Roadmark reconstruction from stereo-images acquired by a ground-based mobile mapping system

Bahman Soheilian

To cite this version:

Bahman Soheilian. Roadmark reconstruction from stereo-images acquired by a ground-based mobile mapping system. Automatic. Université Paris-Est, 2008. English. NNT : . tel-00296601v1

HAL Id: tel-00296601
https://theses.hal.science/tel-00296601v1

Submitted on 13 Jul 2008 (v1), last revised 1 Apr 2010 (v2)

THÈSE

présentée pour obtenir le titre de docteur de
l'Université Paris-Est

Spécialité : Sciences de l'Information Géographique

Bahman SOHEILIAN

Roadmark reconstruction from stereo-images acquired by a ground-based mobile mapping system

Soutenue le 1 avril 2008 devant le jury composé de

Dr. Patrick RIVES	Président
Dr. Simon LACROIX	Rapporteur
Dr. Michel DHOME	Rapporteur
Pr. Bertrand MERMINOD	Examinateur
Dr. Jean-Philippe TAREL	Examinateur
Dr. Didier BOLDO	Examinateur
Dr. Nicolas PAPARODITIS	Directeur
Pr. Jean-Paul RUDANT	Directeur

BibTeX document entry / fiche bibliographique BibTeX :

```
@PhdThesis{Soheilian:PhD,
  author = {Bahman Soheilian},
  title = {Roadmark reconstruction from stereo-images acquired by
           a ground-based mobile mapping system },
  school = {Universit{\'{e}} Paris-Est},
  year = 2008,
  address = {Marne-la-Vall{\'{e}}e, France},
  month = apr
}
```

SOHEILIAN, Bahman: Roadmark reconstruction from stereo-images acquired by a ground-based mobile mapping system. Marne-la-Vallée, France, Université Paris-Est, Ph.D. thesis, April 2008.

Remerciements

Je tiens en premier lieu à remercier M. Jean-Paul Rudant, mon directeur de thèse pour son soutien et la confiance qu'il m'a toujours accordée.

Je veux exprimer ici mes sincères remerciements à M. Nicolas Paparoditis qui m'a permis de travailler sur un sujet qui me tient à cœur, pour son encadrement et pour ses conseils toujours judicieux. Je souhaite le remercier aussi pour son optimisme, ses encouragements et pour sa grande qualité humaine. Je tiens à remercier M. Didier Boldo le directeur du laboratoire MATIS de m'avoir accueilli au sein du laboratoire, pour son encadrement et d'avoir pris le temps de répondre à mes questions techniques. Je n'oublierai jamais sa bonne humeur.

Je souhaite remercier M. Patrick Rives qui m'a fait l'honneur de présider mon jury de thèse. Je remercie également M. Simon Lacroix et M. Michel Dhome qui ont accepté d'être rapporteurs de cette thèse et pour le temps qu'ils ont consacré à la relecture du manuscrit. Je remercie par ailleurs M. Bertrand Merminod et M. Jean-Philippe Tarel pour m'avoir fait l'honneur de bien vouloir participer au jury.

Je remercie également tous les membres permanents ou stagiaires du laboratoire MATIS pour leur soutien. J'aimerais tout particulièrement remercier Olivier, pour les nombreuses discussions enrichissantes que nous avons eues pendant la thèse, pour ses sacrifices dans nos travaux collectifs et pour m'avoir aidé sur plusieurs problèmes informatiques, sans oublier sa sympathie. Je n'oublie pas mes anciens collègues de bureau : Lionel qui était toujours prêt à résoudre avec moi de nombreux problèmes de mathématiques et de géométrie ; Matthieu pour sa gentillesse et ses relectures de mes papiers ; merci à tous les deux. J'aimerais aussi remercier Corina, Mahzad, Erwann, Bertrand et Jean-Pierre pour leur sympathie et leurs aides dans les moments parfois difficiles de ma thèse. Je remercie aussi Grégoire de m'avoir mis à disposition le formidable outil de mise en correspondance d'images. Un grand merci à Fadi Dornaika pour les relectures de ce manuscrit et les remarques pertinentes qu'il m'a faites. Je n'oublie pas de remercier Patrice Bueso chef du service de la recherche d'IGN pour sa gentillesse. Je remercie par ailleurs, Alain Sombris pour son suivi administratif. Mes remerciements vont ensuite à François Boyero pour sa disponibilité et sa patience devant les problèmes techniques, sans oublier sa gentillesse.

Je n'oublie pas de remercier Pardis, Venus, Sylvain et Fabrice pour les aides précieuses qu'ils m'ont apportées à mon arrivée à Paris pour commencer mes travaux de recherche. Un très grand merci à ma sœur Ladan qui m'a poussé à faire des choix courageux. Je la remercie pour toutes les aides qu'elle m'a apportées. Enfin, je fais un petit clin d'œil à Behrouz, ainsi qu'à tous mes amis qui ont été vraiment formidables.

Je termine en remerciant mes parents, mon frère et encore une fois ma sœur pour leurs encouragements et leur soutien indispensable pendant toutes ces années d'études.

A ma famille.

A la mémoire de mon grand-père.

Abstract

Despite advances in ground-based Mobile Mapping System (MMS), automatic feature reconstruction seems far from being reached. In this thesis, we focus on 3D roadmark reconstruction from images acquired by road looking cameras of a MMS stereo-rig in dense urban context. A new approach is presented, that uses 3D geometric knowledge of roadmarks and provides a centimetric 3D accuracy with a low level of generalisation.

Two classes of roadmarks are studied: zebra-crossing and dashed-lines. The general strategy consists in three main steps. The first step provides 3D linked-edges. Edges are extracted in the left and right images. Then a matching algorithm that is based on dynamic programming optimisation matches the edges between the two images. A sub-pixel matching is computed by post processing and 3D linked-edges are provided by classical photogrammetric triangulation. The second step uses the known specification of roadmarks to perform a signature based filtering of 3D linked-edges. This step provides hypothetical candidates for roadmark objects. The last step can be seen as a validation step that rejects or accepts the candidates. The validated candidates are finely reconstructed. The adopted model consists of a quasi parallelogram for each strip of zebra-crossing or dashed-line. Each strip is constrained to be flat but the roadmark as a whole is not planar.

The method is evaluated on a set of 150 stereo-pairs acquired in a real urban area under normal traffic conditions. The results show the validity of the approach in terms of robustness, completeness and geometric accuracy. The method is robust and deals properly with partially occluded roadmarks as well as damaged or eroded ones. The detection rate reaches 90% and the 3D accuracy is about $2-4$ *cm*.

Finally an application of reconstructed roadmarks is presented. They are used in georeferencing of the system. Most of the MMSs use direct georeferencing devices such as GPS/INS for their localisation. However in urban areas masks and multi-path errors corrupt the measurements and provide only 50 *cm* accuracy. In order to improve the localisation quality, we aim at matching ground-based images with calibrated aerial images of the same area. For this purpose roadmarks are used as matching objects. The validity of this method is demonstrated on a zebra-crossing example.

Keywords: *Photogrammetry, Computer vision, Mobile mapping system, Edge matching, 3D reconstruction, 3D modelling, Image-based georeferencing, Road, Roadmark, Zebra-crossing, Dashed-line.*

Résumé

Malgré les récentes avancées des Systèmes de Cartographie Mobile, la reconstruction automatique d'objets à partir des données acquises est encore un point crucial. Dans cette thèse, nous nous intéresserons en particulier à la reconstruction tridimensionnelle du marquage au sol à partir d'images acquises sur le réseau routier par une base stéréoscopique horizontale d'un système de cartographie mobile, dans un contexte urbain dense. Une nouvelle approche s'appuyant sur la connaissance de la géométrie 3D des marquages au sol est présentée, conduisant à une précision de reconstruction 3D centimétrique avec un faible niveau de généralisation.

Deux objets de la signalisation routière horizontale sont étudiés : les passages piétons et les lignes blanches discontinues. La stratégie générale est composée de trois grandes étapes. La première d'entre elles permet d'obtenir des chaînes de contours 3D. Les contours sont extraits dans les images gauche et droite. Ensuite, un algorithme reposant sur une optimisation par programmation dynamique est mis en œuvre pour apparier les points de contours des deux images. Un post-traitement permet un appariement sub-pixellique, et, les chaînes de contours 3D sont finalement obtenues par une triangulation photogrammétrique classique. La seconde étape fait intervenir les spécifications géométriques des marquages au sol pour réaliser un filtrage des chaînes de contours 3D. Elle permet de déterminer des candidats pour les objets du marquage au sol. La dernière étape peut être vue comme une validation permettant de rejeter ou d'accepter ces hypothèses. Les candidats retenus sont alors reconstruits finement. Pour chaque bande d'un passage piéton ou d'une ligne discontinue, le modèle est un quasi-parallélogramme. Une contrainte de planéité est imposée aux sommets de chaque bande, ce qui n'est pas le cas pour l'ensemble des bandes formant un marquage au sol particulier.

La méthode est évaluée sur un ensemble de 150 paires d'images acquises en centre ville dans des conditions normales de trafic. Les résultats montrent la validité de notre stratégie en terme de robustesse, de complétude et de précision géométrique. La méthode est robuste et permet de gérer les occultations partielles ainsi que les marquages usés ou abîmés. Le taux de détection atteint 90% et la précision de reconstruction 3D est de l'ordre de 2 à 4 cm.

Finalement, une application de la reconstruction des marquages au sol est présentée : le géoréférencement du système d'acquisition. La majorité des systèmes de cartographie mobile utilisent des capteurs de géoréférencement direct comme un couplage GPS/INS pour leur localisation. Cependant, en milieu urbain dense, les masques et les multi-trajets corrompent les mesures et conduisent à une précision d'environ 50 cm. Afin d'améliorer la qualité de localisation, nous cherchons à apparier les images terrestres avec des images aériennes calibrées de la même zone. Les marquages au sol sont alors utilisés comme objets d'appariement. La validité de la méthode est démontrée sur un exemple de passage piéton.

Mots clé : *Photogrammétrie, Vision par ordinateur, Système de cartographie mobile, Appariement de contours, Reconstruction 3D, Modélisation 3D, Géoréférencement à partir des images, Route, Marquage au sol, Passage piéton, Lignes blanches discontinues.*

خلاصه

علاوه بر پیشرفتهای جدید در زمینه سیستم های نقشه سازی متحرک، کماکان استخراج و بازسازی عوارض به صورت خودکار دور از دسترس می نماید. موضوع این پایان نامه بازسازی سه بعدی خط کشی های راهنمائي و رانندگی به صورت خودکار از تصاویر استریوی برداشت شده به وسیله يك سیستم نقشه سازی متحرك در مناطق پر رفت و آمد شهری می باشد. در اینجا يك شیوه جدید ارایه می شود که با استفاده از اطلاعات مربوط به فرم هندسی خط کشی ها، مدلی سه بعدی با تعمیم جزئي و دقت سانتیمتر فراهم می نماید.

بازسازی دو نوع از علامات مورد مطالعه قرار گرفته است: خطوط غیر ممتد و خطوط عابر پیاده. استراتژی کلی مشتمل بر سه بخش می باشد. قسمت اول شامل فراهم نمودن لبه ها به صورت سه بعدی است. ابتدا لبه های تصاویر راست و چپ استخراج شده و سپس به وسیلهء يك الگوریتم تناظر یابی که بر اساس بهینه سازي با برنامه نویسی دینامیك می باشد لبه های متناظر استخراج می شوند. بعد از محاسبهء تناظر نقاط با دقت زیر پیکسل، با بهره گیری از يك مرحله پس پردازش، زنجیره هایی سه بعدی از لبه ها بوسیلهء مثلث بندی کلاسیك فتوگرامتری مهیا می گردند. بخش دوم الگوریتم براي فیلتر کردن زنجیره های سه بعدی مذکور از خصوصیات هندسی معلوم خط کشی ها استفاده می کند. این قسمت از الگوریتم زنجیره های سه بعدی ای که احتمالا متعلق به لبه هاي خط کشي ها هستند را استخراج می نماید. آخرین قسمت الگوریتم وظیفه معتبر سازی کاندیدا های استخراج شده را بر عهده دارد. کاندیدا های تأیید شده به صورت دقیق بازسازی می شوند. مدل انتخاب شده برای هر باند خط کشی عابر پیاده و خط کشی غیر ممتد شامل يك شبه لوزی می باشد. هر باند يك صفحه فرض شده ولی کل شيء (شامل چند باند) روی يك صفحه نمی باشد.

الگوریتم پشنهاد شده بر روي ۱۵۰ زوج عکس گرفته شده در منطقه ای پر رفت و آمد شهری مورد ارزیابی قرار گرفته است. نتایج بدست آمده دلالت بر معتبر بودن الگوریتم از نظر استحکام، کامل بودن و دقت هندسی دارد. این روش بدرستی اشیائي را که به صورت جزئي پنهان بوده یا تغیر شکل جزئي داده اند استخراج می نماید. میزان استخراج اشیاء به ۹۰ % و دقت سه بعدی هندسی آنها به ۲-٤ سانتیمتر می رسد.

در نهایت کاربردی از علامت های راهنمائي و رانندگی استخراج شده ارایه می گردد. این علامات برای مکانیابی سیستم استفاده می شوند. اغلب سیستم های نقشه سازی متحرك از دستگاه های مکانیابی مستقیم همچون سیستم موقعیت یاب جهانی و سیستم موقعیت یابی جاذبه ای برای مکانیابی خود استفاده میکنند. ولی خطا های ناشی از دریافت نامناسب امواج ماهواره ای مانند دریافت غیر مستقیم امواج یا پنهان بودن ماهواره از دید آنتن به دلیل ارتفاع زیاد ساختمان ها در مناطق شهری باعث می شود خطای تعیین موقعیت به ۵۰ سانتیمتر برسد. به منظور بهینه سازی دقت مکانیابي، پیشنهاد مي نمائیم عکس های زمینی با عکس های هوائي مثلث بندی شده همان منطقه تناظر یابی شوند. به این منظور خط کشی های خیابان به عنوان عوارض مورد تناظر یابی استفاده می شوند. اعتبار این الگوریتم روی مثالی از خطوط عابر پیاده نشان داده می شود.

کلمات کلیدی : فتوگرامتری، ماشین بینائي، سیستم نقشه سازی متحرك، تناظر یابی لبه، بازسازی سه بعدی، مکانیابی تصویر مبنا، جاده، خط کشی خیابان، خط عابر پیاده، خطوط ممتد.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

1.1 Global context

Nowaday there is a growing need for 3D city and road models for very various applications such as city planning, electromagnetic wave propagation analysis, architectural state documentation, realistic gaming, autonomous navigation, traffic monitoring or virtual tourism projects. Buildings and roads are the main features in 3D city models. Since the beginning of the 90s, MATIS [1] laboratory of IGN[2] (French national mapping agency) has been conducting research in the field of fully automatic 3D building reconstruction from digital aerial images. Production results with the developed tools are very promising especially from an automation rate point of view. Flamanc *et al.* [2003] present results of these researches. In order to provide a more complete description of city in parallel to building modelling, road reconstruction have been investigated. The research work presented by Ruskoné [1996] and Tournaire [2007] deals with automatic road reconstruction in urban and semi-urban areas from aerial images. Nevertheless aerial imagery does not provide sufficient geometric and radiometric accuracy for reconstruction of all urban objects. Indeed due to its limited resolution and their special orientation in relation to the terrain, some interesting object such as facades, urban furniture, vegetation and roads can not be modelled with adequate accuracy. In this scope, the *Archipolis* research project that is developed in ARCHI [3] working group of MATIS laboratory is in charge of looking for very high scale ground based future products to complete the description of city models produced by aerial means.

1.2 *Archipolis* project

What should an ideal 3D city and road model contain? Indeed, the needs of users are very different in terms of database contents, level of detail, and relative and absolute accuracies. To fulfil all possible applications, the *Archipolis* project aims at providing a database of georeferenced terrestrial stereopairs covering the whole city. This would mean that users, with adapted softwares that already exist at low prices (e.g. PhotoModeller®, RealViz©, *etc.*), could then reconstruct a model of the scene adapted to their needs and budget. Nev-

---

[1] *Méthodes d'Analyses pour le Traitement d'Images et la Stéréorestitution*

[2] *Institut Géographique National*

[3] Architectural and close range photogrammetry working group

ertheless, some specific objects such as pavements, roadmarks and facade geometry can be extracted in a systematic way [Paparoditis *et al.*, 2005].

In order to provide the georeferenced image database a Mobile Mapping System (MMS) called *Stereopolis* is developed at the MATIS laboratory. The *Stereopolis* system as shown in Figure 1.1(a) is equipped with three stereo-rigs of CCD cameras and georeferencing device (GPS). The CCD cameras provide high resolution images (4000 $\times$ 4000 pixels $\equiv$ 3 mm GSD [4] at 10 m of distance) with high quality ($SNR = 300$)[5].

The two rear looking cameras (horizontal stereo-base in Figure 1.1(a)) provide stereo-images of the roads and urban furniture (See Figure 1.1(b)). The two across viewing stereo-bases (Vertical stereo-base in Figure 1.1(c)) acquire stereo-images of buildings' facades (See Figure 1.1(c)). The system provides good imaging geometry and coverage of object space. With 28 mm focal length lenses, the field of vertical base covers a six storey building high at a distance of 6 meters. All 6 cameras are perfectly synchronised. Interior calibration of cameras and also relative orientation between all 6 cameras are *a priori* estimated and assumed rigid.

Most of the MMSs are based on direct georeferencing devices like GPS/INS. However in dense urban areas GPS masks and multi-paths do corrupt the quality of measurements and provide only 50 cm to 1 m absolute positioning accuracy. In order to satisfy the needs of different users in georeferencing accuracy point of view, the *Archipolis* project aims at providing a fine georeferencing by post-processing at the best possible accuracy (5 to 10 cm for absolute positioning).

1.3 Specific context of the thesis

This PhD thesis is held at ARCHI group of MATIS laboratory and integrates the previously mentioned *Archipolis* project. The goal is to provide a tool for fully automatic 3D roadmark reconstruction from stereo-images acquired by *Stereopolis* system. Interests of 3D roadmarks reconstruction within *Archipolis* project are in:

- Fine georeferencing of the system;
- Road inventory generation;
- Providing landmarks for autonomous navigation systems;

1.3.1 Interest of roadmarks in fine georeferenced image-database generation

Georeferencing of data means its translation and orientation in relation to a fixed coordinate system. Most of the MMSs [6] adopt an absolute coordinate system. The absolute georeferencing of most of the systems is obtained using GPS[7] receivers that are integrated with INS[8] [El-Sheimy, 1996; Schwarz *et al.*, 1993; Lapucha *et al.*, 1990; Novak and Bossler, 1995]. However in urban areas, masks of high buildings and multi-path of GPSs' signals do corrupt

---

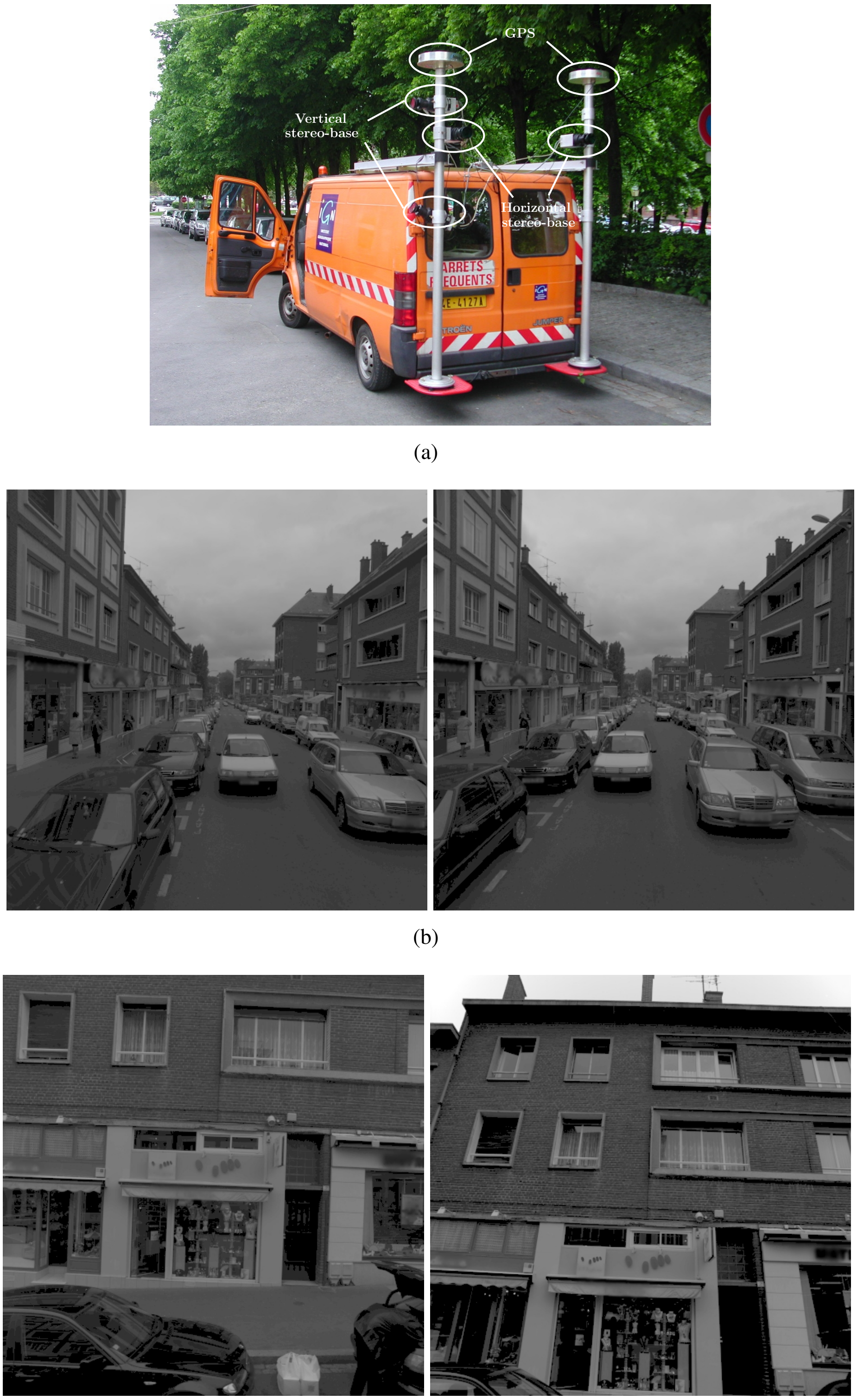
[4] *GSD*: Ground Sample Distance

[5] *SNR*: Signal to Noise Ratio

[6] MMS: Mobile Mapping System

[7] GPS: Global Positioning System

[8] INS: Inertial Navigation System



(a)

(b)

(c)

Figure 1.1: a) The *stereopolis* mobile mapping system. b) Stereo-pair provided by rear looking horizontal stereo rig, c) Stereo-pair taken by side looking vertical rig.

the measurements. Even thought INS can interpolate between GPS interruptions, internal drift of INS accumulates soon and provide only 50 *cm* to 1 *m* absolute georeferencing accuracy for gaps of two minutes [APPLANIX].

In order to improve the georeferencing of *Stereopolis*, the idea is to use its own images as a sub-system of positioning. In this purpose Bentrah [2006] uses images to estimate relative orientation between successive image nodes. It can improve relative orientation locally but due to error accumulation, absolute positioning remains of poor quality.

The idea for improving absolute positioning of the system is to take benefit from exterior control points to perform a classical photogrammetric bundle adjustment. In photogrammetry control points are often measured by topographic operations which is very costly and also time consuming. In order to generate needed control points we use the results of Tournaire [2007] that reconstruct roadmarks from oriented aerial images (20 *cm GSD*). The same roadmarks are then reconstructed automatically from horizontal stereo-pairs of *Stereopolis*. The reconstructed roadmarks are then repositioned with respect to those reconstructed from aerial images. Finally these roadmarks are used as Ground Control Objects (GCO) that can integrate a bundle adjustment for positionning ground-based images. Indeed the interest of roadmark is that their shapes and sizes are constrained by strict specifications [Transport Ministry and Interior Ministry, 1988] and constitute high contrasted objects. These two properties simplify considerably the detection of roadmarks from both systems and also matching process between two systems. This concept is demonstrated in Figure 1.2.

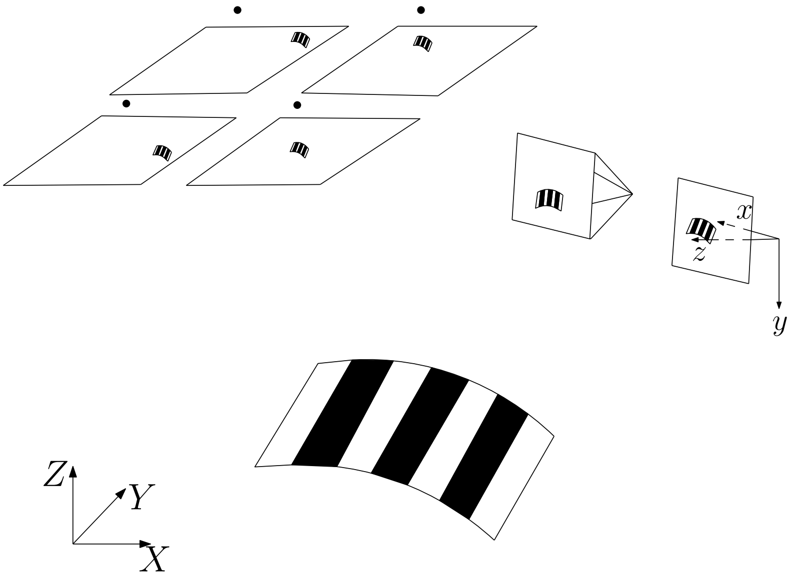


Figure 1.2: Absolute georeferencing of *Stereopolis*: Matching aerial-based and ground-based imagery using roadmarks

1.3.2 Interest of roadmarks in road inventory generation

The resolution of images has a great influence on the representation of roads. Indeed within algorithms developed for road recognition and reconstruction from low resolution satellite images, road is considered as homogeneous bands of different lengths and orientations

[Stoica *et al.*, 2004]. This can be an optimal representation of road considering the low resolution of images. However it does not provide any details and gives just a global representation of road. Within higher resolution images [Zhang, 2003] use aerial images and uses multiple cues such as road's borders, roadmarks, shadows and possible position of road in relation to known buildings to provide a higher level of details for road recognition. In higher resolution level images such as ground-based ones, the level of presented details is much more important. In this case road can no more be considered as an homogeneous area. Moreover in urban areas road's borders are most of the time occluded by cars and can not be used as a road cue. In this scenario roadmarks become more interesting. On the one hand their known specification can simplify considerably their recognition, on the other hand their precise 3D reconstruction provide useful information such as number of lanes and their width.

1.3.3 Interest of roadmarks in autonomous navigation systems

The most important component of autonomous navigation is permanent and reliable localisation. Indeed for autonomous navigation, a robot needs to understand its environment. More precisely robot need to know where it is located and where to go. BODEGA is an example of autonomous navigation projects [Bonnifait *et al.*, 2006]. This project aims at developing vision-based and sensor-based methods for the autonomous navigation of mobile vehicles moving around an urban environments. It is realised through a collaboration between several French universities and research centers . One of the main component of this project consist in providing a landmark GIS in urban area. The landmarks are extracted and georeferenced in an off-line mode. During navigation, robot recognises the landmarks and localise itself by photogrammetric resection. Several kind of landmarks such as features on facades and roads can be considered in urban area. However landmark must be:

- Available in the whole area of navigation;
- Easy to extract;
- Localised accurately;

Roadmarks constitute excellent landmarks because in urban areas they are available most of the time. Furthermore, thanks to their strict shape and size they are quite easy to extract as well as accurately measurable.

1.4 Organisation of the manuscript

This thesis is built over eight main chapters. After this introduction, in Chapter 2 we will first introduce the input data and needed reconstruction quality. Then an overview of existing methods is provided. Finally our adopted global strategy is introduced.

The three following Chapters explain the three main steps of our roadmark reconstruction algorithm.

Chapter 3 describes the procedure of 3D linked-edges reconstruction. A brief state of art in stereo-matching methods is presented. Then adopted method is explained in details. Finally, the results that consist in 3D linked-edges are presented.

In Chapter 4, we present how 3D linked-edges resulting from reconstruction process are used in roadmark detection. Detection step takes benefit from particular 2D signature of zebra-crossing and dashed-line. Some basic tools of filtering are introduced. Then we explain how these tools are used in detection of each roadmark. The intermediate results are demonstrated on one example for each roadmark. Finally, results of detection step that consist in a set of 3D line-segments as hypothetical strips are presented.

Chapter 5 describes modelling step. A perceptual grouping and also a line-segment grouping algorithm are developed and used in recognition of strip from hypothetical line-segments provided by detection algorithm. Then a method of fine reconstruction is presented. This method provides a planar parallelogram model for each detected strip. This is the final stage in roadmark reconstruction algorithm.

The reconstruction algorithm is tested on a large set of real images. Chapter 6 presents first some qualitative results that demonstrate robustness of the method to different problems such as unfavoured traffic conditions, partial occlusions and damaged roadmarks. Completeness and geometric accuracy of the method is evaluated and quantitative results are presented on a set of 150 stereo-pairs. Then the sensibility of the algorithm to resolution of input images is studied. Finally computation times are presented for each resolution.

Chapter 7, presents application of 3D reconstructed roadmarks in absolute positioning of *Stereopolis* with centimetric accuracy. First a brief state of art in image-based positioning methods of land vehicles is presented. Then our methodology of matching ground-based and aerial-based roadmark is explained. Zebra-crossing is used as roadmark object. The results are presented on real images.

Finally, Chapter 8 provides a synthetic analysis of this work. In order to guide the future researches, we propose a global schema of a roadmark reconstruction strategy that deals with different kind of roadmarks.

Chapter 2

Input data, state of the art and general strategy

As explained in Chapter 1, the goal of this thesis is to provide an algorithm for automatic reconstruction of roadmarks from ground-based images. These reconstructed roadmarks can be used for absolute georeferencing of the system, road database generation and also as landmarks in autonomous navigation projects. Since the quality of reconstruction from representation, generalisation and precision points of view depends largely on characteristics of input data, we present first the input data. Then needed characteristics of reconstruction are discussed. The Sections 2.4, 2.5 and 2.6 present a state of art on road detection and reconstruction from ground-based data. Section 2.7 describes our adopted strategy for roadmark reconstruction.

2.1 Roadmark specifications

Roadmarks are painted on the roads according to strict specifications (See Figure 2.1). In each country, their size and shape are normalised for all roadmarks including zebra-crossings (crosswalks), dashed-lines, stop-lines, arrows, etc. In France specification of roadmark are published by ministry of Interior and the ministry of Transport [Transport Ministry and Interior Ministry, 1988]. In Sections 2.1.1 and 2.1.2, we present the geometric specification of zebra-crossings and dashed-lines in France.

Figure 2.1: The road worker lays out pre-manufactured shapes in the design required. Once in place, a blow torch is used to melt the plastic on the surface of the road.

2.1.1 Zebra-crossing specification

A zebra-crossing is constituted of parallelogram form strips of 50 *cm* width and 2.5 *m* minimum length. However accurate length (ℓ) and exact shape (θ) of the parallelogram are unknowns. Distance between the strips varies between 50 *cm* to 80 *cm* (See Figure 2.2).

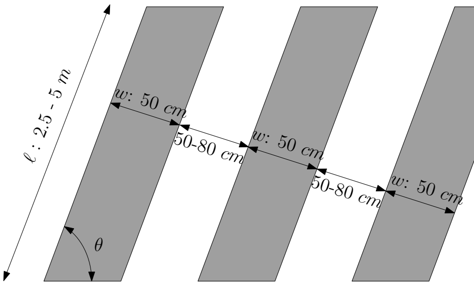


Figure 2.2: Specification of zebra-crossing

2.1.2 Dashed-line specification

Dashed-lines are always constituted of repetitive rectangular shape strips. The shape (size and form) of each strip and the spaces between the strips are conformed to a conventional standard. Dashed-lines are applied to specify lane boundaries, parking lots, stop lines etc. Geometric specifications of each dashed-line vary by their application and location (urban/rural). Figure 2.3 on the next page shows different types of dashed-line defined in [Transport Ministry and Interior Ministry, 1988] as a French standard. Length and the distance between the strips are fixed. Strips width varies following the location and is a multiple of $u \in \{3\ cm, 5\ cm, 6\ cm, 7.5\ cm\}$. So u varies in relation to location and importance of road but it is constant along a same road. Table 2.1 denotes different width values as a function of u. Exceptionally dashed-lines of type $T'2$ can take the width values other than multiple of u. For each type of dashed-line, their strips' length, width and also their inter-distance are specified (see Figure 2.3).

Type of dashed-line	Width
$T1$	$2u$
$T'1$	$2u$
$T3$	$2u$, $3u$, $5u$
$T2$	$3u$, $5u$
$T'3$	$3u$
$T'2$	$2u$, $3u$, 50 *cm*, 15 *cm*, 10 *cm*, 25 *cm*

Table 2.1: Width of different kinds of dashed-lines

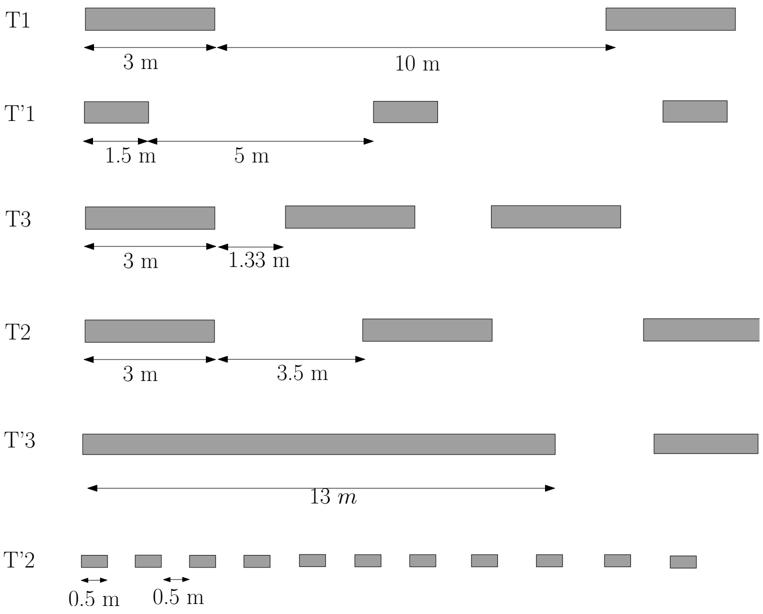


Figure 2.3: Specification of different kind of dashed-lines

2.1.3 Road surface model

Road surface is rarely flat but slightly curved or sloped with the two sides a bit lower to drain raining water from road surface (See Figure 2.4). As we aim at reconstructing roadmarks in the most precise way as possible, we suppose only that each strip of zebra-crossing or dashed-line is a flat object. However the totality of zebra-crossing or dashed-line is not necessarily flat and follow the road surface topography. In other words we have to reconstruct the road surface after the final modelling phase of dashed-line or zebra-crossing reconstruction.

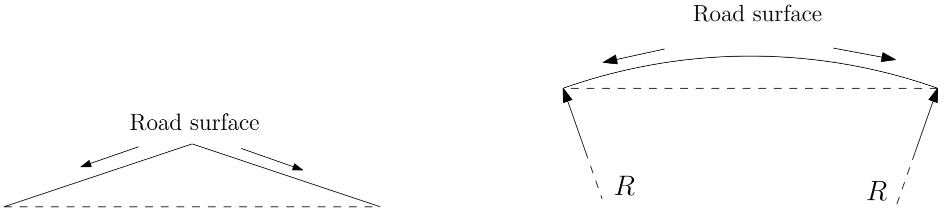


Figure 2.4: Two kinds of road section: sloped and curve surfaces are applied to guide raining water to the borders.

Even if the so called surface is constrained by some technical specifications (slope, curvature radius etc.), on the one hand their geometric shapes depend on many parameters (width, speed limit, etc.) which are rarely known *a priori*, on the other hand the specifications are not perfectly respected when the roadmarks are painted or glued onto the road. This is the reason why we do not use any constrained road model in our roadmark reconstruction algorithm.

2.2 Input image data description

As explained in Chapter 1, our input data consists in rear looking stereo-pairs provided by a horizontal stereo rig of *Stereopolis* system (See Figure 1.1(a)). Data set has been acquired in the center of Amiens city in France. Figure 2.5 shows the trajectory of *Stereopolis* on an ortho-image of the city center of Amiens.

2.2.1 Remarks on images of road and roadmarks in real urban areas

Our surveying is performed in real traffic conditions. In this scenario, presence of cars and pedestrians cause some occlusions. Figure 2.6 demonstrates such problem within three stereo-pairs. In certain cases some of dashed-line's strips are occluded by cars. This is the case in Figures 2.6(a) and 2.6(b). Sometimes only one strip of a dashed-line is visible in stereo-images. The dashed-line's strip assigning parking-lot in Figure 2.6(a) is an example of this case. In these scenarios dashed-lines can not be recognised only thanks to repetition and alignment of their strips (looking for repetitive patterns). Figure 2.6(c) demonstrates zebra-crossing's strip which is partially occluded by a pedestrian.

Besides occlusion, another main problem in real urban area is that roadmarks can be partially damaged or deformed. Figure 2.7 shows such case for a zebra-crossing's strip deformed by a man-hole cover and a dashed-line's strip that is worn out by the passage of cars.

2.2.2 Interior calibration of *Stereopolis* cameras

The *Stereopolis* system is equipped with cameras produced at LOEMI[1] laboratory of IGN for aerial surveying. They provide very high resolution and quality images. The interior calibration parameters are estimated *a priori* using calibration targets. The residuals of the calibration are usually under 10^{th} of a pixel. Our model is composed of a principal point of autocolimation, a principal point of symmetry and distortion coefficients. Table 2.2 present the specifications of images provided by this camera.

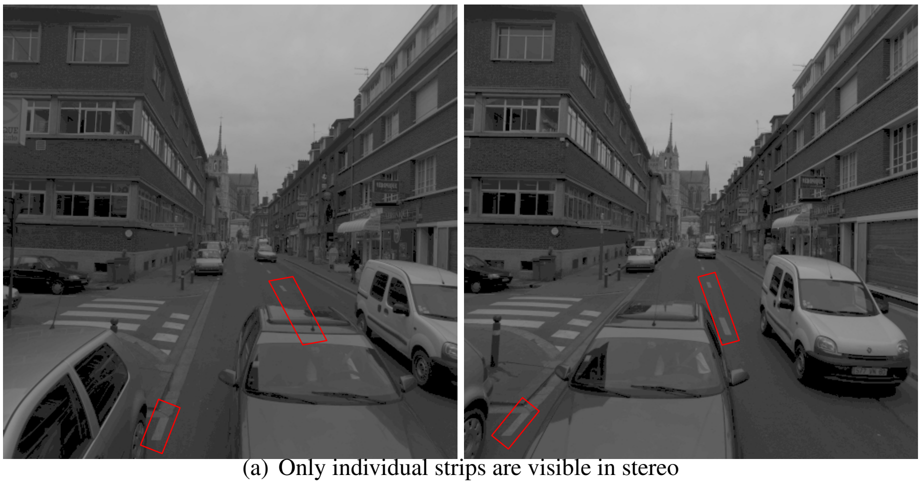
Parameter	Value
Image size	4096×4096
Pixel size	$9\ \mu m$
Camera focal length	$29\ mm \equiv 3222.2$ pixels
SNR	~ 300
Radiometric resolution	12 bits
Number of channels	1
Distortion parameters	Known
Principal point	Known
Ground resolution at distance of 10 m	$\sim 3\ mm$

Table 2.2: Interior calibration of *Stereopolis* cameras.

---

[1] Laboratoire d'Optique Et de Micro-Informatique

Figure 2.5: Trajectory of *Stereopolis* system in the center of Amiens city in France



(a) Only individual strips are visible in stereo

(b) Repetition of strips is violated by occlusions

(c) Strips are partially occluded

Figure 2.6: Samples of different traffic conditions within stereo-images provided by *Stereopolis* in real urban area.

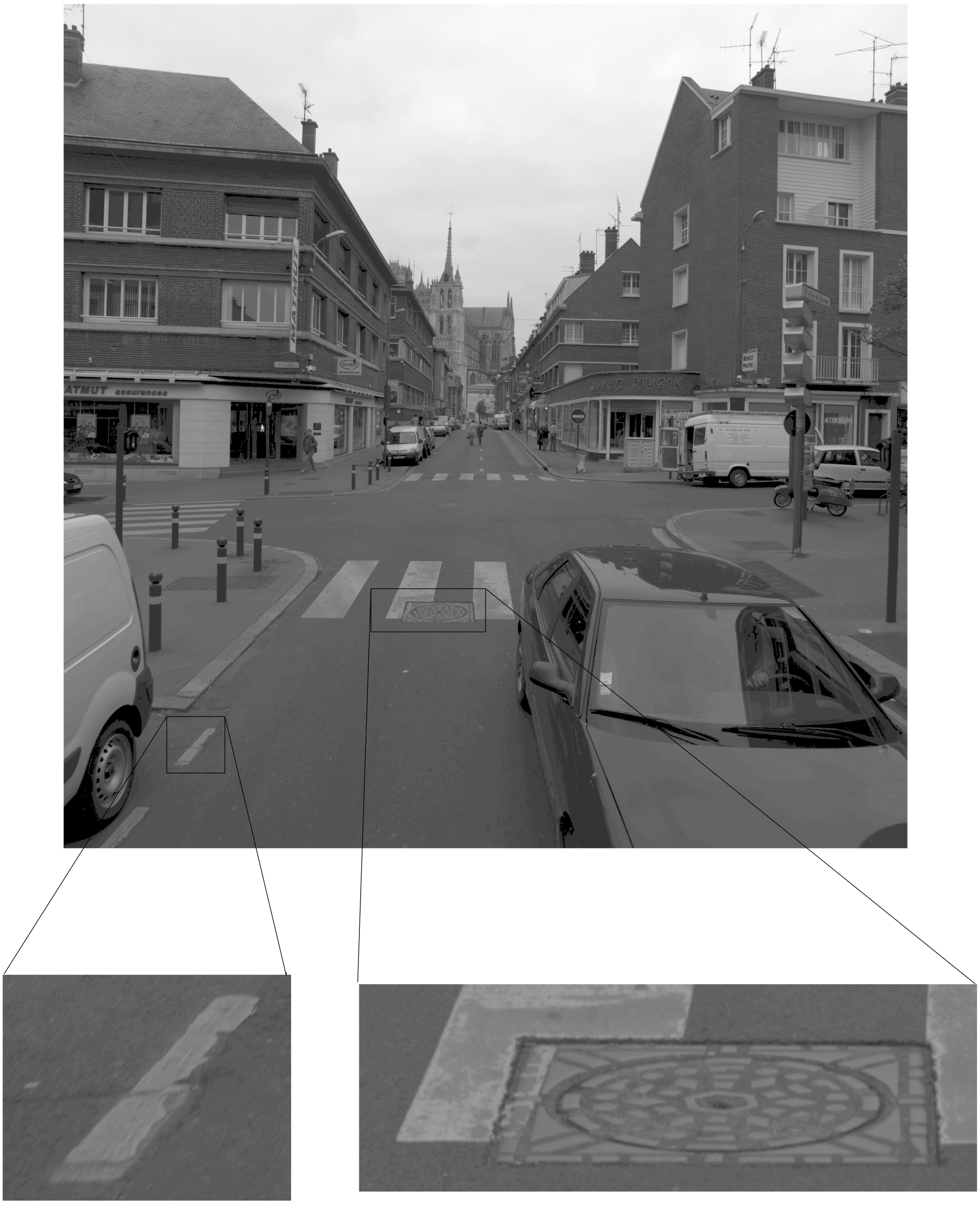


Figure 2.7: An example of eroded and covered roadmarks.

2.2.3 Calibration of the stereo rig

As mentioned before the rear looking horizontal stereo-rig of *Stereopolis* system provides stereo-images of road areas. In this thesis only the so called stereo-rig is used for road reconstruction. The relative orientation between the left and right cameras (translation and rotation) is estimated using manually measured tie points. Since the stereo-base is supposed to be rigid the relative orientation will be the same for all acquisition node. The height of stereo-base on the vehicle is about 2.2 m and change slowly during surveying operation. Figure 2.8 demonstrates the calibration of stereo-rig and its orientation in relation to road surface. Given relative orientation of stereo-rig, an at-scale 3D model of features can be obtained by matching homologous points between left and right images. However the resulting 3D coordinates are expressed in a relative system (often in the coordinate system of the left camera). In order to integrate the reconstructed model in model of whole city, the coordinates must be transformed to an absolute system. In this purpose the exterior orientation of stereo-rig at each exposure point is needed.

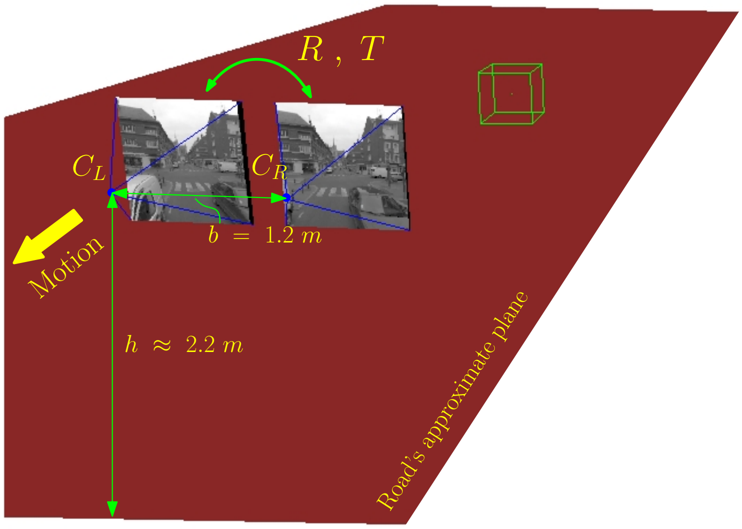


Figure 2.8: Calibration of *Stereopolis*'s horizontal stereo-rig, (R,T): Parameters of relative orientation, h approximate height of stereo-rig in relation to road surface, b Length of stereo-rig, In green, a one meter cube gives the scale.

2.2.4 Exterior calibration of the stereo-rig

Exterior calibration of stereo-rig consists in estimating its orientation (translation and rotation) at each exposure point in an absolute system. As mentioned in Section 1.2 most of the MMSs use GPS/INS for absolute orientation but it provides insufficient accuracy in urban areas. In our project, exterior calibration of stereo-rig is unknown. So the reconstruction must be performed using only one stereo-pair for each exposure point and strategies that are based on multi-image (successive stereo-pairs) can not be applied. However after estimating absolute orientation (by matching aerial and terrestrial images using roadmarks), it can be applied in a validation or rejection strategy to improve the completeness of results.

2.3 Objectives

2.3.1 Which kind of roadmarks to reconstruct?

Different kind of roadmarks are used on paved roads to provide guidance and information to drivers and pedestrians. Countries categorise roadmarks in different ways. They vary from the most common ones such as zebra-crossing (crosswalk) and dashed-lines to more particular marks such as arrows, stop lines and words (BUS, SCHOOL, etc). We aim at reconstructing zebra-crossings and different kinds of dashed-lines because on the one hand they are available on most of the roads and on the other hand they provide useful information about number and width of the lanes.

2.3.2 What level of generalisation?

As mentioned in Section 2.2.1 in real urban areas, white strips are often deformed in relation to their theoretical shape. In order to simplify the representation, we aim at retrieving their theoretical shape. We suppose that each strip is a planar object. The four vertexes of a strip are then computed at the intersection of its long sides and transversal sides on the strip's plane (See Figure 2.9).

2.3.3 How should we represent them?

Our reconstruction has to provide a 3D vector model for each strip of zebra-crossing or dashed-line. What we mean by a vector model is a set of four 3D points that are linked together to provide a quasi-parallelogram shape. Vector representation compress the information. Moreover it is more easy to integrate such information in a Geographic Information System (GIS). Let's recall that in real urban areas the unfavoured traffic conditions occlude often some of strips within a dashed-line or a zebra-crossing (See Figure 2.6). So modelling each strip independently instead of zebra-crossing or dashed-line in its whole, let to extract information even if most parts of the roadmark object is occluded.

2.3.4 What accuracy should we achieve?

As mentionned before in Section 1.3 the roadmarks are used for road database generation and also for sub-decimetric localisation of the system. Nowadays for urban feature database generation only a decimetric accuracy would be sufficient. However in our system, roadmarks will be used as control points to integrate a bundle adjustment of the ground-based images. This is the reason why our reconstruction algorithm must be as precise as possible. Since our localisation strategy consist in matching aerial and terrestrial roadmarks and knowing that aerial images are oriented with 2 *cm* accuracy, we aim at providing same accuracy in ground-based reconstruction.

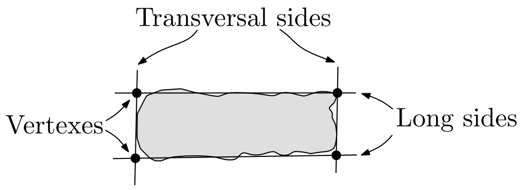


Figure 2.9: Generalisation of a deformed strip to a planar quasi-parallelogram shape.

2.4 Review of previous work on road extraction and reconstruction from ground-based data

Until now, many research works in computer vision and photogrammetry have investigated road detection and reconstruction from aerial and satellite images [Heipke *et al.*, 1995; Ruskoné, 1996; Mayer *et al.*, 1997; Zhang, 2003; Tournaire *et al.*, 2006]. These methods can cover wide areas; however their efficiency is limited by spatial resolution. Ground-based imagery can provide higher resolution for large scale road modelling. Road extraction from ground-based data has been investigated mainly by two communities: photogrammetry and robotics.

- **Robotic oriented approaches**

  These systems use the road geometry for their positioning. Road geometry is extracted and then position of the system in relation to road is estimated. The need of positioning at each instant implies an automatic real time road extraction algorithm. This is why, contrary to cartographic approaches in autonomous navigation and robotics the accuracy of extracted object is less important than computation time.

- **Cartographic oriented approaches**

  In this category of approach the final goal consists often in generation of a geometric database that integrates large scale 3D city models. Accuracy and reliability of results have priority over computation time. So if semi-automatic algorithms only reduce production time, they can be acceptable.

