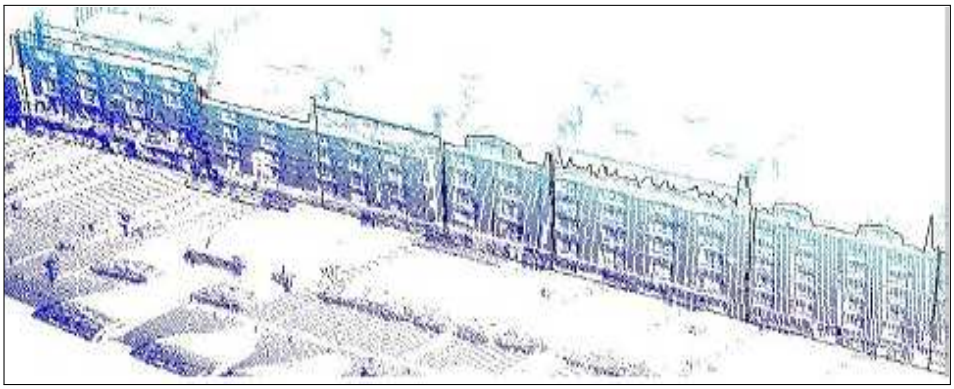
(a) Part of an external panoramic image illustrating the facade tops under study.



(b) Linking of the 3D extracted points corresponding to the facade boundaries. Various type of facade top are observed. 3D top points belonging to the attic windows, chimneys or top stages.



(c) The extracted 3D top points are projected onto the estimated plane. The points are finally linked providing a polygonal representation of the facade.



(d) Classification of the facade according to the top shape.

Figure 5.10: Generated 3D Laser-based polygons of facades. Facades with small discontinuities in height are represented by quadrilaterals. Otherwise, the facades are represented by detailed polygons.

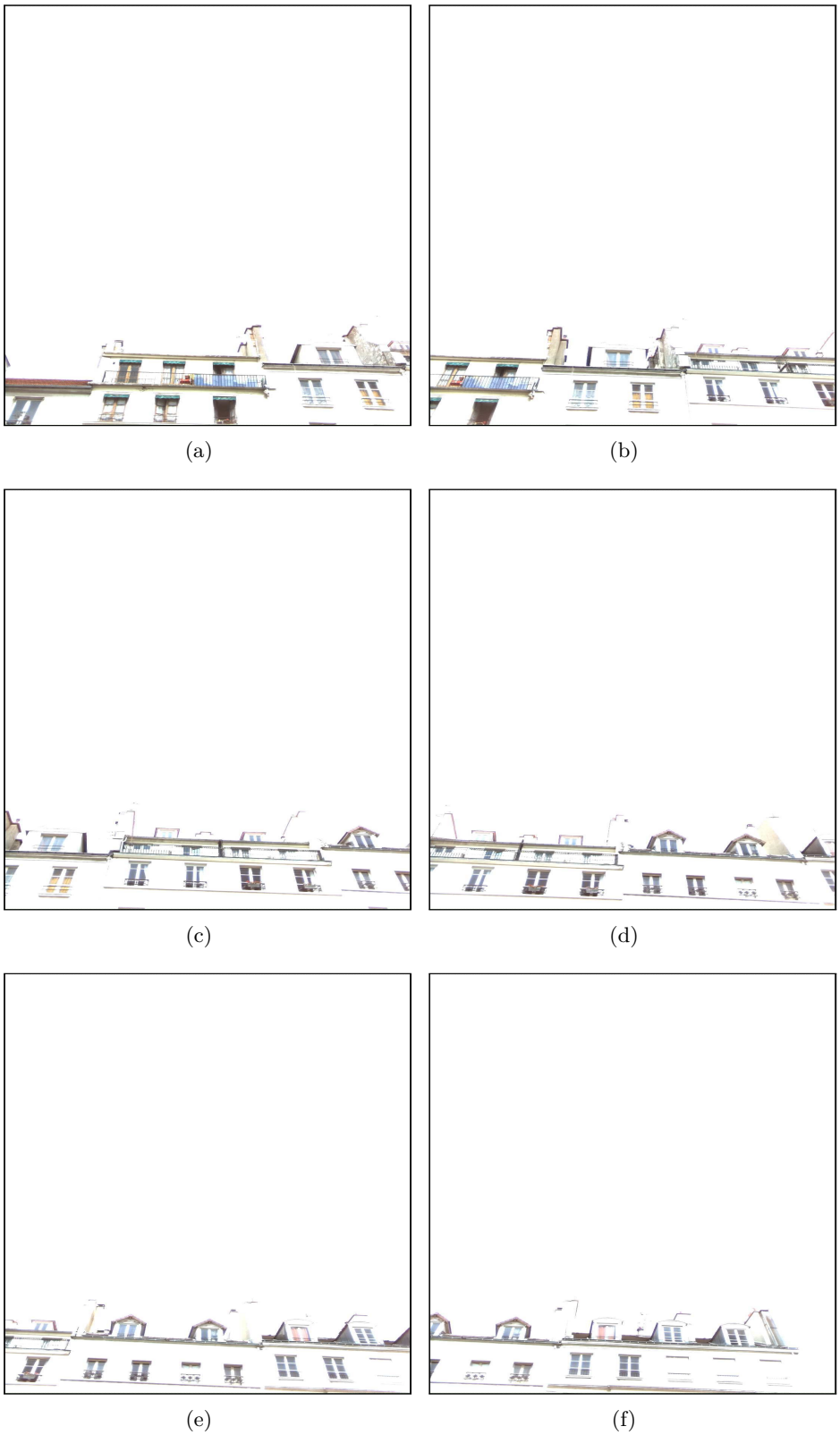


Figure 5.11: Set of facade images with overlaps acquired under gray sky condition. The size of the original images is 1920×1080.

5.4.4 Image-based Skyline Detection

As previously mentioned, the MMS allows to acquire georeferenced and calibrated multi-source terrestrial data. In addition to the Laser sensor, the MMS is equipped with ten cameras mounted at the vehicle roof. Eight cameras are oriented horizontally in a way that their field of view covers the full surrounding scene at 360°. More precisely, the cameras have been oriented to the cardinal and ordinal directions. Moreover, two cameras are laterally located to the vehicle trajectory and obliquely oriented to the facade top direction. These two cameras acquire RGB images in Full HD resolution (see Figure 5.11).

As illustrated in Figure 5.12, the skyline has been detected in very complex facade top image (cloudy condition). Indeed, the selected facade image depicts in Figure 5.12(a) exhibits a gray and white sky. After the binarization step (shown in Figure 5.12(b)), we can observe that the facade image in Figure 5.12(c) is finely segmented into sky and building region by the detected boundary of pixels.

Finally, the facade top details are exhibited in Figure 5.12(d) by superimposing the sky mask of Figure 5.12(c) with the original raw image of Figure 5.12(a). The image skyline can be converted into 3D by using the perspective line of sight passing by the detected pixels and the estimated facade planes. These stages can be achieved since the facade planes and the associated images are calibrated and georeferenced. We have not carried out experiments on converting the 2D skylines into 3D lines due to the time limitations of the thesis.

Local imperfections are observed at the level of the attic window located of the right facade (see Figure 5.12). Besides, in spite of high complexity of the selected facade image, the detected skyline reasonably well delimits the real relief for the totality of the left and central facade parts (visually). The facade top detection slightly fails at the right facade in reason of the texture of the facade top that is sometimes merged with the sky texture. Nevertheless, we argue then that the image skyline detection will be in all likelihood more robust by using set of facade images acquired under a cloudless blue sky.

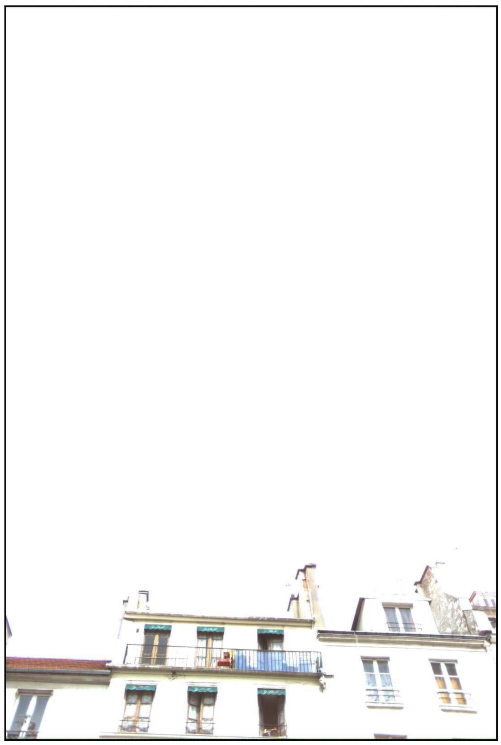
Moreover, the use of multiple facade images can reduce the perspective effect associated with the obtained 2D skyline. Indeed, from one viewpoint parts of recessed structures will be included in the estimated 2D skyline which in turn add some 3D points to the estimated 3D skyline.

5.4.5 Generated Occlusion-free Texture

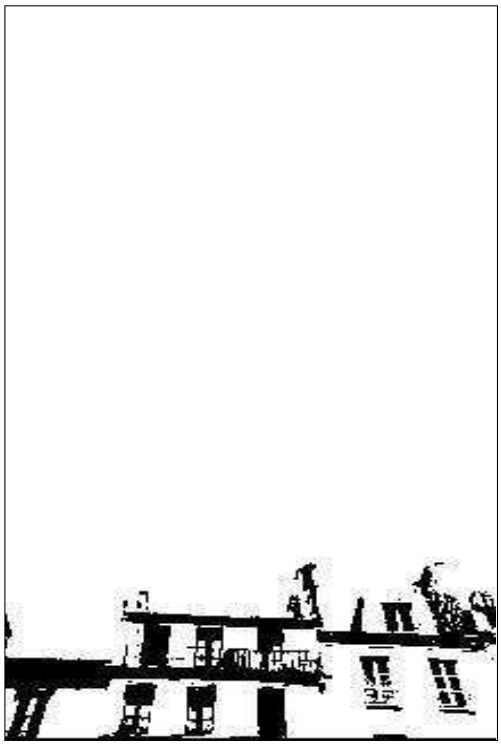
Figure 5.13 illustrates the texturing of a wire-frame model of a facade strongly occluded by the foliage of trees. A frontal image has been shown in Figure 5.13(a) in order to be aware of the problem complexity. Figure 5.13(b) depicts the occlusion-free facade texture reconstructed from a set of multiple facade images acquired by the MMS in displacements along the street.

5.4.5.1 Performance in Generating Masks for Occluding Objects

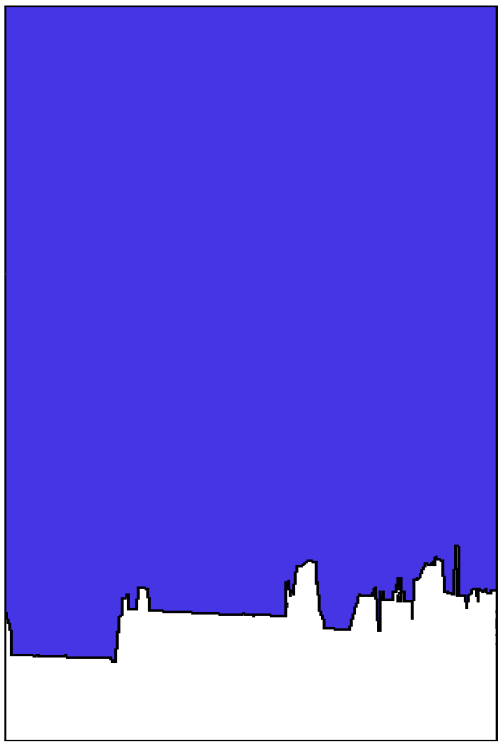
Figure 5.14 illustrates some intermediary results associated with the texturing stages. More precisely, the texturing mode is depending on the presence of urban occluding objects located between the vehicle and the facades.



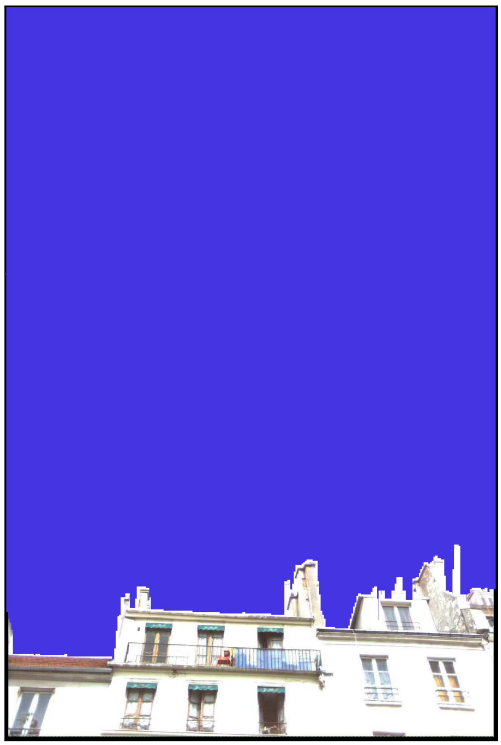
(a) Original raw image acquired in cloudy sky condition



(b) Binarized image in the blue channel



(c) Image segmented by the boundary pixels after the skyline detection (detailed relief) as proposed in (Liao et al., 2007)



(d) Artificial sky mask (background) of image (c) added to image (a) brings out the facade and the top details (foreground).