An ideal method of road extraction from ground-based data is a combined method that takes benefit from both the automation ideas used in autonomous navigation systems and the accuracy provided by cartographic approaches. This chapter reviews a large number of existing methods in each of the two mentioned categories.

2.5 Review of robotic or computer vision oriented approaches

Approaches developed in autonomous navigation systems for road geometry extraction are quite different due to difference in strategy, input data, type of feature used for road identification and precisions. We classify these systems in two major categories of monocular and binocular systems.

2.5.1 Road extraction from monocular systems

The goal of these systems is to automatically extract road features in 2D image space. Different road features such as lane markings, lanes and boundaries have been the focus of extraction in these systems. They do not provide 3D reconstruction. Considering detection, we can identify four types of methods:

- Road feature from pixel classification,
- Road feature from region classification,
- Road feature from structure,
- Road feature from frequency analysis.

Road feature from pixel classification

In these methods pixel of images are analysed for classification in one of the road feature classes such as white lines, asphalt or boundaries. In different systems colour, grey scale and texture information are used as classification criteria.

colour-based classification

Charbonnier *et al.* [1998] present a method for detection of road borders using colour images. It starts with computation of chromatic saturation for all pixels in a predefined region of interest. Then the histogram of chromatic saturation is analysed until two clusters are presented. Chromatic saturation express the proportion of colors in each pixel. Road regions being often of gray color have chromatic saturation values that are near 0. This value is near 1 for pure colored objects (red,green and blue objects). The advantage of chromatic saturation is that it is quite insensitive to shadows and pavement variability. So the lower saturation in saturation histogram correspond to the pavement and the other one to the non-road class. The last step consists in extracting road borders from segmented regions. A robust regression technique is then applied to discard the outliers. Borders are presented as line-segments.

Chiu and Lin [2005] propose another segmentation technique that is based on colour images. It is used for lane boundary detection. The algorithm uses region of interests to find out a threshold using statistical methods in colour image. The threshold is then used to retrieve possible road boundary and a quadratic function is used to approximate it.

The work presented in [Hu *et al.*, 2004] use multiple features (color and edges) to segment road area. This is the reason why the method can be classified as road from pixel or region classification and also road from structure. Considering color information image is classified in 3 classes. They suppose that road regions are gray with a little blue and never green or red. Considering color information, pixels are classified into three classes of road, non-road and uncertain road. Then the coherence of uncertain road class with edge information is analysed to decide if it can be re-classified in road class.

Intensity-based classification

McCall and Trivedi [2004] present filters called *steerable filters* that are based on second derivatives of two dimentional Gaussians to detect pixels belonging to painted lane-marking and circular reflectors on road surface. This filter is applied to gray level images and is robust to lighting changes and shadows. It can be applied to extract the lane-marking with a given direction. This property is used in a tracking mode. Predicted lane direction defines the steering angle and an adaptive angle filter is applied to retrieve the exact pixels of the lane-marking.

In most of the vision based navigation systems, position of lane-marking in relation to the camera is used as navigation information. So the reliability of detected lane-marking affect directly the reliability of positioning. Illumination condition and intrinsic properties of lane-marking are the main influencing characteristics in lane-marking detection methods. Kluge and Johnson [1995] propose a system for characterizing the visual characteristics of painted lane markings. The system consists of a camera that takes the fronto parallel images of the road surface. The image area is fairly small so only asphalt and lane-markings are visible (See Figure 2.10). A histogram-based threshold is then selected and the images are binarized using this threshold. The pixels belonging to the left and the right sides of lane-markings are extracted and approximated with line-segments. For each detected lane-marking some characteristics such as width, mean intensity of inside and outside pixels, starndard deviation and contrast are computed.

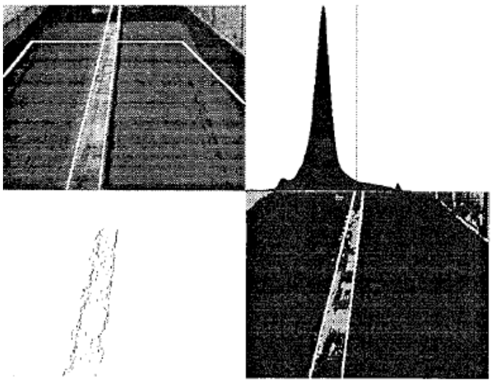
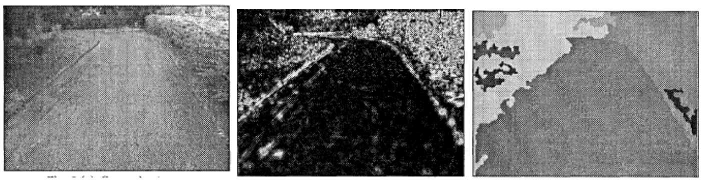


Figure 2.10: According to [Kluge and Johnson, 1995]: up-left: original image with corresponding boundaries and strip edges drawn in white, up-right: intensity histogram with threshold marked by vertical black line, bottom-left: left-/and right most strip pixel in each row, bottom-right: thresholded image with strip edges drawn in white.

Texture-based classification

In addition to gray level and colour, some authors use texture as an additional feature to robustify road segmentation. Jeong and Nedevschi [2005] propose a combined method that takes benefit from gray level as pre-classification. The final classification is performed as a weighted combination of pre-classification and texture-based classification. Two types of rural and highway roads are studied.

In the roads without lane-markings and colour difference between road and un-road areas like rural roads only texture information can be useful. Zhang and H.H. [1994] propose a new technique of classification that is only based on texture information. Strength of texture anisotropy is used as classification criterion supposing that road regions are strongly anisotropic. A measurement of texture anisotropy is defined and computed in image space. Classification is then performed in two phases. First an initial unsupervised classification is performed. The resulting classes are then analysed in a supervised classification that compare anisotropy of each detected class againts the stored statistical model of each class. Figure 2.11 shows results of unsupervised classification.



(a) Gray value image (b) Strength of texture anisotropy (c) Segmentation result

Figure 2.11: According to [Zhang and H.H., 1994]: The calculated strength of texture anisotropy in a road image and the unsupervised segmentation of the image in 3 classes.

Remarks on road feature extraction approaches from pixel classification

Regarding our objectives (mentioned in Section 2.3) the disadvantages of pixel-based road extraction methods can be summarised as follows:

- The mentioned methods are applied in rural images supposing a free space in front of camera. Often some thresholds are estimated within a region of interest defined in the supposed free area. In real urban area, unfavourable traffic conditions can disturb these kind of hypothesis and as a result the detection will fail.
- Results of such methods is a binary pixel classification of road or roadmarks. It is still too low level compared to object extraction that we look for.

Road feature from region classification

In these kind of methods, image is segmented into connex pixels that form regions corresponding to different road feature classes such as asphalt, lane-marking and non-road areas.

Gonzalez and Ozguner [2000] uses the fact that lane marking are brighter than road background and propose an histogram-based segmentation method. The method is basically a region growing segmentation, where the growing parameters are function of histogram characteristics. The lower narrow horizontal strip (5 pixel) of image is considered. They suppose that the dominant object in this area will be the road. Under this hypothesis the histogram of this area is going to be unimodal and even in some cases a Gaussian. The minimum, maximum and mean values of the histogram are extracted. These values are used for region growing supposing that all pixels with a value higher than maximum value belongs to lane-marking regions. The values between the minimum and the maximum values correspond to road regions and finally the values smaller than the minimum value are considered as potential obstacles. The found road-markings are then analysed by a decision tree testing their relative geometric relations such as distance, orientation and width to validate candidates and cope with outliers. Regarding this last step of the algorithm, the approach could be classified also in road feature from structure, however structure information (orientation, length, interactions) are applied in a validation phase and not in detection phase.

Broggi [1995] suggest an image resampling in a plane which is parallel to the road surface to reduce the perspective deformation. Then given the width of a lane-marking a low level filtering is performed supposing higher intensity on a lane-markings than its outsides. A simple threshold seldom provides a complete detection due to shadows and different illumination conditions and provide disconnected pixels along lane-marking. An enhancement is provided by *geodesic morphological dilation* that use correlation of pixels along lane-marking. The final result is a set of connected pixels classified as lane-markings in image space. The Figure 2.12 shows the results of different steps of this approach.

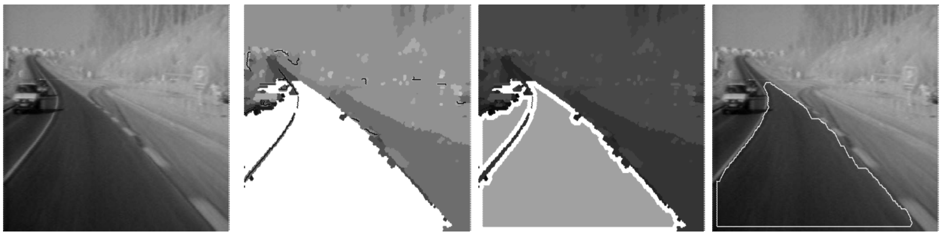
(a) Original (b) Remapped (c) Filtered (d) Enhanced (e) Binarized

Figure 2.12: According to [Bertozzi and Broggi, 1998]: Sequence of images produced by the low-level lane detection phase.

Still using mathematical morphology another research work presented in [Beucher *et al.*, 1990] aims at classifying lanes in a road image. A morphological regularized gradient followed by a watershed is used to segment road region (See Figure 2.13(b)). A top-hat transform is used to detect lane-markings (See Figure 2.13(c)). The extracted road-markings are then used for segmenting the road region into lanes (See Figure 2.13(d)).

Remarks on road feature extraction approaches from region classification

Comparing to pixel classification approaches, region classification methods are more interesting since they provide objects instead of pixels. Extraction of objects, instead of isolated pixels make possible validation tests such as size,direction, interaction between objects, etc. This can robustify considerably the extraction algorithm. However the geometric precision of extraction remains quite poor (> pixel). Moreover like for the pixel classification algorithm these approaches are applied in highway and rural roads with a large free space in front of camera. The frequent occlusions in real urban areas (See Figure 2.6) can reduce considerably the efficiency of these algorithms.



(a) Original (b) Ground layout detection (c) Lane markings detection (d) lane segmentation result

Figure 2.13: According to [Beucher *et al.*, 1990]: Lane segmentation by mathematical morphology

Road feature from structure

These approaches use different kinds of geometric models such as lines and curves for approximating road features. The extraction problem is often transformed into detection and estimation of given geometric models (structures) within edges points in image space. Three main techniques can be mentioned considering applied structure.

Linear structures

Some of road features such as zebra-crossings and in certain cases lane-markings are constituted of linear features. Due to perspective deformation parallel lines in 3D intersect each other at vanishing point in image space. Some authors use this property to detect road features via vanishing point detection.

Se and Brady [2003] use vanishing point constraints to detect zebra-crossings. The Canny edge detector is used to detect the edge points. A Hough transform is then applied to estimate line-segments. A group of intersecting line-segments are then found by RANSAC as hypothesis for some structures containing parallel lines. However in urban areas there are a lot of structures with parallel lines. In order to eliminate the false ones a constraint is added to the detection algorithm. The constraint is that the edges of zebra-crossing can be partitioned into two sets of equally spaced parallel lines corresponding to light-to-dark and dark-to-light transitions. Figure 2.14 summarises different steps of this algorithm and obtained results.

In a very similar work presented by Schreiber *et al.* [2005] lane markings are used to segment road lanes. The line segments are detected and vanishing point is estimated very similarly to [Se and Brady, 2003]. Within the line-segments passing trough the vanishing point, lane markings are detected. Left and right sides of each lane-marking are then distinguished.

The efficiency of these two methods depends largely on detection of vanishing point corresponding to roadmarks (zebra-crossing and lane-markings). In real urban areas, occlusions due to unfavoured traffic conditions can reduce the efficiency of vanishing point detection and as a result roadmark detection will fail.

In a visual traffic surveillance application presented in [Lai and Yung, 2000] a video camera is used to control permanently the frequency of vehicles lane changes. Complete

(a) Original image (b) Line detection results (c) Detected edges (d) short sides detection (e) Final detected edges after removing short sides

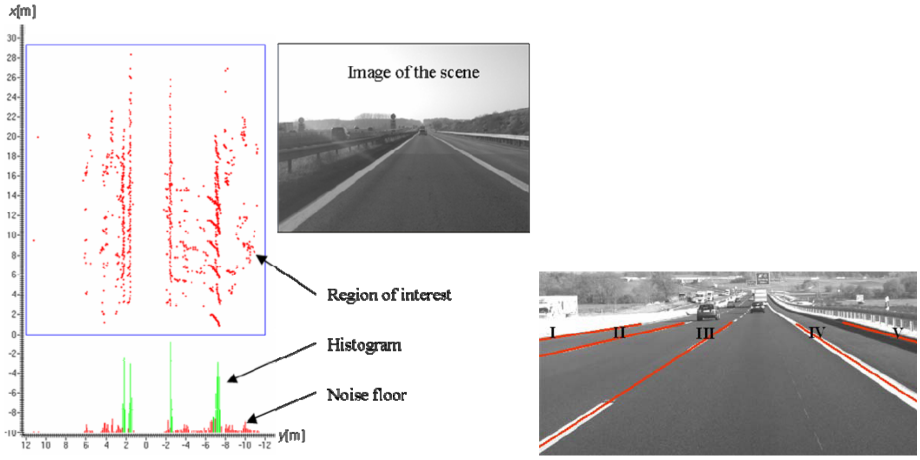
Figure 2.14: According to [Se and Brady, 2003]: Zebra-crossing detection: long sides are detected by vanishing point segmentation, short sides are detected in the orthogonal direction and removed to perform a more accurate Hough transformation for estimating long sides. Enhancement can be compared between (a) and (e)

multiple lane information is needed. The algorithm starts with background extraction by averaging over a set of video sequence taken with an unchanged camera position. The averaging provides an image of road area without any obstacle due to cars. After that Sobel edge detector is applied to background image to detect the edge points. The edge points are then approximated with straight lines. Supposing a planar model of the road and known orientation of the camera, the line segments are projected onto a 3D plane that approximates the road surface. They are then analysed by their prominent orientation and length to reduce the number of false-alarms.

The unchanged position of camera in relation to road surface simplifies one the one hand extraction of image background (road area) and on the other hand 3D estimation of line-segments. In our application the MMS moves permanently along the road surface. Moreover the position of camera in relation to road changes at each image node time and thus is unknown.

In a driver assistant system proposed by Dietmayer *et al.* [2005] a forward looking laser scanner is used to detect roadways and lane-markings. Thanks to difference in the reflectivity of asphalt and lane-marking, the receiver channels of scanner can be adapted to receive only signals coming from lane-markings. So the initial data is a set of 3D points corresponding to lane-marking surface and some noises that are caused by other objects on the asphalt with reflectivity similar to lane-markings. Figure 2.15(a) shows an example of initial 3D points. In order to identify lane-marking reliably histogram of points occurence is evaluated. In this histogram noises cause small values. In contrast, lane-markings thanks to their alignment with road direction, generate higher peaks. The measurements corresponding to higher peaks of histogram are considered as potential lane-markings. Hypothetical lane-markings are then evaluated by lane's width information to filter out false alarms. Finally line-segments are estimated using a Hough transformation (See Figure 2.15(b)).

This histogram based approach takes benefit from real 3D geometry of road features to handle noise. The advantage of 3D space is that information known *a priori* such as lane's width can be used to filter out false positive and identify lane-markings. It is only laser-based, so measurements are acquired directly in 3D space without any processing. However the precision of extraction is limited to the resolution of laser scanner. Moreover, it provides a global model for aliened discontinues lane-markings. We believe that the same strategy can integrate an image based approach for more accurate reconstruction.



(a)

(b)

Figure 2.15: According to [Dietmayer *et al.*, 2005]: (a) Initial data of a highway scenario and corresponding histogram. (b) final line-segments projected in an image of scene .

Curves structures

As vanishing point constraint is not applicable in curved road scenarios, some autonomous navigation systems use circular or parabolic curves as a road model. These methods consist often of an edge point detection followed by a robust curve fitting [Risack *et al.*, 1998; Kluge, 1994; Jung and Kelber, 2004; Ieng *et al.*, 2003]. However in the case of highly noised edge points (in presence of shadows), the curve fitting can fail.

Another technique adapted specially for noisy edge point scenarios is proposed in [Kluge and Lakshmanan, 1995]. The algorithm is based on deformable templates. Instead of thresholding the intensity gradient, the Metropolis algorithm is used to maximize a function that evaluates how well the image gradient supports a set of template deformation parameters.

[Ieng *et al.*, 2004] propose a curve fitting algorithm together with some confidence measures for extracted curves. The method benefits from robustess of *M-estimators* and considers the covariance matrix of the curve estimation as a confidence measure. The proposed method is robust to different weather conditions. Figure 2.16 shows an example of obtained result.

Snake structures

Snakes are known as deformable features to model curves. So as opposed to previous studied curve structures the snakes can be deformed to adapt themselves with more sophisticated curves. Often the deformability power is modelled as the sum of an internal and an external energy that depend on the application context. In opposite to external energy an

Figure 2.16: According to [Ieng *et al.*, 2004]: Curve fitting result under foggy conditions (green curve) and uncertainty curves (red curves).

internal energy tends to preserve a particular shape. A trade-off is obtained by minimizing the total energy that is the sum of internal and external energies. This technique is used in some research works for road lane detection [Borkowski *et al.*, 1999; Wang *et al.*, 2004; Sawano and Okada, 2004].

Use of curve and snake models for lane-marking detection robustifies recognition algorithms. However it is only applicable in highway scenes in which there is a sufficient free space (without obstacles) in front of the camera. In real urban areas the presence of many occlusions and obstacles (See Figure 2.6) reduces consequently the efficiency of such methods. Moreover curve models introduce some approximations that do not take into account local irregularities.

Road features from frequency analysis

Some authors investigated in road feature detection using frequency analysis. Kreucher and Lakshmanan [1999] present a frequency domain feature extraction combined with a deformable template technique in order to detect the lane-markings.

Another research work uses edge filters in the spatial domain to detect edges to which a straight line is fitted. Fourier analysis is then used to recognize continuous from discontinuous lane-markings [Collado *et al.*, 2005].

The main interest of these methods is in their speed. However, dealing with geometric constraints such as orientation and size of features in frequency domain is not straightforward.

2.5.2 Road extraction from a binocular system

The use of stereo images enables us to eliminate the assumption of flat road that is always used in monocular systems for lane detection and provides the real 3D road surface. In addition, it can be applied for obstacle detection using disparity analysis within stereo-images.

3D Lane-marking and lane reconstruction

As mentioned before curve models are used in monocular images for detecting the lane-markings. In these methods road are supposed to be flat and only the horizontal curvature of road is taken into account. Stereo-vision allows to eliminate the flatness assumption and to estimate the vertical curvature of lane-marking too. In this context [Nedevschi *et al.*, 2004; Luong *et al.*, 1995] propose a clothoid model for detecting continuous and discontinuous lane-markings in a tracking mode. The 3D information is used to filter out vertical features before the estimation of curves. This operation robustifies considerably detection algorithm. The Figure 2.17 shows the back-projection of a reconstructed lane using the mentioned method in image space.

In a more advanced work Nedevschi *et al.* [2006] propose a contour based subpixel matching technique that improves considerably the accuracy of reconstruction. Subpixel matching is inevitable in the case of short baselines which is the case for most of the autonomous navigation and mobile mapping systems.

Most of the efforts for lane-marking reconstruction from binocular robotic systems have gone into 3D curve modelling. The benefit of curve estimation from binocular system is the elimination of obstacles before the estimation process. This operation removes some noises from input data of curve estimation. However, as mentioned before, curve models are not well adapted to our specific application.

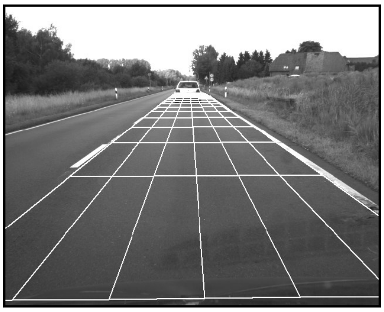


Figure 2.17: According to [Nedevschi *et al.*, 2004]: reconstruction of lanes in presence of obstacles.

Road surface and obstacle detection

Road surface provides interesting information about localisation of autonomous vehicle. Localisation is performed by estimating the measurements such as height of vision system and its orientations in relation to the road surface. Beside for localisation purpose, road surface is interesting for obstacle detection. The existing methods are based on two main hypotheses of flat and non-flat road model.

planar road surface hypothesis

Simond and Rives [2004a] propose an algorithm for road plane detection in urban area. The algorithm consists in two steps. First dominant vanishing point of the scene is estimated. Second coplanar line-segments passing from dominant vanishing point are used to define a region of interest corresponding to road plane. Then interest points are estimated within this area. Finally line-segments and interest points are used to estimate road plane's homography. The Figure 2.18 shows the results of road plane segmentation and reconstructed textured model.

Sappa *et al.* [2007] use a stereo-vision system that provides dense 3D points in real time. This 3D point cloud is then used to estimate road plane. They suppose that road is the dominant feature in the point cloud. A pre-processing group 3D points into cells. Road plane estimation is carried out by an efficient RANSAC operating on the barycentres.

The majority of the obstacle detection techniques apply stereo inverse perspective mapping (IPM) algorithm [Bertozzi *et al.*, 1998a]. The algorithm supposes a flat road and knowing orientation of stereo-pairs in relation to road. Using a planar homography (based on road plane) the right image of a stereo-pair can be resampled in geometry of the left image. So all the pixels belonging to road surface are correctly mapped in the other image but the obstacles are mapped incorrectly. The difference between the remapped right image in the geometry of the left one and the original left image correspond to obstacles. Within the method applying this strategy we can state Okutomi *et al.* [2002]; Bertozzi and Broggi [1998]. An extension of IPM algorithm is presented in [Bertozzi *et al.*, 1998b] to cope with non flat roads and presence of bumps and hills. Other authors have investigated obstacle detection algorithm in non-flat roads scenario.

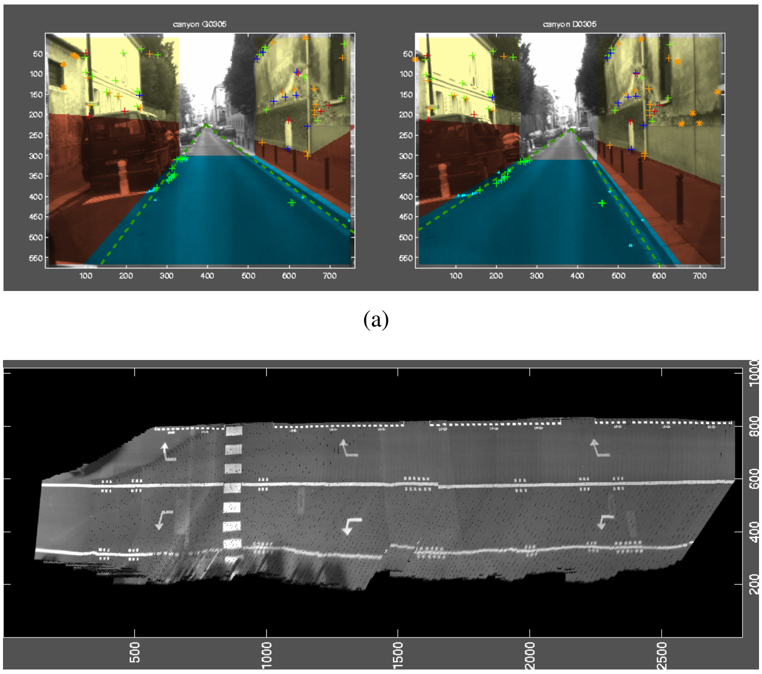
Beside interesting information provided by these methods about road position in relation to camera system, their main benefit is in their rapidity. The main inconvenient is in the planar road hypothesis that disregards longitudinal and transverse curvature of road surface. In [Bertozzi *et al.*, 1998b] the authors present an extension of IPM algorithm to cope with non flat roads and presence of bumps and hills. Other authors have investigated obstacle detection algorithms in non-flat road scenarios.

non-flat road surface hypothesis

Yu *et al.* [2003] do not use flatness assumption and supposes only smoothness of road surface. So the surface can be locally non-flat with an accepted tolerance. A normal vector is computed for each pixel using a disparity map provided by stereo-matching. All the vectors belonging to the road pixels will be around a main direction with an angular tolerance and all other deviated vectors correspond to obstacles. This method is easy to implement. However, quality of results depends largely on the quality of disparity information. Moreover, angular tolerence regulation can be critical.

Labayrade *et al.* [2002] propose another method which supposes that obstacle surfaces are often perpendicular to road surface. Road surface is approximated by a succession of flat parts. The disparity that correspond to road surface is computed on each part via *V-disparity* space. Then on each part, all the pixels with a disparity different from road disparity are considered as obstacle pixels.

Recently a novel method of road reconstruction is proposed by Tarel *et al.* [2007] that



(a)

(b)

Figure 2.18: According to [Nicolas Simond]: (a) the images are segmented in 3 parts : the road plane (in blue) and 2 others (in yellow) assumed as vertical planes, (in red) parts of the scene where the feature points are not coplanar. (b) reconstruction of the road plane.

takes into account the longitudinal variations of road profile. The method uses stereo images and supposes that they mostly contain road. A disparity model that takes into account unknown longitudinal variations of road is defined. In this way 3D road reconstruction is formulated as a matching problem under an unknown parametric road model. These unknowns are computed by an exact optimisation in a multi resolution framework. This method provides more accurate road surface than planar models. Introduction of road model constraint in the estimation, make the method robuster than the method based on *V-disparity* space.

These three methods provide more accurate results than those supposing planar road model. However, they do not still reach perfect road surface. Indeed transverse curvature of road is not taken into account. As a result at each transverse profile of road, the estimation provides an approximation. The resulting error can be significant for narrow streets in urban areas with high transverse curvature.

2.6 Review of Cartographic oriented approaches

The automation of ground-based road data-collection system was motivated first by the need for highway surveying and pavement performance evaluation in long distances. The first generation of these systems consists of a vehicle equipped with film cameras, inertial tools, gyroscopes and odometers. The photos are tagged with time and a geographic position is computed for each exposure [Birge, 1985]. In this stage, acquired images are used as pictorial information without any geometry extraction.

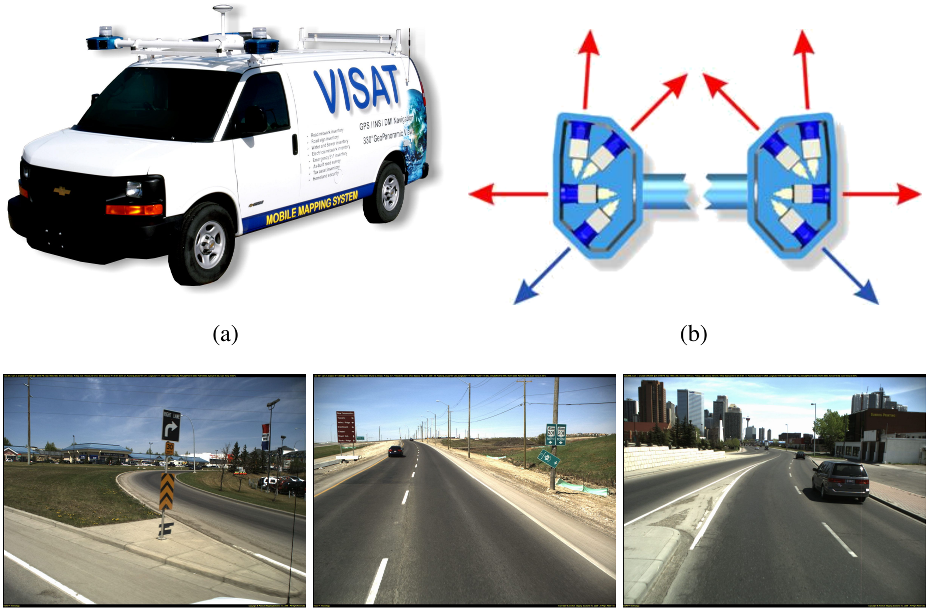
An important advance of such systems took place with the invention of the civil GPS antennas and video cameras. On the one hand video cameras simplify image sequence recording, on the other hand each sequence can be positioned in an absolute coordinate system. The result is a set of georeferenced image sequence along a highway [Schwarz *et al.*, 1990; Lapucha *et al.*, 1990]. Poor image quality and lack of stereoscopic sequences made precise measurement difficult. However some information such as height and offset can be extracted [Tao, 1997].

The most important evolution in data-collection for road took place with the generation of mobile mapping systems thanks to the effort made by the mapping center of Ohio State University and the Department of Geomatics Engineering at Calgary University [Schwarz *et al.*, 1993; Novak and Bossler, 1995]. The main advantage of MMS is that it provides georeferenced digital data that can be analysed directly with computers. Moreover these systems are often equipped with more than one camera allowing stereo imaging and Laser scanners. So accurate 3D measurement is possible. The table 2.3 shows some data-collection systems used for road surveyings.

System	Positioning	Acquisition system	Reference
MHIS	GPS,INS,Gyro.	Video	[Lapucha *et al.*, 1990]
GeoVan	GPS	8 video	`www.geospan.com`
Kronos	GPS, Odometer	1 video	`www.geo-3d.com`
Digital Imaging	GPS	CCD, Video	`www.intlcybernetics.com`
ARAN©	GPS,INS	1 VHS, 2 CCD, Laser	`www.roadware.com`
GPSVision	GPS,INS	Stereo CCD	[He, 2002]
VISAT™	GPS,INS	Stereo CCD	[El-Sheimy, 1996]
Photobus	GPS,INS	1 CMOS	[Gontran *et al.*, 2003]
IRCAN	Odometer	1 CCD	`www.lcpc.fr`
—	GPS,INS	Video, Laser	[Yu *et al.*, 2007]
LARA-3D	GPS,INS	Laser	[Abuhadrous *et al.*, 2004]
—	Indirect methods	2 Lasers ,1 CCD	[Früh and Zakhor, 2005]
GeoMaster	GPS,INS, Odometer	6 line CCD 3 Lasers	[Zhao and Shibasaki, 2003]

Table 2.3: Some mobile mapping systems used for road inventory.

VISAT™(See Figure 2.19(a)) was developed at the University of Calgary in the early 1990s [El-Sheimy, 1996]. It was one of the first terrestrial MMS. Recently, an improved version is developed by Absolute Mapping Solutions Inc in Calgary, Canada. This system integrates a cluster of digital cameras (See Figure 2.19(b)), receiver of the Global Positioning System (GPS), an Inertial Navigation System (INS), and Distance Measuring Instrument to collect panoramic views along roadways throughout the project area. The Figure 2.19(c)



(a) (b)

(c)

Figure 2.19: According to [Absolute Mapping Solutions Inc.]: (a) VISAT™ MMS, (b) Vision component, (c) samples of provided georeferenced images.

shows some georeferenced images provided by this system.

An MMS called GPSVision is proposed in [He, 2002]. This MMS provides georeferenced stereo-pairs. A software is provided for visualising and processing the stereo-pairs. The user can click on the interest points in the right and left images and a 3D position is computed in an absolute coordinate system by photogrammetric triangulation. After adding some semantics information to the reconstructed feature, it can be directly integrated into GIS.

The very large amount of data collected by these systems implies a high degree of automation of feature extraction and reconstruction. Li *et al.* [1996] propose an object-directed and user-guided strategy for semi-automatic feature reconstruction. The user recognizes a feature and activates an object-directed algorithm to automatically reconstruct the feature. A number of object-directed algorithms such as road centerline, road boundary and vertical-linear objects are proposed.

In the following, we give a classification of existing methods in the literature and industrial solutions for road inventory. The classification is performed in relation with the feature of interest.

2.6.1 Road signs extraction

The georeferenced image provided by mobile mapping systems proved to be useful sources of data for road features inventory generation. Among these features, road signs are important for many applications such as autonomous navigation projects. These applications motivated automatic extraction of road signs from ground-based images. In the following we give some references of related works.

Tao [2001] proposes an algorithm for reconstruction of vertically structured objects such as road signs, light poles, etc. The algorithm is based on detection of vertical edges in image space, followed by line-segment estimation and line matching. The reconstructed vertical lines are then intersected with ground model to obtain the footprint of vertical object. This approach is quite generic for extraction of vertical objects, but does not recognise kind of road sign.

Habib *et al.* [1999] propose an algorithm based on classification of color images for hypothesis generation. This step takes benefit from known colors of road signs. Then during a validation step, the geometric characteristics of the hypothesis is analysed to see how well they fit the standard geometric model of road signs. The result consists in boundary of detected road signs in image space.

Dahyot *et al.* [2000] propose a robust unsupervised pattern recognition system which is based on M-estimators. This method is generic and not limited with any particular application. Performance of the method in recognition of road signs are demonstrated under adverse conditions such as occlusions and noises.

2.6.2 Road surface inspection

These methods provide information about geometry and condition of road surface.
Some commercial imaging systems include ARAN©, GeoVan and Digital Imaging and Profiler Systems (See Table 2.3).

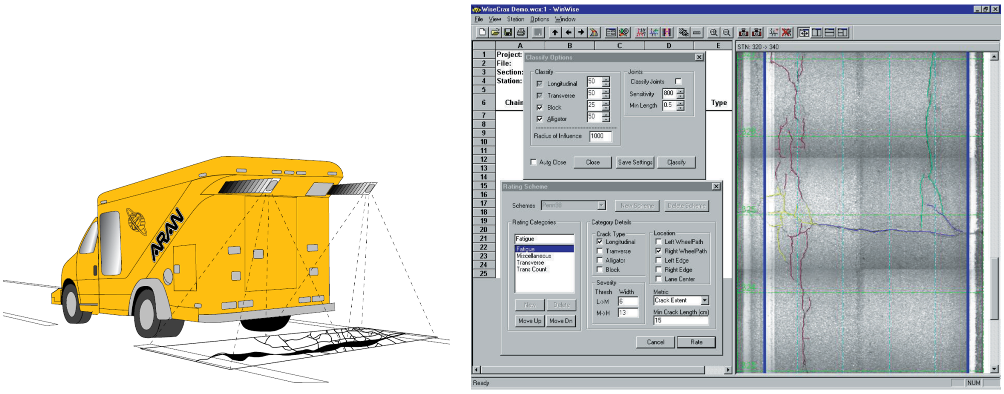
The ARAN© system is developed at Roadware Group Inc. for collecting pavement condition. It uses laser scanners and video camera to provide geometric road profile and rut measurements. Figure 2.20 shows the data collector system and a screenshot of proposed software for crack measurement.

GeoVan is an MMS developped at Geospan Corporation for collecting road distress information and pavement condition from video sequences.

The International Cybernetics Corporation propose the Digital Imaging System for distress measurements and Profiler System for flatness measures and road slope draining system checking.

Within scientific papers Yu *et al.* [2007] uses video sequences and laser scanners to collect road data. A 3D surface reconstruction is proposed that process the raw data provided by a laser scanner. The provided accurate 3D model is textured with higher resolution 2D image. The output is used for characterisation and visualisation of cracks.

Laurent *et al.* [1997] uses two synchronized laser scanners in order to obtain transverse 3D and intensity profiles. An algorithm is proposed for rut and crack detection. The output of algorithm is a crack map of the pavement.



Classify:
☑ Longitudinal 50
☑ Transverse 50
☑ Block 25
☑ Alligator 50
Radius of Influence 1000

Joints:
Classify Joints ☐
Sensitivity 800
Min Length 0.5

☐ Auto Close

Schemes: Penn98

Rating Categories: Fatigue
Fatigue
Miscellaneous
Transverse
Trans Count

Category Details

Crack Type:
☑ Longitudinal
☐ Tranverse
☐ Alligator
☐ Block

Location:
☐ Left WheelPath
☑ Right WheelPath
☐ Left Edge
☐ Right Edge
☐ Lane Center

Severity: Thresh Width
L->M 6
M->H 13

Metric: Crack Extent
Min Crack Length (cm) 15

(a) (b)

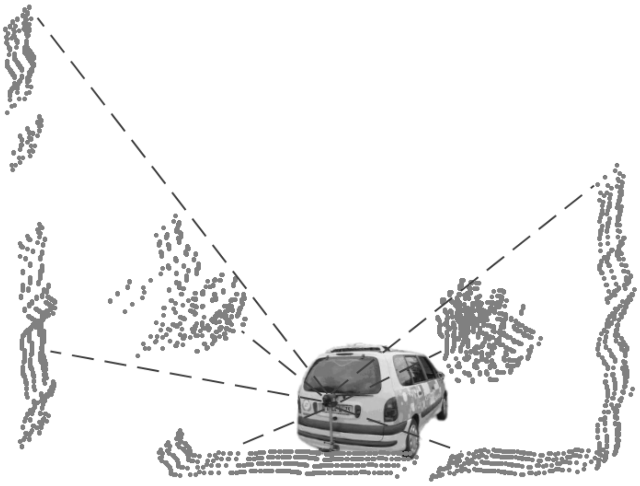
Figure 2.20: According to [Roadware Group, Inc.], (a) Roadware's Automatic Road Analyzer (ARAN©), (b) WiseCrax© system screen: automatically detected cracks overlaid on pavement image.

LARA-3D is developed at *Ecole Nationale Supérieure des Mines de Paris* for digitising urban areas (See Figure 2.21(a)). It is equipped with a 2D laser scanner for data collecting and GPS/INS and odometer for positioning [Abuhadrous *et al.*, 2004]. They propose some real-time algorithms for segmentation and modelling of different features in the city such as facades, roads, trees, etc. Segmentation is performed only on the geometry of the scene considering direction (vertival, horizontal, etc.) and the shape of objects (planes, spheres, etc.). Each segmented object is then modelled independently. Road's point cloud is processed by a decimation algorithm that detects road borders. Figures 2.21(b) and 2.21(c) shows the results before and after decimation algorithm. Some higher level information such as road curvature and width are computed with the provided data.

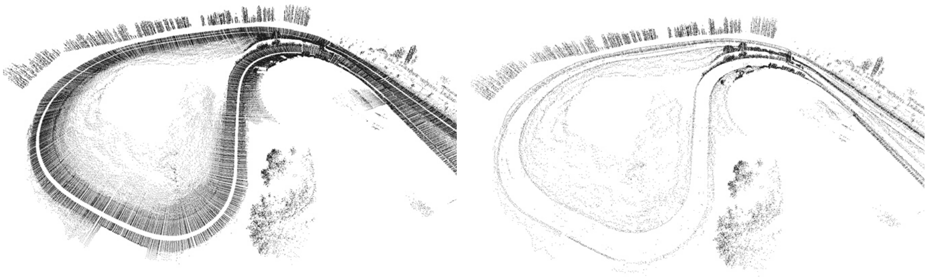
2.6.3 Road lane-markings and boundary extraction

In cartographic applications lane-marking and boundaries give useful information about number of lanes and road width.

Gontran *et al.* [2006] propose a MMS that consists of GPS/INS for georeferencing and a CMOS camera that is oriented vertical to road surface. The goal in this system is to provide a cartographic database of road centerline in the trajectory of the system. Due to specific orientation of the camera in relation to the road, the lane marking detection in this system is based on binarization of the image using a radiometric threshold. In each image sequence the barycentre of all the pixels passing the threshold is calculated as the center of lane-marking. The absolute positions of detected elements are then computed using the known position of the vehicle and the orientation of camera in relation to the vehicle. This method approximates lane-markings with its barycenter at different positions of the vehicle along the road. This is too generalised compared to what we try to reconstruct (See Section 2.3).



(a)



(b) (c)

Figure 2.21: According to [Goulette *et al.*, 2006]: (a)LARA-3D, on-board laser range sensing prototype, (b) Extracted laser points, (c) decimation result on road scene.

The first trial in lane-marking extraction was performed by He and Novak [1992]. The input of the algorithm consists of georeferenced stereo-images provided by a MMS and the output is a set of 3D line-segments that represent road edges and centerline. The algorithm starts with line-segments detection on each image and only the line-segments passing trough an approximated vanishing point are saved. Then the line-segments of similar slope and offset are grouped to longer line-segments. Finally line-segments of two images are matched and provide 3D line-segments as road boundary and centerline. This method supposes that the longest line-segments represent the roads lane-markings. This hypothesis can be violated in an unfavourable traffic condition and cause false road boundary or centerline.

A more complete and robust algorithm is proposed by Tao *et al.* [1998]. Input of the algorithm is a set of georeferenced stereo-images and the output is a 3D curve that represents centerline of the road. The method uses a predicted road centerline that is provided by GPS/INS tragectography. The predicted centerline is projected in the left and right images. An oriented edge detector is then applied in the predicted centerline direction and edges are extracted in its neighbourhood. The fact that centerlines are often lighter than their background, opposite gradient direction of their two sides, their approximate width and direction as well as their continuity are the criteria that are used to recognize a centerline in each image. The detected centerlines in stereo-images are then matched in a rectified geometry in relation to the road and provide a 3D centerline curve. This curve is then optimized to minimize an energy according to external and internal forces. The method is tested on a large number of image sequences and provides accurate results under different conditions.

Habib [2000] proposes an algorithm for road edge matching within stereo-images. The method starts with an edge detection followed by straight-line extraction by Hough transform in each image. After a matching process the candidate 3D straight-lines are computed by a photogrammetric triangulation. The candidates are projected in the second stereo-images. In order to validate candidates a statistical method called Multiple Hypothesis Approach (MHA) is used.

Roncella and Forlani [2006] propose a semi automatic method for road width extraction from stereo-images acquired by a MMS. In order to compute road width, the algorithm tries to detect the lane-marking at the two boundaries of road surface. It starts with rectifying each stereo-images on a horizontal plane in relation to road. Edge points are then extracted on these resampled images. As lane markings are the linear features, a Hough transform is used to obtain straight lines. The presence of other linear features can create ambiguities during the extraction. In order to reduce these ambiguities the Hough transform is applied on a restricted area, which is initially defined by a user on the first stereo-pair and tracked on the following ones. The results of width computation are compared with manual measurements and maximum of difference reaches 10 *cm*. Figure 2.22 presents an example of obtained results.

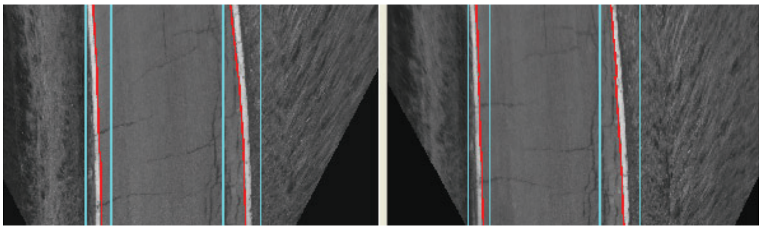


Figure 2.22: According to [Roncella and Forlani, 2006]: Example of curved path. The circular boundaries are approximated by several straight edges.

Cheng *et al.* [2007] propose ARVEE (Automated Road Geometry Vectors Extraction Engine), that is a road geometry extraction system. The input data consists in a set of georeferenced stereo-images which are provided by the VISAT™system. The algorithm consists in four steps. The first step extracts the linear features in each image. Second, the linear features are filtered and grouped into lane line-segments. Third, the 2D line-segments are matched and provide 3D line-segments. Fourth, the lane line-segments are classified into different types by a synthetic analysis. Figure 2.23 shows an example of provided results. As can be seen on this figure algorithm is efficient in free areas but generates some false positives in presence of occlusions.

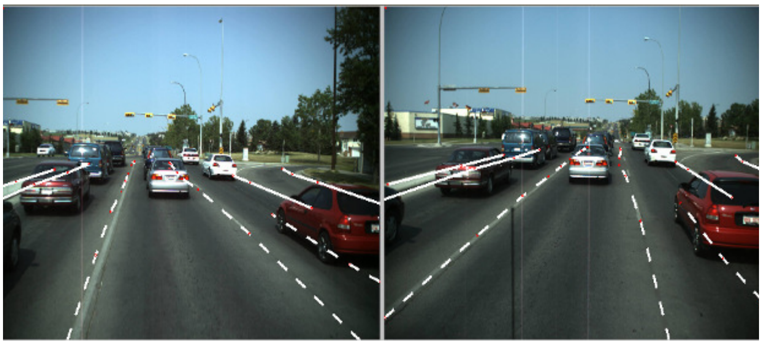


Figure 2.23: According to [Cheng *et al.*, 2007]: Example of correct reconstruction in the center of image and false detection in heavy occlusions.

The four mentioned algorithms suppose linear and curve models for lane-markings. The extraction is so an approximation of a curve or line along entire strips of a dashed-line. On the one hand the efficiency of these algorithms is reduced by occlusion (cars, pedestrians, etc.), on the other they generalise dashed-lines with a global model (curve or line) and are not able to delimit different strips within a dashed-line. As discussed in 2.3 this level of generalisation is not acceptable in our application.

2.7 Conclusion and our strategy

Large number of existing algorithms of roadmark extraction from ground-based systems are mentioned in Sections 2.5 and 2.6. These systems are different from data collection, data analysis and applied assumptions points of view. Efficiency of these methods in presence of problems such as occlusions, illumination variation, perspective deformation vary widely. The reader is invited to refer to comparison of the most common methods which is provided in Table 2.4.

As explained before, we look for a fully automatic method for 3D reconstruction of roadmark objects. Required properties of method are as follows:

- Robustness to illumination variations: This is why radiometric thresholding is avoided.
- Robustness to partial occlusions: For example occlusion of some strips of a dashed-line must not stop reconstruction of other strips. The methods supposing global parametric curve or straight-line models for dashed-lines can suffer from partially occlusions and as a result cause false detections or miss detections (See Figure 2.23). At a higher level, the algorithm must allow to reconstruct a strip even if it is partially occluded.
- Vector representation : This representation simplifies integration of data in a GIS and also use of data in linked applications. This is why pixel-based and region-based strategies are not adopted. For this purpose, edges and line-segments seems to be particularly interesting.

- Capacity for integrating *a priori* knowledge (size and shape): As mentioned before, the shape and size of roadmark objects are known. It is useful to introduce this knowledge in the roadmark recognition process. However, objects are often deformed by perspective effect in image space. This deformation reduces considerably efficiency of metric constraints. This is the reason why 3D methods such as stereo vision are of great interest.

- High geometric accuracy: As explained before we need to reach a 2 *cm* geometric accuracy. At this precision road surface can no more be approximated by a plane. This requirement justifies a precise 3D reconstruction using stereo matching techniques.

In order to fulfil all the above requirements, we propose an algorithm that consists in three main steps of 3D linked-edge reconstruction, roadmark detection and roadmark modelling.

1. The first step provides 3D linked-edge by matching left and right image (Chapter 3).
2. Second step consists in filtering of reconstructed linked-edges. It provides hypothetical roadmark objects (Chapter 4).
3. The last step provides a fine model for the validated hypotheses and reject the other ones (Chapter 5).

Flowchart of Figure 2.24 depicts the global strategy. Next part of the manuscript explains this roadmark reconstruction algorithm.

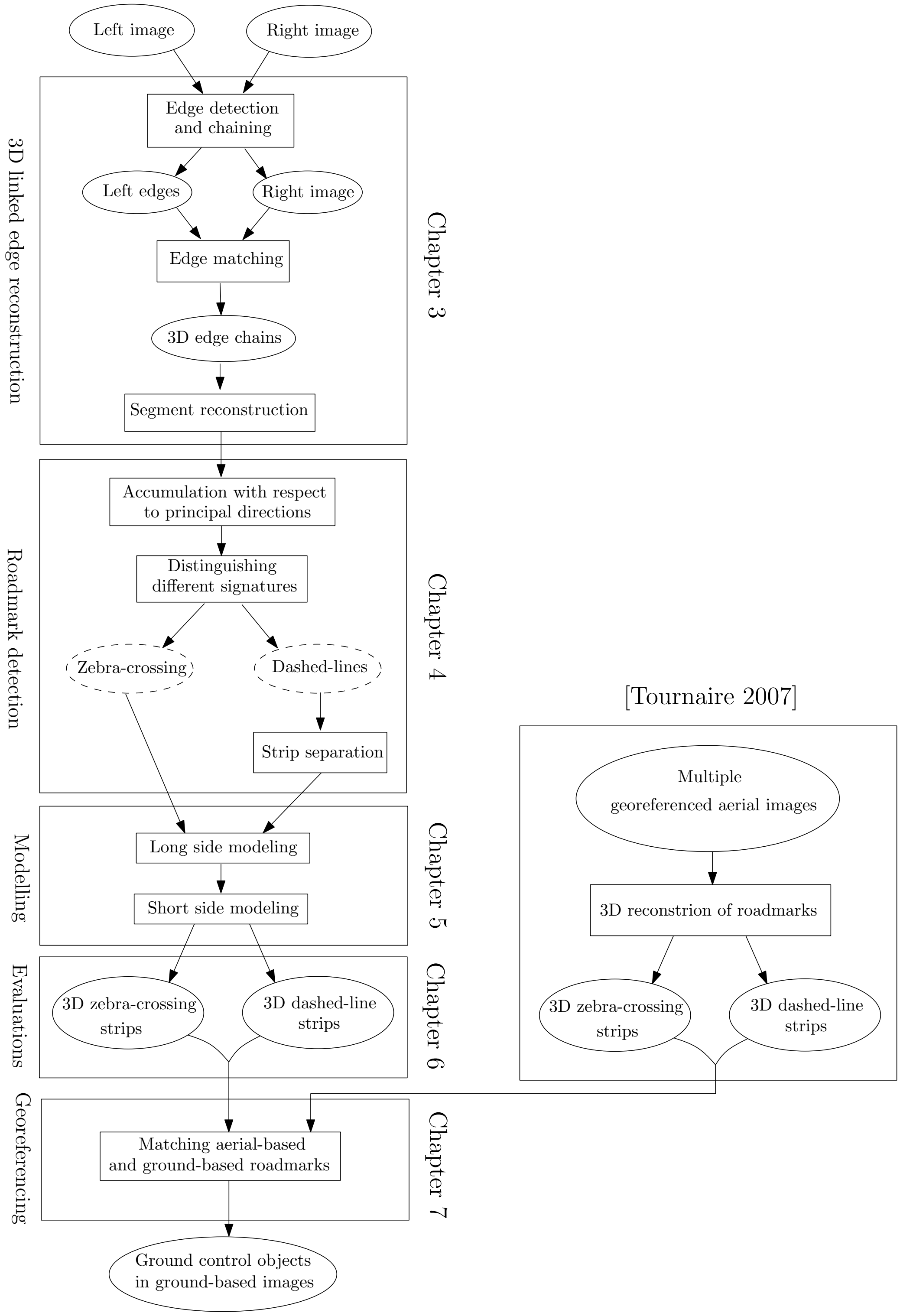


Figure 2.24: Our roadmark reconstruction and georeferencing strategy.

	Pros	Cons
Radiometric thresholding	Easy and rapid	choice of threshold is critical
Region of interest (ROI)	Fast processing, Less false positive	Choice of ROI is critical
Known size and shape of roadmark strips	Enhancement of the search criterion	In image space, shape and size of objects are deformed due to perspective effect
Global parametric models (curve and straight-line) for continuous and discontinuous lane-markings	simplifies the detection problem, Robustify detection to false positive, well adapted to highway scenes	generalisation of discontinuous lane-marking to continuous one, unable to detect isolated strips .
A priori knowledge on the road shape and its position in relation to vision system	Simplifies mapping between image space and word coordinate system	Hypotheses (e.g., flat road, position of vision system in relation to road) are not always met in real cases
Single image analysis	Simple and fast algorithmic solutions	Perspective deformation, less of information about real position and size of objects in word coordinate system, high sensibility to occlusions.
Stereo vision	Provide 3D reconstruction, allows application of detection criteria in the pertinent space (word coordinate space) without any deformation due to perspective effect. Low sensibility to occlusions	Computationally complex, imply to know the exact relative orientation between left and right cameras.
Laser scanning	Provide 3D reconstruction without any complex computation, low sensibility to occlusions, weather and illumination condition	High discretisation effect
Multi stereo vision	All the pros of stereo vision + lower sensibility to occlusions,	All cons of stereo vision + implies to know the exact absolute orientations of all stereopairs.

Table 2.4: Comparision of different approaches for roadmark extraction.

Part II

3D reconstruction of roadmarks

Chapter 3

Contour matching and 3D linked edge reconstruction

3.1 Stereo-restitution principles

3D automatic reconstruction from stereo-image is widely applied in different domains of cartography, inverse engineering, machine vision, etc. It can be defined as the technique providing 3D information from two images of an object taken from different points of view. It consists of three main steps:

- Sensor pose estimation,
- Stereo matching,
- Triangulation.

The result of stereo-restitution is a set of 3D coordinates of points describing the matched features.

3.1.1 Geometric caracterisation of stereoscopic vision system

Figure 3.1 demonstrates a stereoscopic vision system. C_L and C_R represent the projection centers of the left and right camera respectively. The distance between two projection centers is called Baseline (B). The Projection of 3D point A in left and right images are a_R and a_L. Two pixels in two images are called homologue points if they correspond to the same 3D point. The variation of homologue points within stereo-images is called disparity ($parallax_A = a_L - a_R$). Knowing the geometry of stereo-images the parallax can be transformed to depth information (D). Equation 3.1 indicates the relation between *parallax* of a point (represented by ρ) and its depth [Moffitt and Mikhail, 1980]. For simplification, it is supposed that the two image planes are parallel. f is the focal distance.

$$D = \frac{Bf}{\rho} \tag{3.1}$$

In a stereoscopic system, the projection centers (C_L , C_R), the 3D point (A) and its projections in the left and right images (a_L , a_R) are belonging to a plane called the *epipolar plane*.

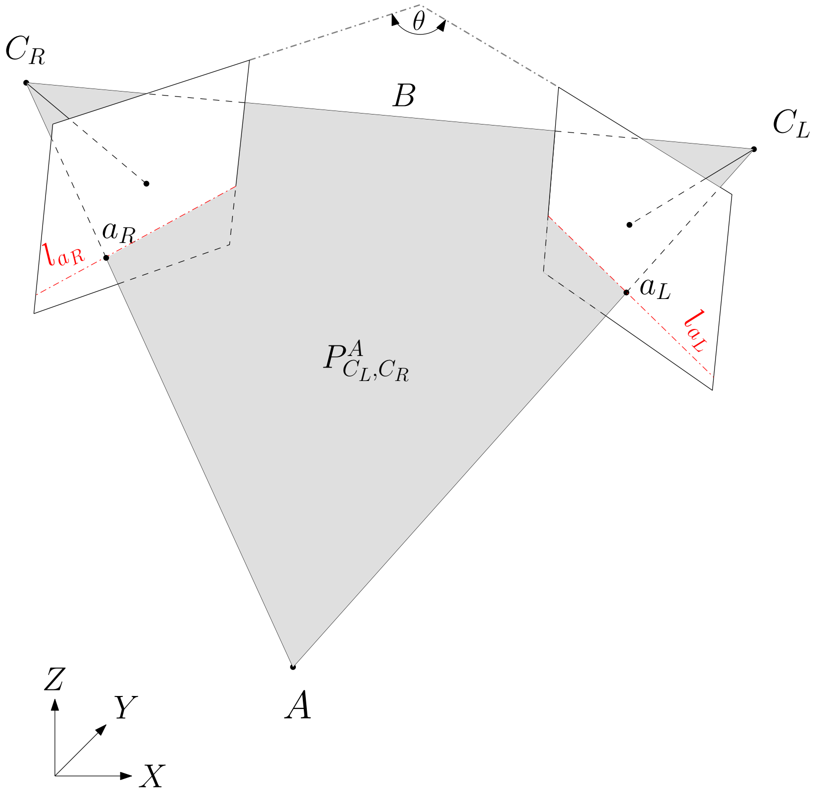


Figure 3.1: Photogrammetric stereoscopic geometry.

The intersection of the *epipolar plane* with the image plane are called *epipolar lines* (l_{a_R} , l_{a_L}). The homologue point of each point on l_{a_R} is on l_{a_L} and vice versa. Given a point in the left image (p_L) and knowing the geometry of stereo-images, the corresponding *epipolar line* in the right image (l_{p_R}) can be estimated. Thus, search space for homologue point can so be reduced to a 1D space. The position and orientation of the *epipolar lines* in each image is a function of the relative orientation of two images. However the two images can be resampled in order to align the *epipolar lines* with image rows. This resampling can simplify considerably the implementation of matching algorithm.

An important factor in stereoscopic systems is the relation between the baseline and objects' depths (B/D). The longer the base the better the depth discrimination. Equation 3.2 expresses the relation between depth precision (σ_D) and B/D [Moffitt and Mikhail, 1980]. σ_ρ indicates the precision of the *parallax* (disparity) measurement. Often the stereo-base is adapted to the distance from object to provide the required accuracy.

$$\sigma_D = \frac{D\,\sigma_\rho}{(B/D)f} \tag{3.2}$$

3.1.2 Stereo matching

As explained in Section 3.1.1, knowing the orientation of stereo-images, features can be reconstructed in 3D by photogrammetric intersection using corresponding points. Finding point correspondence is often the most complex problem in vision-based 3D reconstruction algorithms. Thinking of a matching algorithm consists in answering two main questions: 1)What to match? 2) How to match? The answer of both questions depends on the context and objectives of application in terms of shape of the objects, accuracy, level of details, computing time, type of surface representation etc. Many researchers have investigated the problem of image matching. The proposed algorithms are different as the contexts and applications diversify. Differences between the matching algorithms are often due to the three following choices:

- Type of primitives,
- Type of matching constraints,
- Type of global matching strategy.

The first aspect stands for the question of «what to match » and the two others for the question of «how to match ». In this Chapter we give a brief review of state of art in employed primitives, applied constraints and global strategies in matching algorithms. Then we present and justify our algorithm in relation to state of art and our needs in term of accuracy, level of details and computation time.

Type of primitives

Matching is often performed between feature sets belonging to the left and right images. The choice of matching primitive should depend mainly on the application and characteristics of the scene. However some important criteria must be considered. In other words, matching primitives must be invariant, stable and possible to localise accurately within left and right images.

Many types of primitives are found in the literature. They vary from low-level ones such as pixels [Wood, 1983; Panton, 1978] to intermediate level ones such as interest points [Moravec, 1977], linked-edges [Han and Park, 2000] and line-segments [Schmid and Zisserman, 1997]. More complex primitives also can be extracted in each image. Then matching process is performed on the extracted complex features. Chehata *et al.* [2003] propose an algorithm for region matching in a stereo-image scenario.

Often more complex primitives reduce matching ambiguity and are easier to match. However their extraction is more difficult. In each application the choice of matching primitive should provide an equilibrium between extraction complexity and matching complexity.

Local constraints

Similarity constraint: In a matching problem, constraints are needed to retrieve true matches within all matching candidates. The most important constraint is the similarity constraint. Two corresponding primitives must be similar. A similarity measurement that is invariant to stereo projection must be defined. Often photometric and geometric attributes of features are used to define similarity measurements. The photometric measurements are based on comparing the radiometric characteristics of 3D objects when projected in different images. This is performed usually by comparing gray scale or color values, gradient, contrast, etc. Geometric properties such as size and orientation seem to be more stable. However they can variate from one image to the other because of different image orientations. Figure 3.2 shows such scenario in our ground-based images of urban area.

Figure 3.2: Variation of geometric properties within a stereo-image. A set of 90 pixels of the left images corresponds to only 45 pixels in the right one. In this scenario even correlation measurement fails if geometric deformation is not taken into account.

Search area constraint: The matching problem within two views is an ill-posed problem [Poggio *et al.*, 1985], because the solution is not unique. Complexity and probability of the error in a matching problem are increasing with size of search area. In order to decrease search area, local photogrammetric constraints are often applied. In stereo-matching scenarios, *epipolar geometry* and credible parallax interval are applied as local constraints to decrease the search space. However, even after applying local constraints a unique solution is not still available. In order to provide a unique solution, global constraints that are deduced from physical properties of the scene are applied during a global strategy.

Global constraints

Uniqueness constraint: Marr and Poggio [1977] suppose that a given point on a physical surface has a unique position in space and variation of surface in relation to the viewer is small. They deduce that each pixel in one image can correspond at most to one pixel in other image and disparity varies smoothly in image space. However different viewing directions between two images together with object slope, can set up a configuration in which the region corresponding to one pixel in one image is projected in more then one pixel in the other image.

Ordering constraint: This constraint states that relative positional order of objects are conserved within the images. So moving forward in one image during matching process, it is forbidden to move backward in other image. It can reduce the search space during matching and help to avoid false matches. However this hypothesis is not always verified. Specially in an urban scene the order of objects can changes from one image to other when seen from an stereo camera system. Figure 3.3 shows an illustration of such an order instability within images.

Continuity constraints: This constraint supposes that the scene is constituted of continuous patterns such as continuous patches of surfaces, edges, etc. It follows that the disparity is a continuous function [Marr and Poggio, 1977]. But, often surfaces are only locally continuous.

Pollard *et al.* [1985] applied disparity gradient limit that relies on surface slope limit. It reduces considerably search space and the number of false matches. The continuity of surface is applied successfully in small scale images. However the presence of many occlusions together with several different direction of objects in terrestrial stereo-poses reduces considerably the efficiency.

Other research works have investigated the *figural continuity* which supposes that the contours in an image correspond to 3D continuous curves. So the variation of disparity along a contour in image space is continuous [Mayhew and Frisby, 1981]. This constraint can be violated by occlusions and perspective deformation, that are frequently present in terrestrial contexts.

Global strategy

As mentioned before local constraints do not provide a unique matching solution and several matching candidates are plausible. In order to reduce matching ambiguities, the global constraints should be applied. However the so called global constraints are not absolutely true constraints and can be violated by the perspective deformation, discontinuities, occlusions, etc. So they can not be applied directly in the matching process. This is the reason why a global strategy is needed to administrate the constraints and provide an unique solution. Proposed strategies are quite different due to differences of application, extracted primitive, needed accuracy and computation time. A very rich states of art can be found in Scharstein and Szeliski [2002]. Here we give a brief review on the global matching strategies often used in feature matching.

Winner-take-all: In this technique, the only discrimination criterion is the similarity measurement. Indeed for a primitive in left image, all the candidate primitives in right image are considered and the match corresponding to maximum of similarity measure is chosen as a local optimum solution [Mohammed *et al.*, 1979]. The so called strategy is used by Schmid and Zisserman [1997] for line matching within stereo-images. The main advantage of this method is its speed. Its disadvantage is that the uniqueness of solution is only enforced in first image whereas a primitive of the other image can correspond to multiple primitives of the first image.

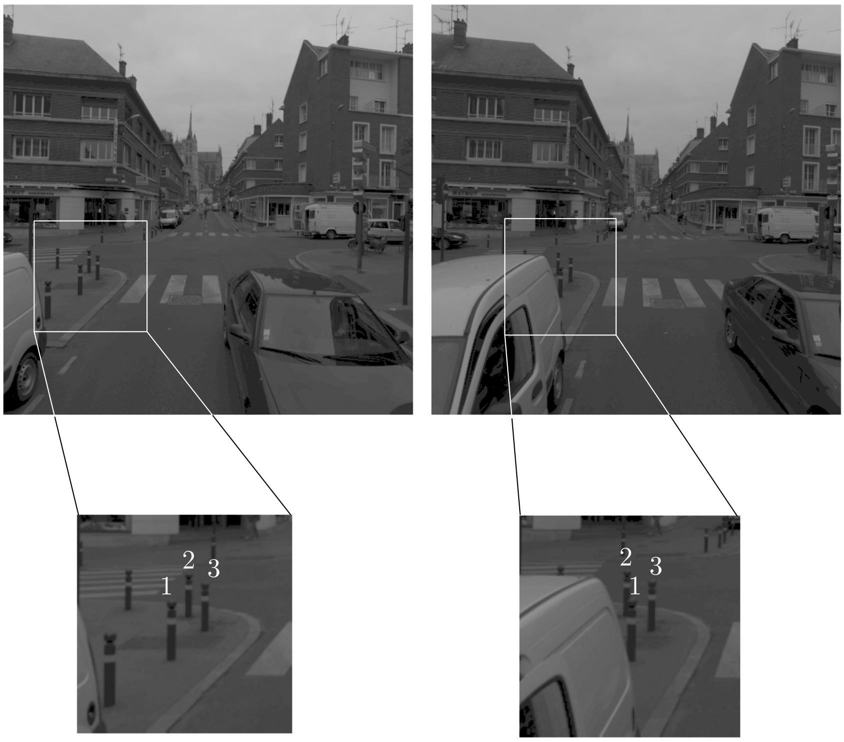


Figure 3.3: Order of stanchion changes from 1-2-3 in the left image to 2-1-3 in the right image.

Reciprocity matching In most of strategies, the matching is performed from left image to right or *vice versa.* Bijective strategies propose to change the roles of two images and accept only the matches that are stable in the two modes. It provides good results in presence of occlusions and can be implemented simply for feature matching Faugeras *et al.* [1996]; Ma and Thonnat [1993].

Coarse-to-fine matching: It consists in under-sampling each image in a multi-scale framework. The matching starts at the coarsest level. The idea is to guide the search space of each level using the matches established at a coarser level. Scale reduction, decreases the search space and so the probability of false matches. Moreover computation time is reduced because of smaller number of candidates to test. The so called strategy can integrate all other matching strategies. In dense pixel matching it consists of image under-sampling in different scales and matching from coarsest level to the finest level [Hannah, 1989]. In the primitive based matching it consist of hierarchical primitive extraction on the same image with different parameters [Chen and Medioni, 1990].

Graph search: This strategy is applied by Horaud and Skordas [1989] for line-segment matching within stereo-images. Each line-segment is characterised in each image by its position, orientation and its relationship with the nearby line-segments. So a *relational graph* is build for each image. For each line-segment a set of potential assignment in other image is determined by taking into account *epipolar*, position and orientation constraints. A *correspondence graph* in which the nodes represent assignments and the arcs express the compatibility between assignment is defined. The problem of line-segment matching is so transformed to the problem of searching for a set of reciprocally compatible assignment. This is performed by looking for maximal cliques. The maximal clique which represents the maximum sum of similarity measures is chosen as the final solution.

Dynamic programming: It can find the global minimum of constrained optimisation problems in polynomial time. It is based on separation of a large problem into smaller problems. It was first used by Baker and Binford [1981] for edge matching in a stereo context. It is based on primitive matching along conjugated *epipolar lines*. Primitives are extracted in each image. The sequence of primitives on left and right *epipolar line* form a 2D correspondence graph in which each node represent a matching possibility. To each node a matching cost is associated. The problem of matching can be considered as finding a minimum cost path in the correspondence graph by taking into account ordering, uniqueness and disparity gradient constraints. Dynamic programming provides an optimized solution for each *epipolar* line. This is why it can be considered as a one dimensional optimization method. The main problem of matching with this technique is to administrate the constraint between nearby *epipolar lines*. For this purpose an interesting work in performed by Baillard and Dissard [2000] by taking into account *figural continuity* along contours (linked edges).

2D optimisation methods: These methods are formulated in a global energy minimisation framework. The objective is to find a disparity function that minimise a global energy. This energy consists in a weighted sum of two terms. The first is called data-term that measures how well a disparity function agrees with input image pair. The second is called smoothness-term that encodes smoothness assumptions made by algorithm. The latter term control the smoothness of disparity function between neighbouring pixels. A global minimisation of the energy can be obtained by *max-flow* and *graph-cut* methods Roy and Cox [1998]. The advantage of these methods in relation to dynamic programming is that smoothness constraint deals implicitly with both directions and not only with *epipolar* line direction.

3.2 Specific context of the thesis

The selected primitive and global strategy in this thesis rely on the properties of available data-set (ground-based stereo-pairs) as well as the context of our application and objectives.

Regarding the specifications of our input image data (See Section 2.2), the across track Ground Sample Distance (GSD) at a distance of 10 *m* from stereo-base is 3 *mm*. Indeed the stereo-base is short. Using Equation 3.2 and supposing a matching accuracy of one pixel ($\sigma_\rho = 1$ *pixel*) it follows that depth accuracy will be 2.5 *cm* and reaches approximately the objective accuracy for 3D roadmark reconstruction (2 *cm*). However 3D reconstruction is only an intermediate step and does not provide the final result. The provided accuracy with pixel quality matching is already slightly higher than our objective. Moreover feature modelling (roadmarks) introduces often some approximations in the process and accumulate errors. This is the reason why at this stage a subpixel matching quality is required.

Considering our application and the needed accuracy, edge and corner point features appear to be interesting matching primitives because:

- Thanks to efficient algorithms they can be extracted accurately in image space [Canny, 1986; Harris and Stephens, 1988; Devernay, 1995].
- Corner points map the roadmark features vertexes and edges represent their borders (See Figure 3.4).

In real urban scenes, the presence of cars and pedestrians, can occlude roadmarks partially. In such scenarios the occlusion of each roadmark object vertex can fail the pattern recognition step whereas occluded vertexes can be estimated by intersecting non occluded or even partially occluded sides (See Figure 3.4). This is the reason why edges are chosen as matching primitives rather than corner points.

3.3 The matching approach

As explained in Section 3.2 edges are well adapted features to be used in our roadmark reconstruction application. Edge matching is investigated in the literature under two main forms:

- Line-segment matching,
- Linked-edges (contours) matching.

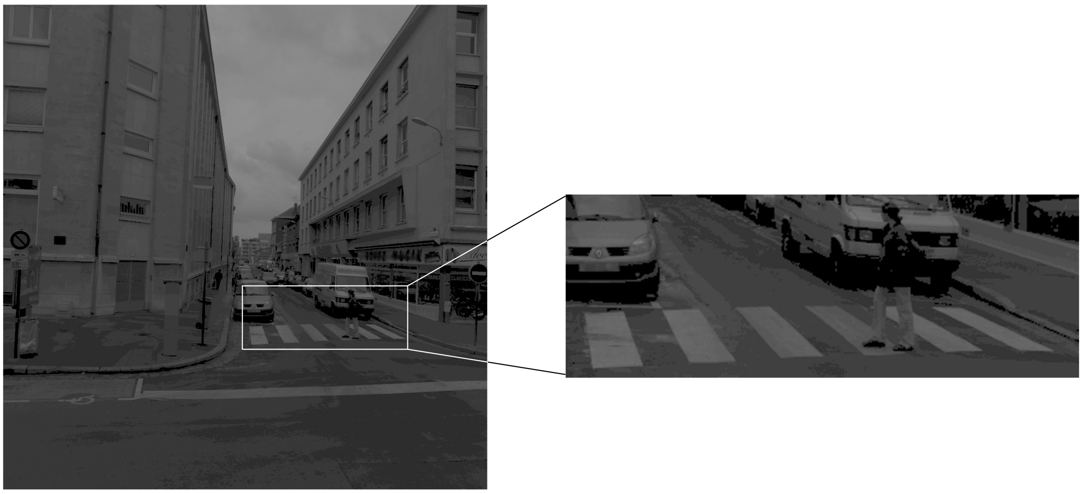


Figure 3.4: Partially occluded zebra-crossing strip by a pedestrian.

3.3.1 Line-segment matching

Within the first approach, edges are extracted with an edge detector [Canny, 1986; Deriche, 1987; Sobel and Feldman, 1968], they are then linked and approximated by line-segments. These line-segments are then used as matching primitives. Horaud and Skordas [1989] present a graph search method that match a group of line-segments in a stereo framework by taking into account the relationship between nearby line-segments and geometric similarity measures. Schmid and Zisserman [1997] propose a correlation-base similarity measurement followed by a winner-take-all strategy to find a unique solution within a stereo-matching framework.

The main problem of these approaches is that the approximation of linked-edges by line-segments often cuts out contours into multiple fragments. Moreover it is a difficult task to find optimal approximation parameters that gives a trade off between approximation error and fragmentation effect in the image space.

This is the reason why we aim at matching directly linked-edges without any approximation in image space. However reconstructed 3D linked-edges can be approximated to 3D line-segments. The advantage of 3D space is that parameters of approximation can be selected regarding real conditions without any perspective deformation.

3.3.2 Linked-edges matching

Within the second strategy the linked-edges are directly used as matching primitives. Han and Park [2000], propose a correlation-based matching. The correspondence between two linked-edges is deduced from the proportion of edge point correspondences. Serra and Berthod [1994] propose a geometric-based matching by dynamic programming approach for sub-pixel edge matching. Edges are not stable from one image to the other. In the general case, linked-edges can be deformed, fragmented or can totally disappear from one image to the other and edge chain matching will failed.

Instead of working with linked-edges and trying to match the totality of a contour with corresponding one in the other image (we recall that in the other image, it can be fragmented and so there is not only one corresponding contour), Baillard and Dissard [2000] propose to match each individual edge point. In order to respect contour continuity, during the matching process edge points that belong to the same contour in one image are encouraged to match a same contour in the other image. The algorithm is based on dynamic programming optimisation along conjugate epipolar lines in aerial stereo-image framework. An initial matching cost function is defined as a combination of intensity and contrast direction similarities. A *figural continuity* constraint is then implicitly introduced into the final cost function to ensure inter line continuity. Minimization is then performed on the total cost along all *epipolar lines*. The advantage of this method is its robustness to fragmentation of edge chains between images.

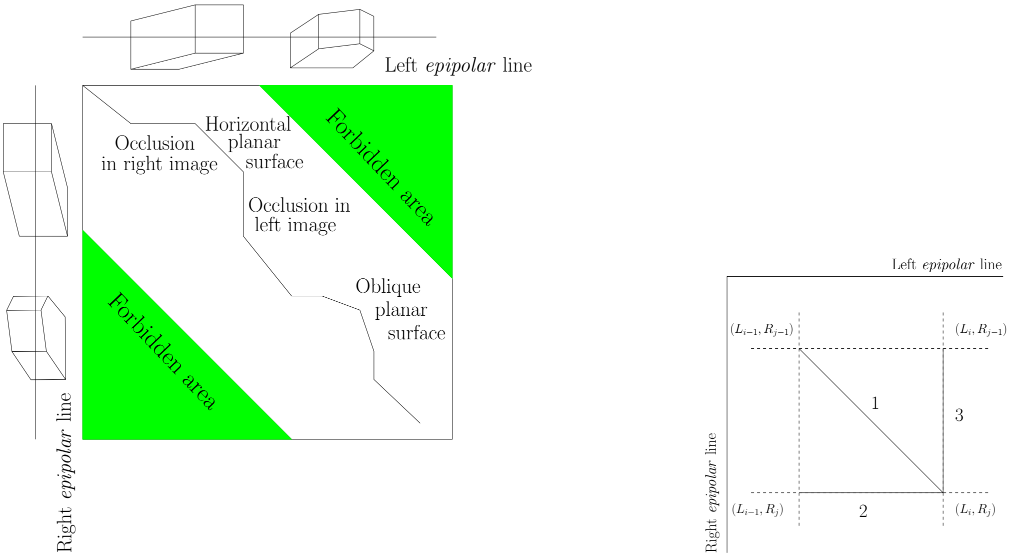
We adopt a similar strategy in order to be robust to contour fragmentation. Section 3.4 gives a brief description of Baillard's contour matching algorithm. Section 3.5 discusses the main difficulties in the algorithm in our ground-based context. Finally Section 3.6 gives some solutions to overcome these difficulties and provides a subpixel quality edge matching.

3.4 Review of Baillard's contour matching algorithm

Baillard and Dissard [2000] propose a hierarchical matching algorithm dedicated to aerial images of complex urban scene. The algorithm begins with matching edge points using *epipolar geometry* and taking into account *figural continuity* along linked-edges. This provides piece wise continuous 3D edges. Then interval pixels between matched edges are matched using radiometric and geometric constraints to provide a dense elevation model. We focus only on the edge matching algorithm. Given a calibrated stereo-pair, the images are re-sampled in *epipolar geometry*. Contours are extracted in each image and provide left and right binary images of edges. The matching is performed using a dynamic programming optimisation.

3.4.1 Dynamic programming optimisation

Dynamic programming is an efficient method for solving constrained non linear optimisation problems. Baillard and Dissard [2000] formulate edge matching as a cost minimization problem along *epipolar* lines by taking into account the constraints such as similarity, order and *figural continuity* along linked-edges. In each *epipolar* line, edges constitute an ordered primitive sequence. The problem is to match the primitives of left and right sequences. This is performed in what is called correspondence graph (See Figure 3.5(a)). Supposing that $(L_0, L_1, \cdots, L_n)$ and $(R_0, R_1, \cdots, R_m)$ constitute left and right edge sequences (horizontal and vertical axis of correspondence graph), the node (L_i, R_j) represents a potential match between L_i and R_i edge points. The goal is to find the path with minimal cost which links (L_0, R_0) to (L_n, R_m). For each node(L_i, R_j), $(C(i, j))$ expresses the cost of elementary path that links it to (L_0, R_0). It is calculated recursively within a disparity range by Equation 3.3.



(a) (b)

Figure 3.5: According to [Baillard and Dissard, 2000], (a) correspondence graph and matching path, (b) Elementary paths for dynamic programming

$$C(i,j) = \min \begin{cases} C(i-1,j-1) + match(i,j) & (1) \\ C(i-1,j) + occ(i) & (2) \\ C(i,j-1) + occ(j) & (3) \end{cases} \tag{3.3}$$

According to order constraint a path going trough a node (L_i, R_j) can come from only three directions (See Figure 3.5(b)).

- Path 1 denotes a match between two edge points with the matching cost $match(i,j)$. So the total cost for going from origin to (L_i, R_j) is the sum of the costs for going from origin to (L_{i-1}, R_{j-1}) and the cost for going from (L_{i-1}, R_{j-1}) to (L_i, R_j).

- Path 2 denotes a point which is visible in left image but occluded in right image with occlusion cost $occ(i,j)$. So the total cost for going from origin to point (L_i, R_j) is the sum of the costs for going from origin to (L_{i-1}, R_j) and the cost for going from (L_{i-1}, R_j) to (L_i, R_j).

- Path 3 is reversed compared to path 2.

The disparity range is computed considering known range of object position in relation to the stereo-base (See Figure 3.6). Nodes which are located out of disparity range are avoided by associating the maximal cost ($C(i,j) = +\infty$). The green regions in correspondence graph of the Figure 3.5 represent such forbidden areas. This constraint reduces considerably search space and computation time. The reduction of search space decrease the probability of false matches.

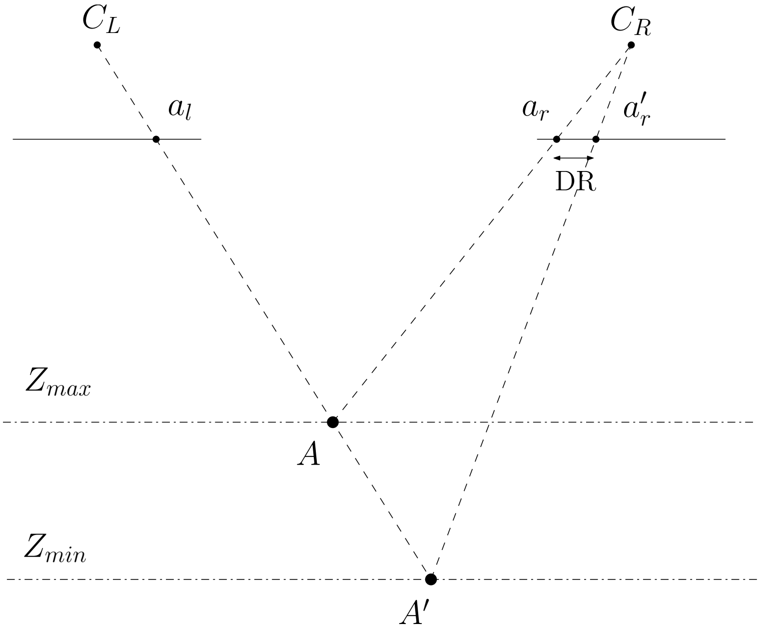


Figure 3.6: The relation between disparity range and object space. Aerial applications often suppose that all 3D objects are included within Z_{min} and Z_{max}. So the point a_l corresponds to a point in $\overline{AA'}$ range. Its projection in the right image is called disparity range $(DR = \overline{a_r a'_r})$.

3.4.2 Matching costs

Matching cost of a potential match (L_i, R_j) is defined in two stages. First a similarity-based function is used to compute an elementary cost within a disparity range and propose potential candidates applying a threshold on the similarity measure. For the proposed potential matches a final matching cost is then calculated by taking into account the elementary cost and the *figural continuity* along linked-edges.

Elementary matching cost

The variation of intensity and gradient direction are used following Equation 3.4 to define an elementary matching cost between two edge points.

$$match_0(i,j) = \alpha \frac{min[\Delta I_l(i,j), \Delta I_r(i,j)]}{\Delta I_0} + \beta \frac{\Delta\theta}{\Delta\theta_0} \tag{3.4}$$

Where $\Delta I_l(i,j)$ and $\Delta I_r(i,j)$ are grey level difference at left and right of the edge points. ΔI_0 is its normalisation factor (255 for 8 bits images). $\Delta\theta$ is the gradient direction variation between two edge points. It is normalised by $\Delta\theta_0 = \pi$. α and β are weight values and satisfy $\alpha + \beta = 1$. A pair of edges (L_i, R_j) constitutes a potential matching if $match_0(i,j) < C_{max}$.

Figural continuity and final matching cost

The final match cost for a potential match is calculated by introducing implicitly the *figural continuity* constraint into computed elementary matching costs. The contours are partitioned in linear segments using a polygonal approximation. Then the distribution of potential matches is analysed supposing that disparity varies continuously along a segment. More precisely the partitioning step provides two set of segments in each images (P_L and P_R). A consistency notion is defined between two potential matches (L_1, R_1) and (L_2, R_2). If L_1 and L_2 belong to the same segment in left image S_L, they said to be consistent if they satisfy the following property:

$$L_1 \in S_L \wedge L_2 \in S_L \tag{3.5}$$

$$\forall d \in \mathbb{N}, d_1 \le d \le d_2, \exists (L,R) \in P_L / L \in S_L \text{ and } d = disp(L,R) \tag{3.6}$$

where:

S_L : a line-segment in left image

$(L_1,R_1),(L_2,R_2)$: two potential matches.

$d_1 = disp(L_1,R_1), d_2 = disp(L_2,R_2)$

$disp(L_i,R_j)$: disparity value for potential match (L_i,R_j)

if (L_i,R_j) denotes a potential match with L_i belonging to a segment in left image S_L and R_j belonging to a segment in right image S_R. The corresponding final matching cost ($match(i,j)$) is computed as follows:

$$match(i,j) = \frac{1}{\ell_{max}} \sum_{L_k \in S_L} cmatch_{(i,j)}(k) \tag{3.7}$$

where:

ℓ_{max} : $\max(\ell(S_L), \ell(S_R))$

ℓ : is the segment length

$$cmatch_{(i,j)}(k) = \begin{cases} match_0(k,l), & \text{if } \exists R_l / (L_k,R_l) \text{ consistent with } (L_i,R_j) \\ C_{max}, & \text{otherwise} \end{cases}$$

3.4.3 Occlusion cost and optimisation

The matching is performed in a correspondence graph (See Figure 3.5(a)). The cost of a node is computed regarding configuration of Figure 3.5(b) and using Equation 3.3. The final matching cost is defined by Equation 3.7. The occlusion cost of a point ($occ(i)$) is defined as a function of maximal matching cost (C_{max}) by Equation 3.8.

$$occ = \frac{C_{max}}{2} \tag{3.8}$$

Equation 3.8 expresses that if $match(i,j) < 2occ$ it is preferred to accept the match between L_i and R_j (path 1) than occlusion of both L_i and R_j (paths 3 and 2). The matching in correspondence graph is started with a predefined (L_0, R_0) in the disparity domain. Other extremity of path is free and is constrained to minimise the total cost along *epipolar* line.

3.5 Remarks on the application of the Baillard's algorithm to our terrestrial scene

The inputs of Baillard's contour matching algorithm consist of a calibrated aerial stereo-pair and an altitude range (Z_{min}, Z_{max}). In our terrestrial stereo-pairs acquired under real urban conditions, the algorithm can not be applied in the same way. This is due to specific acquisition geometry and higher complexity of urban scene when imaged at high resolution. More precisely due to following reasons the algorithm is not efficient and needs some adaptations.

Disparity domain

As mentioned in Section 3.4.1 and seen in Figure 3.6 the search space in matching process is limited considering an altitude range. This causes two forbidden areas in correspondence graph (See Figure 3.5(a)) and restricts the disparity search area. The more the search area is restricted the less complex the matching process is. In our terrestrial stereo-pairs (See Figure 3.7(a)) the large depth field $(0, \infty)$ causes a very large disparity domain. It is translated by the reduction of forbidden areas in correspondence graph and decreases considerably the efficiency of this constraint. In other words, for a given edge point in the left image the corresponding edge point in the right image can be situated from the most left to the most right extremities of corresponding epipolar line. It can cause matching failure in the case of pattern repetition together with occlusions.

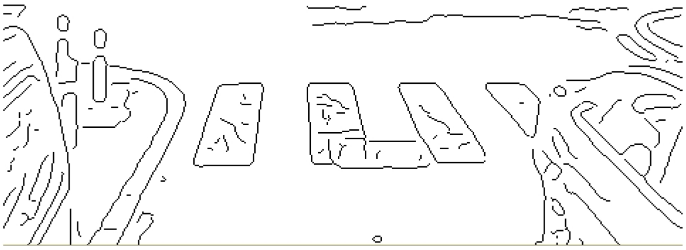
Perspective deformation

As seen in Figure 3.7 edges that correspond to zebra-crossing strips do not have similar orientation. This is due to significant perspective deformation on the road surface. This deformation is not constant everywhere in the image space and depends on the direction and position of features in relation to the stereo-base. Perspective deformations are minimal on the facades that are nearly parallel to images plane and maximal on the road surface which is perpendicular to the image plane. This scenario can disturb matching cost computation of Baillard's algorithm in two ways.

Firstly, the use of edge point gradient direction in the elementary matching cost function (See Equation 3.4) increases wrongly matching cost of corresponding points.

Secondly, the segment inclination caused by perspective effect can disturb the final matching cost. Perspective deformation increases with distance from image's center. An important deformation can incline the segments to a direction near *epipolar* line. This is the case for the line-segment S_L in Figure 3.8(a). In this scenario more than one pixel of S_L are situated on the same *epipolar* line. Baillard proposed to conserve only the first and the last pixel on each *epipolar* line (represented as black pixels in Figure 3.8(a)). Table 3.1 represent potential disparities on each *epipolar* line for matching S_L to S_{R_1} or S_{R_2} by taking into account only the first and the last pixel on each *epipolar* line. Figure 3.8(b)

(a) Left and right images of a stereo-pair acquired by the *Stereopolis* system



(b) Edge extracted from the left image.

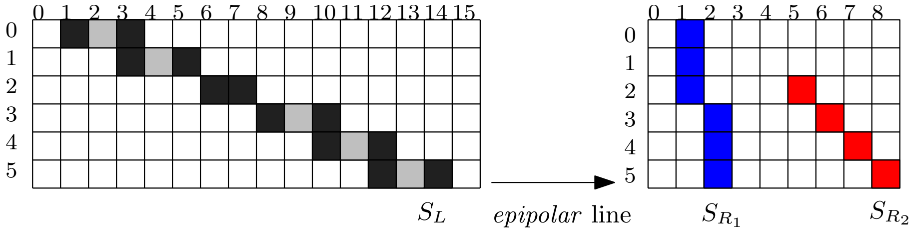
(c) Edge extracted from the right image.

Figure 3.7: Different orientation of homologous contours caused by perspective deformation.

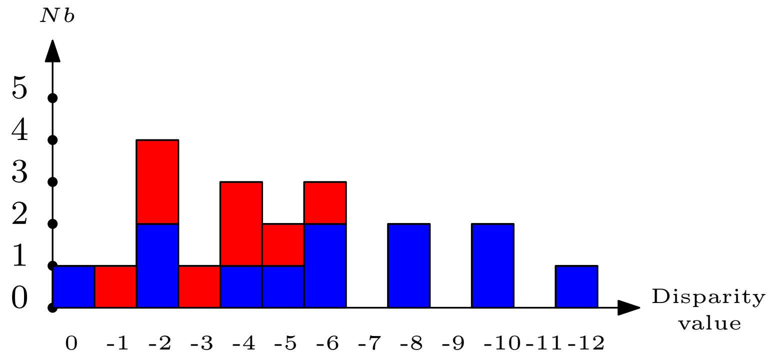
shows potential disparity histogram. Regarding this histogram, Baillard's *figural continuity* constraint is violated in two ways.

- Potential disparities relating S_L to S_{R_1} (blue bars) comprise disconnected elements in the histogram. So even if these elements correspond to the same segment, regarding Equation 3.6 they are not consistent and so penalised in the final cost computation (See Equation 3.7).

- Lets suppose that S_{R_1} is the good match for S_L. Regarding the histogram, a connected element is formed by potential disparities driven from S_{R_1} and S_{R_2} (the red and blue bars in Figure 3.8(b)). So the pixels of S_{R_2} being consistent with one part of S_{R_1}, are wrongly favoured in the final matching cost computation.



(a) Potential match between a segment in left image (S_L) and two segments in right image (S_{R_1} and S_{R_2}). Orientation difference is caused by perspective deformation.



(b) Histogram of potential disparity for pixels of S_L. The blue bars express potential disparities relating S_L to S_{R_1} and the red ones represent those relating S_L to S_{R_2}.

Figure 3.8: Inclined contour matching problem.

Disparity *Epipolar* Line *Nb*	S_{R_1}	S_{R_2}
0	0,-2	—
1	-2,-4	—
2	-5,-6	-1,-2
3	-6,-8	-2,-4
4	-8,-10	-3,-5
5	-10,-12	-4,-6

Table 3.1: Potential disparity values along the segment S_L of the Figure 3.8(a).

3.6 Adaptation of Baillard's algorithm to our specific application

As explained in the previous Section (3.5) the two main problems of the contour matching algorithm are the too large search space and the matching cost computation. In fact, urban features when seen from ground-based systems represent variety of features with different position and orientation such as facades, road surface, vehicles and moving objects. This is why in contrast to aerial images, it is not possible to use a unique parameter set for the whole image. However the search space and matching constraint must be adapted to match each of these features (road, facade, etc.) by taking into account their orientation and position. Our object of interest being road surface, we explain in Sections 3.6.1 and 3.6.2 how search area and matching cost are adapted for road features.

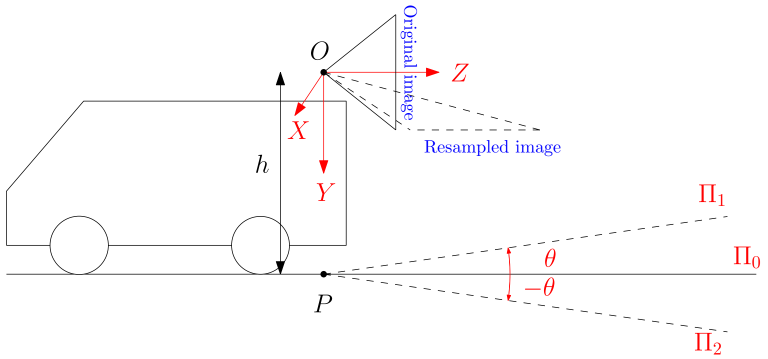
3.6.1 Search area constraint

We aim at providing more accurate matches in a satisfactory time. This is performed by limiting the search area to an area around the approximate road position. Figure 3.9(a) shows the geometric configuration of the system in relation to road surface. Cameras are mounted in a nearly horizontal position, at height h above road surface. We define a 3D coordinates system with left camera center as origin (O). The X and Y axes are defined to be parallel to the X and Y axes of left image. The Z axis is then defined as cross product of X and Y axes. The foot-print of point O (the projection center of the left camera) is called $P(0,-h,0)$. The search area is limited between two planes (Π_1 and Π_2) passing through P with normal vectors that are deviated to $-\theta$ and θ from road approximate plane (Π_0). This angular tolerance is adjusted to take into account the eventual pitch angle of the stereo-camera in relation to the road surface. It can be estimated for each exposure by estimating the camera deviation in relation to the 3D architectural scene via vanishing point detection [Cipolla *et al.*, 1999]. It can be also easily set comparatively greater than the maximum pitch of the vehicle. Figure 3.9(b) shows the search space region in the correspondence graph which is a diagonal region with a fix width between two parallel lines denoted by π_1 and π_2. These two lines correspond respectively to Π_1 and Π_2 planes. The width of this region depends directly on the value of θ. The correspondence path passes within this search area and all obstacles (vertical objects, pedestrians, cars, etc.) situated out of this search area are forbidden. The road area being nearly flat, the path is close to a straight line when seen at a small scale. However due to the transverse curvature of road the path at higher scale is not a straight line. Its shape depends on the curvature of the road and the presence of nearby obstacles (parts of obstacles that are situated within search area such as pedestrian feet, cars' wheel etc.).

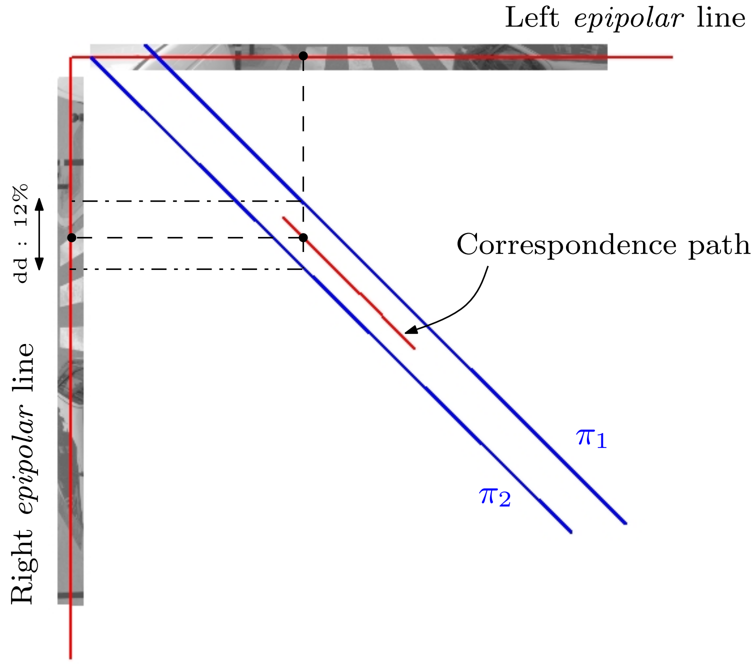
3.6.2 Matching cost constraint

The similarity function should provide a maximal response for real corresponding pair of points. As discussed in Section 3.4.2, Baillard and Dissard [2000] propose a weighted function that uses radiometric similarities on left and right of two edge points and their gradient direction similarity. In our application this function does not provide optimum response on homologous points due to two following reasons:

- A pixel of our images corresponds to a very small region in object space ≈ 3 *mm*. Due to the granularity of the surface, radiometric similarity computed on a window of only one neighbouring pixel is not significant. In this case radiometric similarity can be affected by noises and so not stable within two images.
- The gradient direction is not stable due to perspective deformation.



(a) Configuration of the vehicle and its imaging system in relation to road surface. Search space is limited between the planes Π_1 and Π_2 passing trough $P(0,-h,0)$ with vertical deviation of θ and $-\theta$ around X axis.



(b) The correspondence graph related to the search space defined between planes $\pi_1$and π_2 with $\theta = 8°$. For a given point on an *epipolar* line of left image the disparity domain comprises 12% of *epipolar* line length (490 pixels in an image with 4096 columns).

Figure 3.9: Restriction of search area around road surface.

Our high resolution images provide good texture information around each edge point that can be applied in order to increase the discrimination power of the similarity function. Moreover, on-road features comprise locally continuous and flat regions. This is the reason why a Normalised Cross Correlation (NCC) may be useful. It is preferred to compute this coefficient within a relatively large size window to use maximum texture information and increase its discrimination power.

In practice we use zero-mean normalised cross correlation (*ZMNCC*). It is defined in Equation 3.9. *ZMNCC* is chosen to cope with small variations of radiometric values between left and right images. However the *ZMNCC* can disregard also important variation of radiometry and produce false matches. A match between two radiometrically homogeneous regions of black and white colours is an exaggerated example. In order to prevent this effect the elementary matching cost($match_0$) is set to C_{max} when the grey level variation is greater than 30%. This threshold is regulated empirically by testing on a large set of images. Equation 3.10 denotes the relation between elementary matching cost and zero-mean normalised cross correlation. The *ZMNCC* being limited between -1 and 1, the matching cost in the case of a match is bounded between 0 and 1.

Another possible way for coping with radiometric variation instead of *ZMNCC* is to reduce this effect in a pre-processing step. It can be performed by a linear histogram transformation. This transformation is computed by taking into account mean and standard deviation of left and right images in their overlapping area.

$$ZMNCC(i,j) = \frac{\sum_{c,l\in W}(I_i(l_i,c_i)-\mu_i)(I_j(l_j,c_j)-\mu_j)}{N^2 s\sigma_i\sigma_j} \in [-1;1] \tag{3.9}$$

where:

W is the correlation window,

N is the size of correlation window,

$I_k(l_k,c_k)$ is grey level at pixel (l,c) of W in image $k \in \{i,j\}$,

$\mu_{k\in\{i,j\}}$ is mean of grey levels within W in image k,

$\sigma_{k\in\{i,j\}}$ is root mean squares of grey levels within W in image k,

$$match_0(i,j) = \begin{cases} \frac{1-ZMNCC(i,j)}{2}, & \text{if } \frac{|\mu_i-\mu_j|}{max(\mu_i,\mu_j)} < 0.3 \\ C_{max}, & \text{otherwise} \end{cases} \tag{3.10}$$

The main interest of using *ZMNCC* for elementary matching cost computation is to take benefit of texture within a large correlation window. However important perspective deformation disturb this measurement when using a large correlation window. A solution is to deform the correlation window in the same way in order to absorb the perspective deformation effect. In our case the deformation can be modelled by an affine transformation. The affinity parameters can be computed using known approximate pose of images in relation to road surface (See Figure 3.9(a)). As explained in Section 3.4.2 the final matching cost for a potential match is computed by taking into account the *figural continuity* along edge chains. As discussed in Section 3.4.2 perspective deformation not only disturbs elementary matching cost value but also the final matching cost. So even if the perspective deformation effect can be reduced in elementary matching cost estimation using adaptive shape windows, it is still troubling in the final matching cost computation. This is why we prefer to reduce significantly this effect from the images. This is performed via image re-sampling in a rectified *epipolar* geometry (See Section 3.6.3). Furthermore image rectification rather than adaptive shape windows reduces significantly the computation time.

3.6.3 Rectified *epipolar* re-sampling

We recall that the goal of our image rectification is only to reduce perspective deformation of road features. So rectification consists of rotating the images about its X axis up to 90° (See figure 3.10). Only the half down part of the image is rectified. This part represents the nearby road regions that are situated between D_{min} and D_{max}. D_{min} corresponds to the most down-ward line of images. The D_{max} is computed as the most distant point at which depth precision is acceptable. Regarding Equation 3.2 this distance can be estimated considering the needed depth precision σ_{Depth} and matching accuracy σ_{ρ}. So image lines which are situated between d_{min} (projection of D_{min}) and d_{max} (projection of D_{max}) must be rectified. Rectified line coordinate of line d in original image is called $d^{'}$ and computed according to

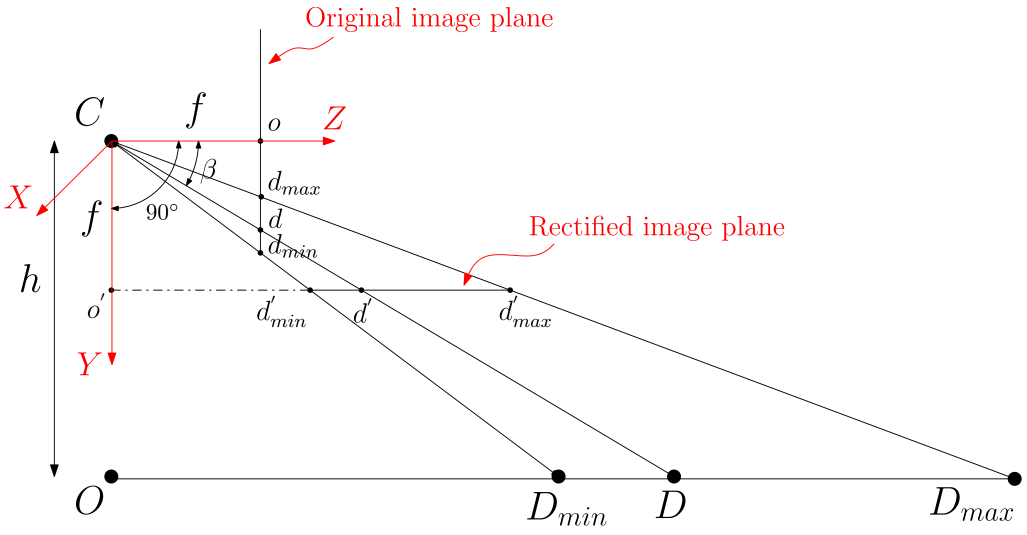


Figure 3.10: stereo-image re-sampling in rectified *epipolar* geometry using approximate exterior orientation.

the following equations:

$$\tan\beta = \frac{\overline{od}}{f} = \cot(\frac{\pi}{2} - \beta) = \frac{f}{\overline{o'd'}}$$

$$\Rightarrow \overline{o'd'} = \frac{f^2}{\overline{od}} \tag{3.11}$$

$$\Rightarrow \delta\overline{o'd'} = \frac{f^2}{\overline{od}^2}\delta\overline{od} \tag{3.12}$$

Supposing : $\delta\overline{od} = t_o$ pixel size of original image

$$\Rightarrow t_r(\overline{od}) = \frac{f^2}{\overline{od}^2} t_o \tag{3.13}$$

where :

$t_r(\overline{od})$ is the size of a pixel after rectification

By using Equation 3.13 for each line, rectified pixel size varies with $\overline{od}$. It is larger for smaller values of $\overline{od}$ and vice versa. In order to avoid oversampling we chose the largest pixel size ($t_r(\overline{od_{max}})$) for re-sampling in rectified geometry. In order to obtain radiometric value at each pixel of rectified image, the four corners of each pixel are projected in original image and define a region. The radiometric value is computed as mean value of the pixels in this area if the area is sufficiently large. Otherwise a bicubic interpolation determines the pixel value.