Figure 5.12: Results of skyline detection from a complex facade image. The size of the original images is 1920×1080.

(a) Frontal image facing the facade top strongly occluded by the foliage of a tree.

(b) Multiview reconstructed occlusion-free facade texture

Figure 5.13: Occlusion-free facade texture Reconstructed from a set of multiple facade images acquired by the MMS in displacements along the street. The image in Figure 5.13(a) has been used in the occlusion-free texture computation and provides an idea order on the problem complexity in case of direct facade texturing.

The detection of this occluding object is achieved by using the estimated 3D facade plane presented in Subfigure 5.14(a). If no occluding objects are detected in front of the facade model, then the texturing mode is the direct use of segmented panoramic as illustrated in Subfigure 5.14(d). Otherwise, the texturing mode employs a mask-based mosaicking. The performance in the detection of facade occluding objects is shown in Subfigure 5.14(e). We stress the fact that the sensor was oriented in direction of the facade top and the sensor opening does not cover the ground here. In Subfigure 5.14(f), a set of images with various orientation have been assigned to the estimated facade plane according to a visibility criterion. The whole of these images and the detected occluding objects will then be used to the generation of an occlusion-free textures.

Figure 5.15 illustrates the generated masks of occlusions as well as the resulting occlusion-free textures. Figures 5.15(a) and 5.15(b) depict the projection of points of occluding objects that have been previously dilated into 3D. The 3D dilation that should slightly improve the quality of the texturing is relatively time consuming since nine points are projected onto the facade image for each occluding point. Figures 5.15(c) and 5.15(d) illustrate the direct projection of the occluding points on another two images without usage of the 3D dilation (the case employed here). Figures 5.15(e) and 5.15(f) show the binary masks generated by projecting the 3D points in the associated images and passing morphological operators (images of 5.15(c) and 5.15(d)). The occluding objects are in majority covered by overlapping the binary masks to the facade images as shown in Figure 5.15(g) and 5.15(h).

Finally, the estimated facade plane is finely discretized into 3D cells and the intensities included in the non-masked facade parts are exploited as follows.

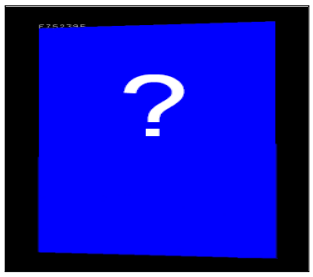
For each image associated to the facade model, the 3D points associated to the facade cells are projected onto the image including the masks (georeferenced and calibrated data). For each 3D point, if the corresponding pixel is a non-occluded facade pixel (i.e., pixel out of the masks), then its intensity (from bilinear interpolation with neighbouring pixels) is mapped onto the associated 3D cell. Otherwise, its intensity is not employed. The non-occluded facade pixels of each image directly vote in the cells and thus redundant non-occluded facade portions often are cumulated from one image to the next. The occlusion-free facade textures illustrated in Figure 5.15(i) has been generated after an elementary raw mosaicking. As it can be observed, the texture reconstitutes with success the facade by removing in majority the occluding facade objects.

As previously mentioned, a criterion of facade visibility has been initially employed in order to assign images acquired to a facade. More precisely, this criterion is the projection of the facade model 3D delimitations onto the surrounding images and the selection of the images having a sufficient intersected surface with the model are retained. This surface represents the facade part potentially meaningful for the mosaicking.

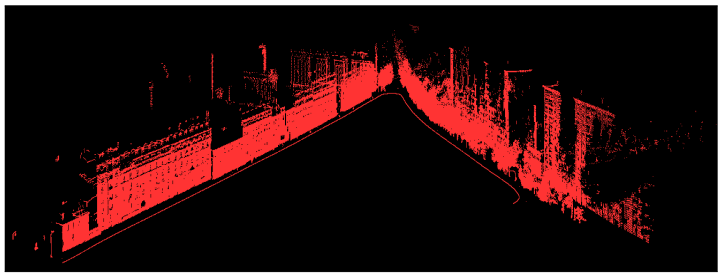
The visibility criterion has been enforced here by crossing this surface with the mask information. Hence, both of the facade model delimitations and detected occlusions are considered for the mosaicking. Finally, the images with a potential facade surface sufficiently meaningful are used.

Nevertheless, the quality of the generated occlusion-free texture can be enhanced. The number of junctions due to the overlaps should be reasonably reduced by optimizing the number of images that participate to the texturing. Also, the redundancy of non-occluded facade parts common to several images can be exploited. Moreover, the quality of the produced masks could be improved since these masks sometimes do not cover the totality of the occlusions. Indeed, the occlusions can be missed by the laser beam in reason of the laser resolution and in reason of the laser sensor orientation.

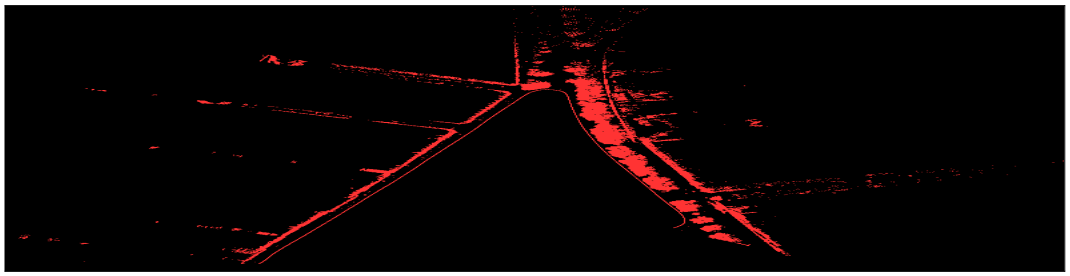
Furthermore, several on-board cameras are employed for the mosaicking. The texture quality can be strongly improved by equalizing the radiometry of the acquired images. Then, it should reduce the contrast effect (see at top-right angle in Figure 5.15(i)) between the recovered facade parts. Also, the geometry of the facade images should be improved since some misalignments lie at the junctions of the recovered facade parts (see at the bottom in Figure 5.15(i)). Finally, the effects could be attenuated by carrying out a blending at mask junctions that participate to the texture reconstitution.



(a) Estimated 3D facade model either occluded or non-occluded (hypothetically).

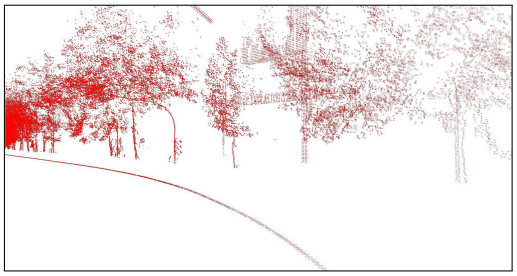


(b) Typical cases of street under study (perspective view). Case 1: street without occlusions (left side of the scene). Case 2: street with strong occlusions such as foliage of trees that cause elliptic holes at facade level (right side of the scene).

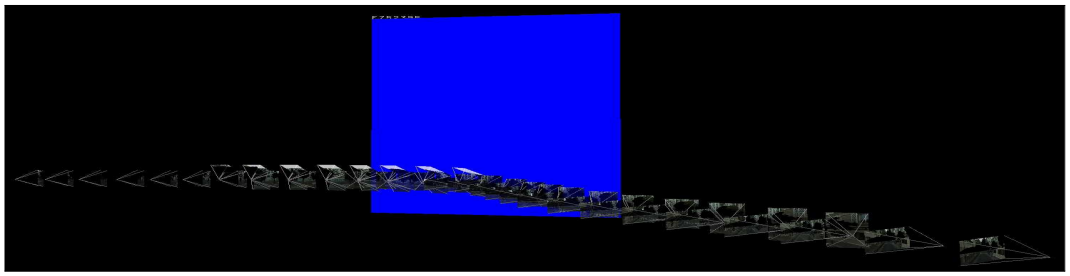


(c) Typical cases of street under study (Nadir view). Case 1: street without occlusions (left side of the scene). Case 2: street with strong occlusions (right side of the scene). Footprints of occluding trees are exhibited.

(d) Case 1: No presence of significant occlusions detected in front of the facade model. The facade image should be quasi-clean. A direct texturing can be achieved by using a segmented panoramic image.



(e) Case 2: Presence of significant detected occlusions in front of the plane illustrated in Figure 5.14(a). The facade image should be occluded. A multi-view image mosaicking should be envisaged in order to generate occlusion-free facade textures.



(f) Case 2: A subset of facade images is extracted of the facade images acquired by the MMS and assigned to the estimated 3D facade model according to a criterion of facade visibility. More than 25 points of view with various orientations for each. These images should be employed for the mosaicking.

Figure 5.14: Intermediary results of the texturing stages. The texturing mode (simple or mask-based mosaicking) is determined according to the detection of facade occluding objects.

5.5 Conclusions

5.5.1 Contribution

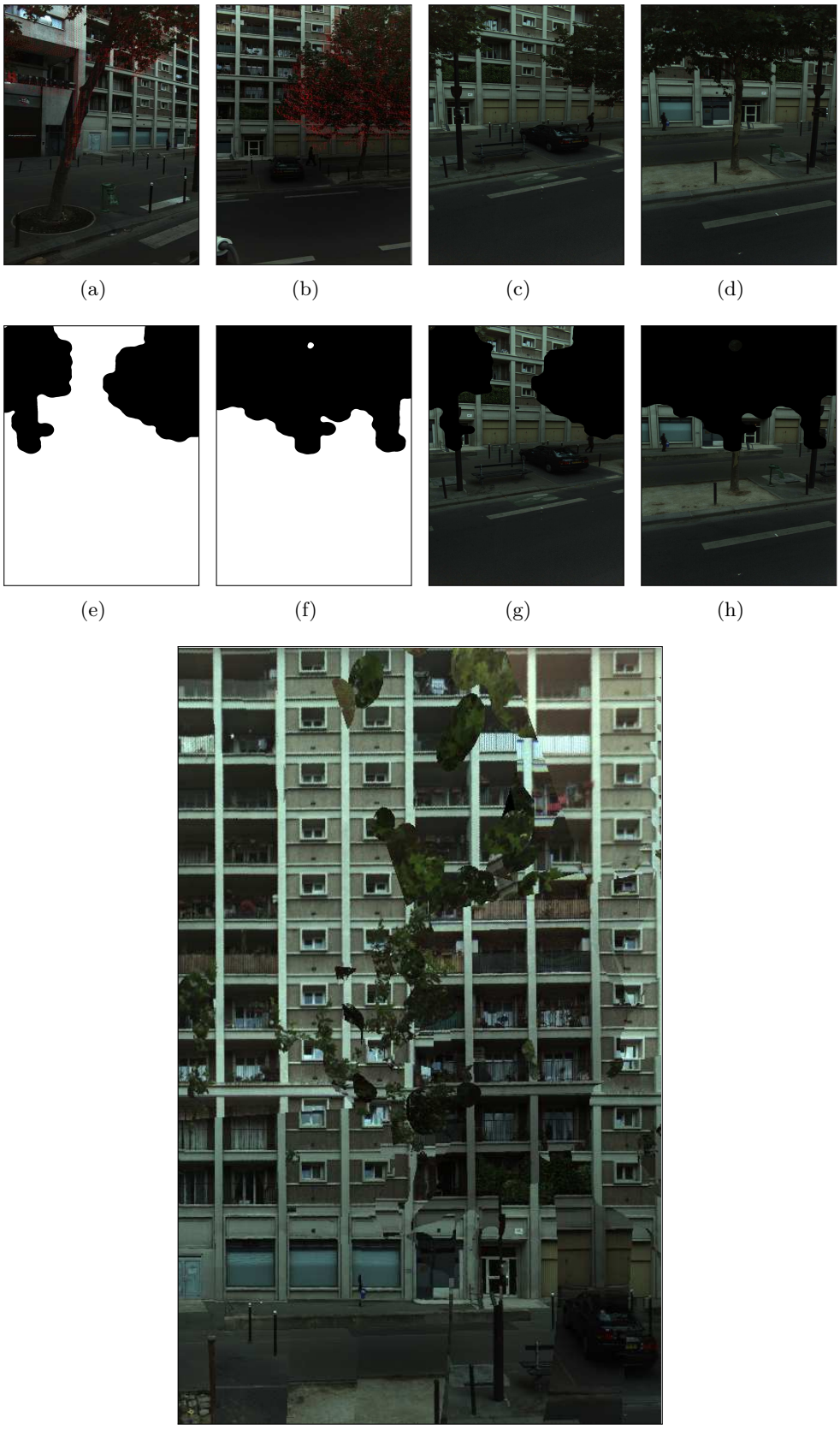
The proposed approach employs the facade clusters of points that have been extracted in Chapter 4. A methodology is proposed in order to generated a wire-frame 3D model of street facades by exploiting the terrestrial multi-source data. The quadrilateral level of detail has been reached by using the terrestrial laser data and the cadastral map. Limitations of the terrestrial data towards enhancing the level of detail of the street facades model have been exhibited. The laser data have been analysed in altimetry and are employed as a decision criterion to classify the facade according the complexity of their facade top shapes.