In order to simplify the future processes, images are positioned in *epipolar* geometry and lens distortion is corrected in the same resampling process. Figure 3.11 demonstrates such a rectified stereo-pair. These rectified images are used each time that image information is required.

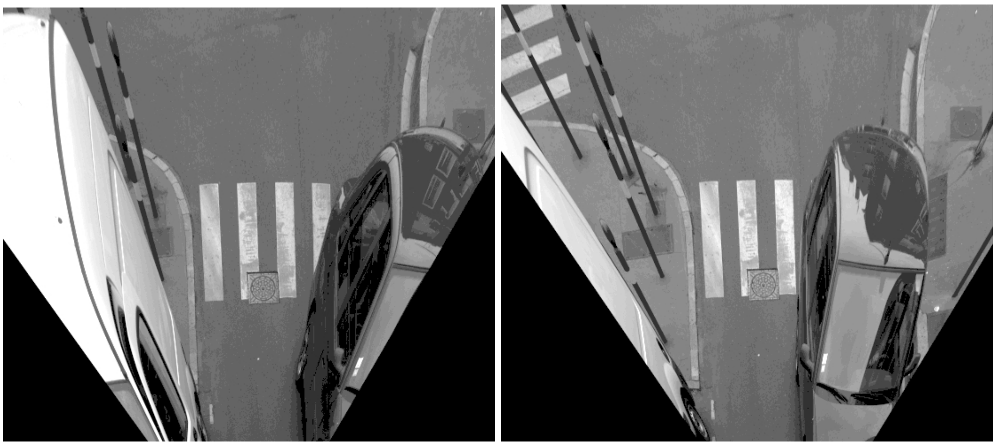


Figure 3.11: Rectified image with $D_{max} = 15m$. The original images are shown in Figure 3.7(a)

3.6.4 Global matching optimisation

The Canny-Deriche edge detector [Deriche, 1987] is applied on each re-sampled stereo-pair to extract edges. The edge points on conjugated *epipolar* lines organise ordered edge sequences and set up a correspondence graph (See Figure 3.9(b)). The search area is limited around road surface. Matching cost is defined according to Equation 3.10 using rectified images. Matching cost is computed for edge points within the search space. The optimisation is performed according to the configuration demonstrated in Figure 3.5(b) and the Equation 3.3. The main difficulty is to regulate occlusion cost (*occ*). Regarding the discussions of Section 3.4.3, occlusion cost (*occ*) depends on maximal matching cost ($occ = C_{max}/2$). C_{max} is related to similarity measure by Equation 3.14. So the relation between occlusion cost and similarity measurement is obtained by Equation 3.15.

$$\begin{aligned} C_{max} &= \frac{1 - ZMNCC_{min}}{2} && (3.14) \\ occ &= \frac{1 - ZMNCC_{min}}{4} && (3.15) \end{aligned}$$

Regarding above equations, occlusion cost is related to a minimal threshold for ZMNCC. Choice of this minimal threshold must give a trade off between radiometric similarity constraint and geometric constraint. The higher the occlusion cost, the more tolerant we are about matching quality in radiometric similarity point of view. Moreover, in our application, road edges can be occluded by cars or even disappear from one image to other (edge detection problem). This is the reason why we chose a low occlusion cost to allow the algorithm to accept edge occlusion and also to ensure good radiometric similarity between accepted matches. The occlusion cost is regulated empirically to 0.075 that corresponds to a minimum similarity threshold of 0.85.

Optimisation is performed independently on each pair of conjugated *epipolar* lines. The results is registered in a disparity map. The disparity map is in the geometry of left image and for each edge pixel contains the information of corresponding pixel in right image.

3.6.5 3D sub-pixel accuracy reconstruction

The matching process by dynamic programming provides a disparity map with pixel matching quality. So a pixel accuracy 3D reconstruction can be performed. However the stereo-base is short and does not provide the required accuracy. Indeed at a distance of $15m$ from the stereo-base the ratio B/H is 0.1. It provides a $3mm$ across track pixel, whereas along track pixel is about $45mm$. The so called effect is still more significant if lower resolution images are used. Figures 3.13(a) and 3.13(c) demonstrate a pixel accuracy reconstruction of edge chains. The roughness effect is dramatic. It can produce some difficulties in 3D line segment estimation and later in pattern recognition. In order to provide a smooth reconstruction sub-pixel matching quality is achieved by a post processing. Let us suppose (L_i, R_i) a pixel quality match on *epipolar* line ℓ_{L_i}, we aim at providing corresponding sub-pixel matching accuracy $(\acute{L}_i, \hat{R}_i)$ on *epipolar* line $\ell_{\acute{L}_i}$ (See Figure 3.12). It consists in the following steps:

1. Sub-pixel edge point estimation in the left image ($\acute{L}_i$). It is performed by quadratic interpolation of gradient magnitude at a neighbourhood of the pixel [Devernay, 1995].

2. Constitution of sub-pixel quality *epipolar* line passing from $\acute{L}_i$. Let us call this *epipolar* line ($\ell_{\acute{L}_i}$).

3. Sub-pixel quality correspondence of point $\acute{L}_i$ is located at the intersection of $\ell_{\acute{L}_i}$ and the sub-pixel contour in the right image.

4. Sub-pixel contour in the right image is also computed by quadratic interpolation. The segment $\overline{\acute{R}_{i-1}\acute{R}_i}$ shows a portion of this contour (See Figure 3.12).

5. $\hat{R}_i$ is computed as the intersection of $\ell_{\acute{L}_i}$ and $\overline{\acute{R}_{i-1}\acute{R}_i}$.

6. $\acute{L}_i, \hat{R}_i$ constitutes sub-pixel matching. It is used for 3D reconstruction by photogrammetric intersection.

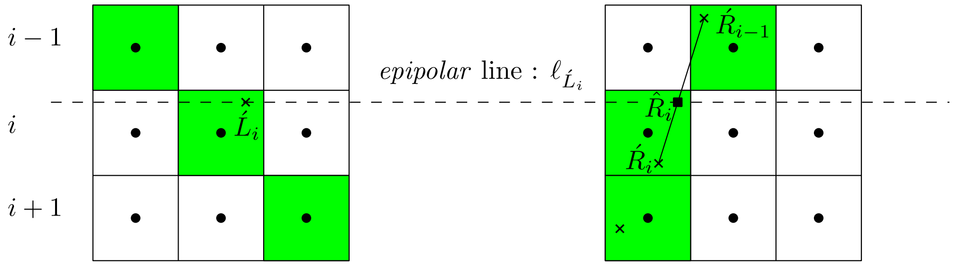
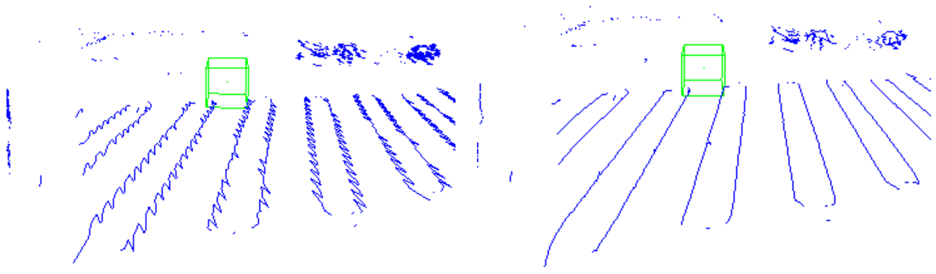
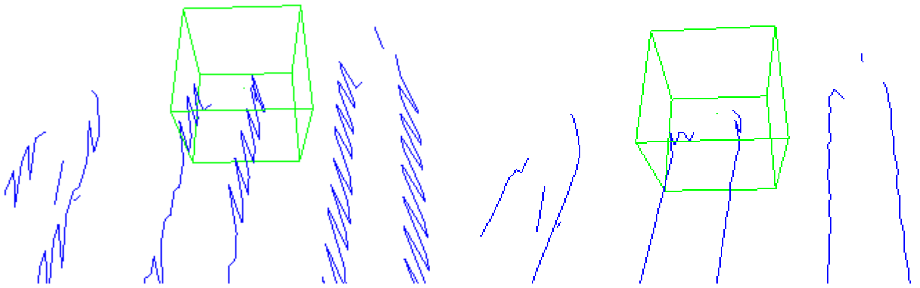


Figure 3.12: Sub-pixel matching $(\acute{L}_i, \hat{R}_i)$ estimation. The sub-pixel matching is achieved by: 1) estimation of sub-pixel position in left image $(\acute{L}_i)$. 2) intersection resulting sub-pixel *epipolar* line $(\ell_{\acute{L}_i})$ with subpixel contour in right image $((\overline{\acute{R}_{i-1}\acute{R}_i}))$.



(a) Pixel accuracy.
(b) Sub-pixel accuracy.



(c) A zoom on pixel accuracy.
(d) A zoom on sub-pixel accuracy.

Figure 3.13: Influence of sub-pixel matching on 3D reconstruction.

3.7 Discussions on parameters setting and tuning

We have presented an algorithm for 3D linked-edge reconstruction. This reconstruction constitutes the first stage of our roadmark reconstruction. As mentioned in Section 2.3 our algorithm needs to be fully automatic. This implies to study the sensibility of the algorithm to the parameters and its robusteness under different conditions. The main parameters of our matching algorithm are related to:

1. Edge extraction,
2. Matching.

Let see how these critical parameters are controlled.

3.7.1 Edge extraction parameters

In our work Canny-Deriche algorithm is applied to extract roadmark borders thanks to their high intensity gradient. This method implies regulation of two sets of parameters:

α of Canny-Deriche:
The edge detection method begins with computing gradient of intensity. This step implies a parameter called α. This parameter controls the accuracy of edge localisation and the sensibility of the filter to noise. The higher the α, the more accurate the edge localisation and the more sensible the method to noise. The majority of the state of the art use $\alpha = 1$. However, higher values are also used to improve the localisation. Lower values are used in noisy images.

In our application, we have high quality images with low noise. So low values of α are not interesting. Beside, size of the objects (roadmarks) we look to extract is much higher than pixel size. Indeed, one pixel in image space correspond approximately to $3 - -5$ *mm* in object space. So pixel size irregularities of object in image space can be discarded. This is why we prefer to use $\alpha = 1$.

Hysteresis thresholds:
Gradient image is computed, edge extraction is performed on this image by hysteresis thresholding. However reflected intensities from objects vary with illumination condition, shadows and texture of objects. Under these conditions, regulation of thresholds is often a difficult task. Inappropriate thresholds cause often over and under-detection problems. It is a crucial task to find a threshold providing trade off between under and over detections.

In an ideal case, these thresholds could be optimaly adapted to each exposure point. This can be performed by analysing illumination condition at each point and taking into account the sun direction and known building models. However it is computationally complex and imply existing 3D model of buildings.

We choose these thresholds empirically on a sufficiently large set of stereopairs. The parameters are adjusted such that they minimise under detection over a large set of images. The values of 2 and 8 are chosen respectively for low and high thresholds. Minimising under detection can cause large number of over detection. These irrelevant edges can be filtered by post processing in the pattern recognition step (See Chapter 4). Figure 3.14 demonstrates the results of edge detection on a stereopair. Irrelevant edges are caused by local textures on zebra-crossing and pavement. As seen in Figure 3.14 they are often correctly reconstructed in the matching step. We apply the same parameters on all stereopairs.

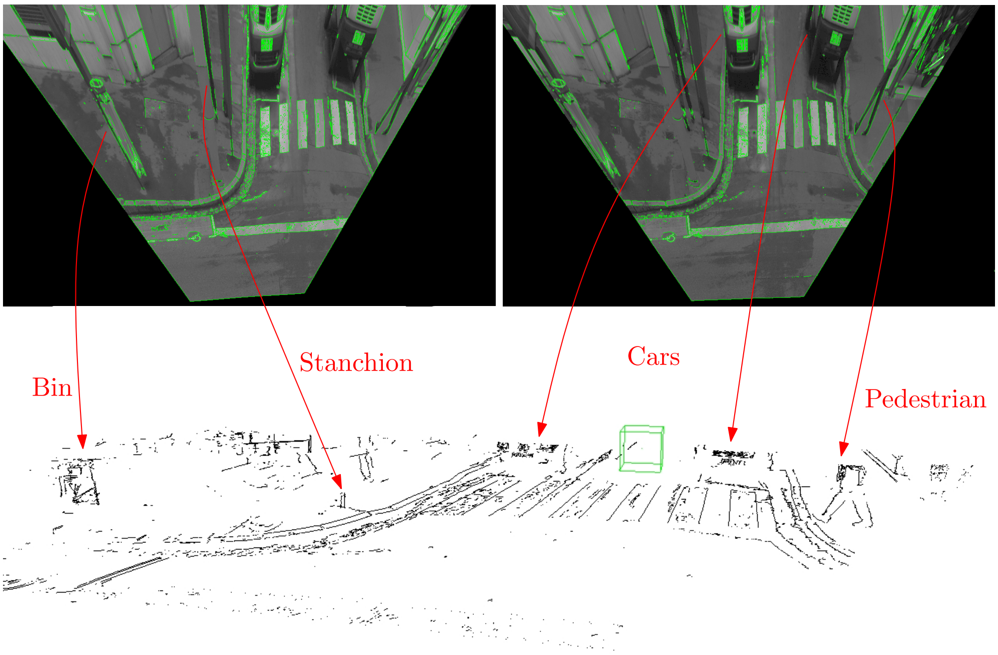


Figure 3.14: Results of 3D reconstruction. Edge extraction parameters: Deriche $\alpha = 1$, hysteresis thresholds = (2,8). Search space parameters regarding configuration of Figure 3.9(a): $h = 2.2\ m$, $\sigma_h = 5\ cm$, $\theta = 6°$. 11×11 correlation window, occlusion cost = 0.075.

3.7.2 Matching parameters

Search space:

As depicted in Figure 3.9(a), a search area is defined by taking into account position and orientation of the vehicle in relation to road surface (h and θ). However due to vibration, acceleration and braking this relative area can be changed. An optimal solution could be estimation of vehicle position in relation to road via vanishing point detection [Cipolla *et al.*, 1999]. Another solution is to define sufficiently large tolerances on h and θ values. We adopt 5 *cm* for variation of h and 6° for θ. The algorithm is efficient on all stereopairs of our test data (150 stereopairs). The only consequence of these large tolerances is that vertical objects such as cars, pedestrians and stanchions appear in the matching process (See Figure 3.14). In the detection step these features are discriminated thanks to their orientation and will be filtered. However vertical edges are useful for reconstruction of other urban objects such as bins, stanchions and road signs [Tao, 2001].

Size of correlation window:

As discussed in Section 3.6.2, we use a correlation based similarity function. Since a large amount of perspective effect is absorbed by rectification, we use a large correlation window (for example larger than 11×11).

Occlusion cost:

As explained in Section 3.6.4, occlusion cost is related to minimum similarity threshold for matching hypotheses. In order to ensure a sufficient radiometric similarity between the left and the right edge points of a matching, we use an occlusion cost that is lower than 0.075. This cost corresponds to a minimal similarity that is higher than 0.85.

3.8 Conclusion

In this Chapter a large number of existing methods for image matching are reviewed. Regarding our specific requirements a global strategy similar to the one proposed by Baillard and Dissard [2000] is adopted. The high complexity of urban area when observed by ground-based systems reduce considerably the efficiency of this algorithm. The algorithm is adapted to our specific ground-based context by taking into account road surface that is our object of interest. A pixel accuracy matching is computed by dynamic programming along *epipolar* lines by taking into account radiometric similarity and *figural continuity* constraints. A sub-pixel matching accuracy is then achieved by post processing. 3D reconstructed linked-edges are geometrically precise but they suffer from fragmentation effect. This is due to edge extraction and edge matching steps. These features constitute the input of roadmark detection algorithm that will be presented in the following chapter (Chapter 4).

Comparing to winner-take-all strategy, the proposed method is much more constraint. Indeed, high level of local textures together with eventual partial occlusions can cause false matches in winner-take-all strategy that is only based on maximum similarity constraint. In addition, taking into account the *figural continuity* improves considerably the smoothness of resulting 3D linked-edges and reduce also fragmentation effect. These two properties are particularly interesting because they simplify the 3D pattern recognition step.

The efficiency of linked-edge matching algorithm is controlled visually on a large set of stereopairs and provide promising results. Indeed the 3D linked edges being unstructured features, quantitative evaluation is a difficult task. However after roadmark modelling we present a quantitative evaluation of the complete chain from accuracy and completeness points of view. This evaluation will be performed on a set of 150 successive stereopairs acquired by *Stereopolis* system (See Chapter 6).

Chapter 4

Roadmark detection

The previous chapter explained how 3D linked-edges are reconstructed from each stereo-pair. This chapter explains the second step of our roadmark reconstruction strategy. This step uses the reconstructed 3D linked-edges as input data. Output consists in hypothetical candidates for each roadmark object. It deals with two class of roadmarks: zebra-crossings and dashed-lines. Recognition is performed thanks to their specific signature in a 2D space that approximates road surface. Some filtering tools are presented. These tools are then used differently in recognition of each roadmark class. Before presenting the mentioned tools, we define in Section 4.1 a geometric model for strips of zebra-crossing and dashed-line. Then, Section 4.2 gives an analysis of input data. The next sections explain our detection strategy.

4.1 Strip model definition

As explained in Section 2.3, we aim at reconstructing strips of zebra-crossing and dashed-lines under the form of a 3D vector model. Fortunately suitable models can be defined since the size and shape of zebra-crossing's and dashed-lines' strip are known in advance (See Section 2.1). This is the reason why a model-based approach is adopted for detection and modelling of strips. However before defining any strip model, the following points should be mentioned about the strips specifications:

- Strips width is precisely known in both cases of zebra-crossing and dashed-lines.

- Only approximate length of zebra-crossing's strips is known and those of dashed-lines even if their length is governed by specifications, sometimes do not respect the specified length and vary slightly.

- Zebra-crossing's strips shape is unknown, but that of dashed-lines is constrained to be rectangular.

The adopted model for strip consists of a planar parallelogram. The model parameters are defined by Equation (4.1).

$$\text{Strip's parameters :} \begin{cases} w \;=\; w_d & \\ \ell \;=\; \ell_d \;\pm\; \varepsilon & \text{In the case of dashed-line} \\ \theta \;=\; 90^\circ & \\ & \\ w \;=\; w_z & \\ \ell \;\in\; [\ell_{min} \;,\; \ell_{max}\,] & \text{In the case of zebra-crossing} \\ \theta \;\in\; [\;\theta_{min} \;,\; 90^\circ] & \end{cases} \tag{4.1}$$

where :

w : strip's width

w_z , w_d : expected strip's width for zebra-crossing and dashed-line

ℓ : strip's length

ℓ_d : specified length of dashed-line's strips

ℓ_{min} , ℓ_{max} : minimal and maximal length of zebra-crossing's strips

θ : angle between long side and transversal side of strip

θ_{min} : the minimum value of θ in the case of zebra-crossing's strips

The parallelogram model can be established by the four corners or the four sides. However the corners can be occluded (See Figure 4.1(a)) or damaged (See Figure 4.1(b)) and so any corner-based detection and reconstruction will fail in these cases. In contrast, the long sides of strip provide sufficient information for strip modelling even if they are partially occluded or damaged. So our detection step consists in looking for parallel line-segments of a given distance. In this purpose the 3D reconstructed linked-edge are used as input data to zebra-crossing's and dashed-line's strip detection. The following section briefly describes the strip detection input data and enumerates some difficulties to overcome.

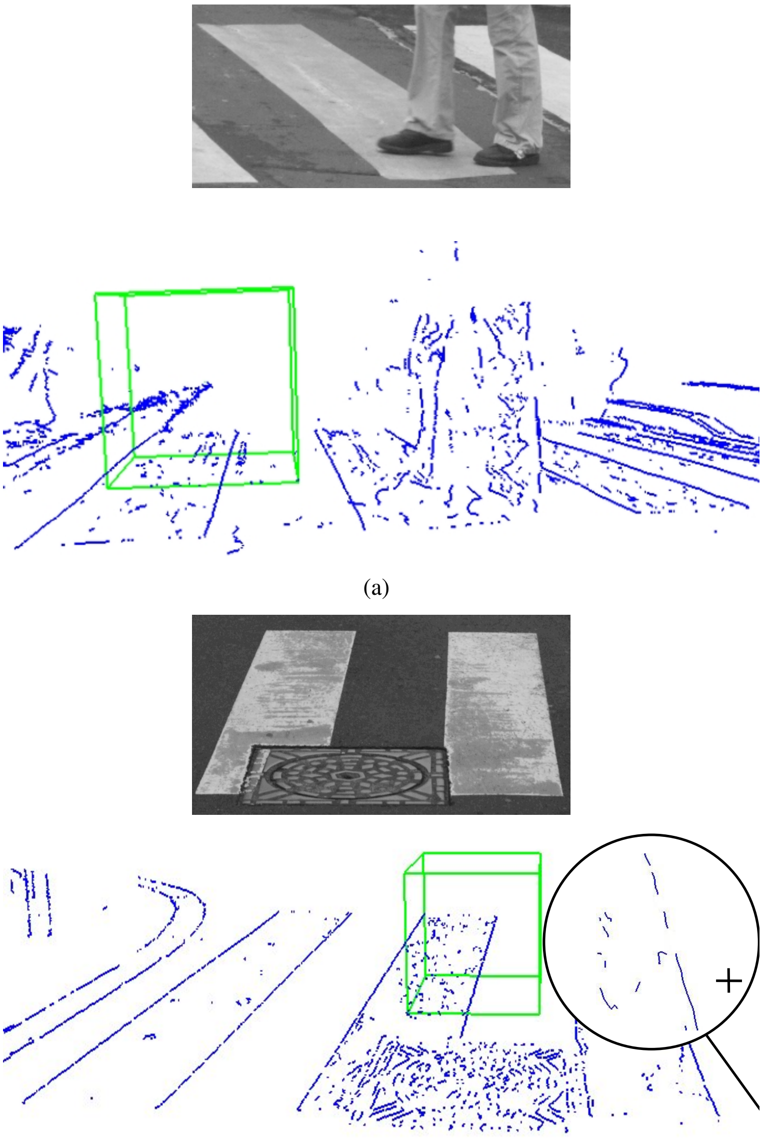
4.2 Input data analysis

As explained in Chapter 3 the results of the reconstruction step is a set of 3D linked-edge. The 3D reconstruction is performed without any flatness hypothesis about the road surface and so represents the real topography of the road.

However the imperfections of the edge detector and the edge matching algorithm turn the detection step into a difficult task. Regarding Figure 4.1, the detection step must handle the following problems:

1. Fragmentation of 3D linked-edges due to image processing operations such as edge detection and matching process.
2. Presence of many uninteresting linked-edges (not belonging to strip boundary). This is due to local textures on bitumen and strip areas or existence of pedestrians and cars in the scene.

3. Non-reconstructed transversal sides of strips. This is due to the fact that transversal sides are often aligned with *epipolar* lines.



(a)

(b)

Figure 4.1: Obstacles for strip detection from reconstructed 3D linked-edge points. (a) Presence of pedestrian, local textures and cracks, (b) Presence of an irrelevant object (manhole-cover), fragmentation of 3D linked-edges (demonstrated by a zoom)

Even if transversal sides of strips are not reconstructed, long sides of strips can be detected taking into account the given width of a strip. The model-based pattern recognition methods are generally based on hypothesis generation followed by a validation step which is based on fitness measurement of the hypothesis in relation to the model.

In our case hypothesis generation consists of selecting a pair of line-segment as candidates of two long sides of a putative strip. They are then analysed to see if they are nearly parallel and at a given distance from each other.

The fragmentation effect causes the line-segments representing two sides of each strip to be segmented into many shorter parts. In this scenario many pairs of nearly parallel line-segments of a given distance can be found. The consequence is a set of redundant hypotheses that describe different parts of a strip and do not provide any suitable hypothesis describing a whole strip. Moreover the large number of irrelevant line-segments can amplify this problem.

The mentioned problems within the input of strip detection method are tackled by performing the following tasks in sequence:

1. **Polygonisation of linked-edges**
   During this operation, initial line-segments are estimated from linked-edges.

2. **Filtering of initial line-segments**
   In order to reduce the number of uninteresting line-segments, initial line-segments must be filtered. The filtering step should recognise interesting line-segments (those belonging to strips sides) despite their fragmentation.

3. **Grouping of pertinent line-segments**
   In order to provide higher level line-segments which are less fragmented in relation to a parallelogram model, line-segments passing through the filtering step are analysed by a grouping process.

The detection step can be summarised as a set of operations performed in order to provide higher level primitives that simplify the model-based strip recognition. Section 4.4 on the next page explains the adopted strategy and data structure for handling the three tasks of polygonising, grouping and filtering.

4.3 Strip detection strategy

As explained in the previous section, model-based recognition method is not efficient using initial data because of the large number of irrelevant linked-edges and fragmentation effect. In order to overcome fragmentation effect and noises due to irrelevant objects, we take benefit from a 1D signature of linked-edges on a 2D plane that approximates the road surface. Figure 4.2 shows the signature of a scene containing a zebra-crossing and noises due to obstacles and local textures. Signature is computed by accumulation of linked-edges in road direction. As can be seen on this figure, even in the presence of noises and fragmentation effect, the resulting signature (in green) is constituted of successive peaks at a specific distances. The distance between the peaks is strip's width and accumulation value is a function of strip's length.

In order to compute this signature all 3D linked edges resulted from matching stages are projected on plane that approximates the road surface. The signature is then computed and guides the detection step to recognise the interesting linked-edges from irrelevant ones and this in presence of fragmentation effect. Line-segments are then fitted to interesting linked edges and provide higher level primitives to be processed by a model-based detection method.

The reduction of the problem to a 2D space reduces considerably the complexity of pattern recognition and geometric operations such as polygonisation and grouping. However, road surface is not flat. As explained in Sections 2.3 and 2.7 in our application this approximation is not negligible for roadmark modelling.

We use this approximation just for strip detection stage. We aim at modelling strips without flat road hypothesis. This is why for each linked-edge we register both set of coordinates (3D and 2D). In this manner, the 2D coordinate set is used to segment linked-edges into line-segments, to recognise interesting line-segments and to decide which line-segments to group. When these decisions are made, the final modelling is performed using the 3D coordinate set.

Section 4.4 presents the adopted data structure that allows to register 2D and 3D information correctly during segmentation, recognition and grouping operations. Section 4.5 discuss about road plane estimation. Section 4.6 explains some filtering tools to recognise interesting line-segments. Then Section 4.7 explains how the introduced filtering tools together with grouping algorithm are applied to propose potential candidates of zebra-crossings and dashed-lines. In order to simplify reading of this manuscript, the grouping algorithm is explained later in Section 5.3.

4.4 Data structure

As mentioned in the previous Section, in our algorithm decisions such as filtering, segmenting and grouping of linked-edges are made regarding 2D approximated coordinates. The final modelling is then performed regarding 3D coordinates. It means that for each point j belonging to a linked-edge e_i, its real 3D coordinates (X_{ij}^{3D}:$(x,y,z)_{ij}^{3D}$) and also its 2D coordinates (X_{ij}^{2D}: $(x,y)_{ij}^{2D}$) on approximate road plane are known.

The adopted strategy leads to the data structure of Figure 4.4. Let's call the 3D linked-edges resulted from edge matching algorithm (See Chapter 3) the Elementary Linked-Edges (ELE). So ELE is a list of 3D linked-edges (e_i). Each e_i is a list of linked points. For each point j belonging to linked-edge e_i, X_{ij}^{3D} and X_{ij}^{2D} are registered.

Let's see how this data structure deals with polygonisation and grouping processes of line-segments.

4.4.1 Dealing with polygonisation of linked-edges

In order to estimate line-segments from linked-edges a polygonisation is performed on each e_i of ELE (See Figure 4.3(a)). The polygonisation is performed on each e_i regarding its 2D coordinates (X_{ij}^{2D}). This polygonisation is performed using the algorithm proposed by Douglas and Peucker [1973]. The maximum authorised error is set to 2 *cm*. An elementary linked-edge (e_i) is segmented to different parts expressing linear segments. For example as seen in the Figure 4.3(b) the linked-edge e_i of Figure 4.3(a) which is constituted of points $\{P_0,\dots,P_j,\dots,P_m\}$ is segmented into two linked-edges e_p and e_q which

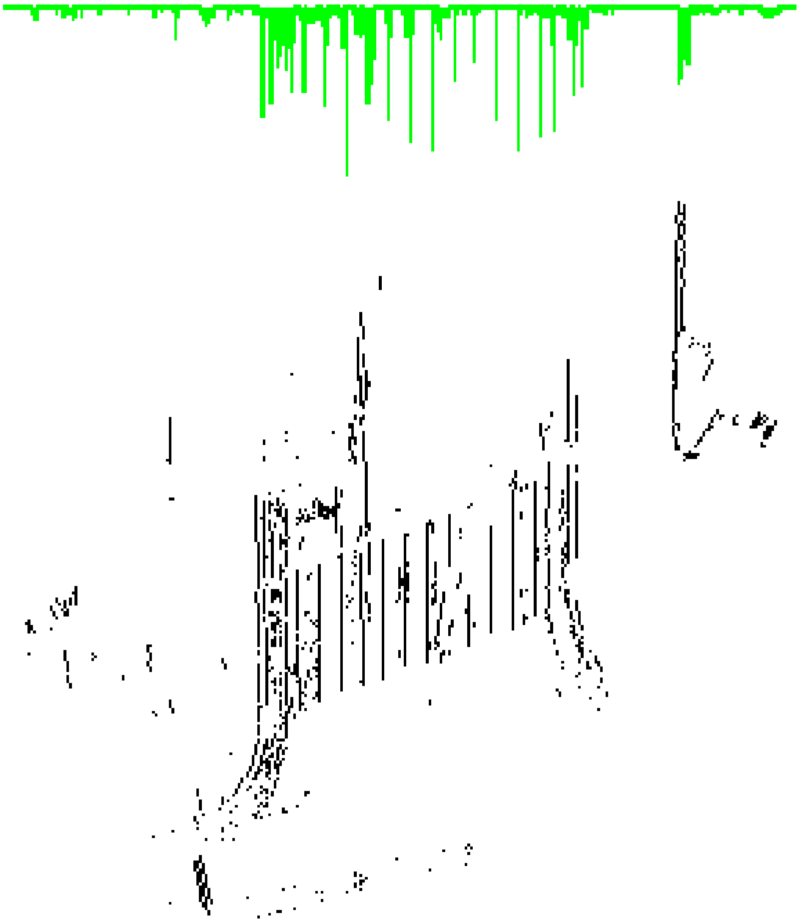


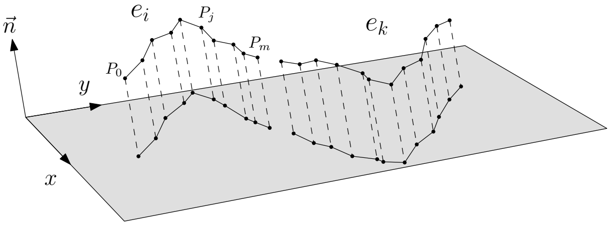
Figure 4.2: 1D Signature on zebra-crossing strips: parallel long sides cause repetitive peaks at specific distances if line-segments' lengths are accumulated in road direction.

express linear segments. The e_p is constituted of points $\{P_0, \ldots, P_{j-1}\}$ and e_q comprised of points $\{P_i, \ldots, P_m\}$. Let us call the set of such segmented linked-edges the *SLE* (Segmented Linked-Edges).

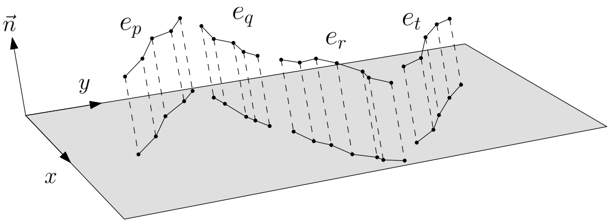
Figure 4.4 demonstrates how the data structure registers properly all points participating in e_p and e_q. For each e_p in *SLE* a 2D line-segment (S_p^{2D}) and a 3D line-segment (S_p^{3D}) can be estimated by linear regression using X_p^{2D} (set of all 2D points in e_p) and X_p^{3D} respectively (set of all 3D points in e_p).

4.4.2 Dealing with grouping of line-segments

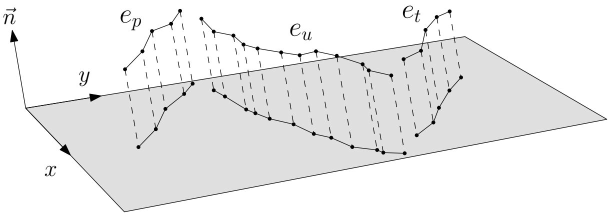
As explained in Section 5.3 *SLE*s can be processed by a grouping algorithm. The grouping algorithm analyses a set of 2D line-segments and proposes subsets of line-segment for grouping. Let's take the example of Figure 4.3(b) and suppose that two *SLE*s such as e_q and



(a) *ELE* set



(b) *SLE* set



(c) *GLE* set

Figure 4.3: Structuring 3D linked-edges.

e_r are proposed to be grouped by the grouping algorithm. In this case a new linked-edge e_u is generated and registered in an array called *GLE* (Grouped Linked-Edges). e_u contains the coordinates of all the points belonging to e_q and e_r. S_u^{3D} and S_u^{2D} are then estimated by a linear regression using X_{ua}^{3D}nd $X_{u(}^{2D}$See Figure 4.3(c) and Figure 4.4).

4.4.3 Dealing with filtering of line-segments

As explained in Section 4.6 the filtering step is based on the recognition of zebra-crossing's and dashed-line's signature in 2D road approximate plane. So given a set of linked-edges $e_q \in SLE$ decisions such as grouping and filtering are taken regarding corresponding 2D line-segment S_q^{2D} in 2D space. The advantage of this data structure is that at each moment for each linked-edge such as e_q all the contributing points are known in 3D and 2D systems. So the corresponding 2D line-segment S_q^{2D} and 3D line-segment S_q^{3D} can be estimated. Moreover thanks to this data structure given a S_P^{3D} or S_P^{2D} all the belonging points are known. Each S_p^{3D} correspond to exactly one S_p^{2D} and vice versa.

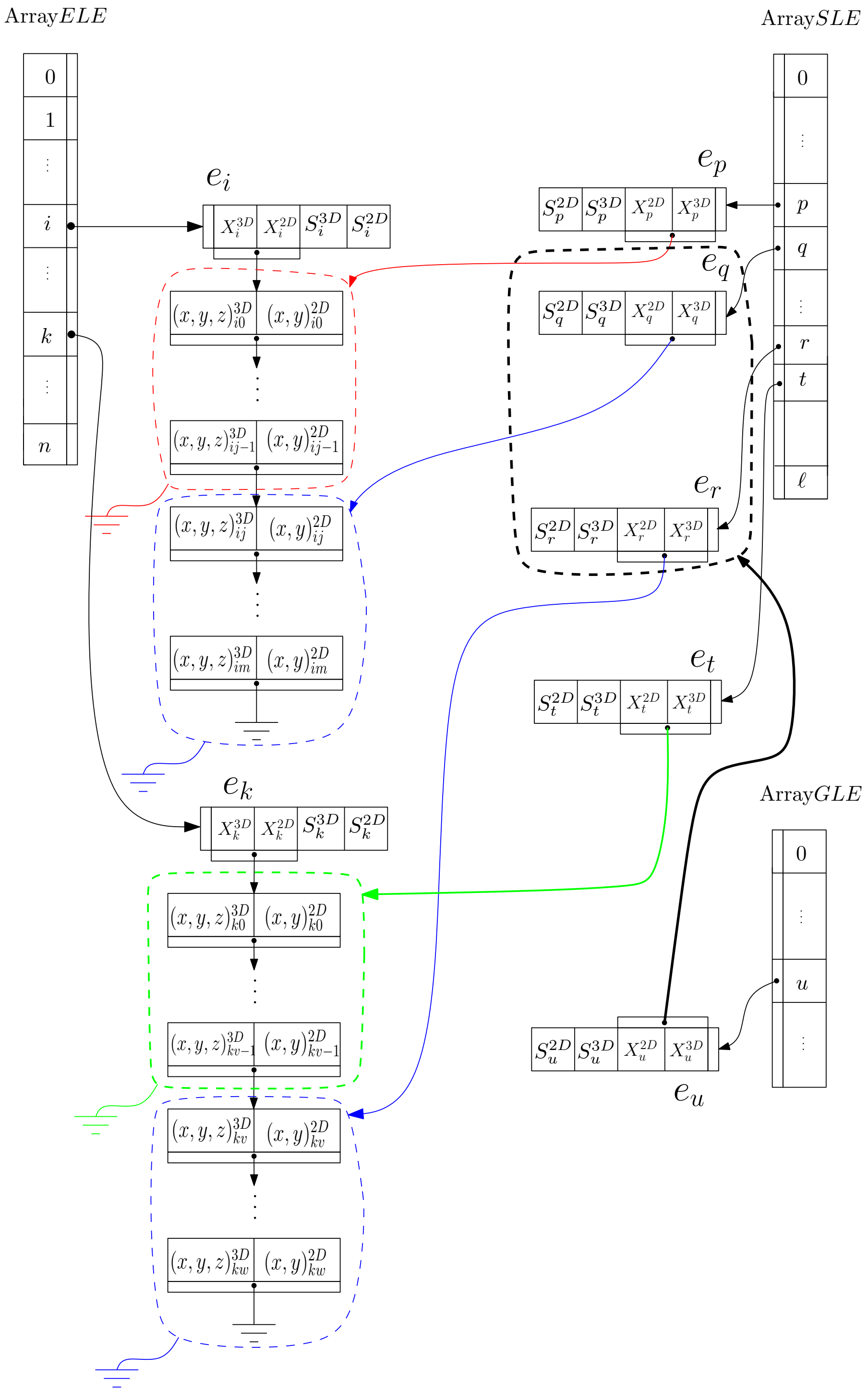


Figure 4.4: Data structure: e_i: a linked-edge, X_i^{3D}, X_i^{2D}: vectors containing coordinates (respectively 3D and 2D) of all points in e_i, ELE: Elementary Linked-Edges, SLE: Segmented Linked-Edges, GLE: Grouped Linked-Edges.

4.5 Road plane estimation?

As explained in Section 4.3, the detection step is performed on the 2D space of the plane that approximates the road surface. An appropriate approximation simplifies the detection step in two ways:

1. It allows to filter out non road features such as cars, pedestrians and vertical features.
2. It minimises geometric deformations during the 2D projection.

As mentioned in Section 2.5.2 many authors investigated road plane estimation from stereo images. Two main approaches exist:

1. Approaches using dense stereo correspondence [Bertozzi *et al.*, 1998a; Okutomi *et al.*, 2002; Sappa *et al.*, 2007]
2. Approaches using feature stereo correspondence [Labayrade *et al.*, 2002; Simond and Rives, 2004a]

As discussed in Chapter 3, dense stereo correspondence is computationally very expensive and difficult. Feature based approaches fails when there are not a sufficient amount of features on the road surface. This is always due to low number of roadmarks, large amount of obstacles and occlusions. This is the case in our edge stereo correspondence.

Since a false estimation of the road plane can wrongly filter out linked-edges and cause failure of roadmarks detection we use approximate position of the vision system relatively to the road surface to define a road plane. Thus the plane Π_0 is considered as the road plane (See Figure 3.9(a)). We suppose that the parameters of this plane in relation to the vision system (h and θ) are constant in all exposure points. In reality these parameters change due to braking, acceleration and vibrations. This is an significant approximation and disables us to filter out non road features. Fortunately the introduced geometric deformation due to 2D projection remains negligible.

4.6 Line-segment filtering tools: accumulation with respect to principal directions

As mentioned before in Section 4.4 in our algorithm zebra-crossings and dashed-lines are recognised thanks to their 1D signature that is computed on a plane that approximates the road surface. So from now on, we suppose that the Elementary Linked-Edges (ELE) are segmented into linear portions that constitute Segmented linked-Edges (SLE) and for each element e_i of SLE the corresponding 2D line-segment (S_i^{2D}) is computed (See Section 4.4.1). Only the so called 2D line-segment are used in detection of zebra-crossings and dashed-lines. So the S_i^{2D} is considered when line-segment is mentioned.

Zebra-crossings and dashed-lines are always aligned with the road main direction. The majority of reconstructed line-segments belong to road structures like roadmarks or road borders. So the road main direction can be estimated as the most occurring direction of line-segment set on the road plane. Sections 4.6.1 – 4.6.6 present some processing tools which are used for recognition of zebra-crossing and dashed-lines.

4.6.1 *X accumulation space*

The road direction being estimated, a 2D coordinate system is defined on the road plane where the Y axis is set to the road direction. The X axis is defined by the orthogonal axis to Y axis (See Figure 4.5). The X axis is then discretised with cells of size Δ_X. Δ_X is chosen to be proportional to the 3D reconstruction precision and irregularity of zebra-crossing or dashed-line strips borders in the real world. In our case it is set to $1cm$ for all the data set. Each cell is then voted by length of all the line-segments' portion that are orthogonally projected in this cell. So the line-segments that are perfectly in principal direction vote by their total length for only one cell and the other line-segments divide their voting power between all the cells in which they are projected. Let us call the sum of all received votes by a cell, the cell's accumulation score. For each cell the accumulation score and the index of all line-segments voting for the cell are registered.

Let us consider the so called discretized X axis as an accumulation space and call it *X accumulation space*. The next Section explains the interest of *X accumulation space* for the recognition of fragmented parallelogram patterns.

4.6.2 Signature of zebra-crossing and dashed-line in *X accumulation space*

Figure 4.5 shows the accumulation of line-segments in the road direction in a scene containing a zebra-crossing and a dashed-line. The *X accumulation space* is represented in green colour. One can see that *X accumulation space* presents repetitive peaks at a fixed distance w_z that is a sign of zebra-crossing. Twin peaks of at distance of w_d can correspond to presence of a dashed-line. Each peak correspond to the presence of aligned line-segments and its value represents the sum of line-segment's length that are projected in this cell.

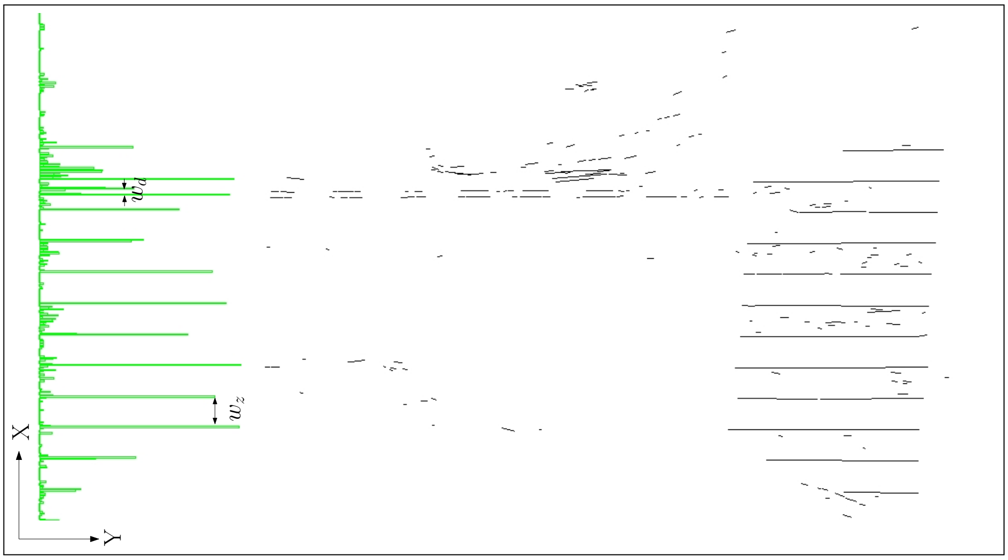


Figure 4.5: Signature of zebra-crossing and dashed-line in *X accumulation space*.

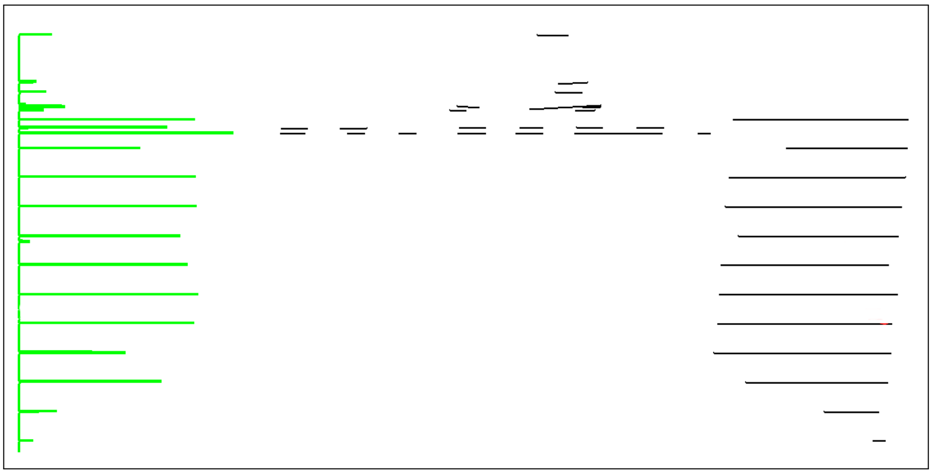
4.6.3 Filtering using *X accumulation space*

Let us take the example of a parallelogram strip. Let us set ℓ as its length and w as its width. The two long sides of the strip vote by their length ℓ in two different cells with a distance of w in *X accumulation space*. So they leave a systematic signature in accumulation space. In contrast to parallelogram patterns, other line-segments that are due to local textures and presence of cars or pedestrians cause random accumulations.

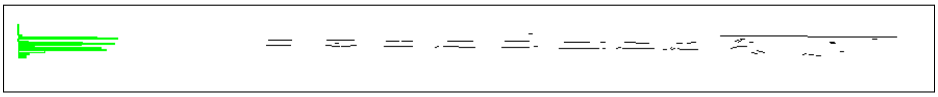
So the detection of strips of a given size consists of separating systematic signatures from the random ones in *X accumulation space*. In this purpose, our global idea is to take benefit from known length (ℓ) and width (w) of strips to filter out line-segment set in the two following stages :

1. The first stage is based on the expected length of strips (ℓ) and threshold the cells with accumulation score smaller than expected length.
2. The second stage is based on width of strip (w). Within the remaining cells it filters out all cells not having any non filtered cell at distance of w.

Figure 4.6 shows resulting line-segments after filtering using the specified the parameters (w, ℓ) of zebra-crossing and dashed-line. This filtering method is based on accumulation score of cells. Accumulation score received by a cell is very sensitive to discretization step. In other words, larger cell size causes higher accumulation score. Inappropriate discretization step reduces the efficiency of such filtering process. Section 4.6.4 explains this problem and gives a simple solution to reduce the sensibility of filtering process to discretization step.



(a) results after filtering with zebra-crossing's parameters (w_z, ℓ_z).



(b) results after filtering with dashed-line's parameters (w_d, ℓ_d).

Figure 4.6: Recognition of zebra-crossing and dashed-line with the *X accumulation space* filtering.

4.6.4 Coping with sensibility of *X accumulation space* to discretization step

Given a strip's length the accumulation value corresponding to each cell of *X accumulation space* can be analysed to see if it represents a strip's long side. In fact if all the fragmented parts of a long side are projected in the same cell in the *X accumulation space*, the corresponding accumulation score approximates the length of a strip's long side. For example an accumulation score of 2 m can be a key to find an approximately 2 m long side of a strip in the principal direction at the position of the so called cell. However the accumulation score of each cell is very sensitive to the discretization step.

As seen in the example of Figure 4.7(a) line-segments resulting from the polygonisation of 3D reconstructed linked-edges are somehow fragmented and irregular. The irregularity can be due to real irregularity of strip borders in the real world or limited precision of 3D reconstruction. In order to find a significant signature of zebra-crossing's and dashed-lines' strips, the discretization step of the *X accumulation space* must be selected near the mentioned irregularity value.

Figure 4.7(a) demonstrates such a *X accumulation space* with the Δ_X discretization step. It leads to an accumulation score of L in a cell. This cell can correspond to a strip side of length L. Now suppose the same line-segment set but a *X accumulation space* that is discretized with $\frac{\Delta_X}{2}$ step conform to the Figure 4.7(b). As a result, line-segments vote in two neighbouring cells and so divide their voting power. It causes an accumulation score of about $\frac{L}{2}$ in each cell. So these cells can no more correspond to a strip side of length L. This example demonstrates the sensibility of the first step of filtering (See Section 4.6.3) to the selection of the discretization step.

Even if the discretization step can be selected by taking into account precision of 3D reconstruction, the irregularity of strip's sides is not constant and so is not possible to take it into account. In order to reduce this effect, instead of analysing each cell independently, we aim at constituting groups of connected cells and analysing the sum of accumulation scores: a hysteresis thresholding is performed on the accumulation scores with high and low thresholds. Both of thresholds are selected proportional to length of strip of interest (ℓ). So groups of connected cells are generated by low threshold ($Tr_l = 20\%\ \ell$) and high threshold ($Tr_h = 50\%\ \ell$). Each connex component constitutes a group. Then each group is accepted, if the sum of accumulation scores reaches approximately the expected strip length (ℓ).

4.6.5 Combined signatures and *Y accumulation space*

Let's consider the global idea of filtering which is mentioned in Section 4.6.3. This strategy simplifies the detection step by reducing a 2D problem to a 1D one. In other words the mentioned strategy supposes that two groups of fragmented line-segments correspond to two sides of a strip ($w \times \ell$) if and only if their projections cause two peaks of distance w with accumulation score of ℓ. However in the general case, such two peaks can be caused by two groups of line-segments that are separated in Y direction (road direction). So such groups are not geometrically neighbours and consequently can not correspond to a strip. As shown in Figure 4.7(c) this effect is still more perturbing in scenarios in which two zebra-crossings which are separated in road direction are present in the scene.

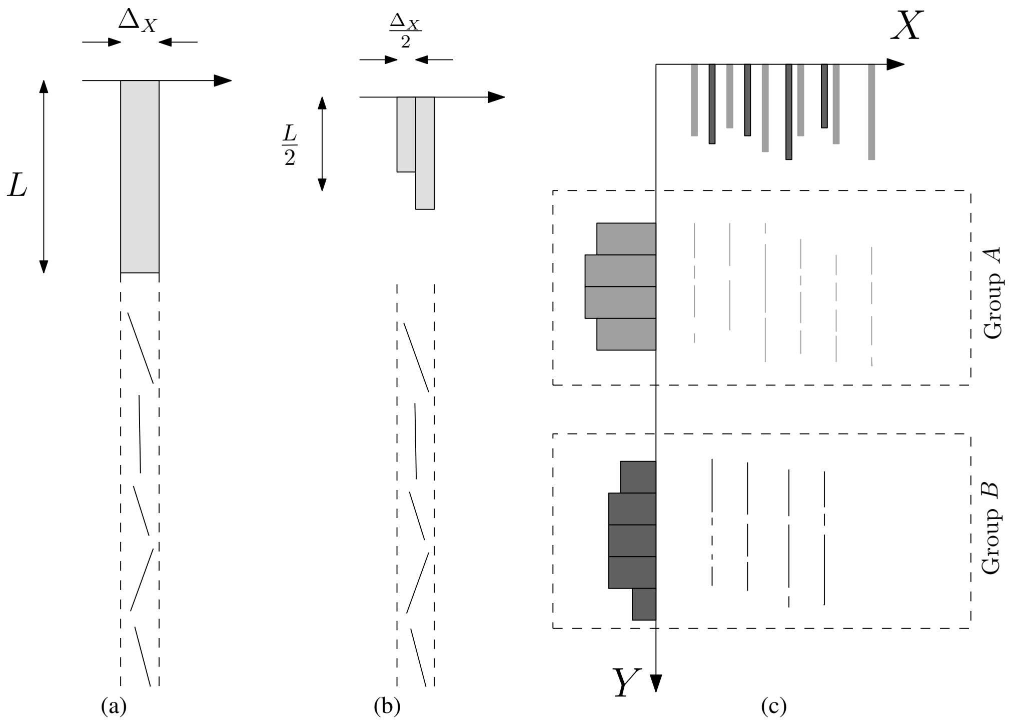


Figure 4.7: Problems of *X accumulation space*.

In order to cope with this problem, the Y axis of the 2D space defined in Section 4.6.1 is also discretized to cells of size Δ_Y which is proportional to strip's length ($n\%$ of ℓ) and constitutes what we call the *Y accumulation space*. Each cell is voted by the number of line-segments that are projected in this cell. The *Y accumulation space* is then decomposed to connex non-zero components. Each component can be analysed independently. For example in Figure 4.7(c) two components A and B are separated. Each group is then re-entered in the *X accumulation space* filtering process. Decomposition of *Y accumulation space* is also inevitable in the case of dashed-lines for separating individual strips. Figure 4.8 shows how this accumulation space is used for the separation of strips. However noises and local texture can lead to irrelevant line-segments that accumulate wrongly in the cells of the *Y accumulation space*. This is the case for three strips in the Figure 4.8. The following Section presents a tool to cope with this problem.

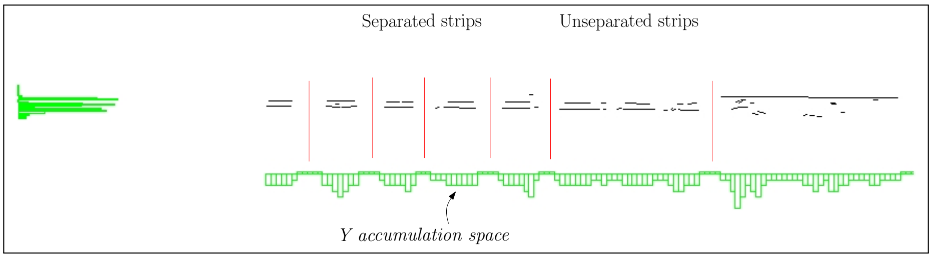


Figure 4.8: Separation of strips thanks to *Y accumulation space*.

4.6.6 Radiometric based accumulation of *Y accumulation space*

In order to decompose a dashed-line into its strips in noisy scenarios, a radiometric accumulation is performed instead of line-segments number. For this purpose line-segments are projected in image space, a radiometric profile is extracted and back-projected in accumulation space. The advantage is that even if local textures are present on the surface of bitumen and strips, the mean radiometric value on strips is higher than on bitumen. In order to decompose a dashed-line to its strips using radiometric base accumulation, the mean value of radiometric profile is used to threshold the profile (See Figure 4.9). This thresholding method is efficient only if the image region in which the radiometric profile is extracted contains exclusively bitumen and strips and so the corresponding accumulation is monomodal. The presence of other objects can cause multimodal accumulations and so non-efficiency of thresholding. In order to limit the region of radiometric accumulation, this method is applied after decomposing a potential dashed-line into smaller parts by an accumulation that is based on number of line-segments as explained in Section 4.6.5.

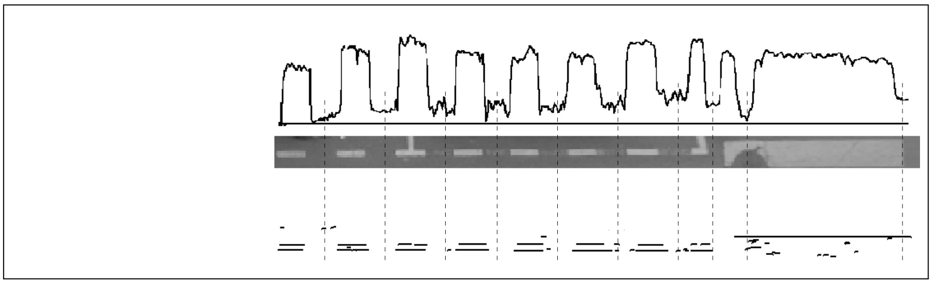


Figure 4.9: Decomposition of dashed-line into strips using radimetric accumulation in Y direction.

4.7 Distinguishing zebra-crossing and dashed-line signatures

This section explains how the introduced filtering tools presented in Section 4.6 and the grouping algorithm developed in Section 5.3 are used to provide potential candidates of zebra-crossing and dashed-line strips. Due to difference of size and geometric configuration of strips in zebra-crossing and dashed-line, the detection step is slightly different. This difference is just in parameters and in the order in which processing tools are applied. Figure 4.10 recall the required pre-processing on the 3D linked-edges resulting from matching step (*ELE*s) before injection in zebra-crossing and dashed-line detection algorithms.

Sections 4.7.1 and 4.7.2 explain respectively algorithms of zebra-crossing and dashed-line detection. For each algorithm, intermediate results are demonstrated on an example.

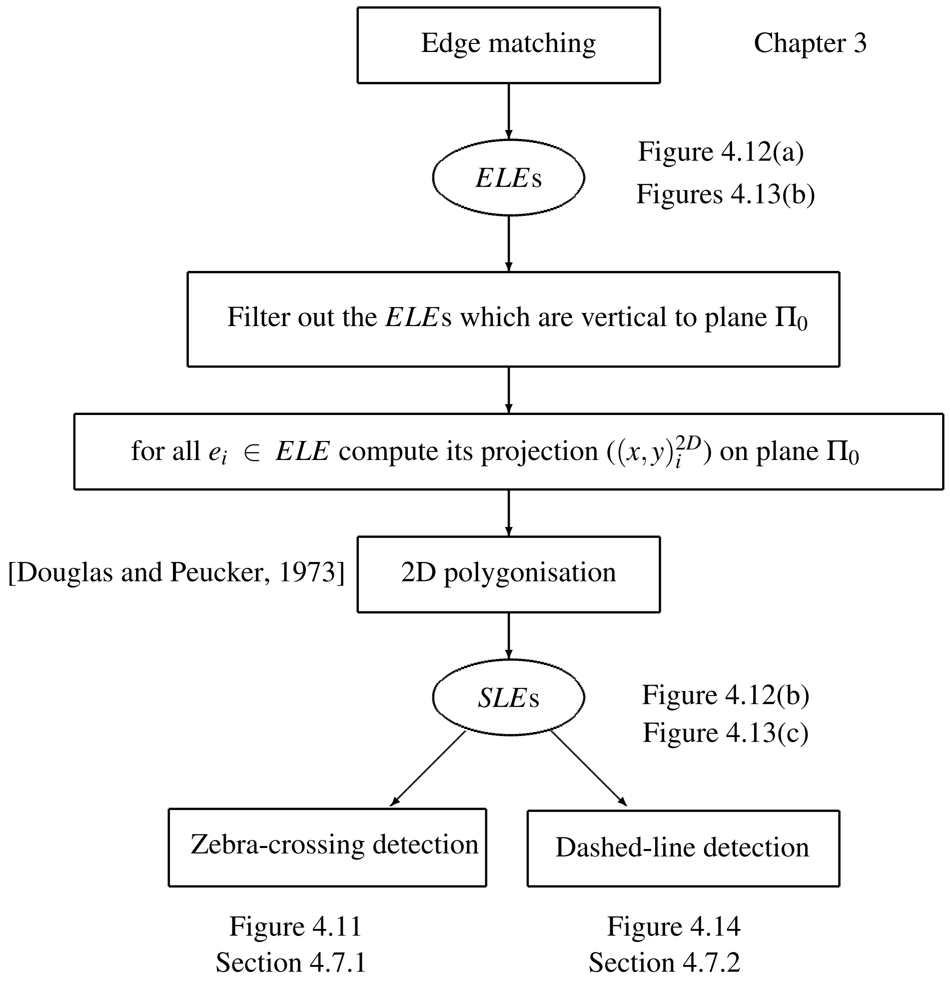


Figure 4.10: Pre-processing performed after matching step and before zebra-crossing and dashed-line detection.

4.7.1 The zebra-crossing detection algorithm

The diagram of Figure 4.11 shows the sequence of operations to perform for zebra-crossing detection. Intermediate results corresponding to each operation are presented in Figure 4.12 for illustration purposes.

Figure 4.12(b) shows the projected 3D line-segments on the 2D space and the corresponding accumulations in X and Y axes. Line-segments are then filtered using the method proposed in Section 4.6.3.

As it can be seen in Figure 4.12(c) a large number of irrelevant line-segments are removed by first step of *X accumulation space* filtering. However the signature of both zebra-crossings are still combined. This can disturb filtering of many uninteresting line-segments.

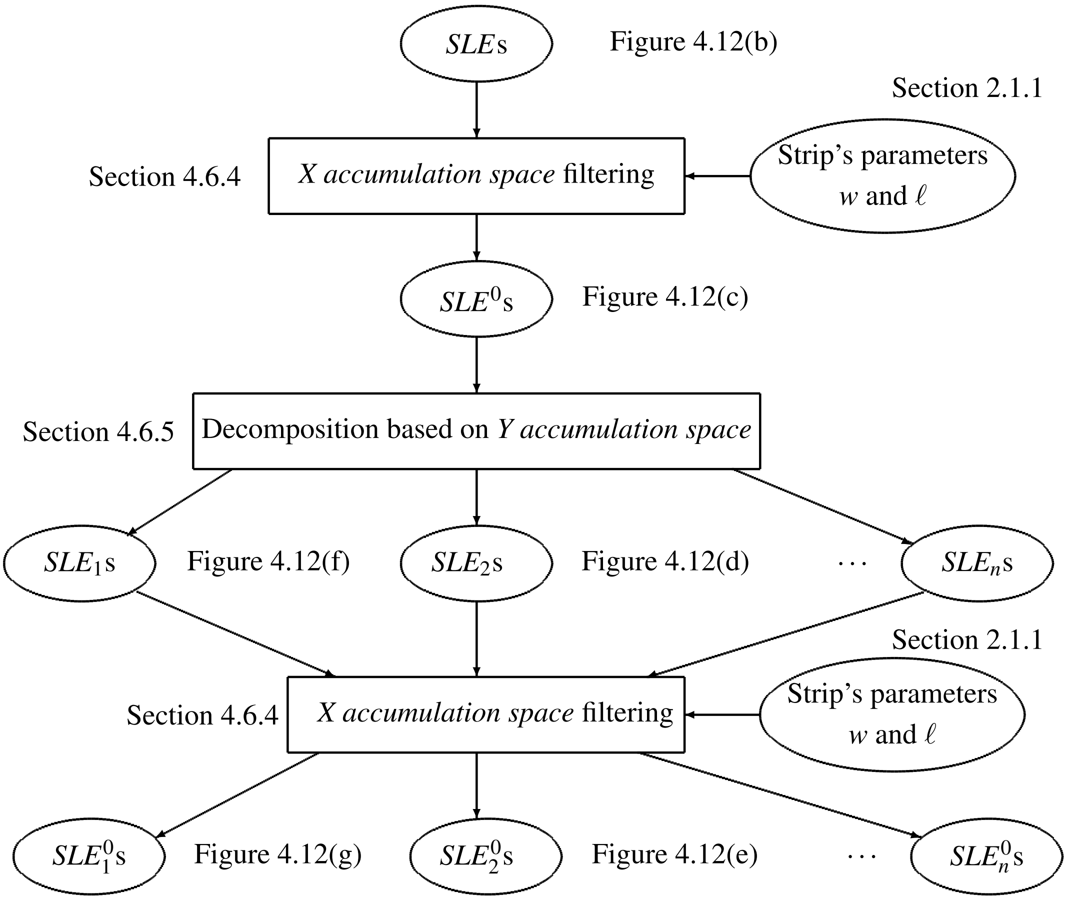


Figure 4.11: The algorithm of zebra-crossing candidates generation from Segmented Linked-Edges (*SLE*s)

In order to cope with this problem line-segments are decomposed into subsets that are separated in *Y* axis. This is performed by analysing *Y accumulation space* as explained in Section 4.6.5. Figures 4.12(d) and 4.12(f) depict the separated zebra-crossings.

Each separated subsets is re-analysed by *X accumulation space* filtering (See Figures 4.12(g) and 4.12(e)). Thanks to this step more irrelevant line-segments are filtered out. Even if there are still some irrelevant line-segments within these candidates, as explained in Chapter 5, they can be filtered in the modelling step by taking into account the gradient directions criteria, and looking for supporting transversal (short) sides. The higher level primitives provided during the detection step simplifies the modelling step considerably.

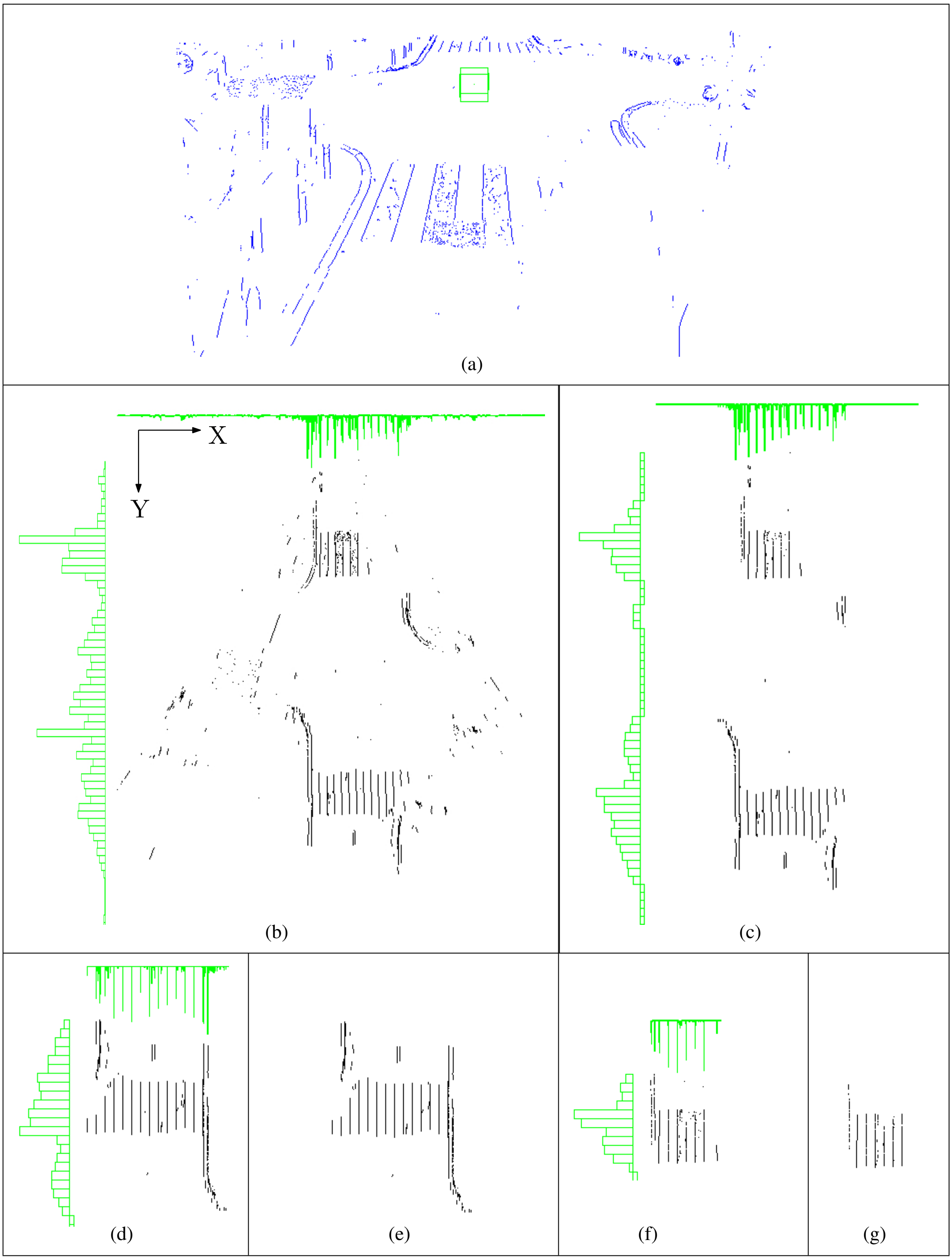
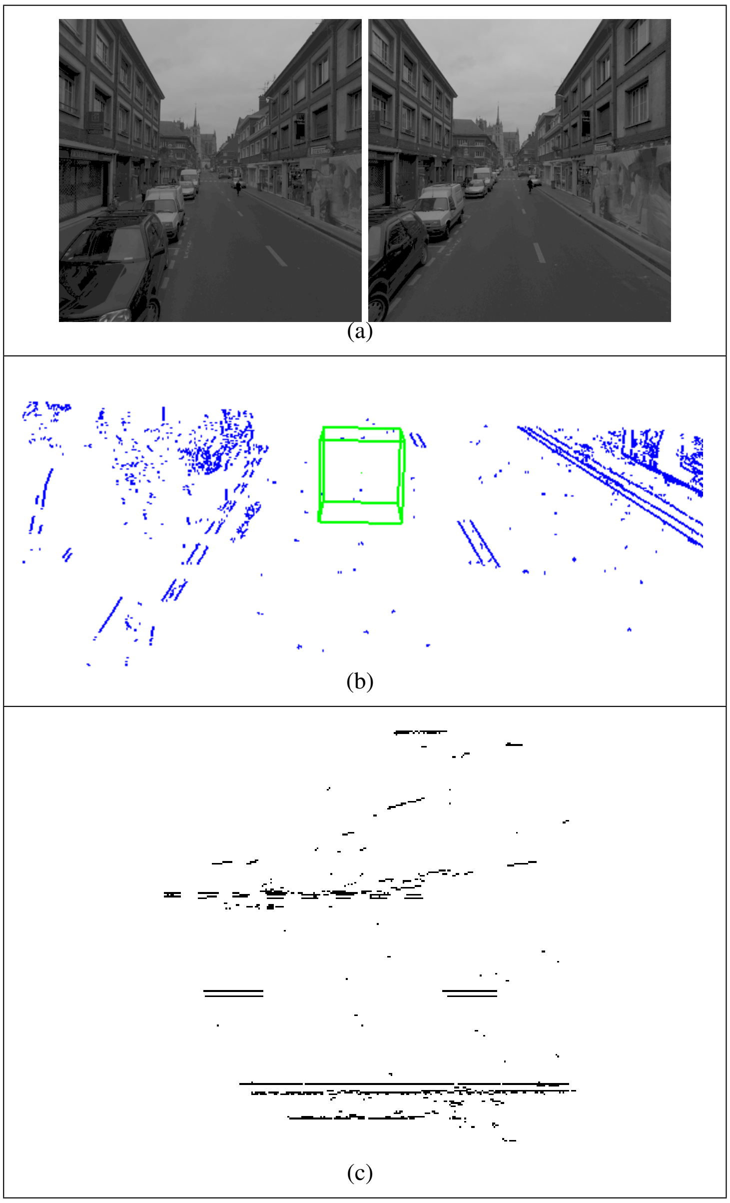


Figure 4.12: Intermediate results of zebra-crossing candidates generation.

4.7.2 The dashed-lines detection algorithm

The diagram of Figure 4.14 expresses the operations to be performed in order to provide potential candidates of dashed-line strips. The input is the same as for zebra-crossing detection (a set of Segmented Linked-edges). The Figure 4.13 shows the stereo-pair on which we present the intermediate results. It contains two kinds of dashed-lines ($T'2$ that corresponds to parking lots and $T1$ that corresponds to road center-line). For each kind of dashed-line, the operations of the diagram are performed independently.


(a)

(b)

(c)

Figure 4.13: a) A stereo-pair containing dashed-lines, b) 3D reconstructed linked-edges, c) projection of 3D segmented linked-edges on the 2D space.

Figures 4.15 and 4.16 depict the intermediate results of each operation for dashed-lines of type respectively $T^{'}2$ and $T1$.

X accumulation space filtering

First a *X accumulation space* filtering is performed by taking into account the parameters corresponding to the dashed-line of interest (w and ℓ). In order to find a dashed-line even if it is constituted of only one strip, the length of only one strip is chosen as the thresholding parameter of the *X accumulation space* filtering (See Section 4.6.4).

Figure 4.15(b) shows the resulting line-segments for $T^{'}2$ type filtering. The center-line of the road (dashed-line of type $T1$) is correctly filtered. What remains is a dashed-line of type $T^{'}2$ composed of 8 strips. Many uninteresting line-segments are present around the so called strips that are due to local textures. There are also some line-segments corresponding to the right border of the road that pass the filter. Presence of noises causes that the border line-segments pass trough the filtering step even if its width is larger than width of $T^{'}2$ strips.

Figure 4.16(b) shows the same scene filtered by parameters of $T1$ type strips. As it can be seen the presence of noises causes that strips of $T^{'}2$ type and the right side border of road pass the filter.

These mentioned filtering problems, are tackled in the two following phases of dashed-line separation and strip separation.

Dashed-line separation

The second step consists in separating remaining components of the *X accumulation space*. In the example of Figures 4.15(b) and 4.16(b) line-segments corresponding to road border, road center-lines ($T3$) and dashed-line delimiting the parking lots ($T^{'}2$) can be separated thanks to their important distance in *X accumulation space*. Figures 4.15(c) and 4.15(d) show the separated component of filtering results with $T^{'}2$ parameters and Figures 4.16(c),4.16(d) and 4.16(e) present those of filtering with $T3$ parameters. The interest of separating components of *X accumulation space* is that after separation each component can be analysed by *Y accumulation space* for separating strips.

Strip separation

In each separated component the signature of line-segments in *X accumulation space* is the combination of all strips' signature. In this case the non alignment of strips and also noise disturb the signature of collection by accumulation of errors. In order to avoid accumulation of errors, we aim at separating the strips within a dashed-line candidate.

Analysing each strip independently simplify considerably dealing with noises and recognition problem. Indeed after separation, each component can be re-analysed by *X accumulation space* filtering. Separation avoid noise accumulation and so filtering is more efficient.

In order to separate the strips within a dashed-line candidate, the corresponding *Y accumulation space* is analysed as explained in Section 4.6.5. First the accumulation based on the number of line-segments is used to decompose the dashed-line into smaller parts.

Then on each separated part radiometric based accumulation is computed and used for final separation.

Figures 4.17, 4.18 and 4.19 shows the use of *Y accumulation space* for separation in the case of $T'2$, $T1$ and road border. In each of the mentioned figures, the left column presents the dashed-line candidate and corresponding Y accumulation using the number of line-segments. The middle column shows the results of separation and also Y accumulation corresponding to each separated part using radiometric accumulation. The right column presents the final separated strips.

Figure 4.17(d) shows the interest of radiometric accumulation when using the number of line-segments is not sufficient for the separation.

The results of separation are good enough for dashed-lines ($T'2$ and $T3$). However as it can be seen in Figure 4.19 the radiometric accumulation cause a random separation. This is due to the choice of mean radiometric value for thresholding. In the case of homogeneous areas like continuous lines the resulting separation is quite random. We recall that the provided strips are just candidates and are re-analysed in modelling step. Such candidates are analysed by their size, gradient directions, and some other parameters and can be rejected in the following steps.

4.8 Conclusion

This chapter presents our strategy for filtering 3D linked-edges resulting from matching step. Filtering is performed regarding 2D signature of each strip type on the 2D space that approximates the road surface. For this purpose the specified shape and size of strips are used. The detection step provides sets of line-segments as hypothetical candidates of roadmark objects.

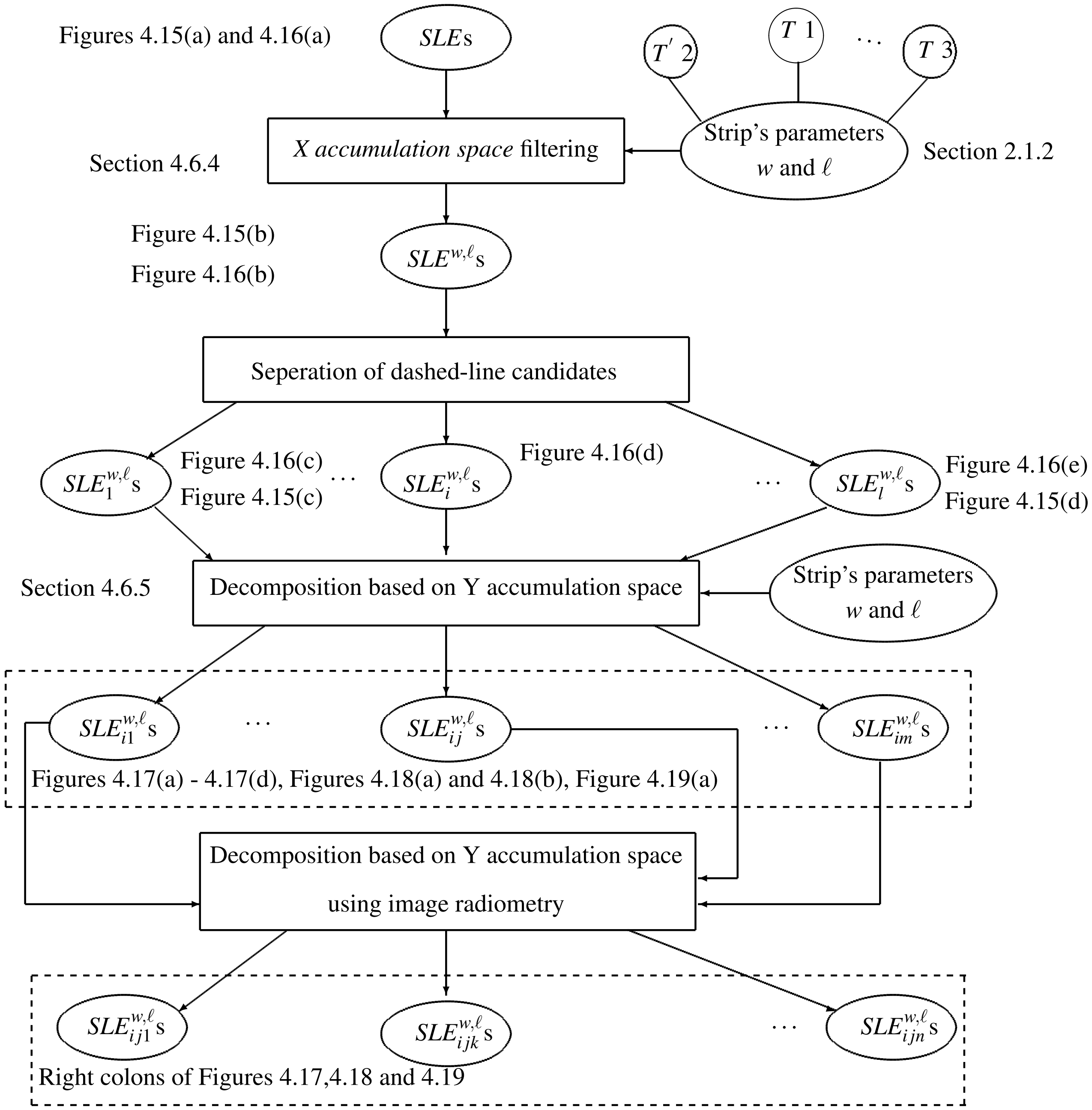
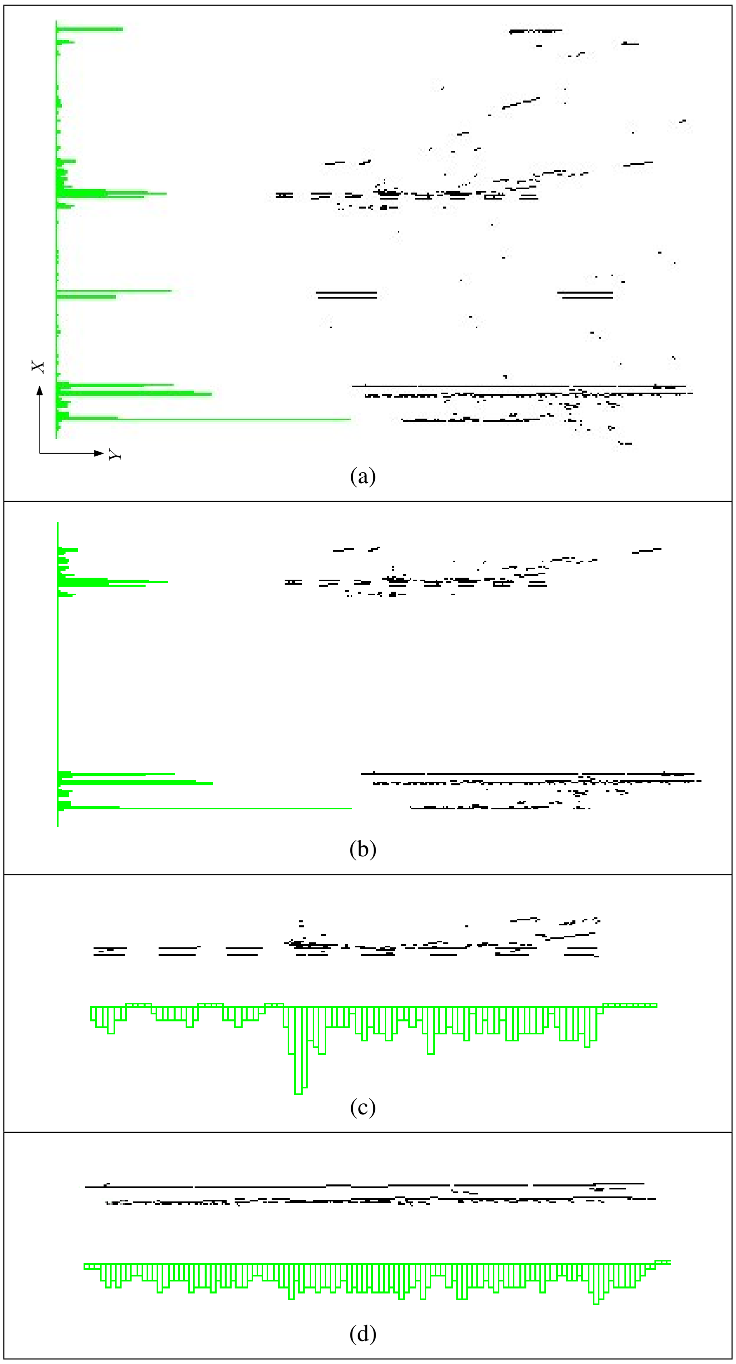


Figure 4.14: dashed-line strip detection algorithm



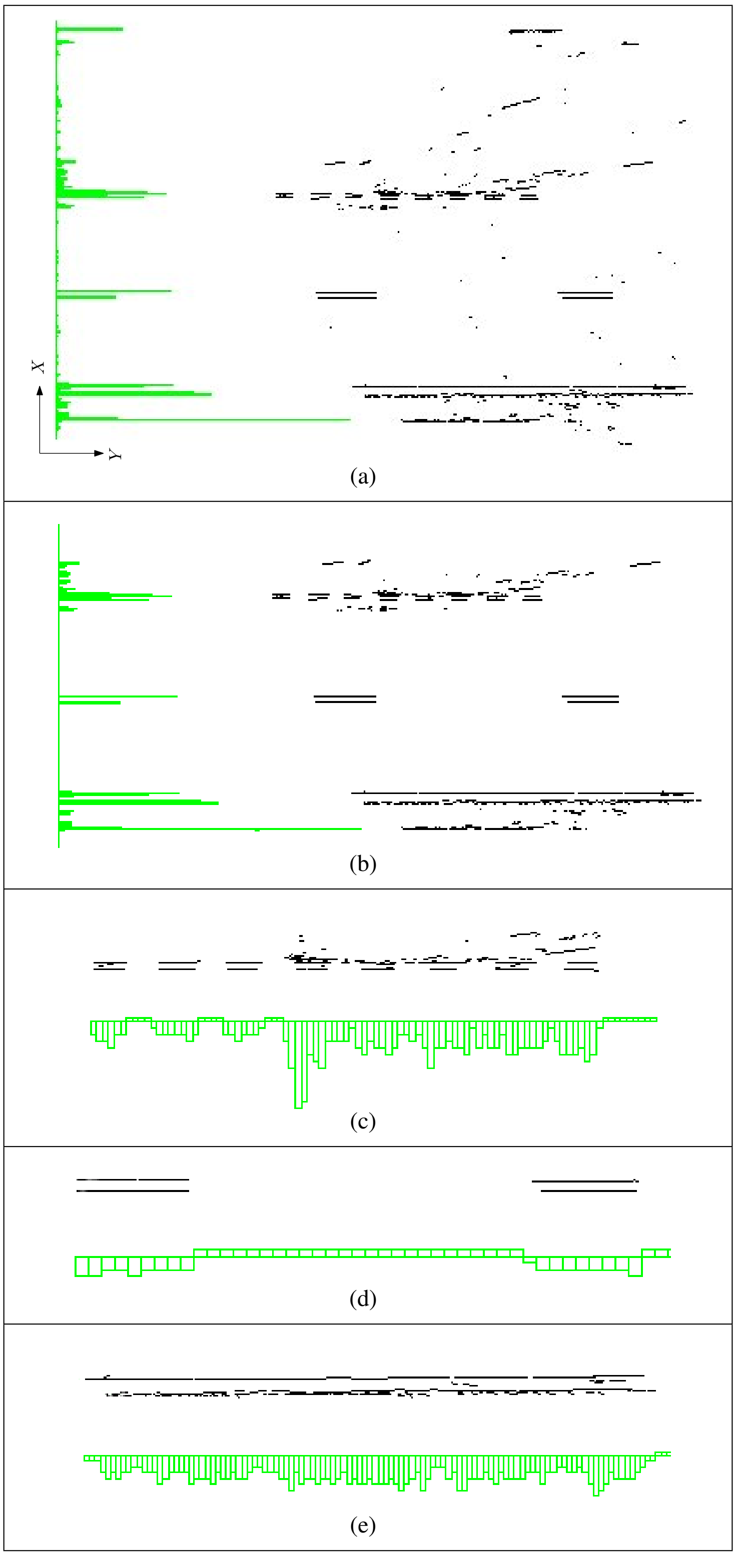
(a)

(b)

(c)

(d)

Figure 4.15: Filtering regarding $T^{'}2$ type strip parameters.



(a)

(b)

(c)

(d)

(e)

Figure 4.16: Filtering regarding $T1$ type strip parameters.

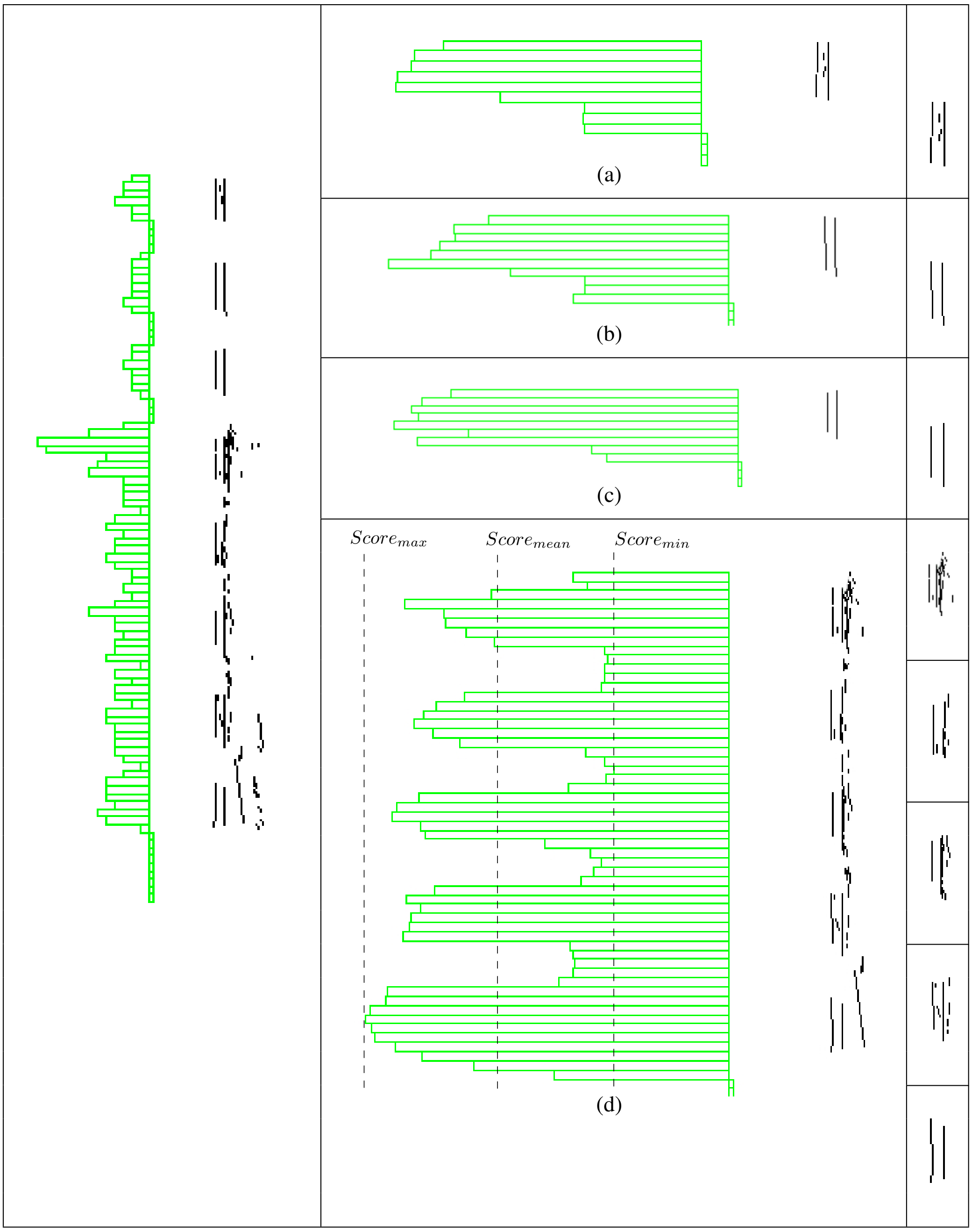


Figure 4.17: Decomposition to strips of a $T^{'}2$ dashed-line.

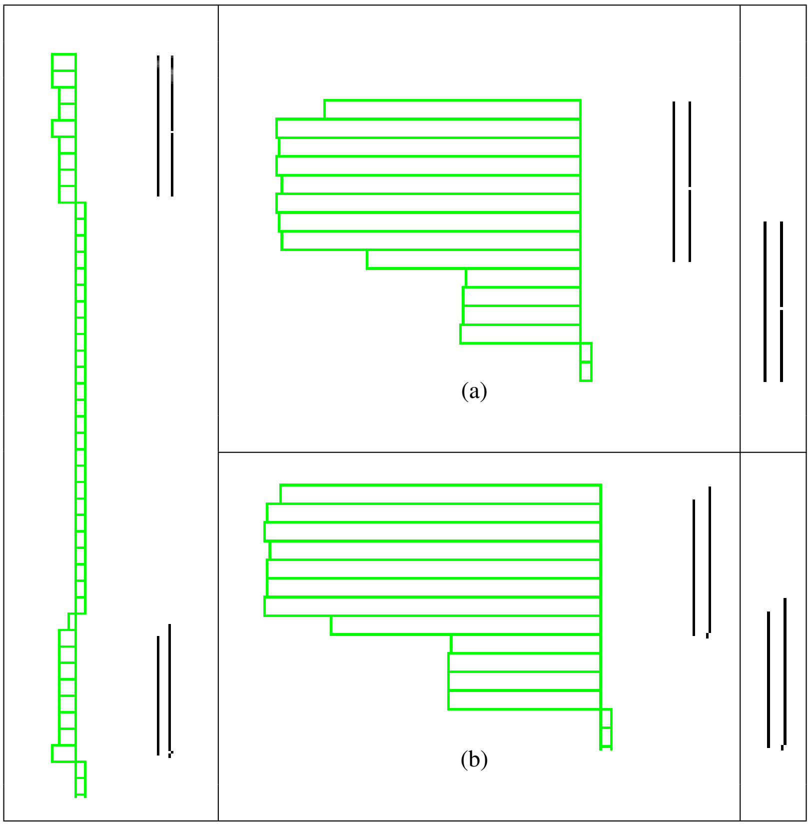


Figure 4.18: Decomposition to strips of a $T3$ dashed-line.

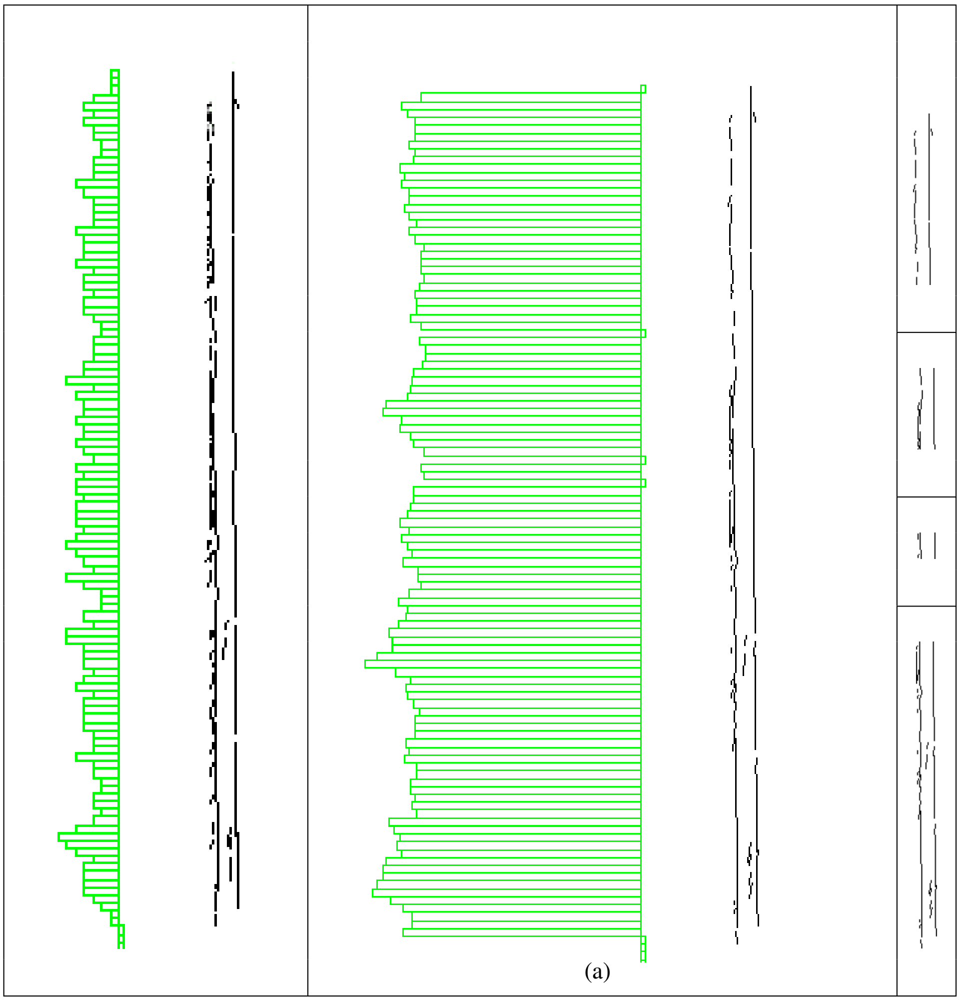


Figure 4.19: decomposition of continuous line.

Chapter 5

Roadmark modelling

The previous chapter discussed recognition of zebra-crossing's and dashed-lines' strips. Result of the detection step is a set of hypothetical strip candidates that are represented by 3D line-segments. In order to generate the roadmark models these line-segments require to be validated and refined with respect to a given model. Both operations will be performed within what we call the modelling step.

In this chapter, we deal with both cases of zebra-crossing's and dashed-line's strips in a same way. According to the specifications presented in Section 2.1 a quasi-parallelogram pattern is adopted for representing a strip. The goal of the modelling is to provide an independent quasi-parallelogram model for each strip.

The modelling step, consists in two steps. The first one, as explained in Section 5.1, tends to recognise a pair of 3D line-segments representing two strip's long sides within the input line-segments. In order to provide a line-segment for each side of the strip, the algorithm must cope with fragmentation effect. For this purpose, we present a line-segment grouping algorithm which is explained in Section 5.3. An initial model is defined using the two long sides.

The second step as explained in Section 5.2, consists in finding the corresponding transversal sides (short sides) for each initially modelled strip to form a quasi-parallelogram pattern.

5.1 Strip long sides modelling

Even if line-segments are filtered in the detection step, some irrelevant line-segments still persist. Moreover line-segments suffer from fragmentation effect. However, line-segments corresponding to each strip are partitioned into a set (one line-segment set per strip candidate). This partitioning simplifies considerably the modelling step.

As mentioned in Section 2.1 some parameters of strip's shape are known in the case of zebra-crossings and dashed-lines. Equation 4.1 summarises the parameters of the strip shape. According to this model and in order to recognise a strip within a set of line-segments, we look for a pair of parallel line-segments that represent left and right sides of a strip. The distance between the so called line-segments is known. Moreover they must be of inverse gradient directions.

Equation 5.1 defines our distance measure between a pair of quasi-parallel line-segments. We suppose that two line-segments are parallel if their orientation difference does not exceed a small tolerance. The tolerance is fixed proportional to the reconstruction precision and also to the eventual geometric anomalies in roadmark set up.

$$Dist(s_i, s_k) = \begin{cases} \frac{\mathcal{H}(\mathcal{M}_i, s_k) + \mathcal{H}(\mathcal{M}_k, s_i)}{2} & \text{if } s_i \parallel s_k \\ \infty & \text{otherwise} \end{cases} \tag{5.1}$$

where:

s_i, s_k : two 3D line-segments

$\mathcal{M}_i, \mathcal{M}_k$: centers of s_i, s_k

$\mathcal{H}(\mathcal{M}_i, s_k)$: distance of point $\mathcal{M}_i$ from line-segment s_k

The gradient of each line-segment is measured by its projection in one of the left or right images (See Equation 5.2).

$$\begin{aligned} \vec{\nabla}(s_i) &= \sum_{p_j \in \tilde{s}_i} (G_x(p_j), G_y(p_j)) \\ \tilde{s}_i &= f(s_i) \end{aligned} \tag{5.2}$$

where :

$\vec{\nabla}(s_i)$: gradient vector corresponding to line-segment s_i

$\tilde{s}_i$: projection of line-segment s_i in one of left or right images

f : the image projection function

p_j : points on $\tilde{s}_i$

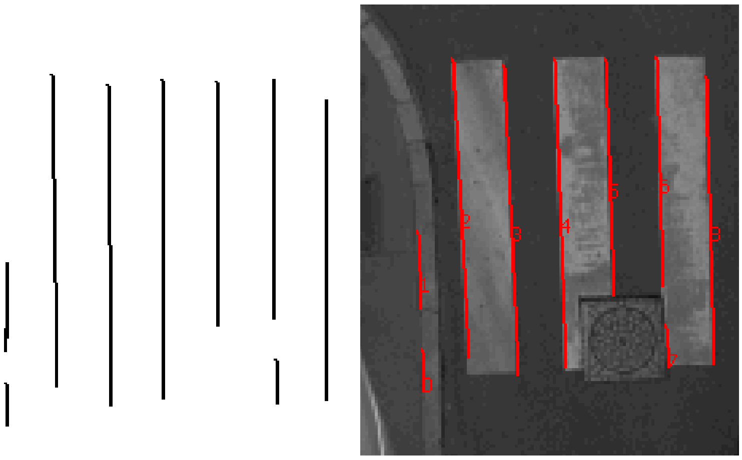
G_x, G_y : the image gradient function in x and y directions

Our long side modelling method is outlined in algorithm 1. The algorithm is the same for zebra-crossing and dashed-lines with a small difference.