To overcome the limitation of the laser data for the producing of detailed facade polygon, the optical images have been integrated in the modeling processing. The detection of the skyline in the facade images have been proposed to enhance the geometry and the realism of the generated model of street facades. The skyline allows to describe in details the top shapes of Parisian facades although very varied. Also, the skyline-based 3D mapping allows the generation of realistic street rendering as visible from the sensor. Finally, a methodology has been proposed in order to texture the facades according to the presence/absence of occluding objects. Hence, the packaging of these approaches should provide a robust mapping method since laser data (plane accuracy), cadastral map (existing separators) and high resolution facade images (top detailed) are combined. Convincing results have been demonstrated on real facade images acquired by the MMS in dense urban street.

5.5.2 Future Works

This part briefly describes key ideas for the final work.

– Detecting occluded pixels in the raw 2D skyline. The process of facade top modeling is described for an occluding free facade top. In some cases, the facade top can be partially occluded (e.g., by a lighting pole). Thus, it is very useful to eliminate the erroneous pixels from the final 3D skyline. We proceed as follows. For the master image associated with each cadastral 2D segment, all the pixels of the detected 2D skyline are mapped to 3D by using the plane equation associated with the 2D segment in question (cadastral map or laser plane). Each obtained 3D point is back-projected onto a neighboring facade image (homologous pixel). A global similarity threshold can be applied to remove pixels that do not belong to the facade support plane. For example, the similarity can be quantified by the Sum of Squared Differences or Sum of Absolute Differences over a small image region. In these ways, the regions associated with the occluding objects can be filled by external pixels belonging to the other images.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i)

Figure 5.15: (top) Generated images with masks occluding the facade occluding objects. (bottom) Reconstructed occlusion-free textures after rectifying images with mask and carrying out the mosaicking.

– Directly injecting the equation planes issue from the cadastral map (existing) or laser plane (estimated) in the matrix of homography induced by a plane (Agarwal et al, 2005) in order to carry out the transfer image to image. The occlusion detection ensures that the skyline-based 3D tracing and the facade model texturing will be reliable even in complex case since a multi-image process is employed to fill holes.

– Individualize facades without the use of cadastral map. Indeed, we mention that potential researches are lead in this sense (Burochin et al., 2009, Hernández et al., 2009, Hernández, 2009, Xiao et al., 2009, Zhao et al., 2010).

– Enhancing the quality of the textures by applying some techniques of image processing as image equalizations, blending, inpainting. Also, the laser sensor orientation can be optimized in order to improve the occlusions covering. The parts of recovered facade images required to be registered the one to the others. The masks that cover the vegetation should be extendible by using region growing techniques. The vegetation could also be detect by using infra-red data. Besides, some efficient image-based recognition technique as face recognition can be crossed with the detected laser occlusions to identify the pedestrians.

Transition

Several topics of research in relation with the fields of computer vision and photogrammetry applied in the context of the city modeling have been presented through the manuscript. The first Part presents an introduction to the addressed topic, the context, the problem statement and the positioning of the topic in the research presented in the literature. The Parts two and three essentially deals with the topic of building modeling using aerial data and terrestrial data, respectively. The next Part aims at bringing a general conclusion and perspectives regarding the whole of the research that has been carried out.

Part IV

Conclusion and Future Works

Chapter 6

Conclusion and Future Works

Abstract

In this manuscript, we were interested in the building modeling by using aerial and terrestrial remotely sensed data acquired in urban environment. We have employed a set of calibrated aerial images with high resolution for the reconstruction of simple polyhedral building models. Also, we have exploited terrestrial laser raw data for the segmentation of dense urban street as well as image and laser data for the generation of facade models and occlusion-free textures. Therefore, various approaches have been developed and each approach aims at bringing a solution at a specific and well-defined problem. Besides, the terrestrial-based approaches that have been developed could be joined the one to the other to produce a terrestrial-based facade modeling pipeline. This last Chapter consists of summarizing the presented contributions as well as proposes future research tracks that could be investigated for the extension of these works.

Contents

6.1 Synthesis of Researches and Contributions

As it has been shown, the building modeling is a rich topic that has been studied under various points of view according to the needs. Besides, we have observed in our study that the terrestrial-based building modeling is few developed in comparison to the aerial-based building modeling regarding the urban environment. In particular, our interest at street level was essentially focused on the facades. Besides, it seems at first glance that the use of a Mobile Mapping System will easily simplified the building modeling problem but it appears

after a more detailed analysis that the wide scale building facade modeling from ground-based urban data is despite everything a hard problem in reason of the complexity of the dense urban environment such as seen from the street and the data specificities.

6.1.1 The Generation of Simple Polyhedral Building Models from Aerial Images

For the reconstruction of polyhedral building models, we have developed a stochastic approach that employs a Genetic Algorithm for the generation of model hypothesis (3D building structure). The model hypothesis have been verified by using a set of calibrated aerial images and the registering of building footprints by comparison with an available and master building footprint. The proposed approach avoids the conventional stages that consists of extracting features and assembling them for generating the building model. The accuracy of the reconstructed building models has been verified by comparing results with a ground truth provided by a building modeling pipeline well-known for its efficiency (a prototype of BATI-3D®). Moreover, this featureless approach has demonstrated its robustness in the generation of building model even in cases of building with roofs that include several superstructures or by using images with partial shadows or reduced resolutions.

6.1.2 The Segmentation of Urban Street in Terrestrial Laser Data

The urban street are very rich in term of urban objects. The acquisition of 3D points can be more or less controlled by calibrating and parametrizing the sensors for enhancing the visibility of specific urban objects in the scene. Nevertheless, the on-drive acquisition generally deals with a huge amount of points that are out of the object of interests. For this reason, we have developed an approach for the global segmentation of a raw street point cloud into surface points, vertical small objects, and vertical planar structures. The approach can be applied independently of the topology of points provided by the sensors (e.g.; concentric circles or 2D frames). Also, only relatively universal a priori knowledge on buildings have been employed (e.g.; verticality, planarity) Therefore, the approach is relatively generic. More particularly, the approach is essentially focused at the facade level (walls and microstructures). Here, the dimensionality of the search space is roughly reduced for fostering the low-level stages of modeling for targeted objects of the urban street scene.

6.1.3 The Facade Modeling from Terrestrial Multi-source Data

In our case, the facade modeling consists of a 3D representation of the urban facade by a 3D wire-frame model. This wire-frame model consists of an aggregation of rectangular facade meshes and provides then an epurated and synthetic 3D representation of the facades at street scale. Moreover, this representation has been chosen for its high compatibility (2D segments) with the actual standard data of urban mapping such as the cadastral map (set of 2D segments). Then, this kind of representation already deserves several existing applications for building modeling. Furthermore, the proposed modeling of urban facades by a rectangular wire-frame simplify the wide scale representation of urban areas (set of facades, street) in the sense that data processing known

for their low computational time have been employed. Besides, this level of representation has been reached by exploiting the laser data of the street scene as well as by carrying an effort in a reduced usage of the conventional cadastral map.

Here, the generated rectangular wire-frame model of urban facades brings a partial representation of building models in reason of the urban ground-based visibility. Nevertheless, the 3D facade models that are generated here are more accurate than the facade models that belong to the buildings produced by combining of aerial data and conventional cadastral map. For this reason, the rectangular representation of facades that is produced from terrestrial data provides a modeling result that are very complementary to the current aerial-based production of building models (e.g.; prismatic or polyhedral building models). The accuracy of the generated facade models have been demonstrated by comparing results with a standard cadastral map of the area under study (ground truth) and also by projecting the 3D model in a platform for the ground-based street visualization (platform iTowns).

Further detailed 3D representations have been proposed and specifically deal with the specific generation of a 3D global model of urban facades such as seen from the ground. These representations are in the spirit of the popular platform for ground-based visualization of street; such as Google Street View. The representation that has been proposed exploits terrestrial laser data as well as the images that have been conjointly acquired. We have demonstrated that the richness in the visibility of facade top details that is observed in the street images could be employed to improve the facade top delimitation. Since the facade shapes are extremely and particularly varied (e.g.; from one facade to the next) in dense urban environments and since generic libraries of parametrized urban facade models do not exist to the best of our knowledge due to the architectural complexity, then we propose the use of the skyline detected in the images as a generic representation of the facade top for the ground-based visualization of urban facades.

The two levels of facade representations (and corresponding generated facade models) respectively propose a solution to a real modeling problem; the one with respect of the physical reality (at a level of defined detail) and the other, in a exclusive context of ground-based visualization and then with possibility of scene enhancement by semi-virtual representations and by rendering. In both cases, the relationship to the current use of 2D cadastral mapping allows continuity between present and future systems.

The obtained results shows that extensions can be achieved in the urban GIS databases by including the segmentation of the street scene into independent and geolocalized urban objects. For example, certain urban objects and facades can be associated to their corresponding cluster of points as well as to their geometric model. Consequently, these categorized geometric features can be employed in various urban applications.

Besides, research has been investigated in the texturing of urban facades. The texturing stage can be seen as one of the final stages in the facade modeling process. The texturing has essentially dealt with facades that include occlusions caused by the accumulation of urban objects located between the image points of view and the facades. The methods consist of masking the occluding objects in the facade images by exploiting the terrestrial laser data. Then, each facade can be mapped by a texture removed of the occluding ob-

jects. These textures could be use for the generated wire-frame model as well as for the external 3D model of existing buildings.

6.2 Future Works

This section presents a set of investigations that could be carried out in the continuity of this work.

6.2.1 An Unified Pipeline for the Facade Modeling

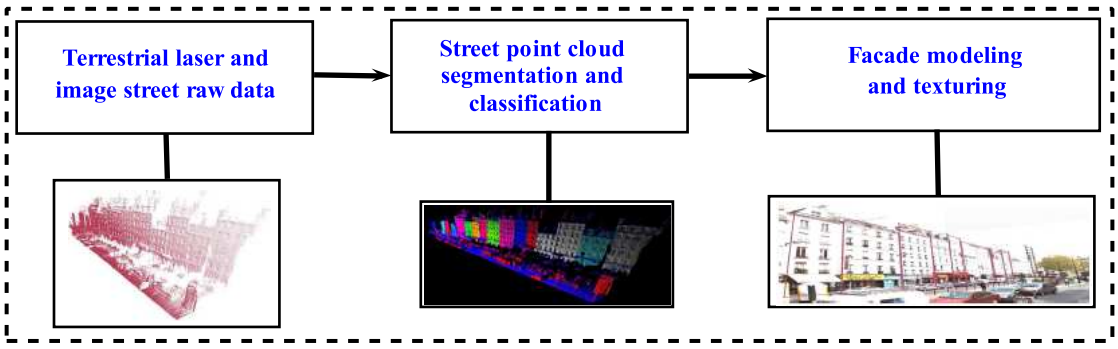


Figure 6.1: Global diagram of a complete prototype pipeline for the urban facade modeling from terrestrial multi-source data.

Some pieces of research have been addressed for the terrestrial-based facade modeling. The facade modeling can bee seen as a global term that regroups a set of elemental modeling stages such as object focusing, the generation of intermediary geometric features, the assembly of features into 3D model, the texturing. In our case, some stages have been carried out often by using data sets that are independent in the sense that the data quality or fields of view have continuously varied or evolved during the thesis. For this reason, a future work will be the unification of these stages in order to establish a complete pipeline for the terrestrial-based facade modeling from terrestrial multi-source data. The global diagram of the unified prototype pipeline for the urban facade modeling is illustrated in Figure 6.1.

6.2.2 The Full Enfranchisement of the Conventional Cadastral Map (Man-made) in the Facade Model Generation

In the terrestrial part of the manuscript, the cadastral map has been used in order to individualize the facades by residential dwelling. Indeed, we have observed that the stage of joined facade individualization was relatively complex by using only the terrestrial laser data. Besides, as it was previously mentioned, some recent approaches in the literature can be exploited for the individualization of the urban facades by using facade images (Burochin et al., 2009, Hernández et al., 2009, Hernández, 2009, Xiao et al., 2009, Zhao et al., 2010). Therefore, it means that high quality facade models could be generated with a total independence of a cadastral maps in the individualization stages. Consequently, it should be possible to generate urban facade models by using a full model-less/map-less terrestrial-based approaches; i.e. full enfranchisement in the use of the conventional Man-made cadastral map.