The input of algorithm for dashed-line strip modelling is a set of line-segments that are candidates for only one strip. In other words strips candidates are already separated in detection step (See Figures 4.17 and 4.18). The line-segments corresponding to each hypothetical strip are entered separately in this algorithm.

In contrast, candidates of zebra-crossing's strips are not still separated (See Figures 4.12(e) and 4.12(g)). However, due to different configuration of zebra-crossing strips from dashed-lines ones, the algorithm can separate them. So the output is a set of separated strips.

In order to simplify the search for conjugated strip's long-sides, the input line-segments are sorted. As all the line-segments are aligned with the road direction, the sorting is performed in the direction orthogonal to the road in relation to each line-segment midpoint. The order is then set in way to represent line-segments from left to right in respect to image space. This sorting is possible because the reconstructed line-segments are in a nearly vertical direction in image space. Let us recall that horizontal line-segments (in image space) are not reconstructed because of the fact that they are parallel to *epipolar* direction. Figure 5.1 demonstrates the sorted line-segments that correspond to a zebra-crossing.



(a) (b)

Figure 5.1: (a) The line-segments resulting from detection step, (b) the same line-segments sorted in image space.

The gradient vector of line-segments is then used to identify if they belong to the left or to the right strip side. We suppose four main directions that are illustrated in Figure 5.2. The gradient direction of segments belonging to the left and right strip side is respectively in $Z_1 \cup Z_4$ and $Z_2 \cup Z_3$. Let us recall that gradient vectors of line-segments are measured in rectified images and so the effect of perspective deformation is negligible.

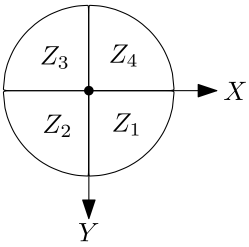
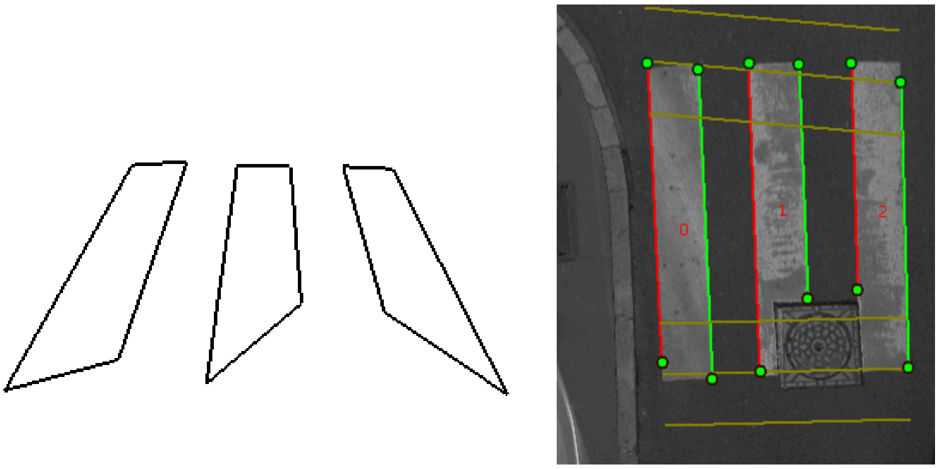


Figure 5.2: Gradient direction zones

The algorithm browses the sorted line-segments from beginning to find a left side candidate. The left side is characterised with its gradient direction $(Z_1 \cup Z_4)$. After founding first candidate for left side, other line-segments that are situated in a distance shorter than strip's width are tested by order. They can be grouped to left or/and right sides' candidates in respect to their gradient direction and also their distance with already found left side's candidate. The grouping is performed using the method presented in Section 5.3.

At the end an hypothetical strip candidate is generated $(Strip_{new})$. Then the algorithm continues to search within remaining line-segments for other strips. Each new strip is registered in a vector of strips $(StripSet)$. Presence of noises and local textures on the road surface can cause some false strips. These false strips overlaps often true strips. However true strips represent often higher contrast on the strip's borders. In order to cope with this problem, before adding a hypothetical strip to the vector of final strips it is tested to see if it overlaps with the last found strip. In the case of overlapping, only the strip with maximum of gradient is accepted. Indeed overlapping is tested only with the last strip, this is logic because line-segments are ordered. Figure 5.3 depicts the obtained results on the input of Figure 5.1.



(a) Initial 3D model

(b) Projection of the initial 3D model in image and search space for short sides of strips.

Figure 5.3: Three strips are generated by associating left sides (in green) to right sides (in red).

5.2 Strip transversal sides modelling

The long side modelling step constitutes a first 3D strip model. However when long sides are damaged, partially occluded or covered with other urban furniture like manhole-covers, they are not sufficient for fine strip modelling. Figures 5.4(a) and 5.4(c) demonstrates insufficiency of long sides in strip modelling. In order to refine the reconstruction we aim at detecting the transversal sides and computing the vertices of each strip by intersecting long sides with transversal sides.

Our strategy for retrieving the transversal sides consists in defining search spaces at bottom and top of each initial 3D strip. The search areas are then projected in image space then candidates are retrieved using the edge map within the search areas.

In the case of dashed-lines, knowing the exact or at least the approximate length of strips (L_{exp}), a search area can be defined around initial model vertexes. It consists of a

```
Input:
$S$ : Set of lines-segments issued from detection step
$W$ : Strip's width
Output: $StripSet$: A set of 3D strips
Sort all $s_i \in S$ from left to right with respect to image space.
for $i = 0$ to $N_s$ do
    $LS$,$RS$ // Left and right strip's sides
    // $\overrightarrow{\nabla}(s_i)$ : Gradient of segment $s_i$
    // $Dist(s_i, s_k)$ : Distance between $s_i$ and $s_k$
    // $\varepsilon$ : Tolerance of strip width error
    if $\overrightarrow{\nabla}(s_i) \notin \{Z_1 \cup Z_4\}$ then
        continue
    end
    $k = i$
    while $k < N_s \wedge (Dist(s_i, s_k) < W + \varepsilon)$ do
        if $Dist(s_i, s_k) < \varepsilon \wedge \overrightarrow{\nabla}(s_k) \in \{Z_1 \cup Z_4\}$ then
            $LS \overset{grouping}{\longleftarrow} s_k$
        end
        else if $|Dist(s_i, s_k) - W| < \varepsilon \wedge \overrightarrow{\nabla}(s_k) \in \{Z_2 \cup Z_3\}$ then
            $RS \overset{grouping}{\longleftarrow} S_k$
        end
        $k++$
    end
    if $\exists\, LS \wedge \exists\, RS$ then
        $Strip_{new}(LS, RS)$
    end
    // Creation of a hypothetical strip's candidate
    if $\exists\, SripSet_{Last} \wedge (Strip_{new}$ overlaps $StripSet_{Last})$ then
        $Strip_{old} = StripSet_{Last}$
        $StripSet_{Last} = \underset{m \in \{old, new\}}{\operatorname{argmax}} \left(\| \overrightarrow{\nabla}(Strip_m^{LS}) \| + \| \overrightarrow{\nabla}(Strip_m^{RS}) \|\right)$
    end
    else
        $StripSet \leftarrow Strip_{new}$
    end
end
```

Algorithm 1: Zebra-crossing and dashed-line strip long side modelling algorithm.

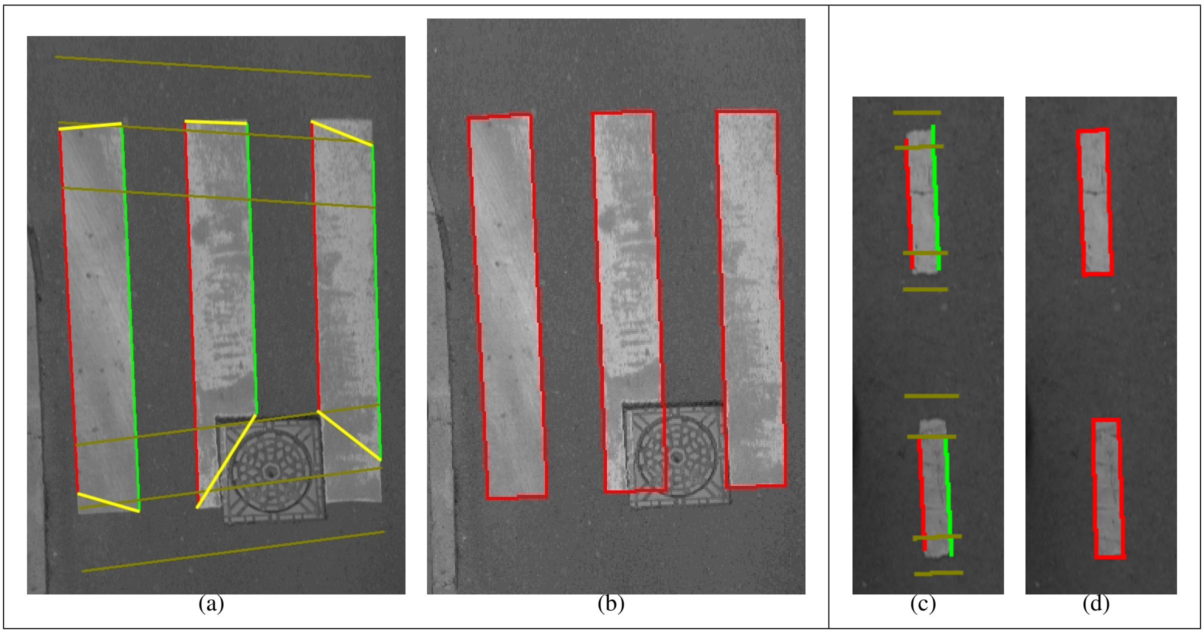


Figure 5.4: The problem of strip modelling only with long sides and fine reconstruction.

rectangular space between two line-segments that are vertical to long sides and shifted to $L_{exp}/4$ on each sides. In Figure 5.4(c) image projection of this area is demonstrated as two line-segments at each extremity of dashed-line strips.

Contrary to dashed-lines the length of zebra-crossing strips is unknown. So it cannot be used to define the search area. In order to define the search space, a global transversal side is calculated for all found strips by taking into account just the longest sides (up to 80% of the maximum length). It is then translated to a sufficiently large distance (40 cm) at each side to define the search area in 3D space. In Figure 5.4(a) image projection of global search area is depicted as light line-segments. The search space for each strip will be the area limited between its two long-sides and this global search area.

A Hough transformation is performed on the edges within the search area to detect the line-segments. In the case of zebra-crossing strip only transversal line-segment candidates with an orientation near the global approximate transversal side are retrieved. In the case of dashed-line strip, candidates are searched only in a direction near vertical to its long sides. The set of local maximum with Hough score higher than 80% of the highest maximum is accepted as transversal sides candidates. A gradient vector is computed then for each putative upward and downward transversal side ($\overrightarrow{\nabla}_{upward,downward}$). Figures 5.2 depict the candidates and their gradient vector for a zebra-crossing and a dashed-line strip. As seen in the Figures 5.5(a) and 5.5(b), some uninteresting line-segments have been found by Hough transformation on the zebra-crossing strip damaged by manhole-cover. In order to find the good pair of candidates we aim at finding the pair of candidates that are of inverse gradient direction and also maximise the gradient norm. The optimal pair of transversal sides is

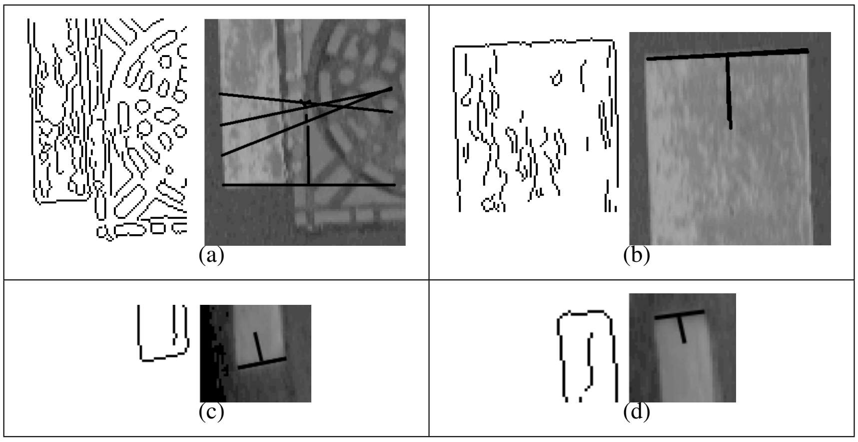


Figure 5.5: Transversal strip sides candidates for a zebra-crossing (top) and a dashed-line (bottom).

found by Equation 5.3.

$$(i,j) = \underset{b \in B, u \in U}{\operatorname{argmin}} (\overrightarrow{\nabla}_b . \overrightarrow{\nabla}_u) \tag{5.3}$$

b, u : Candidates for the downward and upward transversal strips,

B, U : Set of estimated line-segment with Hough transformation for the downward and the upward of a strip.

The final strip vertexes are then computed by intersecting long sides and transversal sides. They are then projected on the 3D plane of initially estimated strip and constitute the final 3D strip model (See Figure 5.7). The image projection of resulted strips are demonstrated in Figures 5.4(b) and 5.4(d). Figure 5.6 depicts the 3D final models of zebra-crossing and dashed-lines.

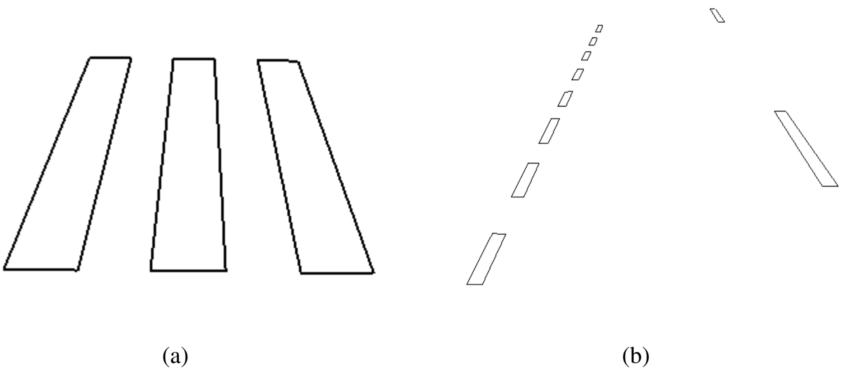


Figure 5.6: Final 3D reconstructed models for zebra-crossing and dashed-lines.

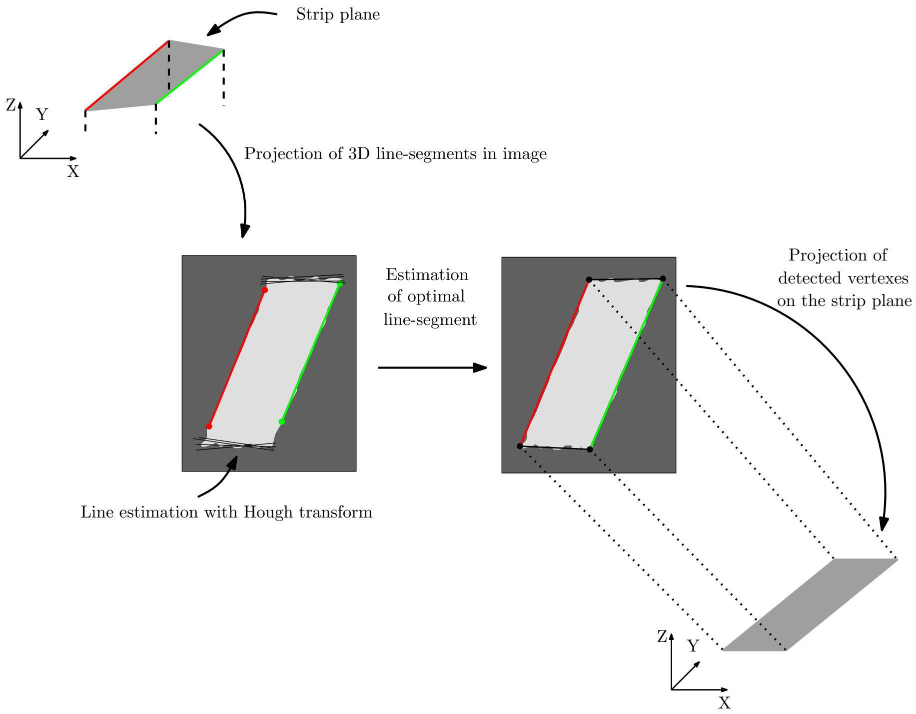


Figure 5.7: Strategy of strip's short sides modelling.

5.3 Line-segment grouping

Line-segment grouping is a crucial task in image-based object detection and reconstruction algorithms. Line segments are extracted by edge point detection in image space followed by a polygonisation step that approximates linked-edge with line-segments. Edge detection imperfection together with polygonisation approximation cause line fragmentation errors. The goal of grouping can be considered as extracting significant line segments by grouping appropriate fragmented ones.

In our specific application the polygonisation is performed on 3D linked-edge resulting from edge detection in a pair of stereoscopic images followed by an edge point matching. In other words the 3D linked-edge are segmented into 3D line-segments. In this case, grouping is a 3D problem.

3D line-segment grouping is a complex problem. However supposing that 3D segments are locally planar, the grouping task can transformed into a 2D problem. This is the case in our specific application because as explained in Chapter 4, we deal with road features that lie on road's approximate plane. In other words, regarding data structure of Figure 4.4, we deduce the grouping of 3D line-segments (S_q^{3D} and S_r^{3D}) from grouping of corresponding 2D line-segment (S_q^{2D} and S_r^{2D}).

We look for a 2D line-segment grouping algorithm that given a set of 2D line-segments nominates subsets as candidates of grouping. The grouping must take into account geometric properties of line-segments such as alignment and proximity.

Suppose that grouping algorithm propose to group n line-segments. Let us see how the data structure presented in Section 4.4 deals with grouping and how the grouped line is estimated. When two segments S_q^{2D} and S_r^{2D} corresponding to e_q and e_r elements are grouped a new element e_u is generated that contains the points belonging to both elements e_q and e_r. We denote $\mathcal{G}(e_q, e_r)$ the function that gather the points of e_q and e_r and compute corresponding 3D and 2D line-segments by regression methods (See the Equation 5.4).

$$e_u = \mathcal{G}(e_q, e_r) \equiv \begin{cases} X_u^{3D} & = X_q^{3D} \cup X_r^{3D} \\ X_u^{2D} & = X_q^{2D} \cup X_r^{2D} \\ S_u^{2D} & = \mathcal{R}^{2D}(X_u^{2D}) \\ S_u^{3D} & = \mathcal{R}^{3D}(X_u^{3D}) \end{cases} \tag{5.4}$$

where:

e_u is the element resulting by grouping.

$\mathcal{R}^{2D}, \mathcal{R}^{3D}$ are 2D and 3D regression functions.

This function has the following properties:

1. Associativity: $\mathcal{G}(e_q, e_r, e_t) = \mathcal{G}(e_q, \mathcal{G}(e_r, e_t)) = \mathcal{G}(\mathcal{G}(e_q, e_r), e_t)$
2. Identity: $\mathcal{G}(e_q) = e_q$
3. Commutativity: $\mathcal{G}(e_q, e_r) = \mathcal{G}(e_r, e_q)$

Given a set of locally planar linked-edges the grouping of elements is deduced from grouping of corresponding 2D line-segments. So from now on, we focus on the 2D line-segment grouping problem.

Section 5.3.1 presents the adopted model for 2D and 3D line. Section 5.3.2 explains our global line-segment grouping strategy. Sections 5.3.3–5.3.5 explain implementation of this approach. Section 5.3.6 explains estimation of 3D line-segment after grouping.

5.3.1 Line-segment estimation

As indicated in Equation 5.4, when n line-segments are grouped, grouping process piles-up their 2D and 3D points $((x,y,z)^{3D}, (x,y)^{2D})$. The $2D$ and $3D$ representations of grouped line-segment are then estimated by regressions on respectively $2D$ and $3D$ point sets. Conform to Deriche *et al.* [1992], we take benefit from two minimal representations of line in two and three dimensions. The regressions are performed using the following models.

2D line representation

We apply the minimal representation of $2D$ line indicated in Equation 5.5. Each line is expressed by two parameters θ, ρ (see Figure 5.8). In contrast to linear representation $(y = mx + d)$ this representation does not suffer from any degenerated configuration. Least squares method is used to fit this model to a point set [Deriche *et al.*, 1992]. After estimation of 2D line, line-segment is computed by projecting the two farthest points on the estimated line.

$$y\cos\theta - x\sin\theta = \rho \tag{5.5}$$

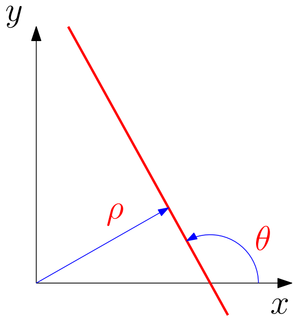


Figure 5.8: 2D representation of line

3D line representation

In order to estimate $3D$ line from a point set, we use the Plücker representation. A line is computed as the intersection of two planes. Three following configurations can be considered regarding line's direction:

$$\begin{cases} \text{Lines non orthogonal to the } Z\text{-axis} & X = aZ + p, \;\; Y = bZ + q, \\ \text{Lines non orthogonal to the } X\text{-axis} & Y = aX + p, \;\; Z = bX + q, \\ \text{Lines non orthogonal to the } Y\text{-axis} & X = aY + p, \;\; Z = bY + q, \end{cases} \tag{5.6}$$

The estimation is optimal in the configuration in which intersecting planes form a right angle. So we estimate line in each of these three conditions. Then, we accept only the configuration in which the intersection angle is closest to 90°.

As mentioned before, grouping of $3D$ line-segments are deduced from corresponding $2D$ line-segments. In this condition, presence of features that do not lie on road approximate plane such as cars and pedestrians causes some outliers to introduce into 3D point set used for line estimation. In order to be robust to these outliers, 3D line estimation is performed in two steps:

1. Fitting a 3D line by minimising L1 norm and detection of outliers,

2. Fitting a 3D line by minimising L2 norm using inliers.

The line-segment is then estimated by projecting the two delimiting 3D points on the estimated $3D$ line.

5.3.2 2D line-segment grouping algorithm

Many authors have investigated 2D line-segment grouping algorithms. Robles-Kelly and Hancock [2000] propose a method of grouping that computes first a grouping field as a function of geometric affinity between individual line segments and then use eigenstructure of the adjacency graph to find the main grouping. Hunt and Nelson [1996] propose a gradient based approach that takes benefit from gradient magnitude and orientation of individual segment lines to provide a coherent grouping. In the method presented in [Jang and Hong, 2002] candidates of grouping are provided first as a function of proximity and orientation between segments. Final line segments are then constrained to respect a global criteria in order to provide the longest line-segments. As the last mentioned method provides a globally optimised result, we chose it as a basis for our algorithm.

This algorithm is explained in the three following Sections. First, Section 5.3.3 presents our geometric criteria for grouping a pair of line-segments. Then, Section 5.3.4 explains how these criteria are iteratively used to generate grouping candidates. Finally, Section 5.3.5 explains how the algorithm deals with redundancy problem within grouping candidates and provides an unique optimum result.

5.3.3 Local constraints for grouping a pair of line-segments

Two segments can be candidates for grouping if they respect alignment and proximity constraints that are defined regarding Figure 5.9. Two segments S_i and S_j are aligned if the condition $\mathcal{A}(S_i,S_j)$ is satisfied:

$$\mathcal{A}(S_i,S_j) = \text{true} \quad \text{if} \quad \max(\theta_i,\theta_j) < \theta_{max} \tag{5.7}$$

where:

θ_i and θ_j are angles between each line-segment with the line-segment formed by mid-points of the two line-segments

θ_{max} is an angular tolerance to be fixed.

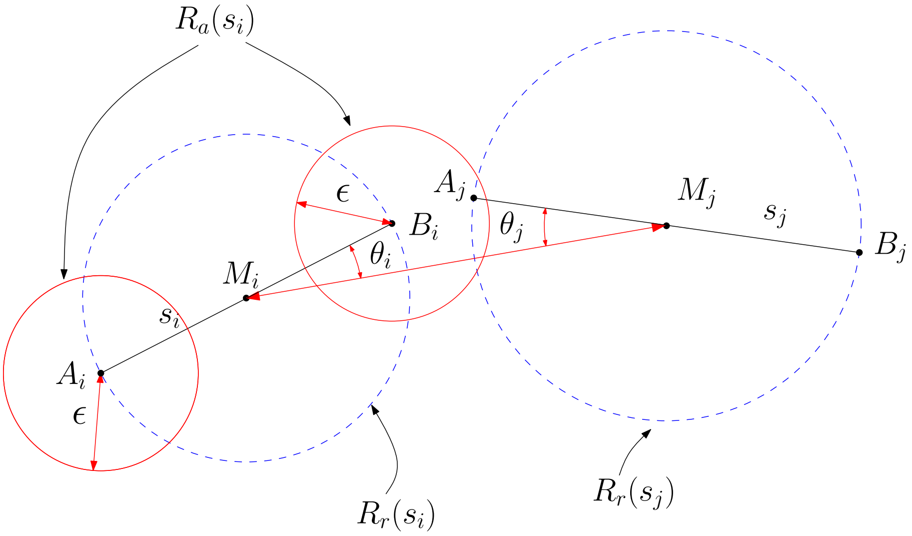


Figure 5.9: Line-segment grouping criteria

In order to define proximity condition between two line-segments ($\mathcal{P}(S_i,S_j)$), we must define attraction and repulsion regions (R_a,R_r) of a segment. According to Figure 5.9 the repulsion region (R_r) is defined by :

$$R_r(S_i) = C_r(M_i, \ell(S_i)/2) \tag{5.8}$$

where:

$C_r(O,r)$ is a disc with a centre of O and radius of r

M_i is the midpoint of S_i

$\ell(S_i)$ gives the length of the segment S_i

The attraction region (R_a) is defined by:

$$R_a(S_i) \ = \ C_r(A_i,\varepsilon) \ \cup \ C_r(B_i,\varepsilon) \tag{5.9}$$

where:

A_i and B_i are the two end points of S_i

ε is a distance tolerance to be fixed

The proximity condition between two segment S_i and S_j is denoted by $\mathcal{P}(S_i,S_j)$. It is satisfied if all the following conditions are verified.

$$\mathcal{P}(S_i,S_j) = \text{true} \qquad \text{if} \qquad \begin{cases} M_i & \notin R_r(S_j) \\ M_j & \notin R_r(S_i) \\ A_i & \in R_a(S_j) \oplus B_i \in R_a(S_j) \\ A_j & \in R_a(S_i) \oplus B_j \in R_a(S_i) \end{cases} \tag{5.10}$$

$\oplus$: exclusive or

This equation expresses that a pair of line-segments are in proximity to each other if:

- none of the line-segments' centers is situated in the repulsion region of other line-segment (two first conditions)
- one and only one extremity of each line-segment is situated in the attraction region of other line-segment (two last conditions).

The conditions presented in Equations 5.7 and 5.10 are necessary for grouping a pair of line-segments. Other conditions such as gradient direction can be considered and added to grouping conditions. Let us see in the following Section how these constraints are applied to group more than two line-segments iteratively.

5.3.4 Strategy for generating grouping candidates

The mentioned local constraints in Section 5.3.3 define the condition of grouping a pair of 2D line-segments. In order to provide grouping candidates , we aim at considering all the combinations of line-segments. Let's call S a set of N_s 2D line-segments, s_i the ith line-segment and a set of grouped line-segments a base-line.

The approach outlined in Algorithm 2 begins by considering a line-segment s_i as a base-line S_i^B. Then all the other line-segments ($S - \{s_i\}$) are tested for grouping with the so called base-line S_i^B regarding the mentioned proximity $\mathcal{P}$ and alignment $\mathcal{A}$ constraints (See Section 5.3.3). If a line-segment s_j satisfy the grouping conditions, it is grouped with the base-line using $\mathcal{G}$ function($S_{i_{cand}}^B = \mathcal{G}(s_j, S_i^B)$). This operation must be repeated until no line-segment group the base-line.

However deviation accumulation when line-segments are grouped iteratively can lead to a curve. In order to avoid this scenario the $S_{i_{cand}}^B$ is accepted only if its residuals of line fitting process (See Equation 5.4) are not greater than a threshold v_{max}. If grouping is accepted the base-line is updated ($S_i^B \leftarrow S_{i_{cand}}^B$).

After each update of the base-line, the so called operation is repeated from beginning of the line-segment list. This repetition is necessary because after grouping with s_j, parameters of base-line $S_{i_cand}^B$ are modified and for some line-segments s_k , $k < j$, the alignment and proximity conditions can become satisfied even if was not the case in previous iterations. So the iteration gives the chance of grouping to all line-segments independently to their order in line-segment set S. Moreover within the iterations a line-segment s_i can contribute in many base-lines $S_l^B, S_m^B, S_n^B, \ldots$. On the one hand this is interesting because it provides a large number of possible grouping without any sensibility to the ordering of line-segments, on the other hand it introduces some redundancy in the grouping process.

In each iteration i, when no more line-segment s_k is grouped with the base-line S_i^B, the index of all contributed line-segment I_i^B is registered and the algorithm continues by initializing S_{i+1}^B with s_{i+1}.

In order to reach final grouped line-segment set the algorithm must cope with the mentioned redundancy problem. It is briefly explained in Section 5.3.5.

```
Input: $S$: a set of $N_s$ line-segments
/* $I_i^B$:  a list of line-segment as candidate of grouping    */
/* $I^B$:  a list of $I_i^B$                                       */
Result: $I^B$
for $i = 0$ to $N_s$ do
    $S_i^B \leftarrow s_i$ // $S_i^B$ :  Baseline segment starting with $s_i$
    Register i in $I_i^B$ // Index list of contributed segment's in $S_i^B$
    Grouping $\leftarrow$ true
    while Grouping do
        Grouping $\leftarrow$ false
        for $j = 0$ to $N_s$ do
            if $(j \notin I_i^B) \wedge \mathcal{A}(s_j, S_i^B) \wedge \mathcal{P}(s_j, S_i^B)$ then
                $S_{i_{cand}}^B = \mathcal{G}(s_j, S_i^B)$
                if Max(Residuals of $S_{i_{cand}}^B$) $< v_{max}$ then
                    $S_i^B \leftarrow S_{i_{cand}}^B$
                    Register j in $I_i^B$
                    Grouping $\leftarrow$ true
                    break
                end
            end
        end
    end
    Register $I_i^B$ in $I^B$
end
```

Algorithm 2: Generation of candidate grouped line-segments.

5.3.5 Strategy for final grouping selection

This part of the algorithm applies the method proposed by Jang and Hong [2002] to select within redundant line-segments in a way to encourage constitution of longer segments. The method is outlined in Algorithm 3. It is based on three following principles:

- Each line-segment is associated to the longest group. So it is unregistered in any other groups.
- Within the resulting group only those conserving $T_r\%$ of initial length are accepted.
- All unaccepted groups are re-entered in Algorithm 2.

In order to perform these operations we define the two following functions:

$$\mathcal{L}(A) = \sum_{k \in A} \parallel s_k \parallel \qquad (5.11)$$

$$\mathcal{F}(A,i) = \begin{cases} \mathcal{L}(A) & \text{if } i \in A \\ 0 & \text{otherwise} \end{cases} \qquad (5.12)$$

where:

A : vector of line-segment's index

$\mathcal{L}(\mathcal{A})$: gives sum of line-segments' length of a list A

The index of the longest base-line containing line-segment s_i is computed using Equation 5.13. Then line-segment s_i is unregistered from all other base-lines.

$$m \; = \underset{j=0 \text{ to } N_s}{argmax}(\mathcal{F}(I^B_j, i)) \qquad (5.13)$$

After modifying base-lines by the unregistering operation, only the base-lines conserving T_r% of their initial length are accepted ($\mathcal{L}(T^B_j)/\mathcal{L}(I^B_j) \geqslant T_r$). Line-segments of the remaining base-lines are re-entered in the Algorithm 2.

```
Input: $I^B$ computed by Algorithm 2
Data: $T_r \in (0,1)$
Result: validation/rejection of grouping candidates
/* $\mathcal{L}$:  defined in Equation 5.11                    */
/* $\mathcal{F}$:  defined in Equation 5.12                    */
$T^B \leftarrow I^B$
for $i = 0$ to $N_s$ do
    $m = \underset{j=0 \text{ to } N_s}{argmax}(\mathcal{F}(I^B_j, i))$
    for $j = 0$ to $N_s$ do
        if $j \neq m$ then
            unregister $i$ from $T^B_j$
        end
    end
end
for $i = 0$ to $N_s$ do
    if $\frac{\mathcal{L}(T^B_j)}{\mathcal{L}(I^B_j)} \geqslant T_r$ then
        accept grouping $T^B_j$ and estimate grouped line-segment by
        regression
    end
    else
        $\forall i \in T^B_j \; S_{ungrouped} \leftarrow s_i$
    end
end
Input $S_{ungrouped}$ in Algorithm 2
```

Algorithm 3: According to [Jang and Hong, 2002] : selection of final groupings between the candidates provided by Algorithm 2.

The example of Figure 5.10 demonstrates different steps of our 2D line-segment grouping. Let us suppose line-segments of Figure 5.10(a) enter in grouping process and let us suppose all line-segments are of length 1. Grouping candidates are constituted by Algorithm 2. Figure 5.10(b) shows the grouping candidates. As mentioned before these candidates suffer from redundancy. For example s_2 is shared between grouping candidates $\mathcal{G}_1$ and $\mathcal{G}_2$ and s_1 is shared between $\mathcal{G}_2$ and $\mathcal{G}_3$. Within all the groups ($\mathcal{G}_1$,$\mathcal{G}_2$,...) a line-segment s_i is associated only to the longest group (in this case $\mathcal{G}_2$ with length of 8). So the $\mathcal{G}_1$ and $\mathcal{G}_3$ lose respectively s_2 and s_1. After eventual loss of some line-segments only the groups that conserve T_r% of their initial length ($T_r \gtrapprox 80\%$) are accepted as final groups and the other candidates are rejected. So $\mathcal{G}_1$ and $\mathcal{G}_3$ are rejected because they conserve only 66% of their initial length (See Figure 5.10(c)). All ungrouped line-segments are then re-inputted in the Algorithm 2 and this operation is repeated until no more group is rejected or is formed.

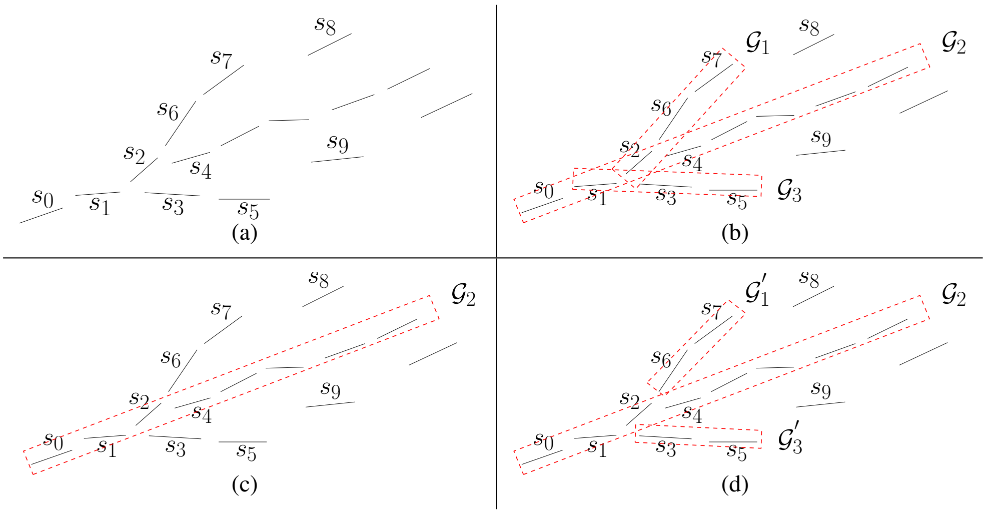


Figure 5.10: According to [Jang and Hong, 2002] (a) A set of line-segments that enters the grouping algorithm, (b) Initial grouping candidates $\mathcal{G}_1$ and $\mathcal{G}_2$ share s_2, $\mathcal{G}_2$ and $\mathcal{G}_3$ share s_1. Some line-segments such as s_8,s_9,... are isolated. (c) $\mathcal{G}_2$ is accepted, the other unisolated line-segments (s_6,s_7,s_3,s_5) are re-entered in the algorithm. (d) The final grouped line-segment without any shared line-segment.

5.3.6 Advantage of our line-segment grouping method

Most of the line-grouping algorithms are developed to group line-segments in image space. Often after finding line-segments that contribute in a group, the final grouped line-segment is computed by a regression applied on end-points and weighted by line-segment length of all the contributing line-segments. In our application the grouping is not performed in image space but in object space. So for the segments of similar length, the uncertainty is raised as distance from the stereo-base increases (See the Equation 3.1). In the same way the along track GSD increases by distance from camera. So line-segments of similar length are constituted of less points as the distance from stereo-pair increases.

So registering initial points of each line-segment in the data structure and performing the regression on contributing points instead of contributing line-segments weighted by length can considerably improve the results.

In fact in our grouping method line-segments of similar distance from stereo-base are weighted by their length that depends directly on the number of points. In contrast, line-segments of same length are less weighted when they are farther from stereo-base because they have less points.

5.4 Conclusion

This chapter presents our strategy for modelling zebra-crossing's and dashed-lines' strips using a parallelogram shape. The candidate line-segments provided by the detection step are analysed to see if they correspond to a parallelogram shape. Each candidate is rejected if it does not satisfy the model. The candidates satisfying the model are geometrically improved by finding the optimal transversal sides. Each strip is supposed to be planar, but there is not any planar hypothesis between strips. It means that a zebra-crossing or a dashed-line can be correctly modelled even if their strips are not all on a plane. This is particularly the case for zebra-crossings that follow the road transversal curvature.

Part III

Results and evaluations

Chapter 6

Evaluation

6.1 Introduction

Quality evaluation of automatic feature extraction approaches is an important issue due to several reasons. First, it provides useful information about limitations of an approach for object extraction and can help to focus future researches. Second, it helps us to compare different proposed algorithms and to convince a user that a specific approach can reach a needed quality in an operational project. Even if an automatic approach does not provide a requested quality directly, it may be interesting to evaluate the needed manual effort for improving the derived results' quality up to a level that satisfies the user. In cartographic applications two aspects of quality are often taken into account: geometrical accuracy and correctness [Schuster and Weidner, 2003].

Correctness evaluation consists often to determine if a detected feature exists in the reference data and symmetrically if a feature that exists in reference data is detected by the algorithm. It is often expressed by measures such as detection rate, false alarm rate and quality rate (Defined in Section 6.4).

Geometrical accuracy expresses the geometrical similarity between a detected feature and its corresponding feature in reference data set. *RMS*[1] difference is consistently applied in this purpose.

In the context of road extraction evaluation, most efforts have concerned aerial-based approaches. They are often based on comparison of the automatically derived data set against a manually compiled high quality reference model. Heipke *et al.* [1997] propose to match the two data sets by defining a buffer around the reference data set and to compute quality measures that consist of correctness and geometrical accuracy. McKeown *et al.* [2000] propose to transform the evaluation of feature extraction into the evaluation of a classification by discretising the space either in 2D pixels or 3D voxels. In this manner a mask of the road is produced by a reference data set and its overlap with the detected road features is analysed for estimating some correctness measures. Recently a very interesting

---

[1] *RMS*: Root Mean Square

effort have been made under the umbrella of EuroSDR[2] in order to compare the existing methods for automatic road extractions [Mayer *et al.*, 2006]. Within this project, completeness and correctness of the extracted data are presented. *RMS* is computed in image space and expressed in pixels.

In contrast to aerial-based approaches, there are only a few reference focusing on quantitative evaluation of road features extracted from ground-based systems. However some efforts have been made by Tao *et al.* [1998] and Cheng *et al.* [2007] for evaluation of road marks that are extracted from stereopairs of ground-based MMSs.

In the work presented by Tao *et al.* [1998], the evaluation consists in the estimation of geometrical accuracy of reconstructed road's centerlines. The relative geometrical accuracy is measured by back-projecting the extracted centerlines in image space and computing their *RMS* differences with image's edges. Then the absolute accuracy is evaluated by overlaying and comparing the extracted centerlines on an existing digital roadmap. However roadmaps suffer often from generalisation. So the difference with roadmaps does not express fine imprecision.

Cheng *et al.* [2007] propose to compare the reconstructed roadmark set with a manually compiled set from the same ground-based images. Detection rate and false detection rate are then computed over a large set of images regarding length of extracted roadmarks.

6.2 Our evaluation criteria

Our evaluation method differ from the mentioned approaches in the previous section in two following ways:

- Completeness evaluation is often performed by comparing length or surface between extracted and reference data. Since our algorithm extract strips, we prefer to express completeness evaluation by the number of extracted objects disregarding their length and surface.
- Geometric accuracy is often expressed in image space (in pixel). However image measurements do not express directly accuracy in object space. 3D accuracy depends on geometric configuration of stereo or multi images. We express geometric accuracy of our algorithm in object space. It is performed by comparing 3D reconstructed data with a 3D reference data that is measured using an independent method (topography).

We present first in Section 6.3 some qualitative results. Section 6.4 describes our criterion for correctness evaluation and present results on a set of 150 successive stereo-pairs. Section 6.5 present our 3D accuracy evaluation method and analyse the obtained results.

6.3 Qualitative results

This section describes some advantages of our method when dealing with real urban areas. Robustness of our algorithm to partial occlusions and damaged or eroded roadmarks is

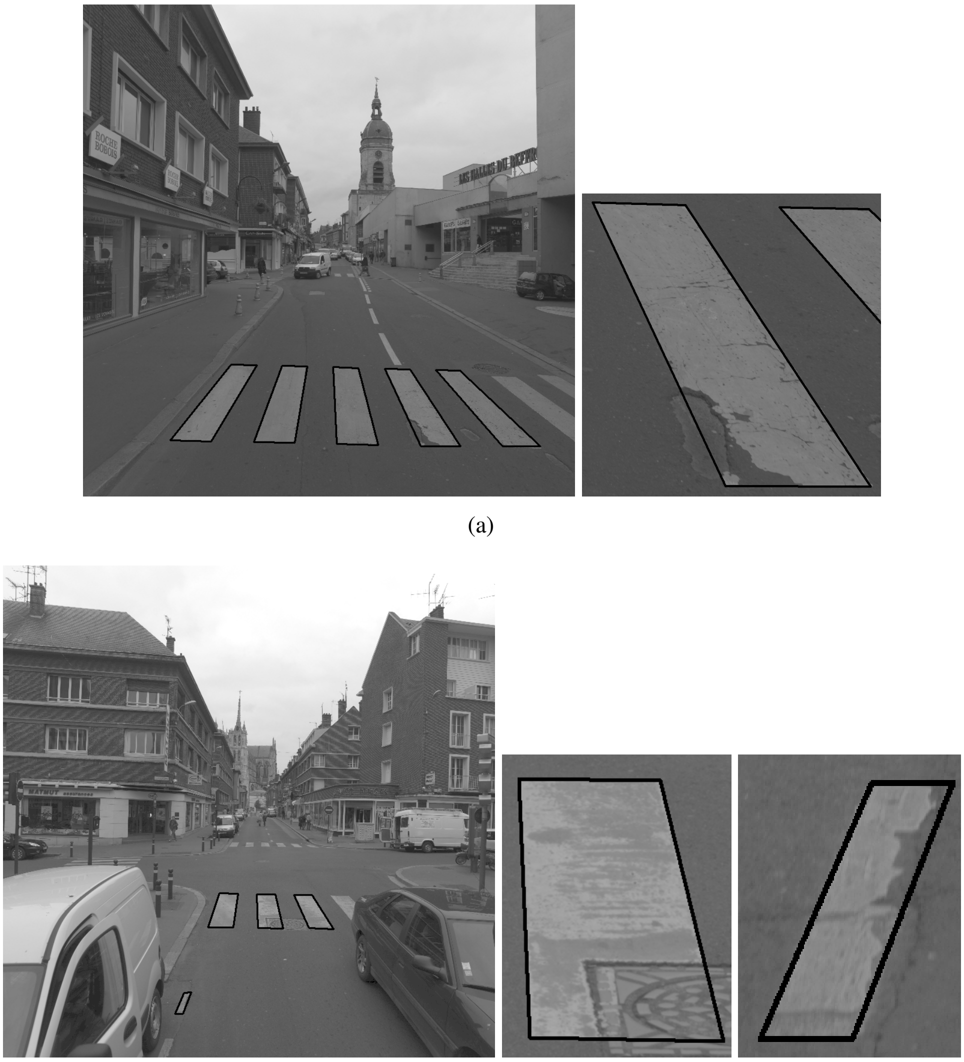
---

[2] EroSDR: European Spatial Data Research

illustrated by several examples.

6.3.1 Robustness to damaged and eroded roadmarks

Even if roadmarks are painted with strict specifications, due to weather and traffic conditions they are often partially eroded. They are also sometimes covered with other urban features such as manhole-covers. In consequence they do not respect their theoretical shapes. In order to be as complete as possible an algorithm must be robust to these conditions. Modelling of strips with their sides instead of corners robustifies considerably the algorithm to these difficulties. This is illustrated in Figure 6.1.



(a)

(b)

Figure 6.1: Robusteness of the algorithm to damaged, eroded and covered roadmarks. Left: re-projection of reconstructed roadmark in left image, Right: zoom on roadmarks.

6.3.2 Independency from repetition of strips

In urban areas, cars often occlude the roadmarks. In this scenario the detection algorithms which are based on alignment and repetition of strips fail. Our algorithm is based on 3D parallelogram shape recognition so it is independent to repetition. Figures 6.1(b) and 6.2 show two detected individual dashed-line strips.

Figure 6.2: Independency of strip recognition to repetition.

6.3.3 Robustness to partial occlusions

In real urban areas roadmarks can be partially occluded by cars or pedestrians. Since our algorithm is based on detection of long and transversal sides, strips can be reconstructed even if some parts of line-segments corresponding to its sides are occluded. An example is shown in Figure 6.3. However the algorithm fails to perform correct modelling when at least one of the sides is completely occluded (See Figure 6.4). In this case the detected strips will not be validated by the modelling step.

Figure 6.3: Extraction under partial occlusion of strip's sides.

Figure 6.4: Extraction under complete occlusion of transversal sides.

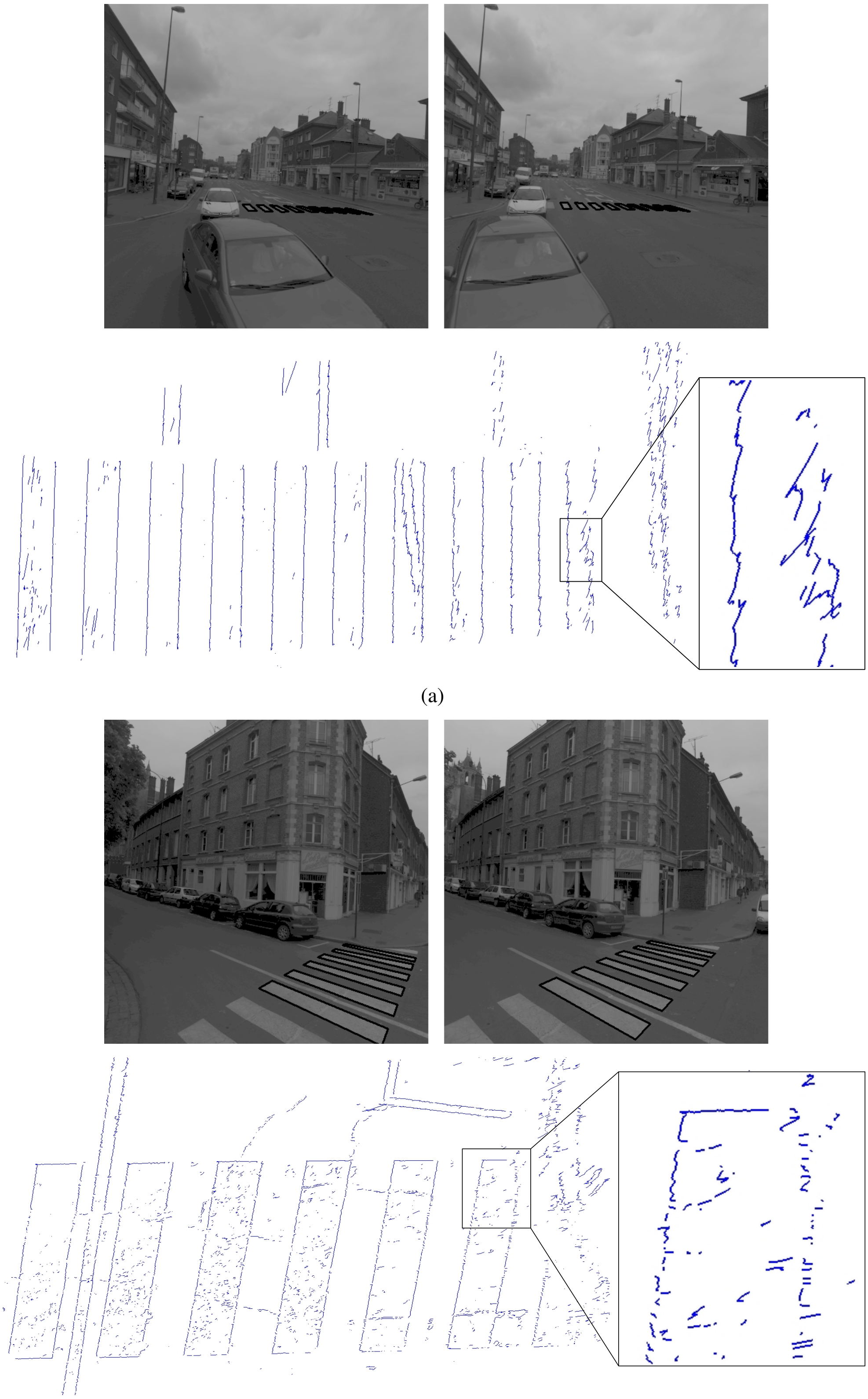
6.3.4 Geometric limitation of method

As mentioned in Chapter 3, 3D reconstruction of linked-edges is influenced by relative pose of stereopair in relation to roadmark object. This is due to two following critical configurations:

- Long distance of roadmark from stereopair
  Since the stereo base is fixed to 1.20 *m*, when distance of object from vision system (H) increases the ratio B/H decreases. This effect causes higher inacuracies in 3D estimation (see Equation 3.2). In depths higher than 20 *m* ($B/H = 0.06$), 3D linked-edges become too noisy. Figure 6.5(a) illustrates an example of zebra-crossing situated at an average distance of 20 *m* from stereopair. As can be seen on this Figure the strip located in extreme right-side of the zebra-crossing is too noisy when reconstructed in 3D. This is the reason why this strip is not detected by extraction algorithm. This effect exists in all strips but it is more significant for extreme right-side one. This is because of smaller intersection angle of photogrammetric triangulation. Despite the mentioned detection problem the 3D accuracy of reconstructed strips is about 20 *mm* (see Table 6.2).

- Orientation of roadmark in relation to vision axis
  As mentioned in Chapter 4 strips are detected thanks to their long sides. This is because long sides are often nearly vertical to *epipolar* lines whereas short sides are nearly parallel to *epipolar* lines. Figure 6.5(b) illustrates a zebra-crossing that form a 45° angle with the vision axis. Long sides are so near to *epipolar* lines in image space. Resulting inaccuracy is illustrated by a zoom on 3D linked-edges. However, thanks to small parts that are correctly reconstructed, the algorithm is still finding the strips. Let's stress that an orientation larger than 45° in object space in relation to vision axis can cause the extraction algorithm to fail.

Both examples illustrated in Figure 6.5 present critical configurations in which our algorithm is still efficient. However regarding the present noise within 3D linked-edges, these two configuration can be considered as efficiency limit of our algorithm in terms of distance and orientation of roadmark object in relation to vision system. Higher level of noise in 3D linked-edges can lead to inefficiency of line-segment estimation and so strip detection.



(a)

(b)

Figure 6.5: Two critical configurations of roadmark strips in relation to stereo rig. a) zebra-crossing at distance of 20 *m* ($B/H = 0.06$), b) a zebra-crossing that is oriented to 45° in relation to vision axis. In this case long sides of strips are nearly parallel to *epipolar* lines and cause some problems in matching step.

6.4 Correctness evaluation

In order to evaluate the correctness of our algorithm, it has been applied to 150 consecutive stereopairs provided by the *Stereopolis* system from the city centre of Amiens in France under normal traffic conditions[3]. The length of trajectory is about 1 Km. It is shown in the Figure 6.6. For each individual stereopair, the output consists in strips of different type of dashed-lines or zebra-crossing. Let us remind here that reconstruction process is completely independent within consecutive stereopairs. There is no merging of the results within the image sequence. So the correctness evaluation is performed in relation to each stereopair. In order to do that, for each stereo-pair and each roadmark object of interest, the number of true-positives (TP_i), false-positives (FP_i) and strips that are visible in stereo which could be reconstructed (S_i) are counted. Let us give a precise definition of these terms:

Figure 6.6: Test trajectory at the city center of Amiens in France (1 *km*)

- Roadmark object of interest: Strip of zebra-crossing or one dashed-line type such as $T^{'}0$,$T^{'}1$, etc. (see Section 2.1).
- True-positive (TP_i): An extracted object from stereopair number i that corresponds to roadmark object of interest.
- False-positive (FP_i): An extracted object from stereopair number i that does not correspond to any roadmark object of interest.
- S_i: Total number of roadmark objects of interest that are visible in both left and right images of stereopair number i. The long sides of these strips can be partially occluded.

The so called values are counted manually in each stereopair. They give useful information about the quality of extraction algorithm in each stereopair. However the number of strips

---

[3] The specifications of the *Stereopolis* system and problem of normal traffic condition are already discussed in Chapter 1.

in each stereopair may not be high enough to be statistically significant. In order to express the performance of the algorithm over a statistically large enough set and under different perturbing conditions (traffic,pedestrians, etc.) the three following measures are defined over whole stereopairs.

- **Detection rate:** expresses the percentage of correctly extracted roadmarks. The optimum value for the detection rate is 1.

$$\rho_d = \frac{\Sigma TP_i}{\Sigma S_i} \in [0,1] \tag{6.1}$$

- **False alarm rate:** expresses the percentage of detection's errors. The optimum value for the False alarm rate is 0.

$$\rho_f = \frac{\Sigma FP_i}{\Sigma S_i} \in [0,\infty] \tag{6.2}$$

- **Quality factor:** this measure is more useful and expresses the "goodness " of detection result. It depends on both TP and FP. The optimum value for the Quality factor is 1.

$$\rho_q = \frac{\Sigma TP_i}{\Sigma S_i + \Sigma FP_i} \in [0,1] \tag{6.3}$$

Table 6.1 expresses these three measures for different types of roadmarks.

Type	Size (cm)	S	TP	FP	ρ_d	ρ_f	ρ_q
			Dashed-lines				
$T^{'}0$	10×50	485	416	20	86%	4%	82%
$T^{'}1$	15×150	51	50	12	98%	23%	79%
$T\,3$	15×300	30	30	2	100%	7%	93%
$T\,2$	22×300	6	6	27	100%	450%	18%
			Zebra-crossing				
—	$50 \times X$	359	329	8	92%	2%	90 %

Table 6.1: Completeness evaluation of roadmark reconstruction.

It can be seen that detection rate is higher than 86%. There are three main reasons for detection problems:

1. Problem in edge detection: This is always due to significant erosion of roadmarks (caused by car passages). Such effect eroded roadmarks are illustrated in Figure 6.7. In this case strip's borders are not detected and so roadmark detection algorithm fails.

2. Significant local textures: Even if the algorithm is designed to be relatively robust to this effect, it is still sensible to significant local textures. It is specially the case for the detection problem related to small roadmarks such as $T^{'}0$.

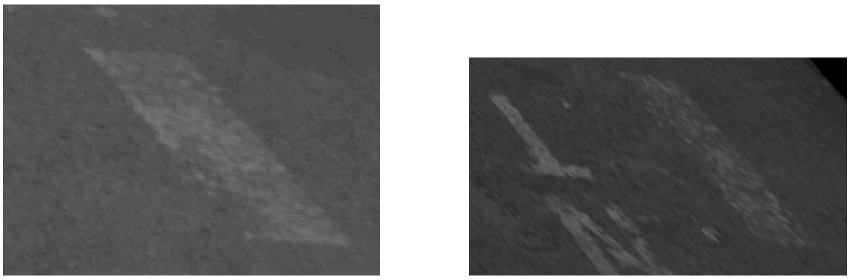


Figure 6.7: In the case of significant erosion, edge detection fails and so roadmark detection.

Last column of Table 6.1 expresses the quality of the algorithm for extraction of each roadmark object. It can be seen that the quality is higher than 80% for zebra-crossing and for dashed-lines' strips of type $T^{'}0$, $T3$. In contrast the quality is decreasing for strips of type $T^{'}1$ and $T\ 2$. Regarding their ρ_d, the algorithm extracted successfully the strips that were present in the images. However their high ρ_f denotes that the algorithm extracts also strips that do not exist in reality. The high false alarm rate of these two strips' type is due to presence of other objects that verify perfectly the geometric constraints used in detection and modelling steps. Figure 6.8(a) and 6.8(b) show an example of this scenario. Exceptionally in this example two strips are detected (See Figure 6.8(a)) and they are closer to each other than the specified distance by the standards (See Section 2.1.2). So they can be filtered. This, however is not often the case. Figures 6.8(c) – 6.8(f) represent individual false strips extracted by the algorithm. In these cases false alarms can no more be filtered by taking into account strips interaction. By interaction, we mean the geometric constraints between successive strips of the same dashed-line.

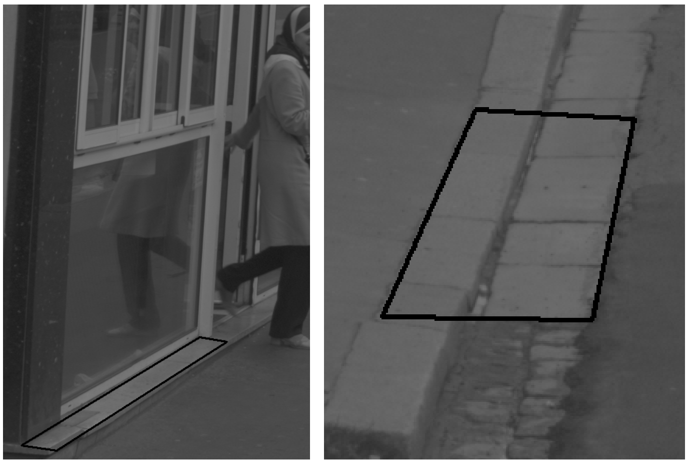
Let us recall that the presented results are obtained using individual stereo-pairs without merging within intra-date stereo-pairs. This is because absolute pose of stereo-pairs are not known precisely enough. In term of completeness, merging intra-date results is interesting because:

- Each strip is present in more than one stereo-pair. This increases the probability of reconstruction. Indeed if a strip is not reconstructed in one stereo-pair, it can be reconstructed from the next or the previous stereo-pair. This can increase significantly the detection rate.

- After merging reconstructed strips from successive stereo-pairs, on can filter out false alarms by testing interaction between strips. This can reduce the rate of false alarms.

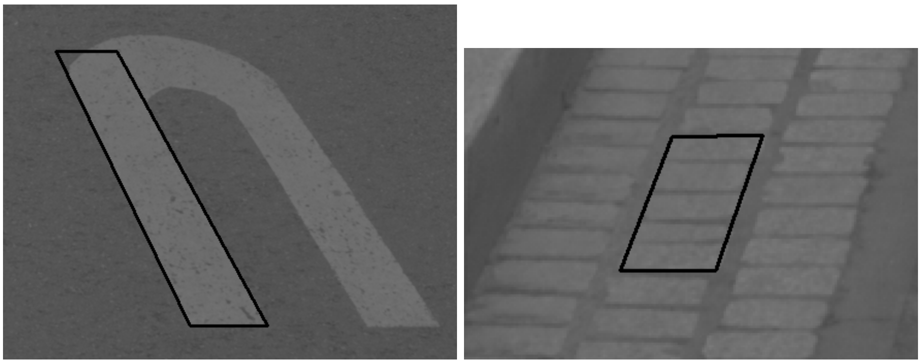
(a) False alarm of type $T^{'}0$

(b) Zoom on false alarm of type $T^{'}0$



(c) False alarm of type T 3

(d) False alarm of zebra-crossing's strip



(e) False alarm of type $T^{'}1$

(f) False alarm of type $T^{'}0$

Figure 6.8: Some examples of what is considered as false alarm in extraction of parallelogram shape strips, the extracted objects are constituted of parallelogram forms in which parallel sides have inverse gradient directions.

6.5 Geometric accuracy evaluation

In order to evaluate the accuracy of our reconstruction algorithm, we aim at comparing the reconstructed objects to the same objects measured by an independent method. For this purpose, a classical topographic operation is planned in the area in which the stereopairs are acquired. In this operation the four vertices of zebra-crossings' strips are measured with millimetre precision. Within the surveyed zebra-crossings ten are seen in stereopairs provided by *Stereopolis*. The so called zebra-crossings consist of 73 strips that are seen in 29 stereopairs (one zebra-crossing can be seen in more than one stereopair). The algorithm is applied to the so called stereopairs and the 3D coordinates of extracted strips are registered.

Evaluation Protocol

As Explained in Section 2.2.4, the extrinsic calibrations of stereopair provided by *Stereopolis* are unknown. The reconstructed roadmarks are so expressed in the 3D relative system of acquisition (origin and orientations are given by the left camera). This is the reason why, difference between reconstructed and reference coordinates can not be used as an error vector. Even if the stereopair could be georeferenced by matching the reconstructed and the reference strips, it is avoided because the georeferencing errors can disturb evaluation measures. We perform a relative comparison by computing the difference of specific distances within a zebra-crossing. Figure 6.9 demonstrates distance measurements applied on each strip of a zebra-crossing. The distances $C_{1,2}$,$T_{1,2}$ and d express the shape of strip. The distances $W_{1,2}$ and $S_{1,2}$ denote the interaction of strip with its neighbouring strip. We define in the Equations 6.4 – 6.8 five quality measures for each strip. Lets remark that the for each strip the interaction measures ($\sigma_{S_{1,2}}$ and $\sigma_{W_{1,2}}$) are computed in relation to its successive strip. So it is not computed for last strip of each zebra-crossing (strips are ordered from left to right in relation to image space).

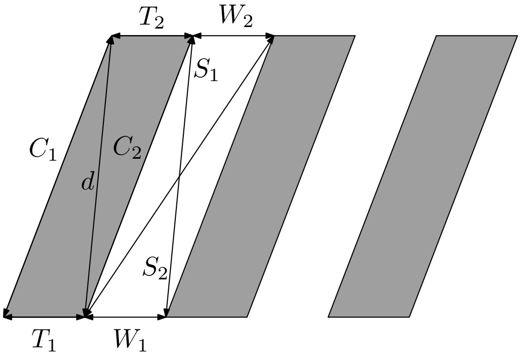


Figure 6.9: Distances used for comparing reconstructed and reference zebra-crossings.

$$\sigma_{C_{1,2}} = C^{ref}_{1,2} - C^{rec}_{1,2}, \qquad \hat{\sigma_C} = \frac{|\sigma_{C_1}| + |\sigma_{C_2}|}{2} \tag{6.4}$$

$$\sigma_{T_{1,2}} = T^{ref}_{1,2} - T^{rec}_{1,2}, \qquad \hat{\sigma_T} = \frac{|\sigma_{T_1}| + |\sigma_{T_2}|}{2} \tag{6.5}$$

$$\sigma_d = d^{ref} - d^{rec}, \qquad \hat{\sigma_d} = |\sigma_d| \tag{6.6}$$

$$\sigma_{W_{1,2}} = W^{ref}_{1,2} - W^{rec}_{1,2}, \qquad \hat{\sigma_W} = \frac{|\sigma_{W_1}| + |\sigma_{W_2}|}{2} \tag{6.7}$$

$$\sigma_{S_{1,2}} = S^{ref}_{1,2} - S^{rec}_{1,2}, \qquad \hat{\sigma_S} = \frac{|\sigma_{S_1}| + |\sigma_{S_2}|}{2} \tag{6.8}$$

Results of precision evaluation

The quality measures are computed for 29 zebra-crossings reconstructed from 29 stereopairs. They constitute 138 strips. The Figures A.1 – A.4 in Appendix A show the reprojection of the reconstructed zebra-crossings by our algorithm in corresponding stereopairs. Table A.1 in Appendix A represents the detailed result on each strip of the so called zebra-crossings. Figure 6.10 demonstrates the histogram of $\sigma_{C_{1,2}}$ on all 138 strips. The error distribution seems to be a mixture of two normal distribution around 0 *cm* and 20 *cm*. The normal distribution around 20 *cm* is suspected to be a bias in reconstruction of some strips.

Table 6.2 summarises the quality measures for each zebra-crossing by averaging the values of corresponding strips (biased zebra-crossings are marked in red). As can be observed in this Table large differences correspond to two reference zebra-crossings H and J. They correspond to the zebra-crossings of Figures A.3 and A.4. So the large errors of difference histogram (See Figure 6.10) are due to only two instances of zebra-crossing.

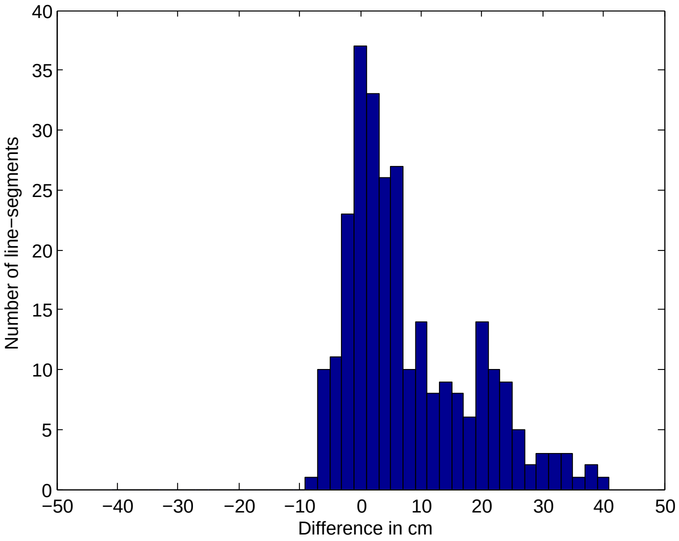


Figure 6.10: Histogram of $\sigma_{C_{1,2}}$ corresponding to 138 reconstructed strips.

Reference zebra-crossing's label	Stereopair's number	number of strips	Distance in *m*	$\hat{\sigma}_C$	$\hat{\sigma}_T$	$\hat{\sigma}_d$	$\hat{\sigma}_W$	$\hat{\sigma}_S$
				in *mm*				
A	71	5	7.6	30	10	18	10	22
	252	4	7.2	24	10	17	13	22
	253	5	11.2	53	12	34	15	30
B	115	4	11.1	40	10	44	6	26
C	117	2	9.4	10	13	3	0	0
D	153	3	8.1	37	13	40	7	28
E	157	4	6.6	25	8	21	5	28
	158	5	12.1	26	10	20	5	33
F	164	2	6.2	14	9	12	14	24
G	165	4	13.1	26	6	25	6	23
	166	8	18.4	22	15	27	7	14
H	236	5	6	16	11	8	8	17
	244	6	15.1	169	29	141	25	164
	246	3	14.2	148	18	139	27	146
	247	6	15	157	26	136	24	153
	249	6	15.6	182	33	153	28	178
	250	6	17.7	154	35	110	27	153
I	94	4	7.1	38	5	32	7	37
	95	6	12.3	32	12	36	15	35
	314	3	6.6	39	7	46	5	36
	315	3	8.2	34	5	36	12	28
	316	4	8.2	19	5	22	8	15
	317	3	10	75	9	86	5	66
	319	4	10	45	8	50	7	46
	320	4	9.9	45	11	53	6	49
	321	5	10	70	14	73	12	61
	322	4	10.7	42	9	48	8	42
J	676	4	7	131	13	116	8	127
	677	8	12.25	232	23	214	20	227
	678	8	17.3	329	31	299	28	314

Table 6.2: Difference of reconstructed zebra-crossings with the reference ones. Large differences are marked in red.

What is the reason of large errors ?

As mentioned before large errors correspond to the reference zebra-crossings H and J. Let us study these two cases separately.

Bias analysis on reference zebra-crossing H:

Regarding Table 6.2, the mean error on long sides of zebra-crossing H is: $\hat{\sigma_C} \approx 170mm$. In order to control the coherence of reconstructed model with stereopairs, the re-projection of the model in image space is studied.

Figure 6.11 demonstrates the re-projection of reconstructed zebra-crossing from stereopair number 244 in corresponding distortion-free left image. As observed in this Figure the re-projected reconstructed line-segments are very close to the long sides of strip (better than 1 pixel). However due to the fact that the strip is not planar, strip's vertices are approached only with 2 pixel accuracy. This is also the case in the right image.

According to Equation 3.2, at distance of this strip from stereo-base (16 *m*) and baseline of 1.2 *m*, with focal length equivalent to 3200 pixels a 2 pixel error in image measurement ($\sigma_\rho = 2$ pixels) cause an error of 130 *mm* in 3D reconstruction. So it can cause an error of $130 * \sqrt{2} = 183$ *mm* in 3D length measurements. This is confirmed with the difference that our algorithm reaches for the strip. Considering distance from object this argument is plausible for the stereopairs number 246, 247, 249 and 250.

To conclude, the error observed on reference zebra-crossing H can be considered as a modelling error. Indeed planar model for strips cause 3D errors when one strip is not planar in reality. The farther the strip pattern is from planar model, the larger is 3D error.

Bias analysis on reference zebra-crossing J:

The image re-projection of reconstructed zebra-crossings in stereopairs number 676, 677 and 678 are coherent with images edges (differences are smaller than 1 pixel). The applied relative orientation and distance between left and right images are suspected to be the reason. In order to verify the rigidity of stereo-rig, a bundle adjustment is performed on stereopair number 677. The processes integrates the manually measured tie points and also control points. We use as control points the J reference zebra-crossing's strips coordinates. All the tie and control points are pointed with subpixel accuracy. However the adjustment results show significant 3D residuals on 2 strips. We believe that the zebra-crossing was repainted between our imaging and topographic operations. As measurements are performed six month after imaging operation it is possible.

If our hypothesis is true, our reconstruction can be correct. Indeed the incoherence between reconstructed zebra-crossing from stereopairs number 676, 677 and 678 can be explained by increasing object distance (from 7 *m* to 17 *m*) and is not in contradiction with the correctness of our reconstruction.

General error analysis on unbiased reconstruction results

The biased results are discussed and justified in the previous paragraphs. Let us analyse the errors within the remained unbiased zebra-crossings. Figure 6.12 demonstrates the histograms of error measurements that are defined by Equations 6.4 – 6.8 (See also Figure 6.9).

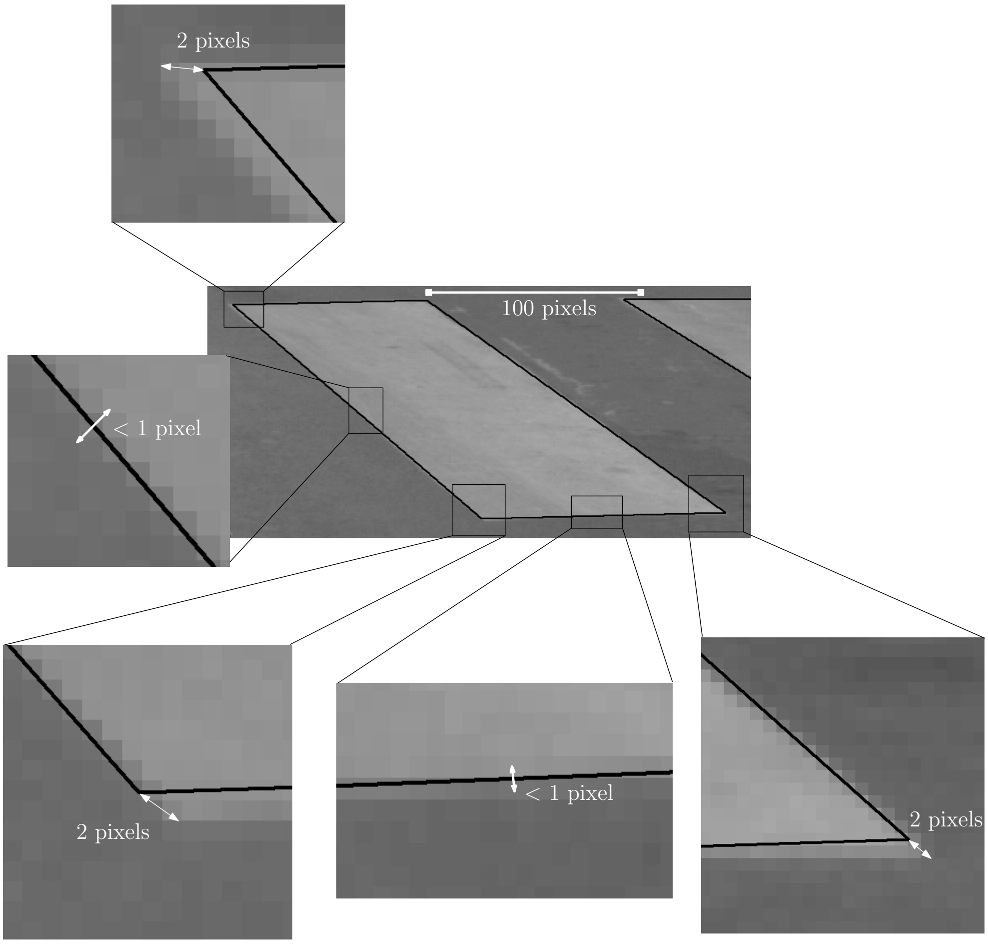
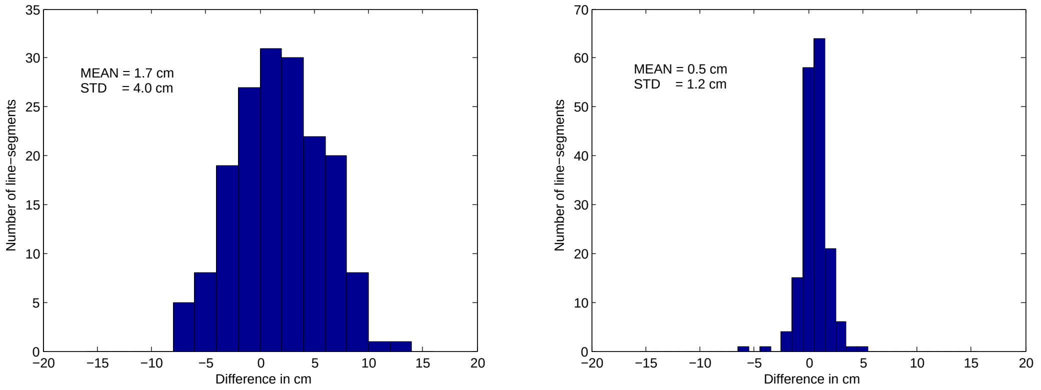


Figure 6.11: The image re-projection of third strip (from left to right) of zebra-crossing number 244. The transversal sides are estimated as a linear regression on all corresponding edge points. In the case of non planar strip image residual are important (> 2 pixels) at the strip vertexes even if in the center of transversal side residuals are smaller than 1 pixel.

It is observed that errors are larger in the direction of long sides than in transversal sides. It can be explained by the fact that long sides are nearly aligned with Z axis of the cameras (Depth direction). This direction is often called along track direction and both other orthogonal directions, across track ones. In photogrammetric intersection the across track accuracy ($\sigma_{x,y}$) and along track one (σ_Z) are related by Equation 6.9. In our specific case D (Depth) being relatively higher than B, uncertainty of along track direction is higher than across track ones.

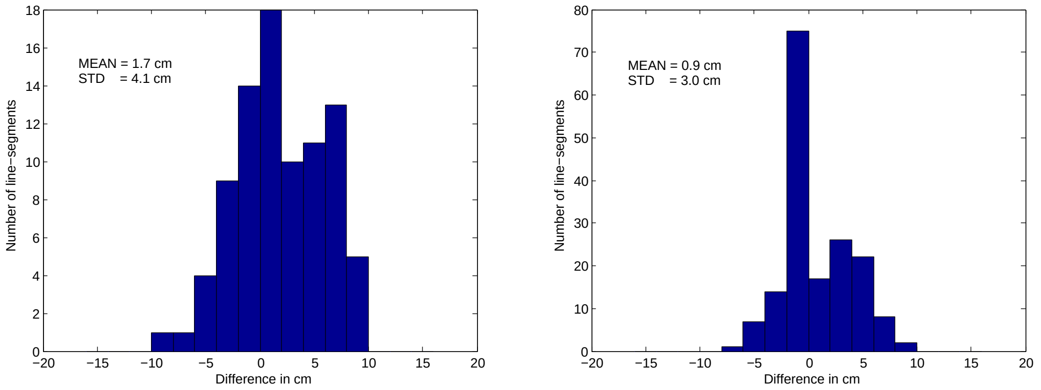
Regarding evaluation results the estimated standard deviation of our 3D roadmark reconstruction algorithm under real conditions is 4 *cm* for an average object distance of 10 *m*. Considering our edge matching algorithm we reach 1 *cm* accuracy at the distance of 10 *m* for 3D linked-edges. The lower accuracy obtained on the zebra-crossings can be explained by error accumulation during reconstruction, detection and modelling steps. Regarding the simplicity of our strip model it is an acceptable accuracy. Indeed roadmark modelling under real conditions with higher accuracy (< 4 *cm*) implies a much more complex model to be capable to deal with damaged or eroded roadmarks. Finally since process of dashed-line reconstruction is similar to this of zebra-crossing the geometric precision of dashed-lines must be similarly 4 *cm*.

$$\sigma_z = \frac{D}{B} \times \sigma_{x,y} \tag{6.9}$$



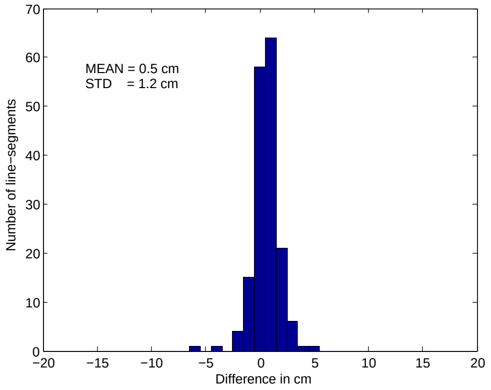
(a) Histogram of $\sigma_{C_{1,2}}$

(b) Histogram of $\sigma_{T_{1,2}}$



(c) Histogram of $\sigma_{d_{1,2}}$

(d) Histogram of $\sigma_{S_{1,2}}$

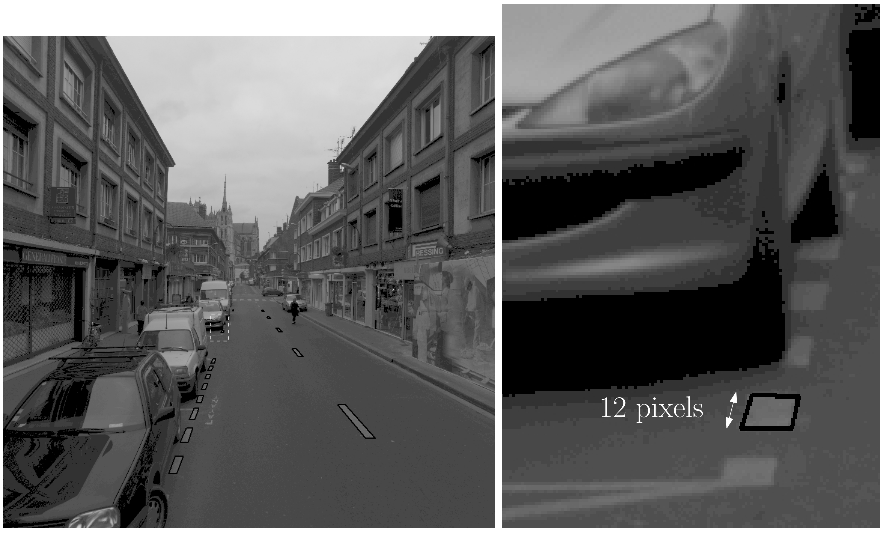


(e) Histogram of $\sigma_{W_{1,2}}$

Figure 6.12: Evaluation of geometric precision according the criteria defined in Equations 6.4 – 6.8.

6.6 Influence of image resolution on the algorithm

The results presented in previous Sections are obtained by using full resolution images of 4096 $\times$ 4096 pixels, with 3260 pixels focal length as input of the algorithm. These images are acquired by IGN's digital cameras that are destined for aerial imagery. Most of the ground-based mobile mapping systems use lower resolution cameras. This is the interest of examining the behaviour of our roadmark reconstruction algorithm in lower resolutions. In order to test the efficiency of the algorithm in lower resolutions, we have used undersampled images as input of the algorithm. Image resolution influence both correctness and accuracy of roadmark reconstruction.



(a) Images resolution: 4096 $\times$ 4096, the reconstruction is efficient until distance of 30 *m* from camera.

(b) Images resolution: 1024 $\times$ 1024, the reconstruction is efficient until 10 *m* from camera.

Figure 6.13: Influence of image resolution in roadmark detection.

6.6.1 Influence of resolution on correctness of reconstruction

As explained during Chapters 4 and 5 detection and modelling of roadmark objects are performed on 3D linked-edges. So the quality of 3D linked-edge in terms of completeness and accuracy influence the roadmark object recognition step. Considering Equation 3.2, accuracy of 3D point estimation increases by first, distance from camera (*D*) and second, accuracy of disparity (σ_ρ). In edge matching the σ_ρ is a fraction of pixel size. This is the reason why, accuracy of depth (*D*) estimation increases, when resolution decreases. This effect causes appearance of roughness along 3D reconstructed linked-edges. Such effect is similar to the one shown previously in Figures 3.13(a) and 3.13(c).

At long distances from camera, this roughness effect is amplified. This is troubling in line-segment estimation that is the first step of detection stage and can fail roadmark recognition stage (detection and modelling). Fortunately, this effect is less significant at shorter distances from camera and the detection algorithm may still be efficient.

In order to give an idea for influence of resolution on the detection algorithm, we test it first on a full resolution stereo-pair (4096×4096 pixels). The algorithm detects correctly all the 10×50 *cm* strips that are closer than 30 *m* to cameras. Then sub-sampled images (1024×1024 pixels) are used. This time, the algorithm reconstructs correctly the same strips until 10 *m* from cameras and fails for farther strips (See Figure 6.13). The algorithm is no more efficient for detection of these strips (10×50 *cm*) at lower resolutions. For larger roadmarks such as zebra-crossing the detection algorithm is still efficient at resolution of 512×512 pixels.

6.6.2 Influence of resolution on accuracy of reconstruction

As discussed in the previous Section, the algorithm is still detecting roadmarks at lower resolutions than our full resolution. However it is interesting to compare reconstruction accuracy at different resolutions. An interest is to choose an optimal camera from resolution point of view that reduce the cost and provide the requested accuracy and correctness.

In order to compare the reconstruction accuracies, a reference zebra-crossing is reconstructed using different resolutions of input images. This zebra-crossing is situated at an average distance of 11 *m* from cameras. Then the differences are measured between what is reconstructed by the algorithm and what is measured by topographic methods. Table 6.3 present the results. As can be seen on this Table, we obtain a difference of 2 *cm* between the model reconstructed from full resolution images and the one reconstructed from images that are resampled with a step of 8 pixels. It means that using 8 times smaller images, causes a lose of accuracy of about 2 *cm*. The loss of accuracy is higher when distance from camera increases.

6.7 Computation time analysis

Computation time analysing is an important issue in any object extraction algorithm. Our roadmark reconstruction algorithm can be interesting in road database generation and also autonomous navigation systems.

Resampling step	Image's size (pixels)	$\sigma_{C_{1,2}}$	$\sigma_{T_{1,2}}$	σ_d	$\sigma_{W_{1,2}}$	$\sigma_{S_{1,2}}$
		in *mm*				
1	4096×4096	47	60	67	70	43
2	2048×2048	46	60	57	70	41
4	1024×1024	60	61	72	69	71
8	512×512	68	65	66	83	60

Table 6.3: Comparison of 3D accuracies resulting from different input image resolution. Difference between a reference zebra-crossing and reconstructed one at different resolutions is expressed in *mm* according to the criteria of Equations 6.4–6.8.

Resolution of input image	Linked-edge matching	Detection and modelling
512×512	1.2 s (87%)	0.17 s (13%)
1024×1024	5 s (68%)	2.3 s (32%)
2048×2048	20 s (73%)	7.5 s (27 %)
4096×4096	78 s (84%)	14 s (16%)

Table 6.4: Computation time.

In database generation systems, computation time must be just reasonable in relation to volume of data to process. So it's not a critical issue. In contrast to database generation systems, autonomous navigation systems require real time processing algorithms.

The algorithm presented in this thesis is designed in a database generation scope. It implies to select accurate and reliable solutions even if they are time consuming. This Section provides an estimation of computation time for each step of the algorithm.

Implementation of the algorithm consists two main parts:

1. Linked-edge matching (Corresponds to Chapter 3),
2. Detection and modelling of roadmark objects (Corresponds to Chapters 4 and 5),

Both parts are implemented in C++. The hardware used for experiments is Pentium IV 2.8 GHz running under Windows XP. Table 6.4 presents computation time of each part for different resolution of input image. Processing time increase with image resolution. It reach 92 *s* at our full resolution (4096×4096) that is an acceptable time for off-line database generation. Computing time for undersampled images (512×512) reaches 1.4 *s*. It is possible to reduce this time by only a code optimisation.

Majority of computing time goes into matching step. So, in order to perform more optimisation, this part can be studied. In term of computing time, matching step is constituted of four main operations:

1. Resampling images in *epipolar* geometry,
2. Extracting edges,
3. Matching edges,
4. Other operations such as result writing, etc.

Figure 6.14 present proportion of each operation for a 4096 × 4096 resolution stereo-pair. As can be seen in this Figure, the image resampling and the edge extraction consume the majority of computation time (71%). In order to reduce the computation time of the both operations, two ways can be considered:

- Applying faster algorithms such as Sobel filter (instead of Canny-Deriche) for edge extraction [Sobel and Feldman, 1968] and a pre-computed grids for *epipolar* resampling. Of course, the use of these solutions can reduce the quality of resulting 3D linked-edges and so the quality of reconstruction.

- In order to save the quality of results, an alternative solution is to take benefit from computational performance that is provided by the latest issues of consumer graphics cards for personal computers. These cards are equipped with a fourth generation Graphics Processing Unit (GPU). Monti *et al.* [2005] present the interest of this technology in term of computation time for real-time image processing.

The computation time of roadmark recognition step (detection and modelling) is relatively lower than edge matching step. It is composed of four following parts:

1. 3D line-segment estimation from matching results,
2. Detection,
3. Modelling,
4. Other operations such as access to files, etc.

Figure 6.15 present the proportion of each operation in computation time. This step of process can be significantly accelerated by code optimisation.

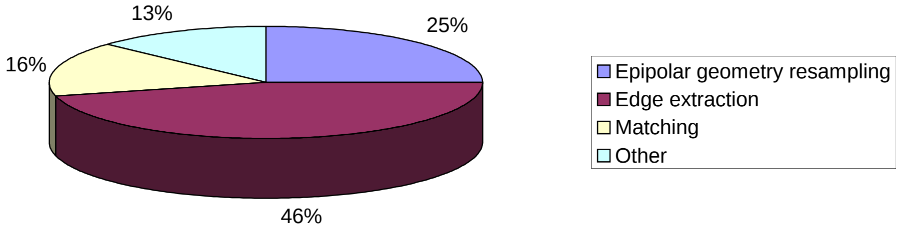


Figure 6.14: Computation time proportion for each operation in edge matching step.

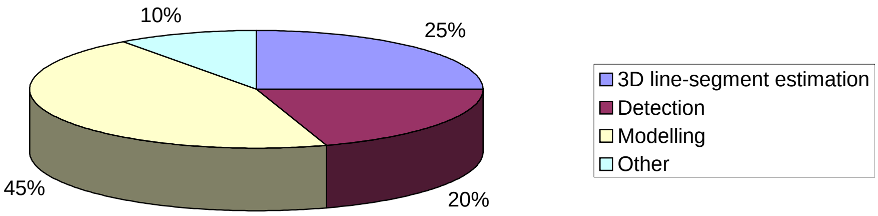


Figure 6.15: Computation time proportion for each operation in roadmark recognition step.

6.8 Conclusion

In this chapter we presented the efficiency of algorithm to cope with different difficulties such as partial occlusions and deformed roadmarks. The completeness of the method is evaluated on a set of 150 stereo-images. It provides a detection rate higher than 86%. The geometric accuracy is also evaluated on a large set of ground truth. It reaches 4 *cm* accuracy for zebra-crossing strips. Supposing a mean length of 2 *m* for strips, the 4 *cm* error correspond to a relative error of 1%.

The influence of image resolution on the reconstruction results in terms of accuracy and completeness is also studied and proved the efficiency of the algorithm in lower resolutions. The computation time is about 90 *s* at full resolution (4096×4096 pixels) and decreases to 2 *s* for 8 times under-sampled images. These computation times can be still decreased by algorithm optimisation and also by faster implementation of the same algorithm on GPUs.

Part IV

Application of roadmark database

Chapter 7

Application of roadmark reconstruction in sub-decimetric georeferencing of ground-based mobile mapping systems in urban areas

This chapter presents an application of 3D roadmark database generation in urban areas: sub-decimetric quality georeferencing of mobile mapping systems. The global idea is to use roadmarks in order to match ground-based images acquired by MMS with calibrated aerial images of the same area. This work is a common work with Olivier Tournaire [Tournaire, 2007].

7.1 Introduction

The pose estimation of mobile mapping systems is most of the time based on direct georeferencing devices such as GPS/INS [El-Sheimy, 1996; Schwarz *et al.*, 1993; Lapucha *et al.*, 1990; Novak and Bossler, 1995]. Even though direct localising devices provide decimetre absolute positioning accuracy in open areas, in dense urban areas such as European cities, GPS masks and multiple paths do corrupt the measurement quality. Even if GPS is integrated with INS for interpolating between GPS interruptions, intrinsic drifts of INS cumulates soon and lead to metric accuracies.

As mentioned in Chapter 1, we aim at providing an image database with sub-decimeter georeferencing accuracy. This database is applied in 3D facade and road modelling. It can also be used in autonomous navigation systems. Most of these systems are based on real time extraction and matching of visual landmarks with the ones collected from an image database that is precisely georeferenced in an off-line mode. The accuracy and reliability of navigation will directly depend on the quality of the data acquired within the off-line mode. So a georeferenced image database with sub-decimeter quality in whole city can provide the reference database needed for autonomous navigation. These are the reasons why subdecimetre positioning of MMS in dense urban areas is of great interest. An alternative positioning technique for MMS is to use its own collected data as a localisation tool. The Section 7.2 provides a brief review on such existing methods and introduces our strategy.

7.2 Previous work and strategy

7.2.1 State of art on indirect localisation of mobile mapping systems

The image acquired by mobile mapping systems have already been used to improve the relative localisation of MMS for example by determining a better relative orientation between intra-date and inter-date images using tie points [Chaplin and Chapman, 1998; Bayoud *et al.*, 2004]. Another author uses line-segments to improve relative orientation of images [Bauer *et al.*, 2002]. Within the frame work of the *Stereopolis* project, Bentrah *et al.* [2004] propose another image-based relative positioning approach. The method is based on estimating intra-date relative orientation within successive exposure points.

Beside the mobile mapping community, image based navigation is investigated by robotic researchers. As explained in Section 7.1, these systems are based on real time matching between image sequences that are acquired during navigation mode with image sequences that are georeferenced in an offline mode. Within this framework, some attempts have been made for image-based georeferencing. In a stereo vision system proposed in Simond and Rives [2004b], the authors use line-segments and interest points for computing inter-date and intra-date homography between images. In another work presented by Royer *et al.* [2005] a mono vision system is presented. The authors aim at estimating camera pose during vehicle motion. In order to do that, interest points are matched within images. The camera poses are then computed by a hierarchical bundle adjustment.

These relative localisation strategies can improve local accuracy but, in absence of control points random errors pile up and provide poor absolute positioning quality.

Within Lidar-based systems Zhao and Shibasaki [2001] propose an approach for improving localisation of the system. The approach consists in registering the laser points by estimating relative transformations between the coordinate system of successive range profiles. In order to cope with error accumulation, conjunctive profiles are detected automatically and registration error is distributed equally between all profiles.

In another Laser-based system proposed by Frueh and Zakhor [2001] a scan-to-scan matching strategy is adopted to register relatively the scan profiles in one coordinate system. An initial absolute localisation of the system is assured by a map matching approach using available road map. The positioning is refined then by matching laser profiles with aerial images or a DSM[1].

Except the last mentioned approach, external data is seldom applied in localisation strategies. Section 7.2.2 explains how georeferenced aerial images are applied to provide absolute control points in the scope to perform a classical imagegrammetric triangulation

---

[1] DSM: Digital Surface Model

on ground-based images provided by *Stereopolis*.

7.2.2 Our global strategy

As mentioned in Section 7.2 most of the image-based localisation strategies attempt to improve relative orientation between images. This is often performed by introducing tie objects (points, line-segments,etc) in bundle adjustment equations. However in absence of control point it is impossible to handle cumulative random errors. In photogrammetry control point consists in point's coordinates that is measured with an external technique. Often topographic operations are applied to collect them. Such techniques is used by the mobile mapping system developed by Belgian GeoAutomation company [GeoAutomation]. However, topographic operations are expensive and also time-consuming. The cost and operation time increase by the number of images. Classical methods are not affordable for a large image database such as the one provided by our *Stereopolis* system. This is the reason why automation is inevitable.

Our global idea is to employ georeferenced aerial images of the same area to provide control points for ground-based images. In order to do that roadmarks are used as matching objects between two systems. The Figure 7.2.2 demonstrates the idea. The main interests of roadmarks are that:

- They are available on the whole vehicle's trajectory and also visible in aerial images.
- They are constituted of contrasted pattern which simplify their matching process between two systems.
- They are less generalised when seen from different points of view and at different scales.
- They are easy to detect from both aerial and ground-based systems.

Roadmarks are reconstructed from aerial images thanks to the work presented in [Tournaire, 2007]. The same roadmarks can be reconstructed from ground-based images applying the algorithm proposed during Chapters 3, 4, 5. Aerial images being precisely georeferenced, matching aerial and ground-based roadmarks provide control points in ground-based images. Indeed accuracy of such control points is limited by the accuracy of aerial images' orientation. Within the explained strategy the main problem is to match aerial and ground-based roadmarks.

7.3 How do we match aerial and ground-based roadmarks?

As explained in Chapter 3, strategy of a matching process is adopted as a function of primitive and constraints of matching. In our case roadmarks are extracted and reconstructed in 3D from multiple aerial images and also from ground-based images provided by *Stereopolis*. They are presented as a vector model. So each roadmark object is presented as a set of 3D points and also topological information that connect the so called points by line-segments. Concerning constraints of matching, direct positioning devices (GPS/INS)

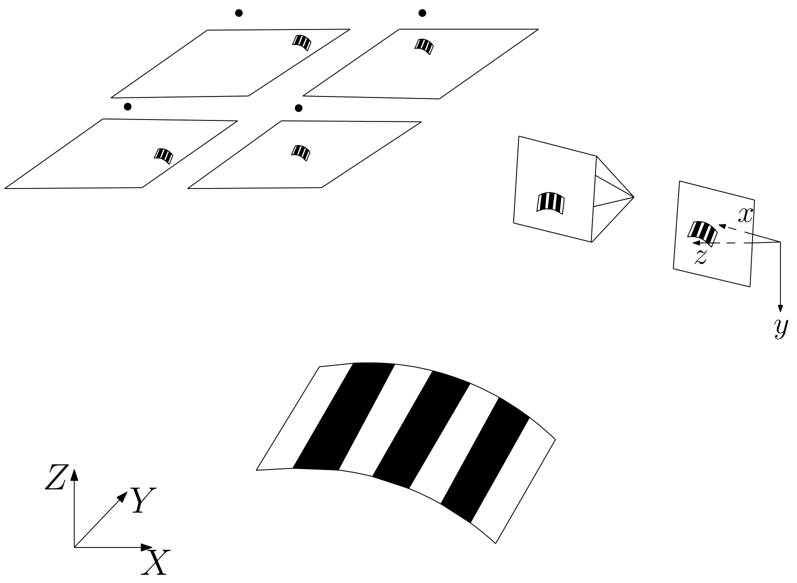


Figure 7.1: Absolute georeferencing of *Stereopolis*: Matching aerial-based and ground-based imagery using roadmarks.

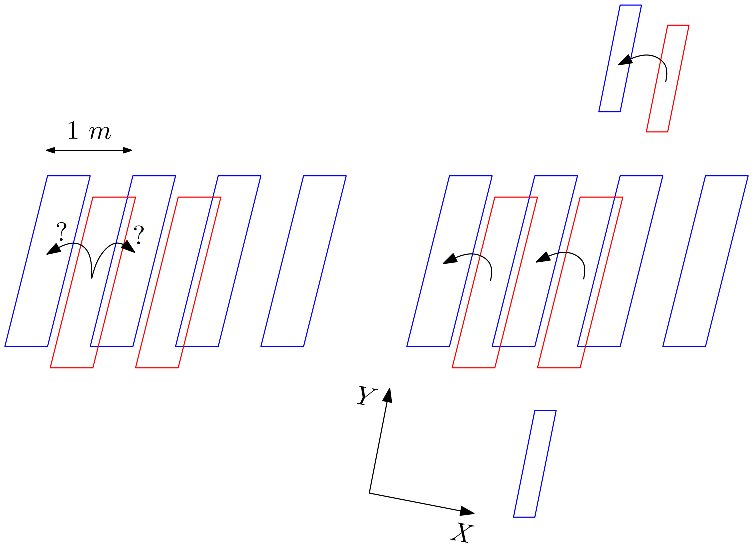
provide an initial absolute position for the ground-based roadmarks. For two minutes of GPS signal outage, the localisation drift reaches 0.04% of distance traveled [APPLANIX]. Under this condition the localisation accuracy reaches 70 *cm* at speed of 50 km/h. The orientation accuracy is about 0.06° for roll and pitch. Heading accuracy is about 0.03°. So during matching process ground-based roadmarks can be displaced to 70 *cm* and rotated to 0.06° in roll and pitch and to 0.03° in heading directions.

Lets think about matching process. Can it be a straightforward 3D graph matching using initial position of ground-based roadmarks in relation to aerial ones? Two main obstacles appear in using graph matching strategy:

1. Ambiguity in presence of repetitive roadmarks. Minimal distance between different strips of zebra-crossings and also those dashed-lines is 50 *cm*. The so called distance is shorter than the accuracy of solution provided by GPS/INS ($\approx$ 1 *m*). Moreover strips are often geometrically quite similar. So vector mode do not provide sufficient distinguishing power to avoid mismatches. Even though performing matching process on whole of a zebra-crossing or a dashed-line can resolve ambiguities. This is the case if they are reconstructed completely in both of the systems. Indeed in real urban area some strips of a zebra-crossing or a dashed-line can be occluded and so not reconstructed from ground-based or aerial images (See Figure 7.2(a)).

2. Besides problems caused by non resolved ambiguities, use of vector models as matching's primitive decrease precision of matching. Indeed errors of both aerial and ground-based modelling processes are introduced into matching process.