6.2.3 The Fusing of Generated Aerial-based Building Models and Terrestrial-based Facade Models

In the aerial part of the manuscript, we have seen that 3D building modeling required the use of a cadastral map for the focusing stage of the aerial-based building modeling pipeline. We remind that the cadatral map provides a set of 2D segments for each building. Besides, it should be also possible to automatically produce online a great part of the cadastral map such as 2D segments of facades that are visible from the street. Therefore, these generated segments could replace the corresponding segments that belongs to the conventional cadastral map. Since these generated segments are more accurate, the replaced segments should be regrouped by building in order to recompose a cadastral map. Thus, the use of this more accurate map should be directly remarked in the enhancement of the quality of the produced building models. This evolution should then mark an advancement towards the generation of full model-less/map-less hybrid aerial- terrestrial-based building modeling pipelines. For the aerial-based building modeling pipelines that avoids the use of the cadastral map, a similar fusing method should be applied in order to replace in part the facades of aerial-based building models by terrestrial-based facade models.

6.2.4 The Enhancement of LoD in the Facade Models

Here, we have generated elemental rectangular urban facade models at street scale. Then, an objective will be to enhance the level of details of the facade models by including the facade microstructures (e.g.; windows, doors, balconies). For each residential dwelling, we have previously extracted in the neighbouring of the facade planes a sets of points belonging to the openings (essentially windows points). Then, these sets of points could be exploited in various way. For example, the contours of the windows could be extracted by using the technique for contour detection in 3D points proposed in (Boulaassal et al., 2007). Then, the windows planes and their delimitations could be computed. Alternatively, the accumulations at the intersections of vertical and horizontal lines passing by these sets of points should generate hypothesis of windows regions. The distribution of these sets of points (location, periodicity) can also be studied to generate hypothetical facade models. Besides, some works such as (Burochin et al., 2009) deals with the segmentation of calibrated facade images into elemental 2D models (i.e., image regions). These 2D models can be swept until fittings with the extracted and identified sets of points. On the one hand, this data fusing should then semantize image regions. On the other hand, the image regions that corresponds to the identified windows could be reconstructed by extruding their image delimitations into 3D since the windows planes could be computed and the dominant facade wall plane have been estimated. By these ways, detailed facade models could be produced.

6.3 Work Outcomes

To put it in a nutshell, we have presented in this manuscript original approaches for the reconstruction of three-dimensional polyhedral building models from aerial images as well as a set of approaches that essentially aims at establishing a prototype of pipeline for generating the urban facade models

by using terrestrial multi-source raw data acquired at street level by a Mobile Mapping System. The presented approaches have been applied onto real world datasets and their efficiency have been demonstrated by comparison with some standard ground truths.

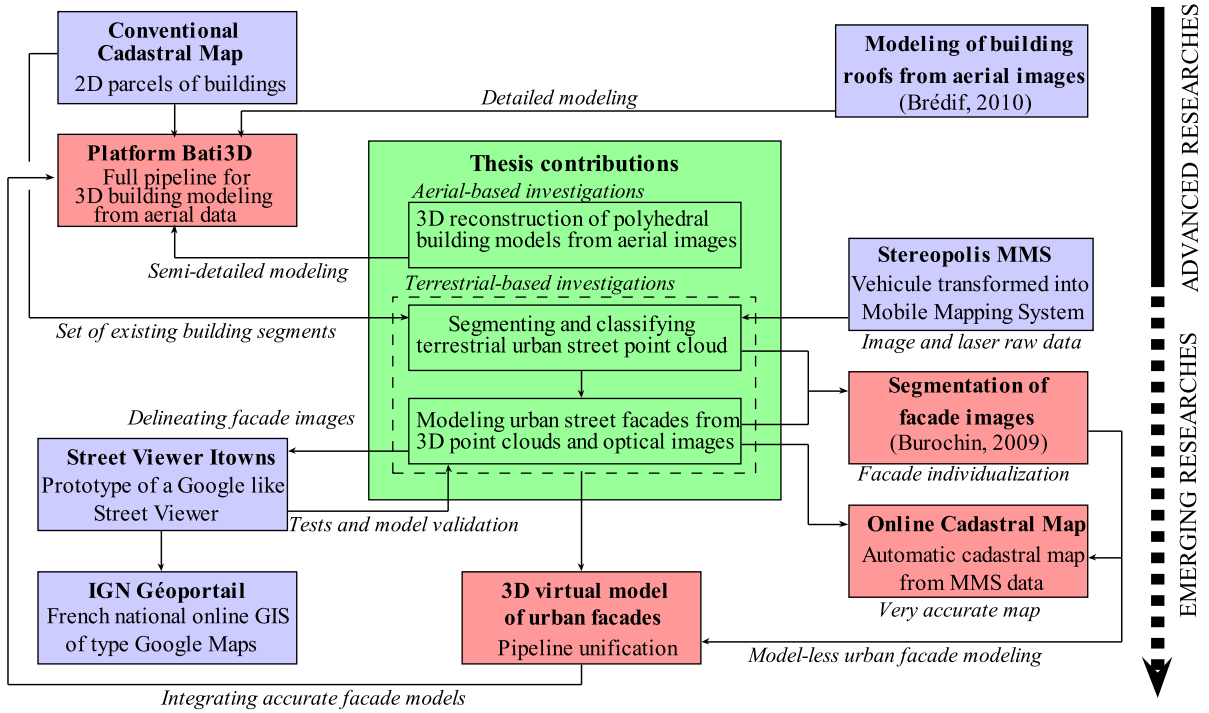


Figure 6.2: Global diagram that summarizes the thesis contributions, the connections with IGN researches, the research convergences (fusing possibilities) and the future works that could be investigated. The green cells correspond to the thesis contributions, the blue cells represent indirectly contributing surrounding works, the red cells represent future works that could be undertaken.

The state of our advancements and several research tracks have been presented for extending the current works. These investigations are represented into a global diagram (Figure 6.2) that summarized the thesis contributions, the connections with IGN researches, the research convergences and the future works that could be investigated. Some objectives could be reached in the short and medium term by following in the future works the guidelines and research tracks that have been previously emphasized.

Part V

Appendix

Appendix A

Extended Abstract in French / Résumé étendu

Abstract

Le texte présenté dans cette rubrique correspond à une retranscription en langue française de passages importants sélectionnés en partie dans les chapitres du manuscrit. Ainsi, le texte fournit une vue d'ensemble des projets de recherches environnants et décrit de manière concise certaines des recherches associées entreprises durant la thèse ainsi que des résultats obtenus. L'objectif ici est de mettre en évidence les tenants et les aboutissants dans les études et les recherches que nous avons menées; entre autres, A quoi sert la génération de modèles 3D de bâtiments? Quels sont les besoins actuels et à venir? Quelles méthodologies avons nous mis en place pour y répondre?

Contents

Depuis maintenant deux décennies, la reconstruction 3D est un sujet de recherche activement traité au sein du laboratoire MATIS de l'Institut Géographique National. Le laboratoire MATIS a pour objectif de satisfaire dans ce champs de recherche deux types de demandes. D'une part, le labora-

toire répond à un ensemble de missions émanant de l'Institut Géographique National lui-même, de missions de son ministère de tutelle (le ministère de l'équipement [1]) et de missions dites régaliennes [2]. D'autre part, le laboratoire vise à conduire des recherches d'intérêt général en étroite collaboration avec des institutions compétentes du domaine. Ainsi, les nombreux travaux de recherches menés sur divers aspects du sujet de la modélisation 3D ont conduit le laboratoire à en constituer son coeur de compétence.

A.1 Projets environnants de recherche finalisée aboutis ou émergents

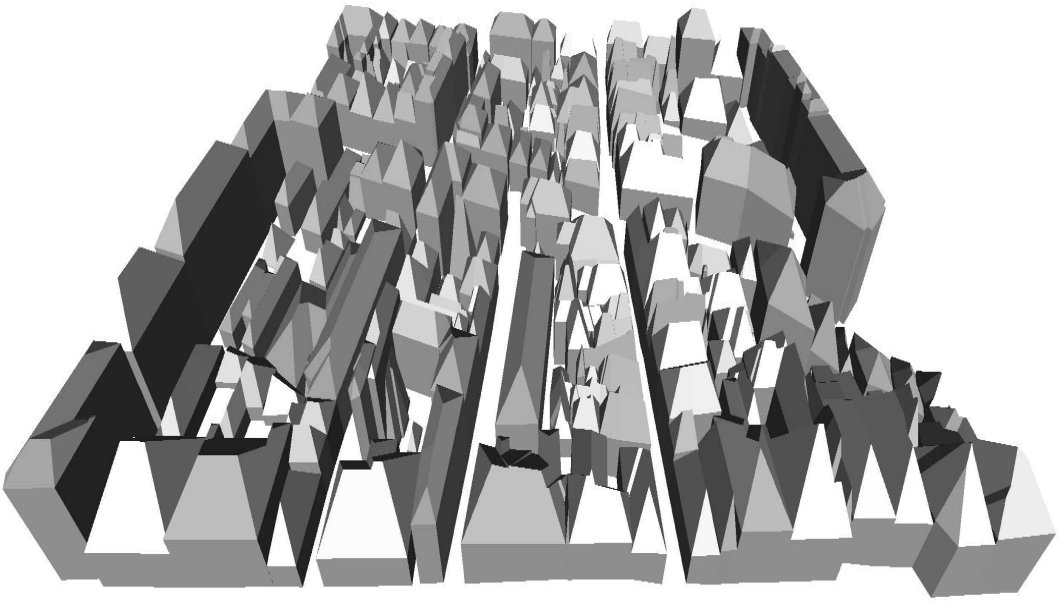


Figure A.1: Illustration de modèles de bâtiments 3D polyédriques produits à partir de données aériennes par un prototype de la chaîne de reconstruction de bâtiment de BATI-3D® (Figure extraite de (Bredif, 2010)).

Cette thèse se positionne autour de trois projets majeurs de l'IGN. Ceux-ci sont décrits dans les sections suivantes dans l'ordre chronologique d'apparition.

A.1.1 Un pipeline de modélisation de bâtiments à partir de données aériennes

En 2006, l'IGN développe une chaîne de modélisation semi-automatique de bâtiments exploitant les données aériennes et qui se nomme BATI-3D® (cf. Figure A.1). Il convient de noter que cette chaîne s'appuie sur une longue expérience en modélisation 3D du bâti; en particulier par le développement

1. communément appelé le ministère de l'équipement, la dénomination actuelle du ministère est *Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement (MEDDTL)*.

2. du latin *rex, regis*, roi est un terme emprunté à l'histoire et qui définit ce qui est attaché à la souveraineté (peuple, roi, selon les régimes politiques) (définition wikipedia). Les *missions régaliennes* désignent ici les missions spécifiques qui ne peuvent être confiées qu'à l'état.

d'une chaîne précurtrice de modélisation 3D de bâtiments mise en production par l'IGN fin 1970 (TRAPU[3]) et permettant entre autres de représenter la volumétrie des bâtiments dans les espaces urbains sous forme de modèles polyédriques. D'autre part, le projet BATI-3D® vise à capitaliser les recherches sur l'extraction automatique de bâtiment qui ont été entreprises au laboratoire MATIS depuis les années 90. Les objectifs fixés sont entre autres l'amélioration de la représentation 3D des bâtiments au niveau de la forme des toits. En particulier, un toit comportant plusieurs pans doit pouvoir être représenté par un modèle non plus à une facette mais multi-facette. Qui plus est, il est souhaité atteindre un niveau de précision des modèles reconstruits qui soit non plus métrique mais décimétrique. Les applications de ces types de modèles 3D de bâtiments sont nombreuses et variées; et on remarque une utilisation fréquente de tels modèles (souvent complétés d'un multitude d'objets du mobilier urbain) en vue de constituer en 3D des zones virtuelles de villes servant au grand public à la visualisation et la navigation ludique urbaine ou encore servant dans le milieu professionnel à la constitution de maquette 3D pour la présentation de projets d'aménagement urbains.

A.1.2 Un véhicule de cartographie mobile

Figure A.2: Système de cartographie mobile de l'IGN baptisé Stéréopolis.

En 2008, l'IGN développe un véhicule de cartographie mobile équipé d'un ensemble de capteurs extéroceptifs[4] (e.g.; caméra optiques, capteur lasers)

3. *TRAPU* désigne Tracé Automatique de Perspectives Urbaines

4. Les capteurs *extéroceptifs* ont des fonctions lié à la finalité recherchée tel que les capteurs de visions, capteurs de détection, capteurs de mesure de distance.

et proprioceptifs [5] (e.g.; capteurs de positionnement, capteurs de mesures du déplacement.). Ce véhicule de cartographie mobile sillonne les rues et permet de collecter des données urbaines terrestres géoréférencées de façon massive. Ce véhicule fournit donc des données essentielles pour mener des recherches en vue d'améliorer la modélisation 3D de villes au niveau des façades urbaines. Ce véhicule a été baptisé Stéréopolis [6] et est illustré en Figure A.2. Il s'agit de la seconde génération du véhicule Stéréopolis initialement développé en 2002. Nous soulignons que nous nous positionnons dans nos études uniquement en tant qu'utilisateur des données acquises par ce véhicule.

A.1.3 Une plate-forme de visualisation avancée de rues au travers des images panoramiques

Figure A.3: Visualiseur de rue IGN de type Google Street-View dénommé Itowns. La figure illustre une scène de rue obtenue par assemblage des images optiques acquises par le véhicule de cartographie mobile et qui constitue des images panoramiques hautes résolutions utiles pour la visualisation et la navigation de rues à 360°.

En 2008, l'IGN amorce le développement d'une plate-forme de visualisation de rues de type Google Street-View qui exploite le flux d'images hautes résolutions et géoréférencées acquis par le véhicule. Plus précisément, les images acquises sont assemblées pour constituer des images panoramiques et fournissent ainsi une visualisation des scènes de rues à 360°appropriée à la navigation immersive urbaine (cf. Figure A.3). Au delà de l'aspect visualisation,

5. Les capteurs *proprioceptifs* recueillent des informations sur l'état du véhicule tel que la localisation ou la vitesse

6. Stéréopolis est la contraction de deux mots grecs; à savoir *stereo* qui fait référence à la stéréoscopie et *polis* qui signifie la ville.

nous soulignons que cette plate-forme intègre en complément un moteur de recherche novateur dans le sens ou; en plus des informations liés aux bases de données d'adresse conventionnellement exploitées; le moteur prend en considération des informations extraites du contenu de la scène à partir des données multi-sources acquises par le véhicule ainsi que des informations déduites par agrégation avec des bases de données existantes variés en vue d'augmenter le niveau d'information délivré à l'utilisateur à travers l'accès à une scène ou dans les indications fournies pour la navigation. Parmi les extensions d'accès; on peut notamment mentionner l'identification des mots dans la scène par des algorithmes d'analyse d'images ainsi que leur référencements dans une base de donnée associée aux objets du mobilier urbain identifiés ou reconstruits. Ainsi, les internautes peuvent accéder plus directement à un lieu précis (e.g.; magasins, commerces) en fonction de la requête textuelle saisie. Concernant les indications de navigation; un enrichissement plus directement lié au thème de notre étude pourrait être l'affichage des cotes d'un bâtiment (e.g.; hauteur de la façade, dimensions des ouvertures). Cette plate-forme de visualisation enrichie est dénommée iTOWNS [7] et a été employé dans le cadre de nos travaux en vue de mener des expérimentations et des évaluations de nos approches au niveau de façades urbaines.

A.2 Synthèse de contributions marquant un avancement significatif des recherches

Les contributions de la thèse ainsi que les interactions entre nos recherches et celles menées dans les différents projets précédemment décrits ont été précisément présentées dans un schéma récapitulatif illustré en Figure 6.2 du Chapitre de conclusion de la thèse. Dans les prochaines sections, nous exposons strictement les contributions majeures; à savoir :

1. La modélisation directe de formes polyédriques simples de bâtiments à partir d'images aériennes calibrées.
2. La segmentation de nuage de points brut de paysages urbains acquis par un système de cartographie mobile.
3. La modélisation 3D de façades urbaines à l'échelle de rues à partir de nuages de points lasers segmentés.
4. Le texturage de façades de rues urbaines occultées à partir de données multi-sources images et laser.

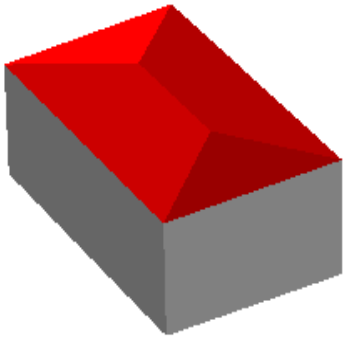
A.2.1 La modélisation directe de formes polyédriques simples de bâtiments à partir d'images aériennes calibrées

Le développement de nouvelles approches de reconstruction 3D de bâtiments à partir d'images aériennes est encore de grand intérêt. De nombreuses chaînes de reconstruction informatisées ont été développées dans le passé mais bien que les degrés de généralisation sont souvent élevés, il arrive que les processus de reconstruction établis dans les approches échouent dans certains cas. Cette déficience est en général lié à la qualité des données introduites en entrées

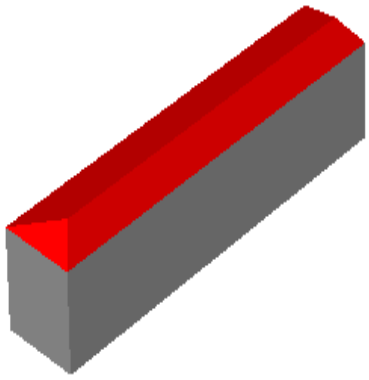
7. iTOWNS est un acronyme signifiant *image-based Town On-line Web Navigation and Search Engine*

(a) Image aérienne visualisée d'un premier point de vue

(b) Image aérienne visualisée d'un second point de vue



(c) Modèle de bâtiment avec un toit à 4 pans.



(d) Modèle de bâtiment avec un toit à 3 pans.

Figure A.4: Modèles 3D polyédriques de bâtiments reconstruits à partir d'images aériennes calibrées sans requérir aux étapes fréquentes d'extraction et d'assemblage de primitives géométriques où à l'utilisation d'un Modèle Numérique d'Elévation.

des approches qui peuvent être relativement bruitées, incomplètes ou bien due à la forme de bâtiments qui peut être complexe. Par conséquent, l'ensemble des chaînes de reconstruction informatisées nécessite encore à l'heure actuelle de la main d'oeuvre qualifiée afin de rectifier manuellement les modèles de bâtiment mal reconstruits. Comme pour beaucoup de chaînes informatisées, c'est également le cas pour la chaîne de reconstruction IGN BATI-3D®.

Ainsi, une nouvelle approche a été proposée pour permettre de reconstruire des modèles 3D polyédriques de bâtiments à partir d'images aériennes calibrées. Dans une grande majorité des cas, les modèles 3D de bâtiments actuels sont produits en appliquant trois grandes étapes qui sont la focalisation qui consiste à sélectionner l'empreinte du bâtiment dans les données, l'extraction de primitives qui consiste à extraire des primitives géométriques intermédiaires à la reconstruction (e.g., segments 2D, droites 3D, plans 3D) et enfin la modélisation qui vise à assembler ces primitives pour constituer un modèle 3D complet de bâtiment.

Par ailleurs, les données insérées en entrée de cette chaîne de reconstruc-

tion informatisée incluent souvent un pseudo cadastre employé pour initialement sélectionner les empreintes de bâtiments dans les images aériennes (étape de focalisation). Plus précisément, la carte actuellement employée résulte d'une transformation de la carte cadastrale standard en une carte vectorielle d'empreintes de bloc de bâtiments. La carte vectorielle employée est le résultat d'un ensemble d'étapes croisées ou enchaînées successivement tel que les relevés terrains de géomètres, la numérisation de cartes papiers en données cartographiques rasterisées, la vectorisation des cartes numérisées (segments 2D de parcelles d'habitations), la simplification de cartes vectorielles (segments 2D de bloc de bâtiments). Il apparaît que cette étape de simplification cause souvent des sur-détections et sous-détections de segments et nécessitent la encore une étape de rectifications manuelle.

Qui plus est, les Modèles Numériques d'Elevation (MNE) sont fréquemment employés comme données supports pour la modélisation 3D du bâti. Le MNE est habituellement généré par un processus de multi-correlation appliqué à des images aériennes. Cependant, le MNE est connu pour être imprécis au niveau des fortes discontinuités (jonction entre le sol et le toit). Dans ce cas, l'utilisation de ce type de données intermédiaires et imprécis affecte inévitablement les modèles 3D de bâtiments résultants.

Dans notre cas, l'approche présentée vise à développer un axe de modélisation polyédrique de bâtiments sans requérir aux étapes d'extraction de primitives et d'assemblages. En particulier, ces étapes délicates peuvent être la cause de nombreuses imprécisions et affecter le modèle 3D de bâtiment résultant. Egalement, notre approche n'emploie pas de données intermédiaires tels que les MNE. Notre approche procède initialement à la paramétrisation d'un modèle générique de bâtiment qui emploie un processus d'évolution différentielle pour déformer le modèle et lui faire épouser progressivement la forme réelle du bâtiment ciblé. Ce mécanisme est alors guidé par un algorithme génétique et par un processus d'optimisation qui fournit une validation globale du modèle en vérifiant directement sa consistance dans les images aériennes calibrées. Les résultats obtenus démontrent le potentiel de l'approche proposée pour la reconstruction de modèle 3D polyédrique de bâtiments (cf. Figures A.4(c) et A.4(d)).

A.2.2 La segmentation de nuage de points brut de paysages urbains acquis par un système de cartographie mobile

Par ailleurs, on dénote encore très peu de travaux qui décrivent des chaînes informatisées de reconstruction 3D hybrides aérien/terrestre. Entre autres, les raisons sont : (*i*) la difficulté d'accessibilité aux données, (*ii*) la complexité des formes de rues (environnements hétérogènes), et (*iii*) la difficulté à fusionner des données qui ont des échelles différentes. Etant donné que les formes de bâtiments apparaissent différentes à des résolutions centimétriques et décimétriques, la fusion de données implique des problèmes de représentation et de généralisation.

Cependant, la fusion de ces données s'avère indispensable si l'on souhaite enrichir davantage la qualité des modèles de bâtiments produits par la chaîne existante de reconstruction informatisée de l'IGN. En effet, celle-ci n'exploite actuellement que des données aériennes et l'exploitation des données terrestres pourrait mener à terme à un enrichissement des modèles 3D de bâtiments au niveau des rues par intégration dans la chaîne de processus efficaces de

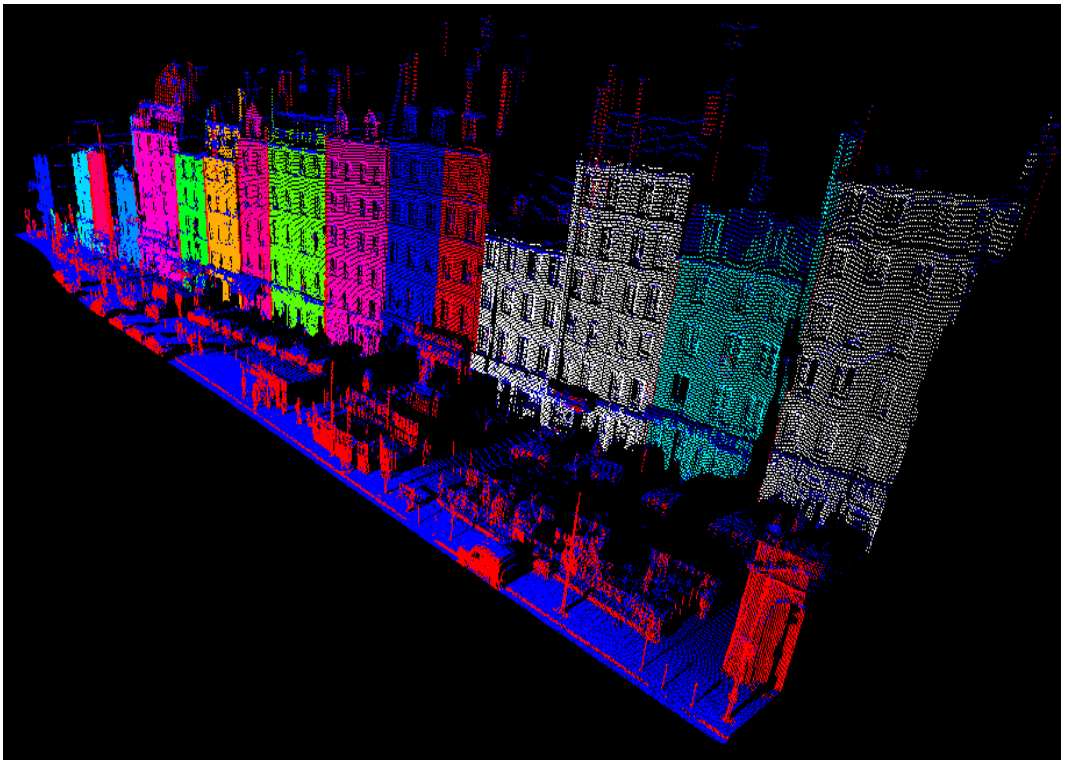


Figure A.5: Nuage de points d'un paysage urbain classifié en points de surfaces (e.g., le sol), de façades de bâtiments, et d'objets urbains verticaux (e.g., lampadaires, poteaux urbains)

modélisation et de texturation des rues. Bien entendu, cet objectif nécessite en amont de développer progressivement les connaissances, les expérimentations et les méthodologies sur la modélisation grande échelle depuis la rue ainsi que leurs validations.

En premier lieu, les recherches au niveau terrestre ont d'abord porté sur la segmentation et la classification des nuages de points bruts de paysages urbains denses afin d'en extraire des sous-ensembles de points identifiés d'objets urbains. Ces étapes qui interviennent à un niveau bas des processus de modélisation sont essentielles en vue de distinguer automatiquement les divers objets présents dans les scènes urbaines. Ainsi, le problème de modélisation peut être subdivisé en plusieurs sous problèmes correspondant à la modélisation d'objets urbains par catégorie. En Figure A.5, on peut notamment observer un résultat de segmentation et classification de données laser d'un paysage urbain acquis dans une zone du $12^{ème}$ arrondissement de Paris. On remarque que le nuage de points d'une rue urbaine dense a été subdivisée en plusieurs sous-ensemble de points correspondant aux catégories suivantes, les points de surfaces (e.g., le sol), les points de façades urbaines individualisés au sens résidentiel, les points d'objets urbains verticaux (e.g., les lampadaires, les poteaux urbains, les véhicules stationnés). Le résultat observé correspond à une des approches qui a été établie qui repose sur la génération et l'exploitation d'une carte d'accumulation de la scène ainsi que sur l'exploitation de données cadastrales. On constate entre autres que la carte cadastrale actuelle conventionnelle usuellement employée dans les processus de modélisation de bâtiment à partir de donnée aérienne est aussi fortement utile et suffisamment fiable pour

permettre de segmenter efficacement des données laser au niveau des façades de rues urbaines. Cette solution fournit un bon compromis dans l'utilisation stratégique des données pour assurer une continuité entre les systèmes de modélisation actuels et futurs. De plus, nous mentionnons qu'une seconde approche de segmentation a été proposé et celle-ci réduit l'utilisation du cadastre et des solutions sont proposés en vue de s'affranchir à terme du cadastre conventionnel par l'exploitation des données images et lasers acquises par le véhicule. Nous soulignons le fait que nous nous sommes volontairement imposés comme contrainte de nous focaliser initialement sur l'exploitation des données laser comme seule source de données cartographiques pour nos recherches.

A.2.3 La modélisation 3D de façades urbaines à l'échelle de rues à partir de nuages de points lasers segmentés

Figure A.6: Modèle en fil de fer 3D de façades urbaines estimé à l'échelle de la rue à partir des données lasers et projetés dans une image panoramique de la scène associée.

En Figure A.6, on observe la reconstruction 3D d'un maillage en fil de fer de façades urbaines à l'échelle d'une rue. Le résultat de modélisation observé a été obtenue en exploitant les nuages de points 3D associés aux façades résultant de la segmentation et de la classification de nuage de points de rues urbaines précédemment décrites. Le maillage 3D représenté au niveau de la rue est composé d'un ensemble de modèle 3D rectangulaires de façades. Cette représentation a été choisi puisque nous cherchons en premier lieu à comparer les délimitations 2D estimés de façades avec les support 2D associés existants appartenant à la carte cadastrale standard. Par ailleurs, étant donné que les modèles polyédriques de bâtiments actuellement reconstruits à partir de données aériennes sont principalement des modèles 3D polyédriques composés de façades de forme rectangulaire, les modèles de façades rectangulaires estimés à

l'échelle de la rue à partir des données laser terrestres conservent de ce point de vue une compatibilité au niveau des représentations. Celles-ci pourrait faciliter à terme la fusion des modèles de bâtiments produits en aérien avec les modèles de façades produits au niveau terrestre en vue de constituer des modèles complets et plus précis hybrides aériens/terrestres.

Par ailleurs, indépendamment à la modélisation hybride aérienne/terrestre, nous avons proposé dans le manuscrit les prémisses d'une seconde méthodologie en vue de générer des modèles de façades terrestres spécifiques à la navigation depuis la rue en exploitant les données terrestres laser et images. Il s'agit donc de modèles 3D de façades davantage appropriés à la navigation au niveau de la rue un peu dans l'esprit de la plate-forme de visualisation Google Street View.

A.2.4 Le texturage de façades de rues urbaines occultées à partir de données multi-sources images et laser

(a) Image frontale d'une façade urbaine fortement occultée par les feuillage d'un arbre.

(b) Texture de façade sans occultations reconstruite par une méthode multi-vue

Figure A.7: Texture de façade sans occultations reconstruite à partir d'un ensemble d'images de façades acquises par un système de cartographie mobile en déplacement le long des rues. L'image visualisée en Figure 5.13(a) a été employée dans le calcul de la texture et fourni une idée sur la complexité du problème en cas d'une texturation directe.

Enfin, des recherches ont été menés au niveau du texturage des façades. Des problèmes majeurs se posent lorsque les façades sont fortement occultés et bon nombre d'approches actuelles plaquent sur les modèles de façades des textures comportant des objets comme la végétation ou les poteaux. Ces occultations viennent du fait que des objets dans la scène sont situés entre le capteurs d'acquisition et les façades. On remarque que ce phénomène est important en milieu urbain dense. Nous avons alors proposé une approche qui consiste à exploiter les données laser pour détecter les objets occultant en avant-plan des façades estimés. Ainsi, les données laser fournissent le positionnement des

occultations dans la scène. Comme les données sont géoréférencés et que les images sont acquises conjointement aux données laser, les points occultant sont projetés dans les images acquises et des techniques sont employées en vue de masquer les occultations visibles dans les jeu d'images qui visualisent au mieux les façades. Finalement, pour chaque façade, les parties visibles sont regroupés par un processus de mosaïquage multi-images permettant ainsi de reconstruire une texture dépourvue d'occultation. Un cas complexe de façade à texturer est illustré en Figure A.7(a) dans le sens où une photographie frontale de la façade montre un arbre feuillu qui recouvre en grande partie la texture réelle de la façade. Un exemple de reconstruction en multi-vue de la texture de la façade correspondante est illustré en Figure A.7(b). Les textures reconstruites ont l'avantage de pouvoir être employé pour texturer les modèles de façades reconstruits à partir des données terrestres ainsi que les façades des modèles de bâtiments existant actuellement reconstruits à partir des données aériennes.

A.3 Conclusion générale

Les recherches à l'IGN visent à améliorer la qualité des modèles 3D de villes en plus des cartes urbaines 2D conventionnelles en consolidant les recherches menées au niveau aérien et en développant les approches au niveau de la cartographie terrestre. Nous avons conçu et développé des méthodologies qui serviront de briques de bases (i.e.; modules) pour la modélisation 3D de façades urbaines. Bien entendu, on parle ici de modélisations urbaines au sens large; c'est à dire recouvrant des étapes pouvant aller de la segmentation et classification de données images et lasers terrestres de scènes urbaines à la modélisation géométrique ainsi qu'au texturage des façades.

Ces modules forment le prototype d'une chaîne de modélisation terrestres des façades de rues urbaines *FACADE-3D* qui pourraient être employés de manière complémentaire en l'introduisant dans la chaîne de modélisation aérienne IGN (BATI-3D®) ou alors en tant que chaîne indépendante pour répondre à des applications externes (e.g.; enrichissements 3D de jeux vidéos, études pour la navigation robotique). Par ailleurs, nous avons prouvé à travers nos travaux l'utilité et l'efficacité des délimitations de la carte cadastral conventionnelle actuelle dans le développement des approches terrestres. Nous soulignons le fait que ce type de carte est très répandue et disponible pour un très grand nombre de villes en France et à l'étranger et nous pensons que l'utilisation ce type de données peut donc être un bonne stratégie pour assurer une continuité et une pérennité entre les systèmes de cartographie actuels et futurs. Enfin, nous avons proposé des pistes de recherches afin que l'utilisation des cartes cadastrales conventionnelles puissent être réduites voir à terme non employées.

Certes, nous mentionnons que les problèmes de calibration et de recalage entre les données aériennes et terrestres qui ont été employées dans ce travail sont en dehors du champs d'étude qui a été défini initialement mais que ces problèmes nécessitent des réflexions au niveau de la compatibilité entre les données et les modèles produits. Toutefois, de nombreuses pistes indiquent que les données terrestres laser et images de très haute qualité collectées au niveau des rues pourraient être combinées afin de générer en ligne pour les parties visibles des bâtiments des cartes proches de celles du cadastre actuel et les recherches en cartographie sont en ce sens très prometteuses.

Appendix B

List of Publications and Scientific Outputs

Abstract

This part references a set of publications and communications dealing with the topic of 3D city modeling that have been carried during the period of the thesis. Conventionally, the aim of these communications was the presentation of objectives, results and led research tracks in the photogrammetry, remote sensing and computer vision communities; exchanges of ideas with researchers for the thesis progression and step-by-step validation of scientific advancements.

Contents

B.1 International Journal with Review Committee

[1] **K. Hammoudi**, F. Dornaika. *A Featureless Approach to 3D Polyhedral Building Modeling from Aerial Images.* In Sensors Journal, 10 Years Sensors - A Decade of Publishing, MDPI publisher, 11(1):228-259., 2011. [5-Year Impact Factor: 1.917 (2010)]
http://dx.doi.org/10.3390/s110100228

B.2 LNCS Book Chapter

[2] F. Dornaika, **K. Hammoudi**. *Estimating 3D Polyhedral Building Models by Registering Aerial Images.* In Advanced Concepts for Intelligent Vision

Systems, Lecture Notes in Computer Science 6475, Springer-Verlag Berlin Heidelberg publisher, ISBN-13 978-3-642-17690-6, pp. 239-248, December 2010.
http://dx.doi.org/10.1007/978-3-642-17691-3_22

B.3 International Conferences with Review Committee

[3] **K. Hammoudi**, F. Dornaika, B. Soheilian, B. Vallet, N. Paparoditis. *Generating Occlusion-free Textures for Virtual 3D Model of Urban Facades by Fusing Image and Laser Street Data.* In Proc. IEEE Virtual Reality 2012 (International conference on), California, USA, March 2012.

[4] **K. Hammoudi**, F. Dornaika, N. Paparoditis. *Generating Virtual 3D Model of Urban Street Facades by Fusing Terrestrial Multi-source Data.* In Proc. IEEE International Conference on Intelligent Environments, Nottingham, UK, July 2011.
http://doi.ieeecomputersociety.org/10.1109/IE.2011.50

[5] **K. Hammoudi**, F. Dornaika, B. Soheilian, N. Paparoditis. *Extracting Wire-frame Models of Street Facades from 3D Point Clouds and the Corresponding Cadastral Map.* In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part 3A, pp. 91-96, Saint-Mandé, France, September 2010.
http://www.isprs.org/proceedings/...

[6] **K. Hammoudi**, F. Dornaika, B. Soheilian, N. Paparoditis. *Generating Raw Polygons of Street Facades from a 2D Urban Map and Terrestrial Laser Range Data.* In Proc. SSSI Australasian Remote Sensing and Photogrammetry Conference, Alice Springs, Australia, September 2010.
http://www.15.arspc.com/proceedings/

[7] **K. Hammoudi**, F. Dornaika, B. Soheilian, N. Paparoditis. *Extracting Outlined Planar Clusters of Street Facades from 3D Point Clouds.* In proc. IEEE Canadian Conference on Computer and Robot Vision, Ottawa, Canada, June 2010.
http://doi.ieeecomputersociety.org/10.1109/CRV.2010.23

[8] **K. Hammoudi**, F. Dornaika, N. Paparoditis. *Extracting building footprints from 3D point clouds using a terrestrial laser scanning at street level.* In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 38 (Part 3/W4), pp.65-70, Paris, France, September 2009.
http://www.isprs.org/proceedings/...

[9] F. Dornaika, **K. Hammoudi**. *Extracting 3D polyhedral building models from aerial images using a featureless and direct approach.* In Proc. IAPR Machine Vision Applications, ISBN 978-4-901122-09-2, Tokyo, Japan, May 2009.
http://www.mva-org.jp/Proceedings/...

B.4 Posters at International Workshops and Selective Summer Schools

[10] **K. Hammoudi**, F. Dornaika, B. Soheilian, N. Paparoditis. *A pipeline for modeling urban street facades from terrestrial laser and image data.* In International Computer Vision Summer School, Sicilia, Italy, July 2011.
http://svg.dmi.unict.it/icvss2011/...

[11] **K. Hammoudi**, F. Dornaika, B. Soheilian, N. Paparoditis. *Building modeling using aerial and terrestrial data.* In International Computer Vision Summer School, Sicilia, Italy, July 2010.
http://svg.dmi.unict.it/icvss2010/...

[12] **K. Hammoudi**, F. Dornaika. *Extracting 3D polyhedral building models from aerial images: a featureless and direct approach.* In Workshop on Models and Images for Porous Media, Paris, France, January 2009.
http://mipomodim.math-info.univ-paris5.fr/...

B.5 Tutorials at International Conferences

[13] **K. Hammoudi**. *Extracting 3D Polyhedral Building Models from Aerial Images using a Featureless and Direct Approach.* In CRV 2010 Tutorial Day in conjuction with the Canadian Conference on Computer and Robot Vision, Tutorial 5, University of Ottawa, Canada, June 2010.
http://www.computerrobotvision.org/...

B.6 Video Presentation at National Research Days (In French)

[14] **K. Hammoudi**. *Extraction de façades à partir de nuages lasers terrestres.* $20^{ème}$ Journées de la Recherche IGN, Vidéo IGN, Saint-Mandé, France, Mars 2011.
http://recherche.ign.fr/...

B.7 National Report, Master Thesis (In French)

[15] **K. Hammoudi**. *Reconstruction 3D de bâtiments à partir d'images aériennes: une approche directe basée image.* Mémoire de Master, Université Paris Descartes, IGN, Archive du Centre de Documentation de l'Institut Géographique National, Paris, France, Juin 2008.
http://gauguin.ensg.eu/

B.8 Media Coverage

B.8.1 International Press

F. De Blomac. *Research Days at IGN: between images and vectors.* In GIS & Geomatics Intelligence, Issue 1, pp. 12-13, June 2011.
http://gis-and-geomatics-intelligence.info/...

B.8.2 National Press (In French)

F. De Blomac. *Journées de la recherche IGN : entre images et vecteurs.* In SIG la lettre, vol. 126, pp. 8-10, Avril 2011.
http://www.sig-la-lettre.com/...

Bibliography

[Abdelhafiz, 2009] Abdelhafiz, A. Integrating digital photogrammetry and terrestrial laser scanning. In *PhD thesis*, Institute for Geodesy and Photogrammetry, Technical University Braunschweig, Germany, ISBN 3-926146-18-4, 137 p., 2009. 49

[Agarwal et al, 2005] Agarwal, A.; Jawahar C. V.; Narayanan P.J. *A survey of planar homography estimation techniques*; In *Technical Report*; International Institute of Information Technology, Hyderabad, India, 25 p., 2005. 76,190

[Anguelov et al., 2010] Anguelov, D.; Dulong, C.; Filip, D.; Frueh, C.; Lafon, S.; Lyon, R.; Ogale, A; Vincent, L.; Weaver, J. Google Street View: capturing the world at street level. In *Computer*, vol. 43(6), pp. 32-38, 2010. 52,164,165

[Arlicot et al., 2009] Arlicot, A.; Soheilian, B.; Paparoditis, N. Circular road sign extraction from street level images using colour, shape and texture database maps. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, Part3/W4, pp. 205-210, 2009. 52

[Baz et al, 2008] Baz, I.; Kersten, Th.; Buyuksalih, G.; Jacobsen, K. Documentation of Istanbul historic peninsula by static and mobile terrestrial laser scanning. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, Part B5, pp. 993-998, 2008. 164

[Boulaassal et al., 2011] Boulaassal, H.; Grussenmeyer, P. Data processing of building facades recorded by vehicle-based laser scanning. In *Geomatics Technologies in the City, International Geomatics Symposium in Saudi Arabia*, 6 p., 2011. 170

[Boulaassal et al., 2007] Boulaassal, H.; Landes, T.; Grussenmeyer, P.; Tarsha-Kurdi, F. Automatic segmentation of building facades using terrestrial laser data. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, Part 3/W52, pp. 65-70, 2007. 130,201

[Bredif, 2010] Brédif, M. 3D building modeling: topology-aware kinetic fitting of polyhedral roofs and automatic roof superstructure reconstruction. In *Ph.D. thesis*, Télécom ParisTech, 250 p., 2010. 50,51,208

[Brenner, 2005] Brenner, C. Building reconstruction from images and laser scanning. In *International Journal of Applied Earth Observation and Geoinformation*, vol. 6, pp. 187-198, 2005. 69

[Bretar et al., 2005] Bretar, F.; Roux, M. Extraction of 3D planar primitives from raw airborne laser data: a normal driven RANSAC approach. In *IAPR Machine Vision and Application*, pp. 452-455, 2005. 130

[Burochin et al., 2009] Burochin, J.P.; Tournaire, O.; Paparoditis, N. An unsupervised hierarchical segmentation of a facade building image in elementary 2D-models. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol 38, Part 3/W4, pp. 223-228, 2009. 190,200,201

[Burtch, 2008] Burtch, R. History of photogrammetry. In *Notes of the Center for Photogrammetric Training*, Ferris State University, 36 p., 2008. 48

[Cahalane et al., 2010] Cahalane, C.; McCarthy, T.; Mc Elhinney; C. (2010). Mobile mapping system performance: an initial investigation into the effect of vehicle speed on laser scan lines. In *Proceeding of the Remote Sensing and Photogrammetry Society*, 8 p., 2010. 119

[Cantoni et al., 2001] Cantoni, V.; Lombardi, L; Porta, M., Sicard, N. Vanishing point detection: representation analysis and new approaches. In *International Conference on Image Analysis and Processing*, 5 p., 2001. 170

[Chen et al., 2003] Chen, J.; Chen, C.; Chen, Y. Fast algorithm for robust template matching with M-estimators. In *IEEE Transactions on Signal Processing*, vol. 51, pp. 230-243, 2003. 79

[Chen et al, 2008] Chen, L.C.; Teo, T.A.; Kuo, C.Y.; Rau, J.Y. Shaping polyhedral buildings by the fusion of vector maps and lidar point clouds. In *Photogrammetric Engineering and Remote Sensing*, vol. 74, pp. 1147-1158, 2008. 70

[Chevrier et al., 2010] Chevrier, C.; Charbonneau, N.; Grussenmeyer, P.; Perrin, J.-P. Parametric documenting of built heritage: 3D virtual reconstruction of architectural details. In *International Journal of Architectural Computing*, vol. 8(2), pp. 131-146, 2010. 106

[Dash et al., 2004] Dash, J.; Steinle, E.; Singh, R.P.; Bahr, H.P. Automatic building extraction from laser scanning data: an input tool for disaster management. In *Advances in Space Research*, pp. 317-322, 2004. 167

[De Blomac, 2011] De Blomac, F. Research Days at IGN: between images and vectors. In *GIS & Geomatics Intelligence*, Issue 1, pp. 12-13, 2011. 48

[Devaux et al., 2010] Devaux, A.; Paparoditis, N. Increasing interactivity in street view web navigation systems. In *ACM Multimedia*, pp. 903-906, 2010. 53

[Dornaika et al, 2009] Dornaika, F.; Hammoudi, K. Extracting 3D polyhedral building models from aerial images using a featureless and direct approach. In *Proceedings of the IAPR Conference on Machine Vision Applications*, pp. 378-381, 2009. 69

[Dornaika et al, 2010] Dornaika, F.; Hammoudi, K. Estimating 3D polyhedral building models by registering aerial images. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science 6475*, pp. 239-248, 2010. 69

[Dorninger et al, 2008] Dorninger, P.; Pfeifer, N. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. In *Sensors*, vol. 8, pp. 7323-7343, 2008. 69

[Douglas et al., 1973] Douglas, D., Peucker, T. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10(2), pp. 112-122, 1973. 173

[Duffaut et al., 2008] Duffaut, J.; Déliot, P.; Vaudelin O.; Blanchard, B. Pelican multispectral airborne system for stereoscopic earth surface imaging. In *Proceedings of the International Conference on Space Optics*, 5 p., 2008. 84

[Durupt et al, 2006] Durupt, M.; Taillandier, F. Automatic building reconstruction from a digital elevation model and cadastral data: an operational approach. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, Part 3, 6 p., 2006. 70,71

[Elberink et al, 2009] Elberink, S.O.; Vosselman, G. Building reconstruction by target based graph matching on incomplete laser data: Analysis and limitations. In *Sensors*, vol. 9, pp. 6101-6118, 2009. 69

[Ellum et al., 2002] Ellum, C.M. and El-Sheimy, N. Land-based mobile mapping systems. In *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 1, pp. 13-17, 2002. 49,52,110

[Fischer et al, 1998] Fischer, A.; Kolbe, T.H.; Lang, F.; Cremers, A.B.; Förstner, W.; Plümer, L.; Steinhage, V. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. In *Computer Vision Image Understanding*, vol. 72, pp. 185-203, 1998. 67,68,71

[Fischler et al., 1981] Fischler, M. A.; Bolles, R. C. Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Communication of the ACM*, vol. 24, pp. 381-395, 1981. 130

[Fuchs, 2001] Fuchs, F. Building reconstruction in urban environment: a graph-based approach. In *International Workshop on Automatic Extraction of Man-made Objects from Aerial and Space Images*, pp. 205-215, 2001. 50

[Gomes et al., 2003] Gloria, G.; Moret Rodrigues, A.; Mendes, P. Wind effects on and around L- and U-shaped buildings. In *International Conference on Urban Climate*, 4 p., 2003. 157

[Haala et al., 2008] Haala, N., Peter, M., Kremer, J. Hunter, G. Mobile LiDAR Mapping for 3D point cloud collection in urban areas - a performance test. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, Part B5, 6 p., 2008. 164

[Haala et al, 2010] Haala, N.; Kada, M. An update on automatic 3D building reconstruction. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, pp. 570-580, 2010. 70

[Habib et al, 2010] Habib, A.F.; Zhai, R.; Kim, C. Generation of complex polyhedral building models by integrating stereo-aerial imagery and lidar data. In *Photogrammetric Engineering and Remote Sensing*, vol 76, pp. 609-623, 2010. 70

[Hartley et al., 2004] Hartley, R.; Zisserman, A. Multiple view geometry in computer vision. In *Cambridge University Press*, 672 p., 2004. 76

[Hernández et al., 2009] Hernández, J.; Marcotegui, B. Point cloud segmentation towards urban ground modeling. In *Proceedings of IEEE URBAN*

GRSS/ISPRS Joint Workshop on Data Fusion And Remote Sensing over Urban Areas, pp. 1-5, 2009. 126

[Hernández et al., 2009] Hernández, J.; Marcotegui, B. Morphological segmentation of building facade images. In *Proceedings of IEEE International Conference on Image Processing*, pp. 4029-4032, 2009. 190,200

[Hernández, 2009] Hernández, J. Analyse morphologique d'images pour la modélisation d'environnements urbains. In *Ph.D. thesis, MINES ParisTech*, 190p., 2009. 190,200

[Hongjian et al, 2006] Hongjian, Y.; Shiqiang, Z. 3D building reconstruction from aerial CCD image and sparse laser sample data. In *Optics Lasers Engineering*, vol. 44, pp. 555-566, 2006. 70

[Hough et al., 1962] Hough, P. V. C. Method and means for recognizing complex patterns. In *U.S. Patent 3,069,654*, 6 p., 1962. 131

[Huang et al., 2010] Huang, X.; Cheng, X.; Zhang, F.; Gong, J. Side ratio constrain based precise boundary tracing algorithm for discrete point clouds. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, Part B3b, pp. 349-354, 2008. 161

[Jaynes et al, 2003] Jaynes, C.; Riseman, E.; Hanson, A. Recognition and reconstruction of buildings from multiple aerial images. In *Computer Vision and Image Understanding*, vol. 90(1), pp. 68-98, 2003. 71

[Jibrini et al, 2000] Jibrini, H.; Pierrot-Deseilligny, M.; Paparoditis, N.; Maître, H. Automatic building reconstruction from very high resolution aerial stereopairs using cadastral ground plans. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 33, pp. 16-23, 2000. 50,67,70,72

[Kada et al, 2009] Kada, M.; McKinley, L. 3D building reconstruction from LiDAR based on a cell decomposition approach. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, pp. 47-52, 2009. 70

[Kang et al., 2007] Kang, Z.; Zhang, Z.; Zhang, J.; Zlatanova, S. Rapidly realizing 3D visualization for urban street based on multi-source data integration. In *Geomatics Solutions for Disaster Management, Lecture Notes in Geoinformation and Cartography, Springer-Verlag publisher*, pp. 149-163, 2007. 165

[Kasser et al., 2002] Kasser, M.; Egels, Y. In *Digital Photogrammetry*, London, Taylor Francis, 351 p, 2002. 48

[Khoshelham, 2005] Khoshelham, K. Region refinement and parametric reconstruction of building roofs by integration of image and height data. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, pp. 3-8, 2005. 70

[Lafarge et al, 2006] Lafarge, F.; Descombes, X.; Zerubia, J.; Deseilligny, M. A parametric model for automatic 3D building reconstruction from high resolution satellite images. In *Research Report of INRIA*, vol. 5687, 29 p., 2006. 70,72

[Lafarge et al, 2008] Lafarge, F.; Descombes, X.; Zerubia, J.; Deseilligny, M. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, pp. 365-381, 2008. 70

[Lafarge et al, 2010] Lafarge, F.; Descombes, X.; Zerubia J.; Deseilligny M. Structural approach for building reconstruction from a single DSM. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 135-147, 2010. 50,67,71,72

[Ledoux et al., 2009] Ledoux, H.; Meijers, M. Extruding building footprints to create topologically consistent 3D city models. In *Urban and Regional Data Management*, CRC Press, pp. 39-48, 2009. 160

[Liao et al., 2007] Liao, H. C.; Yu C. C.; Lo C. H.; Liao, W. L.; Lee P. C. A robust technique for capturing stationary direction of mobile device and its application to scenic point presentation. In *WSEAS Transactions on Information Science and Applications Journal*, 6 p., 2007. 173,184

[Liberge et al., 2010] Liberge, S.; Soheilian, B.; Chehata, N.; Paparoditis, N. Extraction of vertical posts in 3D laser point clouds acquired in dense urban areas by a mobile mapping system. In *International Archives of Photogrammetry Remote Sensing and Spatial Information*, vol. 38, Part 3B, pp. 126-130, 2010. 52

[Madhavan et al, 2006] Babu Madhavan, B.; Wang, C.; Tanahashi, H.; Hirayu, H.; Niwa, Y.; Yamamoto, K.; Tachibana, K., Sasagawa, T. A computer vision based approach for 3D building modelling of airborne laser scanner DSM data. In *Computers, Environment and Urban Systems*, vol. 30, pp. 54-77, 2006. 70

[Matas et al., 1998] J. Matas, C. Galambos, J. Kittler, Progressive Probabilistic Hough Transform, In *Proceedings of the British Machine Vision Conference*, pp. 256-265, 1998. 130,131

[Moosmann et al., 2009] F. Moosmann, O. Pink, C. Stiller, Segmentation of 3D lidar data in non-flat urban environments using a local convexity criterion. In *Proceedings of IEEE Conference on Intelligent Vehicles*, pp. 215-220, 2009. 126

[Ning et al., 2009] Ning, X.; Zhang, X.; Wang, Y.; Jaeger, M. Segmentation of architecture shape information from 3D point cloud. In *Proceedings of ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 127-132, 2009. 126

[Ortner et al., 2007] Ortner, M.; Descombes, X.; Zerubia, J. Building outline extraction from digital elevation models using marked point processes. In *International Journal of Computer Vision.*, vol. 72(2), pp. 107-132, 2007. 160

[Petrie et al., 2010] Petrie, G. Mobile Mapping Systems: an introduction to the technology. In *GeoInformatics*, vol. 13(1), pp. 32-43, 2010. 106

[Poullis et al, 2009] Poullis, C.; You, S. Automatic reconstruction of cities from remote sensor data. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 2775-2782, 2009. 70

[Rau et al, 2002] Rau, J.Y.; Chen, L.C. An interactive system for producing polyhedral and prismatic building models using building outline segments. In *Proceedings of Symposium on Geospatial Theory, Processing and Applications*, 6 p., 2002. 70

[Ridene et al., 2009] Ridene, T.; Goulette, F. Coregistration of DSM and 3D point clouds acquired by a Mobile Mapping System. In *Boletim de Ciências Geodésicas*, vol. 15, pp. 824-838, 2009. 119,136

[Romero et al, 2007] Romero, L.; Calderón, F. A tutorial on parametric image registration. In *Scene Reconstruction, Pose Estimation and Tracking*, I-Tech, pp. 167-184, 2007. 76

[Sampath et al, 2010] Sampath, A.; Shan, J. Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 1554-1567, 2010. 69

[Soheilian, 2008] Soheilian, B. Roadmark reconstruction from stereo-images acquired by a ground-based mobile mapping system. In *Ph.D. thesis*, Université Paris-Est, 208 p., 2008. 52

[Sohn et al, 2008] Sohn, G.; Huang, X.F.; Tao, V. Using a binary space partitioning tree for reconstructing polyhedral building models from airborne lidar data. In *Photogrammetric Engineering and Remote Sensing*, vol. 74, pp. 1425-1440, 2008. 69

[Storn et al., 1997] Storn, R.; Price, K. Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. In *Journal of Global Optimization*, vol. 11, pp. 341-359, 1997. 80

[Studnicka et al, 2011] Studnicka,N.; Zach, G.; Amon, P.; Pfennigbauer, M. RIEGL VMX-250 with modular camera system - combined scan and image data acquisition in mobile laser scanning. In *In SPAR conference*, 10 p., 2011. 164

[Suveg et al, 2001] Suveg, I.; Vosselman, G. 3D building reconstruction by map based generation and evaluation of hypotheses. In *Proceedings of the British Machine Vision Conference*, 10 p., 2001. 70

[Suveg et al, 2004] Suveg, I.; Vosselman, G. Reconstruction of 3D building models from aerial images and maps. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, pp. 202-224, 2004. 70

[Taillandier et al, 2004] Taillandier, F.; Deriche, R. Automatic buildings reconstruction from aerial images: A generic Bayesian framework. In *International Archives of Photogrammetry Remote Sensing and Spatial Information*, vol. 35, 6 p., 2004. 50,68,70,72

[Tarsha-Kurdi et al., 2008] Tarsha-Kurdi, F.; Landes, T.; Grussenmeyer, P. Extended RANSAC algorithm for automatic detection of building roof planes from LIDAR data. In *The Photogrammetric Journal of Finland*, vol. 21(1), pp. 97-109, 2008. 130

[Tournaire et al., 2010] Tournaire, O.; Brédif, M.; Boldo, D.; Durupt, M. An efficient stochastic approach for buildings footprint extraction from digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, pp. 317-327, 2010. 158,159,160,161

[Tseng et al, 2002] Tseng, Y.; Lin, C.; Wang, S. Model-image fitting using genetic algorithm for building extraction from aerial images. In *Proceedings of the Asian Conference on Remote Sensing*, 7 p.,2002. 71

[Vallet et al., 2009] Vallet, B.; Pierrot-Deseilligny, M., Boldo, D. Building footprint database improvement for 3D reconstruction: a direction aware split and merge approach. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, Part 3/W4, pp. 139-144, 2010. 159,160,161

[Verma et al, 2006] Verma, V.; Kumar, R.; Hsu, S. 3D building detection and modeling from aerial lidar data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2213-2220, 2006. 69

[Vestri, 2006] Vestri, C.; Using range data in automatic modeling of buildings. In *Image and Vision Computing*, vol. 24, pp. 709-719, 2006. 70

[Wang et al., 2007] Wang, L.; You, S.; Neumann, U. Semi-automatic registration between ground-level panoramas and an orthorectified aerial image for building modeling. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1-8, 2007. 70

[Wang et al., 2010] Wang, L.; Sohn, G. Automatic co-registration of terrestrial laser scanning data and 2D floor plan. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, Part 1, pp. 158-164, 2010. 159

[Wang et al., 2006] Wang, O.; Lodha, S. K.; Helmbold, D. P. A Bayesian approach to building footprint extraction from aerial LIDAR Data. In *Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 192-199, 2006. 158,159,160,161

[Xiao et al., 2009] Xiao, J.; Fang, T.; Zhao, P.; Lhuillier, M.; Quan, L. Image-based street-side city modeling. In *ACM Transaction on Graphics, SIGGRAPH Asia*, 12 p., 2009. 165,190,200

[Yoo et al., 2009] H. J. Yoo, F. Goulette, J. Senpauroca, G. Lepère, Simulation based comparative analysis for the design of laser terrestrial mobile mapping systems. In *Bulletin of Geodetic Sciences*, Special Issue on Mobile Mapping Technology, vol. 15(5), pp. 839-854, 2009. 115,119

[Zebedin et al, 2008] Zebedin, L.; Bauer, J.; Karner, K.; Bischof, H. Fusion of feature- and area-based information for urban buildings modeling from aerial imagery. In *Proceedings of the European Conference on Computer Vision*, pp. 873-886, 2008. 71,72

[Zhao et al., 2010] Zhao, P.; Fang, T.; Xiao, J.; Zhang, H.; Zhao, Q.; Quan, L. Rectilinear parsing of architecture in urban environment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 342-349, 2010. 190,200

[Zlatanova et al., 2002] Zlatanova, S.; Van Den Heuvel, F. A. knowledge-based automatic 3D line extraction from close range images. In *International Archives of Photogrammetry Remote Sensing and Spatial Information*, vol. 34(5), pp. 233-238, 2002. 174

Short Biography

Karim Hammoudi received the BSc degree in computer science from the University of Picardie Jules Verne (UPJV), France, in 2006. He received the BSc degree in mathematics in 2007. He worked in part during his studies as an Instructor in Computer Science at UPJV. He got a merit scholarship and received the MS degree in computer science, mathematics and image processing from the University of Paris Descartes (UP5) in 2008. From October 2008 to October 2011, he was a PhD candidate at the MATIS Laboratory of the French National Mapping Agency (IGN) and at the Department of Mathematics and ICT of the University of Paris-Est (UPE). His PhD was carried out in collaboration with the Department of Computer Science and Artificial Intelligence of the University of Basque Country (UPV/EHU) and with the Basque Foundation for Science (Ikerbasque) both located in Spain. This PhD research dealt with the 3D building reconstruction by using aerial images and the 3D facade modeling by exploiting terrestrial laser and image street data acquired by a Mobile Mapping System. In parallel, he worked during this period as Assistant Lecturer in Computer Science at the Department of Mathematics and Computer Science of the University of Paris Descartes. Since November 2011, he is working as a StratAG Postdoctoral Research Fellow at the Department of Computer Science of the National University of Ireland Maynooth (NUIM). His current research lies in the areas of MMS data fusing, terrestrial modeling of structured environments as well as information enhancement of real-world scenes.

Abstract / Résumé

The aim of this work is to develop research on 3D building modeling. In particular, the research in aerial-based 3D building reconstruction is a topic very developed since 1990. However, it is necessary to pursue the research since the actual approaches for 3D massive building reconstruction (although efficient) still encounter problems in generalization, coherency, accuracy. Besides, the recent developments of street acquisition systems such as Mobile Mapping Systems open new perspectives for improvements in building modeling in the sense that the terrestrial data (very dense and accurate) can be exploited with more performance (in comparison to the aerial investigation) to enrich the building models at facade level (e.g., geometry, texturing).

Hence, aerial and terrestrial based building modeling approaches are individually proposed. At aerial level, we describe a direct and featureless approach for simple polyhedral building reconstruction from a set of calibrated aerial images. At terrestrial level, several approaches that essentially describe a 3D urban facade modeling pipeline are proposed, namely, the street point cloud segmentation and classification, the geometric modeling of urban facade and the occlusion-free facade texturing.

Keywords: *3D City Modeling, Building Roof Reconstruction, Street Modeling, Facade Modeling and Texturing, Calibrated Aerial Images, Multiscopy, Mobile Mapping System, Terrestrial Laser data, 3D Point Cloud, Geometry, Hypothesize-and-verify Strategies, Computer Vision, Photogrammetry, Remote Sensing.*

L'objectif principal de ce travail est le développement de recherches en modélisation 3D du bâti. En particulier, la recherche en reconstruction 3D de bâtiment est un sujet très développé depuis les années 90. Malgré tout, il paraît nécessaire de poursuivre les recherches dans cet axe étant données que les approches actuelles consacrées à la reconstruction 3D de bâtiment (bien qu'efficaces) rencontrent encore des difficultés en terme de généralisation, de cohérence et de précision. Par ailleurs, les récents développements des systèmes d'acquisitions de rues tel que les systèmes de cartographie mobile ouvrent de nouvelles perspectives d'amélioration de la modélisation des bâtiments dans le sens ou les données terrestres (très précises et résolus) peuvent être exploitées avec davantage de cohérence (en comparaison à l'aérien) pour enrichir les modèles de bâtiments au niveau des façades (la géométrie, la texture).

Ainsi, des approches de modélisation aériennes et terrestres sont individuellement proposées. Au niveau aérien, nous décrivons une approche directe et dépourvu d'extraction et d'assemblage de primitives géométriques en vue de la reconstruction 3D de modèles polyédriques simples de bâtiments à partir d'un jeu d'images aériennes calibrées. Au niveau terrestre, plusieurs approches qui décrivent essentiellement un pipeline pour la modélisation 3D des façades urbaines sont proposées; à savoir, la segmentation et classification de nuage de rues urbaines, la modélisation géométrique des façades urbaines et le texturage des façades urbaines comportant des occultations causées par d'autres objets du mobilier urbains.

Mots clés : *Modélisation 3D des villes, Reconstruction des toits de bâtiments, Modélisation des rues, Modélisation et texturage des façades, Images aériennes calibrées, Multiscopie, Système de cartographie mobile, Données laser terrestre, Lidar, Nuage de points 3D, Géométrie, Génération et vérification d'hypothèses, Vision par ordinateur, Photogrammétrie, Télédétection.*