Regarding these two remarks, 3D graph matching strategy can be straightforward under condition to solve ambiguities. It can be performed by using more than one roadmark at the



(a) (b)

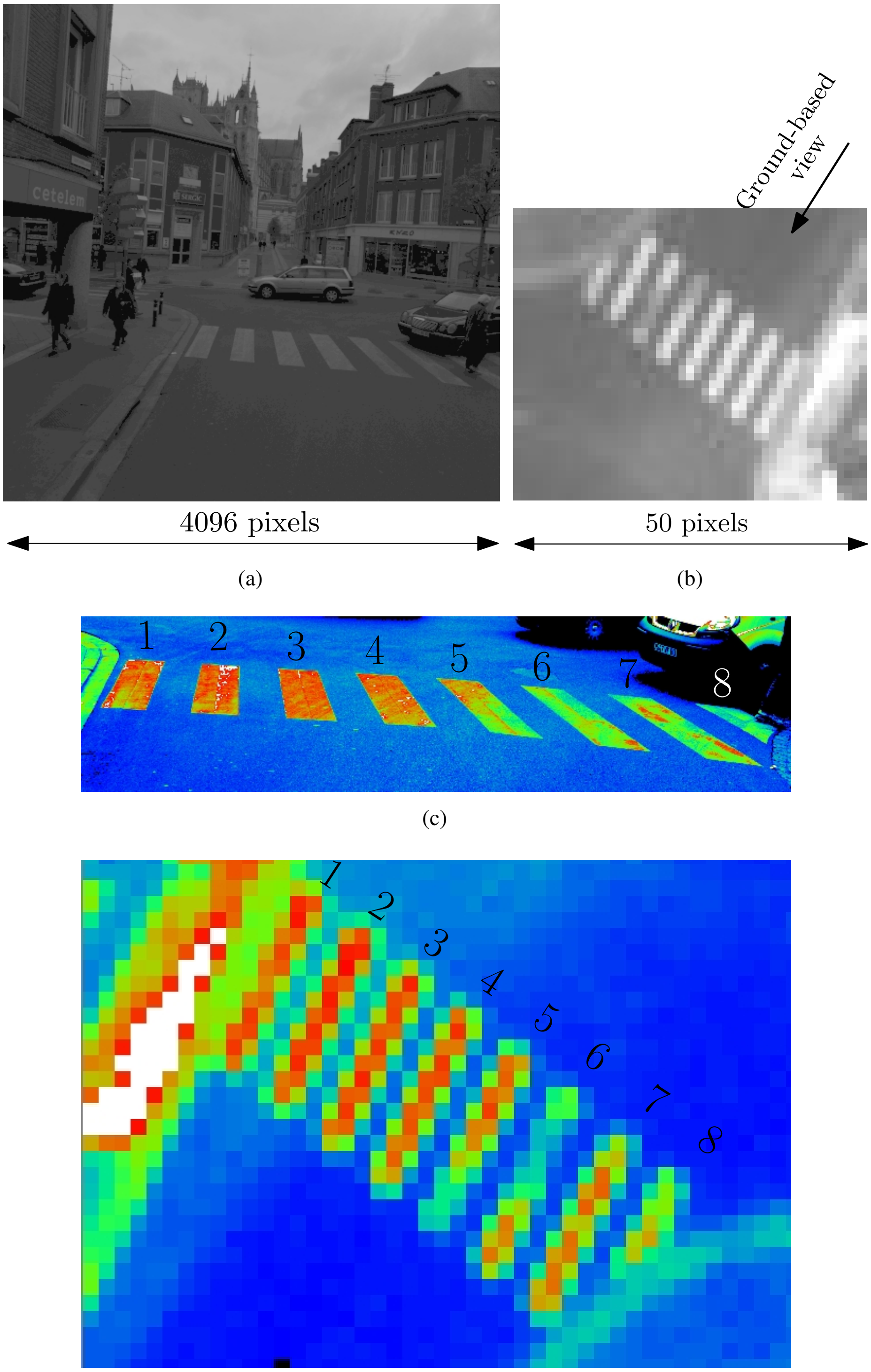
Figure 7.2: (a) Ambiguity of matching. In blue aerial reconstruction of a zebra-crossing with 4 strips. In red ground-based reconstruction of the same zebra-crossing, due to occlusions only 2 strips over 4 are reconstructed. (b) Ambiguity is resolved thanks to presence of two roadmarks. First the X translation is found using dashed-line strip, The Y translation is then found by zebra-crossing's strips.

same time (See Figure 7.2(b)). However the precision is limited by modelling precision.

In order to provide a more robust approach image's radiometry can be applied in matching process. Indeed in aerial and ground-based image spaces, road surface and roadmark objects are not simple dark and light homogeneous regions. Thanks to high radiometric resolution of aerial images (12 bits) the textures are visible even on roadmark objects. Figure 7.3 shows an example of aerial and ground-based images for the same zebra-crossing. As can be seen on Figure 7.3(c) each strip of zebra-crossing is inhomogeneously textured. The textures are also visible in the aerial images (See Figure 7.3(d)). The accordance between aerial and ground-based is specially visible by comparing radiometric variation from strip number 1 to 8. Our adopted strategy is based on the «analyse-synthesis »principle in computer graphics and computer vision [Gagalowicz, 1994]. The approach is based on two main steps:

1. Extraction of roadmark's 3D textured model from ground-based images,
2. Positioning of the extracted model in relation to aerial calibrated images. It is performed by moving the roadmark model in a 3D search space. For each position it is textured with aerial images. Textures coming from aerial images are compared to texture coming from ground-based images. The best position is the position that maximizes the correlation score between aerial and ground-based textures.

Sections 7.4 and 7.5 present these two steps.



(a)

(b)

(c)

(d)

Figure 7.3: (a,b) Aerial and ground-based images of same zebra-crossing, (c,d) zoom on aerial and ground-based images in hypsometric tints.

7.4 3D textured roadmark reconstruction from ground-based images

As explained during Chapters 3 to 5, it is possible to reconstruct roadmarks from each stereo-pair provided by the *Stereopolis* system. Thanks to the work presented in [Tournaire, 2007] roadmarks are reconstructed from georeferenced aerial images. These reconstructed roadmarks can be used to find stereo-pairs of interest within ground-based images. In order to do that, the approximate orientations of terrestrial images provided by GPS/INS are used. The diagram of Figure 7.6 demonstrates the interaction between aerial images in the georeferencing process.

Our matching strategy is based on maximising correlation score between aerial and ground-based texture of a roadmark object. Lets note that aerial-based texture is a function of roadmark's position in relation to the so called images. So the position that maximise the correlation score is considered as the solution. In order to simplify the correlation operation, the 3D roadmark is presented as an ortho-image that contains the roadmark's texture and a DSM[2] that contains the 3D geometry of roadmark. The DSM is needed for generating the ortho-image. In fact given the position of a DSM (roadmark's geometry) in relation to an image (aerial or terrestrial) the corresponding ortho-image (roadmark's texture) can be easily obtained by back-projection in image space.

Roadmark surface model can be easily obtained by applying a triangulation on vertexes of the 3D roadmark extracted by ground-based reconstruction (See Figure 7.4(a)). In order to provide realistic texture, resolution of ground-based DSM and ortho-image is selected proportional to GSD of ground-based images (5 *mm*). Corresponding raster DSM is generated (See Figure 7.4(b)) and used for ortho-image generation using one of the left or right ground-based stereo-images (See Figure 7.4(c)).

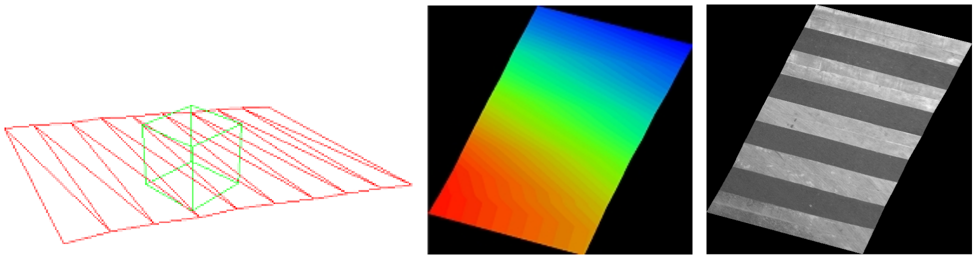

(a) (b) (c)

Figure 7.4: (a) 3D surface obtained by triangulation, (b) corresponding DSM at resolution of 5 *mm*, (c) the ortho-image obtained from ground-based image and the DSM at resolution of 5 *mm*.

As our strategy is to compare ground-based ortho-image with ortho-image coming from aerial images, it is necessary to compute ground-based DSM and ortho-image at the GSD of aerial images. GSD of aerial images being about 25 *cm* and ground-based images of about

---

[2] DSM: Digital Surface Model

5 *mm* at 10 *m* of distance from stereo-pair, there is a high difference of resolution between aerial and ground-based images ($r = 50$). In order to provide reliable ortho-image and DSM at 50 times smaller scale, pixels are integrated as in Equation 7.1. The resulting DSM and ortho-image are shown in Figure 7.5. They are used as reference objects for which we seek the absolute position. Supposing that absolute orientations are known for aerial images, it is sufficient to find the position of reference object in relation to aerial images.

$$\begin{cases} DSM_{ground}^{res._{aerial}}(i,j) & = \frac{1}{r^2} \sum_{k=ir}^{(i+1)r-1} \sum_{l=jr}^{(j+1)r-1} DSM_{ground}(k,l) \\ Ortho_{ground}^{res._{aerial}}(i,j) & = \frac{1}{r^2} \sum_{k=ir}^{(i+1)r-1} \sum_{l=jr}^{(j+1)r-1} Ortho_{ground}(k,l) \end{cases} \quad (7.1)$$

where :

DSM_{ground} : ground-based DSM

$Ortho_{ground}$: ground-based ortho-image

$DSM_{ground}^{res._{aerial}}$: ground-based DSM at the resolution of aerial images

$Ortho_{ground}^{res._{aerial}}$: ground-based ortho-image at the resolution of aerial images

$r : \dfrac{GSD_{aerial}}{GSD_{ground}}$

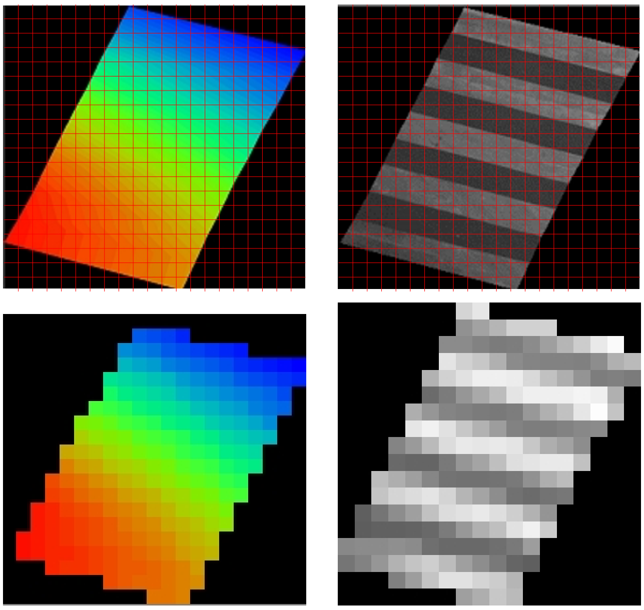


Figure 7.5: Top: ground-based DSM and ortho-image at the resolution of 5 *mm*, Bottom: ground-based DSM and ortho-image after pixel integration at the resolution of 25 *cm*.

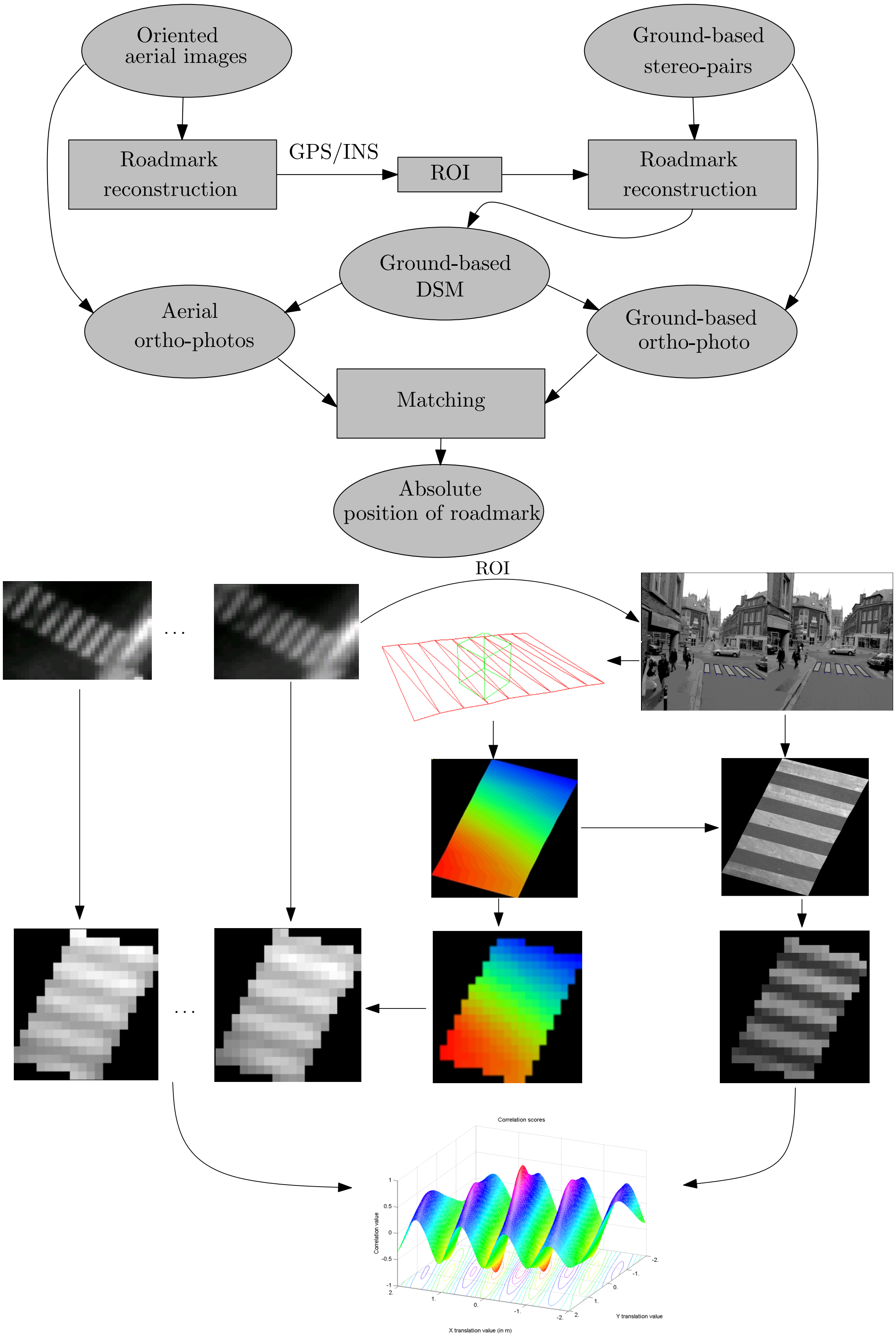


Figure 7.6: Our strategy for georeferencing ground-based images using calibrated aerial images of the same area.

7.5 Matching ground-based and aerial-based roadmarks

As explained in Section 7.4, a ground-based reference roadmark is reconstructed and presented under the form of a DSM and corresponding ortho-image. Thanks to direct georeferencing devices (GPS/INS) an initial absolute position and orientation of the so called roadmark is known. As mentioned in Section 7.3 the localisation accuracy provided by direct georeferencing method varies between 50 *cm* to 1 *m*. The accuracy of rotation estimation is about 0.03° in heading and 0.06° in roll and pitch directions. We are looking for an optimised position and orientation of the so called reference roadmark in the neighbourhood of the initial solution with respect to the oriented aerial images. Thus for one possible instance of roadmark's position and rotation, we simulate what the aerial images looking this roadmark should be (See Equation 7.2). We compare then those images with the real set of images seeing this object (See Equation 7.3).

$$
\begin{array}{rl}
I_k(i,j) & = R_I(i,j) * f_{I_k}(O(x,y,z)) \\
\text{where :} & \\
I_k & : \text{ an aerial image within the set.} \\
O & : \text{ the roadmark object.} \\
R_I & : \text{ the point spread function.} \\
f_{I_k} & : \text{ the image function model.} \\
(i,j) & : \text{ coordinates in image space.} \\
(x,y,z) & : \text{ coordinates in object space.}
\end{array}
\tag{7.2}
$$

The transformation we are looking for is composed of 3 translations (x,y,z) and 3 rotations. In order to cut down the complexity of the problem, the rotation is estimated in the preprocessing. It is performed using 3D line-segments coming from both ground-based and aerial-based reconstruction. The mean orientation of each line-segment set is computed. Then the rotation between these two mean orientations is estimated and applied to ground-based roadmark. The rotation being corrected, the problem is simplified to a $3D$ translation (dx,dy,dz) estimation. To conclude, the translation is estimated by Equation 7.3.

$$
\begin{array}{rl}
\underset{(dx,dy,dz)}{\operatorname{argmin}} \| I_k(i,j) - R_I(i,j) & * f_{I_k}(O(x+dx,y+dy,z+dz)) \| \\
\text{where :} & \\
(dx,dy,dz) & : \text{ translations} \\
O(x+dx,y+dy,z+dz) & : \text{ the translated roadmark object}
\end{array}
\tag{7.3}
$$

In practice the distance minimization proposed in Equation 7.3 is performed by maximising a correlation score between ground-based ortho-image and multiple aerial-based ortho-image. This operation consists in:

1. Displacing the roadmark in 3D space around the initial solution.
   The ground-based DSM represents the roadmark's geometry. So the displacement of roadmark is performed by translating the DSM with a selected step in each of x, y and z directions. Step is arbitrary fixed to 2 *cm*. During displacement, corresponding ground-based ortho-image is not modified but, since the position of roadmark

with respect to aerial images changes, new aerial-based ortho-image are computed for each translation. The Figure 7.7 demonstrates aerial-based ortho-image generated at different offsets from one aerial image. The reader is invited to give attention to difference between ortho-images that is due to applied offset.

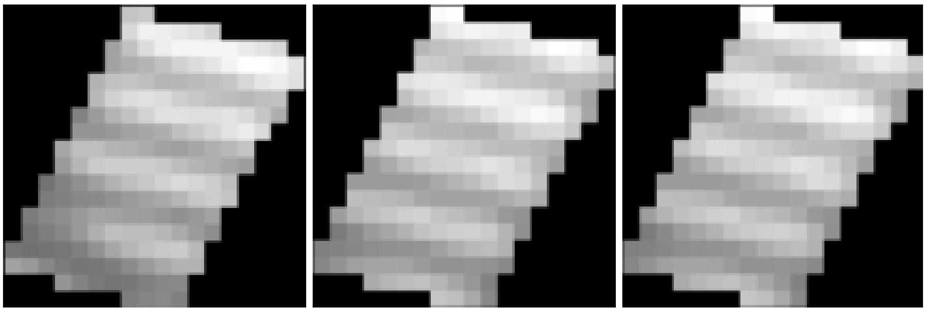


Figure 7.7: One aerial image resampled in geometry of ortho-image for different offsets (dx,dy,dz)

2. Computing correlation score between ground-based and multiple aerial-based ortho-image.
   For each offset the ground-based ortho-image is correlated to aerial-based ones and a correlation score is registered.

$$Score(dx,dy,dy)=\sum_{k=1}^{n}Cov(\mathcal{P},\mathcal{M}_k(dx,dy,dz)) \tag{7.4}$$

$$\mathcal{M}_k(dx,dy,dz)=g_{I_k}(f_{I_k}(\mathcal{S}(x+dx,y+dy,z+dz)))$$

where :

Cov : the normalised covariance

$\mathcal{P}$: ground-based ortho-image at the resolution of aerial images

$\mathcal{M}_k$: aerial-based ortho-image which is re-sampled from image I_k

$\mathcal{S}(x+dx,y+dy,z+dz)$: 3D roadmark's surface translated by (dx,dy,dz)

f_{I_k} : the photogrammetric model of image I_k

g_{I_k} : the fonction that gives the grey-levels of image I_k (7.5)

The computation of this correlation score over the $3D$ search space provides a hypersurface of correlation. The optimum position of the ground-based roadmark in relation to aerial images occurs at the maximum of this correlation hypersurface. Figure 7.8 shows this correlation surface at $dz=0$.

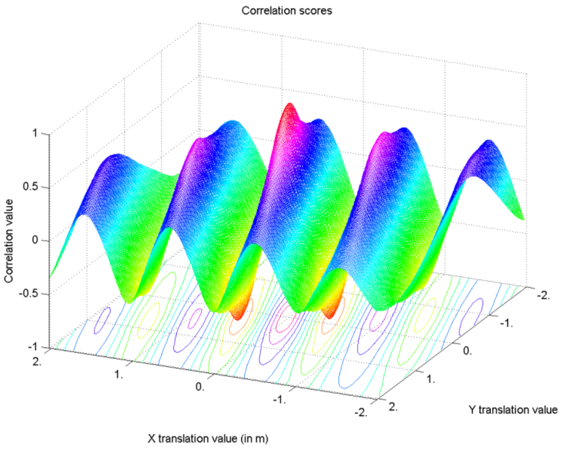


Figure 7.8: Correlation surface at $Z = 0$ on simulated data.

As it can be seen on this figure, the resulting correlation surface is continuous and constituted of periodic peaks. This is due to periodic dark-to-light and light-to-dark pattern of zebra-crossing. Indeed, negative peaks are due to correlation of dark objects with light ones and vice versa. Despite this periodicity, we have no problem of matching ambiguities because the correlation score of the good match is significantly higher than other periodic peaks. This is due to two discriminant properties of strips:

- Texture: as demonstrated in the Figure 7.3 strips are not homogeneous light objects and texture varies from one strip to other.
- Geometry: all strips are modelled by quasi-parallelogram. However their shapes and size can slightly vary between different strips of a zebra-crossing.

In practice, the minimization of Equation 7.3 is performed without taking into account the point spread function (I). This is correct if and only if the point spread function is symmetrical. If not it must be introduced as unknown in the minimization. It increases significantly complexity of the problem. It can be also estimated in laboratory and introduced in the computation.

We have shown our registering methodology on zebra-crossing objects. Of course, it can be generalised to all other roadmarks such as dashed-lines, arrows, etc.

7.6 Interest of the matched roadmarks in absolute georeferencing

We have determined adjusted 3D coordinates of the zebra-crossing object in an absolute coordinate system. These coordinates can be used as control points in ground-based images. These control points together with tie points that can be estimated by existing methods such as those mentioned in Section 7.2.1 can be injected into a global bundle adjustment. This can improve considerably the localisation provided by direct methods such as GPS/INS.

7.7 Conclusion

In this chapter we presented a novel concept and method mixing aerial and ground-based imagery, to achieve what we call GCO-base georeferencing (for Ground Control Objects). The method is tested on a zebra-crossing, but it is quite generic and can be easily extended to use other roadmarks. It is a necessary task for densifying the distribution of GCO. To conclude, in order to provide a complete system of image-based georeferencing, the three following subjects remain to be studied:

1. Application of the registering algorithm to other roadmarks such as arrows, dashed-lines, etc.
2. Tie point generation using algorithms such as those mentioned in Section 7.2.1.
3. Integrating control points and tie points in a global Bundle adjustment.

This approach enables mobile mapping systems to overcome the inefficiencies of direct georeferencing methods (masks and multi-path of GPS signals) in urban areas and to reach sub-decimeter positioning accuracies.

Part V

Conclusion and perspectives

Chapter 8

Conclusion and perspectives

Ground-based 3D reconstruction of urban areas is a complex problem. This thesis has attempted in contributing in automatic road feature reconstruction. High resolution stereo-pairs provided by a ground-based mobile mapping system (MMS) are used as data source. We focused particularly on reconstruction of roadmarks because on the one hand they provide useful information about road's structure, on the other hand they are applied in absolute georeferencing of the MMS by matching ground-based and aerial calibrated images. This Chapter summarises our contributions and proposes some perspectives for future researches.

8.1 Our contributions

8.1.1 3D Roadmark reconstruction algorithm

As indicated in state of art, most of the efforts for road extraction from ground-based images have gone into 2D detection in image space and 3D geometry is obtained by supposing a planar road surface. The methods providing 3D geometry without planar road hypothesis suppose often linear models such as line-segment and curves for roadmarks. Our reconstruction algorithm provides a 3D polygonal model for each individual roadmark object without any planar road hypothesis. Two case of zebra-crossing and dashed-lines are studied.

The algorithm begins with linked-edge matching in a stereo context. It consists in a dynamic programming optimisation along conjugated *epipolar* lines. Search space and similarity measure are adapted to match the edges belonging to an area around road surface. A *figural continuity* constraint is introduced in the optimisation to ensure continuous chains. Finally a subpixel matching quality is provided. Search space and similarity measures are voluntarily constrained to ensure sufficient matching quality for edges that are around road surface which reduce considerably the generality of edge matching algorithm.

Both zebra-crossing and dashed-line detection algorithms use the results of matching step. The main difficulty of detection algorithm is to cope with noises due to local texture and fragmentation of linked-edges. For this purpose a 2D signature-based algorithm has been adopted. Some basic signature recognition tools have been developed. In term of order and parameters, each of the detection algorithms use the basic tools differently. The detection algorithm provides a set of 3D line-segment as candidates of zebra-crossing and dashed-line strips.

The last step consists in rejecting or accepting and providing a polygonal model for each candidate. A quasi-parallelogram shape is adopted as model. A perceptual grouping method together with a line-segment grouping algorithm are used to provide a polygonal shape from line-segments. The presented 2D line-segment grouping algorithm is quite generic and can be used in other applications.

The completeness and geometric accuracy of results provided by both zebra-crossing and dashed-line reconstruction algorithms have been evaluated on a large set of real images and revealed their robustness and accuracy. In term of geometric accuracy, the algorithm reaches standard deviation of 4cm. The detection rate is higher than 85% in image space. However, each strip being seen in more than one stereo-pair, the fusion of results in object space should provide a detection rate higher than 90%. An interpolation between reconstructed strips may allow to reconstruct missed strips and reach higher detection rates.

Both algorithms are relatively rapid. They take about 90 s for full resolution images (4096×9046 pixels). This time is reduced to 1.4 s for under-sampled images of size 512×512 pixels. More than 40% of computation time goes into the Canny-Deriche edge extraction. This portion of time can be reduced by using faster implementations of the same algorithm. Faster edge detection algorithms may also be used instead of Canny-Deriche algorithm. In this case it is very important to study the influence of edge quality on the completeness and geometric accuracy of results.

8.1.2 Sub-decimeter absolute georeferencing of MMS

The reconstructed roadmarks are applied in absolute georeferencing of the mobile mapping system. This is performed by matching ground-based and calibrated aerial images by using roadmarks as matching features. The matched roadmarks are considered as Ground Control Objects (GCO) that can integrate a Bundle adjustment in order to provide an absolute positioning for ground-based images. Thanks to completeness and high accuracy of ground-based reconstruction, this matching process can provide well distributed and also accurate GCOs. The efficiency of the algorithm is demonstrated on a zebra-crossing. However it is completely generic and can be performed using other roadmarks. We have shown the validity of the approach on one example of roadmark. More evaluations must be performed to validate the robustness of the method to occlusions, illumination conditions, etc.

8.2 Some perspectives

We developed two algorithms for reconstruction of zebra-crossing and dashed-lines from calibrated stereo-images. These two roadmark objects are particularly interesting because they provide useful information about the number, width and also intersection of lanes. However, it is important to think about a global strategy of roadmark reconstruction. This global strategy should deal with:

- Reconstruction of more complex roadmarks,
- Geometric and topological relation between roadmark classes,
- Multi-stereo-images as input data,
- Reduction of computation time.

In this scope, we give global outlines of a method for recognition and reconstruction of more complex roadmarks. Then, this method together with those for dashed-lines and zebra-crossing reconstruction are integrated in a multi expert system. A very interesting application of a multi expert system is presented by Giraudon *et al.* [1992] for interpretation of aerial images by extracting buildings, roads and natural features. We use a similar scheme for extracting different classes of roadmarks.

Our global strategy of roadmark reconstruction from multi-stereo-images is depicted in Figure 8.1. It is composed of two main steps:

1. Pre-processing,
2. Multi expert system for roadmark reconstruction.

Inputs of algorithm consist in a set of stereo-images and also the manual of roadmark construction. The output is a 3D roadmark GIS. Section 8.2.1 explains the pre-processing step and Section 8.2.1 introduce the adopted multi expert system. Let us stress that these two components are not still implemented and the presented figures (8.2– 8.5) are just for illustration purposes.

8.2.1 Pre-processing

The goal of this step is to reduce the search space and computation time. It provides suitable candidates for each class of roadmark object.

Top hat filter: The interest of this filter is in reduction of illumination variation in image space. It is widely used for this purpose in much research and has proved to be effective Yu *et al.* [1992]; Kim *et al.* [2006]. A top hat filter is a morphological operator used for extracting the pixels that have higher intensity than that of their neighbours in a given distance. This is the reason why it is particularly interesting in extraction of white roadmarks on dark background of road pavement. Is is based on subtracting an opened image from the initial gray scale image. The size of structuring element of opening operator must be just larger than largest roadmark object in image space. This parameter is regulated by using the given size of roadmark from construction manual.

After applying a top hat filter the bright roadmarks pixels have a much higher value than dark asphalt regions. Even, a sufficiently low threshold is efficient for image binearization. Otherwise a threshold can be selected by histogram analysis [Sahoo *et al.*, 1992]. Figure 8.2 shows a binarized image obtained by applying this method.

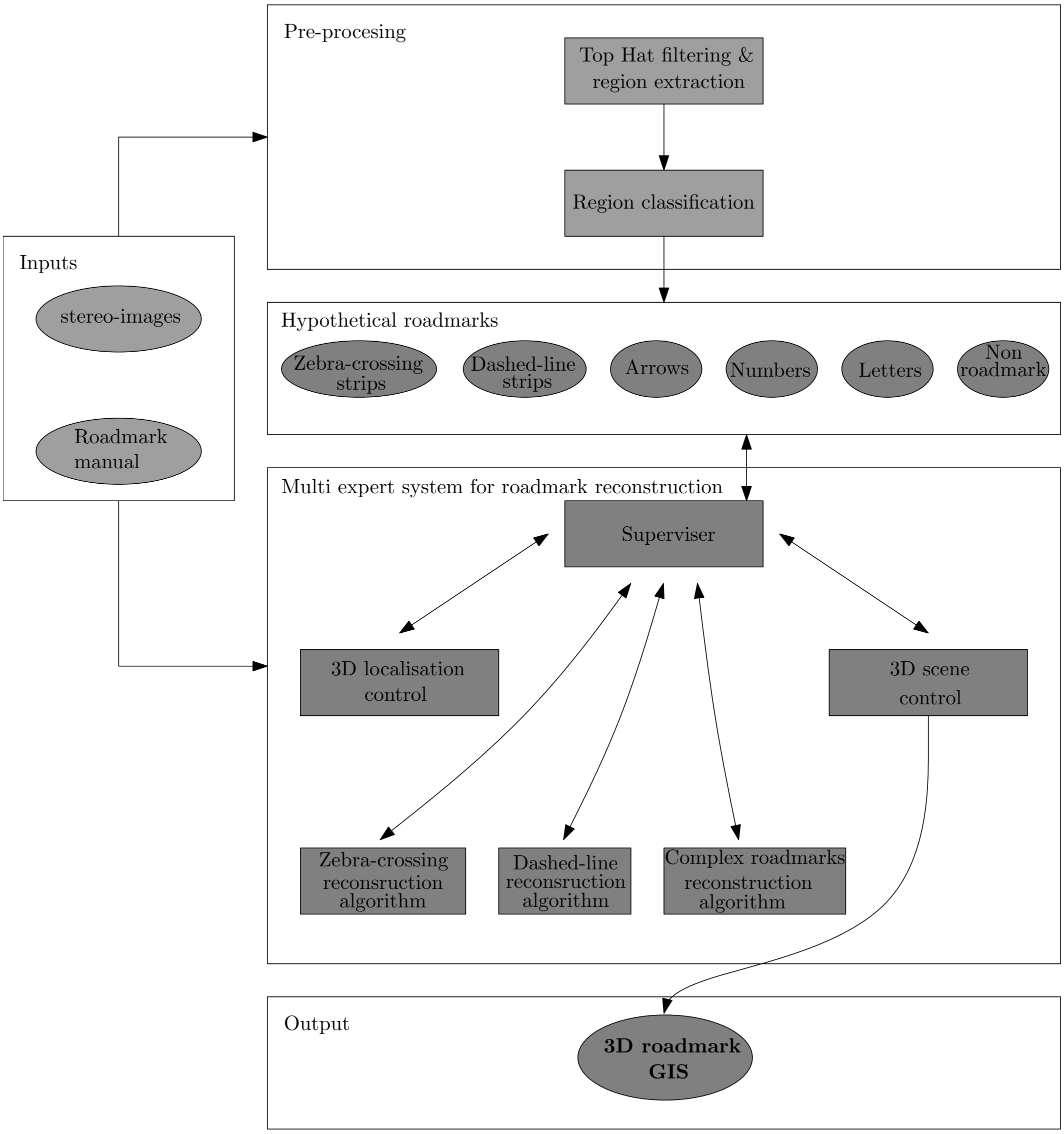
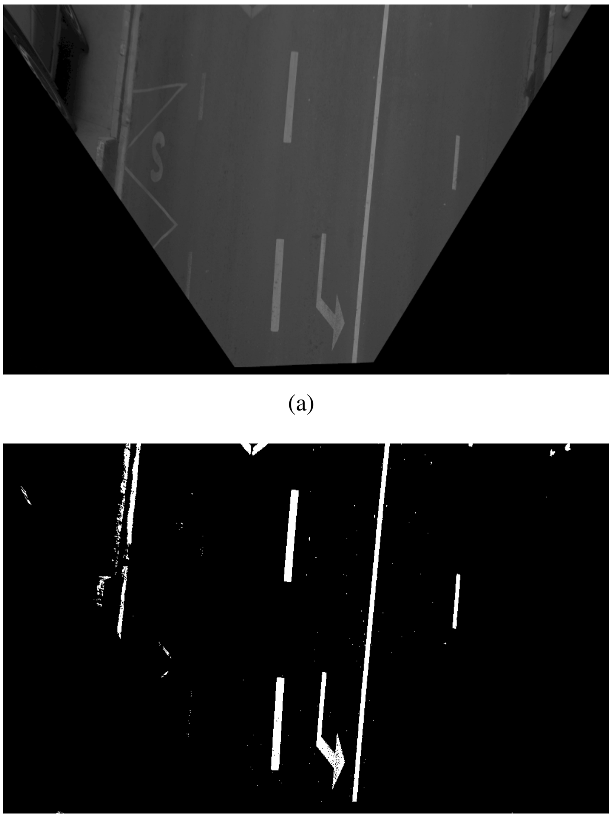


Figure 8.1: Proposed strategy for complete 3D roadmark database generation.

Region extraction: Top hat filter followed by thresholding provide a binarized image in which most of the roadmark pixels are white. An easy way to extract roadmark objects is to segment connected pixels and provide independent regions. Each region can be a putative candidate of roadmark object.

Region classification: The goal of this step is to classify each region that is extracted by the above method to a roadmark class or to non-roadmark class if the region correspond to any class. For this purpose, the roadmark construction manual is used. On the one hand, shape and size of roadmarks are known, on the other hand perspective deformation is reduced by rectification and scale of objects are known sufficiently enough to compare extracted regions with theoretical models. Figure 8.4 depicts theoretical model corresponding



(a)

(b)

Figure 8.2: a) An original rectified image. b) resulting binearized image by applying top hat filter and thresholding.

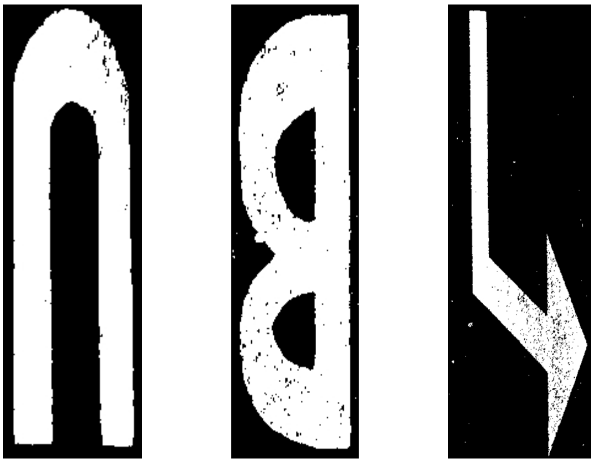


Figure 8.3: Extracted binary regions – each region is presented in its principal direction

to the extracted regions of Figure 8.3.

In order to classify each extracted region, it can be correlated with the theoretical model of all roadmark objects. For this purpose, the extracted region is resampled in a way that its

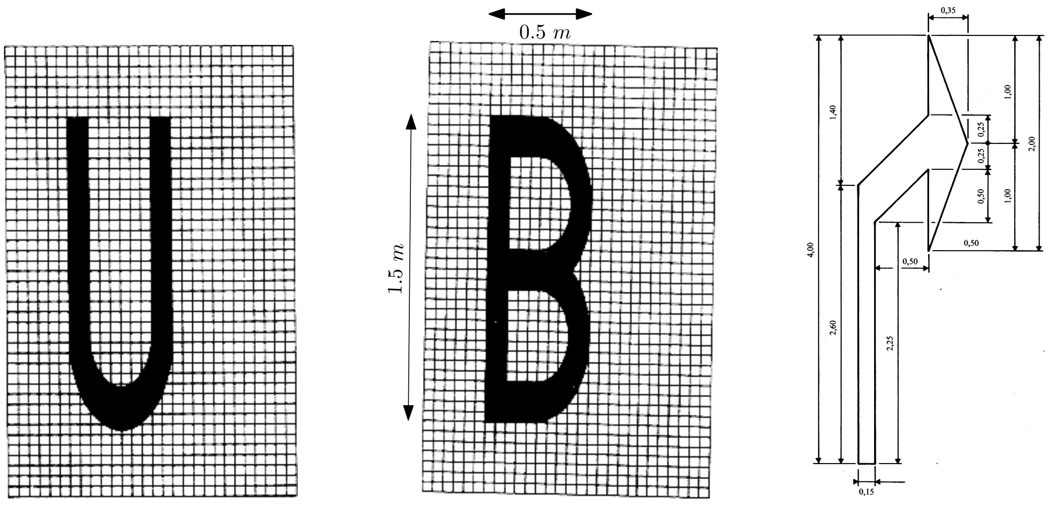


Figure 8.4: According to [Transport Ministry and Interior Ministry, 1988]: Theoretical models of extracted regions of Figure 8.3.

principal axis (often road axis) joins the Y axis of image. Then the correlation coefficient is computed in a bounding box of the extracted region in the resampled image. This operation is performed with all theoretical roadmark objects. The maximum coefficient is accepted if it is larger than a threshold (for example higher than 90%). The region is classified in the class corresponding to maximal coefficient.

Correlation is computationally costing when the number of theoretical objects increases. In order to cope with this problem, a signature based recognition can be interesting. Signature are computed using extracted region in their resampled geometry. For this purpose number of pixels are accumulated in image's X and Y axes. Figure 8.5 depicts signatures of the above regions.

These signatures seems to be sufficiently discriminatory for recognition of roadmark class. Theoretical signatures can be computed in an off-line mode. The signature of an extracted region is compared with signature of all roadmark classes. A class is accepted if corresponding similarity measure is higher than a threshold (for example higher than 90%). If more than one class are accepted, correlation coefficient are computed and determine the good match.

8.2.2 Multi expert system

Disposing one expert algorithm per roadmark class, each candidate provided by the preprocessing step may be introduced in corresponding expert algorithm for final reconstruction. However, high information intensity in images of urban areas does not allow non-hierarchical reconstruction. In fact, all roadmark classes do not provide same quantity of information. For example, it seems important to begin with reconstruction of dashed-lines and then look for arrows within lanes that are determined by dashed-lines. Another good idea is to reconstruct first zebra-crossings and remove the corresponding image area from search space in reconstruction of other roadmarks. Indeed zebra-crossing is a large and particular object so its reconstruction is more reliable than small roadmarks.

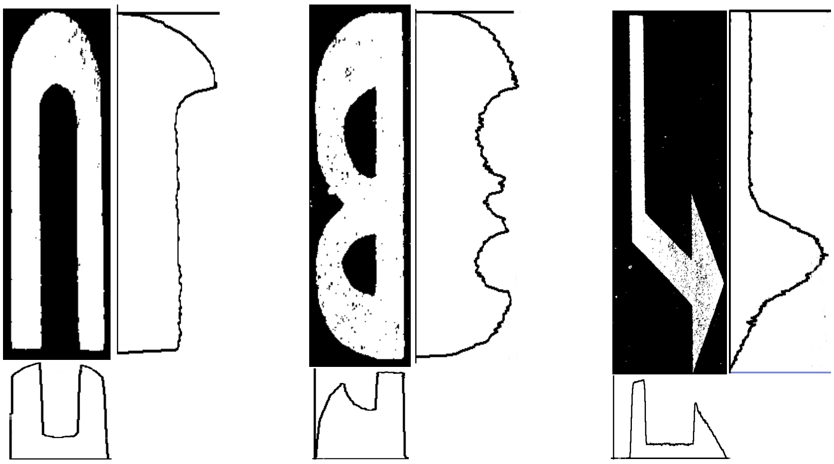


Figure 8.5: Signature of extracted regions of Figure 8.3 in X and Y axis.

This is the reason why a global strategy is needed to supervise the distribution of reconstruction tasks. This strategy must deal with:

- Ordering reconstruction of classes.
- Validating reconstructed objects by taking into account topological relation between extracted objects.
- Administering conflicts of reconstruction.
- Guiding search area by using topological relation between roadmark classes.

Our adopted strategy consists in three following components:

1. Supervisor: It administrates execution of expert algorithms. It defines search area and parameters of each execution by consulting two other components.
2. 3D localisation control: It controls the geometric and topological relations between objects.
3. 3D scene control: It registers extracted objects in a GIS and deal with conflicts.

Supervisor: The importance of the supervisor is in ordering the extractions. It means that it reconstructs first structuring elements such as zebra-crossing and dashed-lines. The interest is in limiting search space for extraction of other roadmarks (explained in next paragraph). Another functionality that can be interesting in multi stereo-image scenario is that before calling an expert algorithm for reconstruction of a hypothetical object, supervisor can send a query to 3D scene control to see if object is already extracted from previous stereo-images or not. If the hypothesis is already extracted it can be used in a validation mode to increase the certainty of object.

3D localisation control: This component controls the geometric and topologic relation between objects. Its main functionality is to define pertinent search spaces for each reconstruction query. For this purpose the objects that are already extracted are used to define search space for new objects. The access to extracted objects in 3D scene control component pass with intermediate of supervisor component. For example it can propose to search for arrows in the region between two consecutive dashed-lines (within a lane). In this manner search space is limited automatically as a function of searched object.

3D scene control: This component register geometric and topological information of extracted objects in a GIS. It deals with conflict of reconstruction such as two different objects at the same 3D position, etc. This component is justified by two complementary activity:

1. In the same way that an expert algorithm can specify if an abject exist is a position or not, 3D scene control can answer this query by looking in the objects that are already extracted. Such information can be requested in order to increase the plausibility of a hypothesis. This is particularly interesting in multi stereo-image scenario because on object can be seen in more than one stereo-image. In this manner objects extracted in stereo-image i help to confirm reconstruction of objects in stereo-image $i+1$.
2. 3D scene control allows to preserve coherence of 3D scene at each addition of new object. This is important for fusion of extracted objects coming from multi stereo-images.

8.3 Conclusion

We have presented an algorithm for 3D roadmark reconstruction from ground-based images. The evaluations revealed the validity of the method from correctness and accuracy points of view. We have also presented a solution for georeferencing the mobile mapping systems when direct localisation methods (GPS/INS) do not provide sufficient accuracy. We believe that the presented work can be integrated into a global strategy such as the multi expert system presented in this chapter and constitute an operational roadmark reconstruction system.

Publications

Conference publications

B. SOHEILIAN, N. PAPARODITIS, D. BOLDO, J.P. RUDANT. *Automatic 3D extraction of rectangular roadmarks with centimeter accuracy from stereo-pairs of a ground-based mobile mapping system.* The 5th International symposium on Mobile Mapping Technology, Padua, Italy. 2007.

B. SOHEILIAN, N. PAPARODITIS, D. BOLDO, J.P. RUDANT. *3D zebra-crossing reconstruction from stereo rig images of a ground-based mobile mapping system.* Proceedings of ISPRS Commission **V** Symposium, Image Engineering and Vision Metrology, Dresden, Germany. 2006.

O. TOURNAIRE, B. SOHEILIAN, N. PAPARODITIS. *Towards a sub-decimeter georeferencing of ground-based mobile mapping systems in urban areas : matching ground-based and aerial-based imagery using roadmarks.* Proceedings of ISPRS Commission **I** Symposium. Paris, France. 2006.

B. SOHEILIAN, O. TOURNAIRE, L. PÉNARD, N. PAPARODITIS. *Different quality level processes and products for ground-based 3D city and road modeling.* In: A. Abdul-Rahman and S. Zlatanova and V. Coors (eds), Lecture Notes in Geoinformation and Cartography, Innovations in 3D Geoinformation Systems, Springer-Verlag, PP. 417- 429. 2006.

O. TOURNAIRE, N. PAPARODITIS, B. SOHEILIAN. *Vers un Géoréférencement Subdécimétrique des Systèmes de Cartographie Mobile en Milieu Urbain.* Journée des Sciences de l'Ingénieur du LCPC. Marne-la-Valée, France. 2006.

N. PAPARODITIS, O. BENTRAH, L. PÉNARD, O. TOURNAIRE, B. SOHEILIAN, M. DEVEAU. *Automatic 3D recording and modeling of large scale cities: the ARCHIPOLIS project.* E.P. Baltsavias *et al.* (eds), Recording, Modeling and Visualization of Cultural Heritage. Taylor and Francis Group, London, UK. 2005.

Oral presentations

B. SOHEILIAN. Reconstruction de marquages routiers à partir d'images terrestres. Journées de la Recherche de l'IGN. Saint-Mandé, France. Mars 2007.

Appendix A

Evaluation results

Table A.1 presents the 3D geometric accuracy measures on each strip. High errors are presented in red color. Corresponding reconstructions are illustrated in Figures A.1–A.4.

Table A.1: 3D geometric accuracy for each strip. Corresponding stereo-pairs are presented in Figures A.1–A.4. High errors are presented by red color

Stereopair's number	Strip's number	Distance in *m*	$\hat{\sigma}_C$	$\hat{\sigma}_T$	$\hat{\sigma}_d$	$\hat{\sigma}_W$	$\hat{\sigma}_S$
			in *mm*				
71	1	7.0	20	10	4	1	7
71	2	7.2	13	7	1	14	18
71	3	7.5	18	9	6	15	20
71	4	7.9	41	10	12	10	44
71	5	8.5	59	11	22		
94	0	7.2	26	7	12	9	31
94	1	7.0	38	2	15	8	43
94	2	7.0	42	6	18	5	38
94	3	7.1	46	4	20		
95	0	12.5	25	6	5	11	18
95	1	12.3	31	3	11	7	46
95	2	12.2	46	7	19	17	42
95	3	12.2	48	13	42	9	48
95	4	12.2	19	33	14	33	19
95	5	12.4	20	13	15		
115	0	11.0	39	6	35	7	8
115	1	11.0	11	3	7	5	43
115	2	11.1	58	14	29		
115	4	11.6	54	17	18		
117	0	9.4	12	5	0		
117	2	9.5	8	21	3		
153	1	7.9	37	22	28		

Continued on next page

Table A.1 – continued from previous page

Stereopair's number	Strip's number	Distance in *m*	$\hat{\sigma}_C$	$\hat{\sigma}_T$	$\hat{\sigma}_d$	$\hat{\sigma}_W$	$\hat{\sigma}_S$
			in *mm*				
153	3	8.2	46	6	24	7	28
153	4	8.5	28	11	8		
157	0	6.5	24	6	8	5	36
157	1	6.5	24	6	14	8	16
157	2	6.6	31	10	17	2	32
157	3	7.0	22	10	3		
158	0	11.9	21	7	11	4	42
158	1	11.9	31	7	12	6	31
158	2	12.0	37	10	23	3	46
158	3	12.1	27	11	3	7	12
158	4	12.4	14	13	0		
164	1	6.2	25	17	5	14	24
164	2	6.2	3	2	7		
165	0	12.5	31	6	20	1	44
165	1	12.8	42	2	24	6	6
165	2	13.3	24	11	4	12	20
165	3	13.7	6	6	2		
166	0	17.0	4	6	4	3	8
166	1	17.2	39	3	5		
166	3	17.8	12	11	5	13	9
166	4	18.1	13	17	8	2	10
166	5	18.6	13	15	6	7	10
166	6	19.1	13	17	18	11	40
166	7	19.6	50	26	32	11	18
166	8	20.2	30	28	29		
236	1	5.9	10	4	3	1	16
236	2	6.3	10	7	6	11	31
236	3	6.8	21	19	4	17	11
236	4	7.5	23	17	4	4	10
236	5	8.3	17	8	4		
244	0	13.7	217	35	72	28	186
244	1	14.2	136	11	72	21	126
244	2	14.7	119	25	46	12	181
244	3	15.2	162	34	72	29	151
244	4	15.9	198	40	80	37	176
244	5	16.7	180	28	80		
246	0	13.7	174	24	66	29	155
246	1	14.2	133	11	69	25	137
246	2	14.7	139	18	73		
247	0	13.7	180	25	68	29	155

Continued on next page

Table A.1 – continued from previous page

Stereopair's number	Strip's number	Distance in *m*	$\hat{\sigma}_C$	$\hat{\sigma}_T$	$\hat{\sigma}_d$	$\hat{\sigma}_W$	$\hat{\sigma}_S$
			in *mm*				
247	1	14.2	132	10	71	20	131
247	2	14.7	123	23	50	10	173
247	3	15.2	147	33	65	29	135
247	4	15.9	193	37	82	34	172
247	5	16.7	168	29	72		
249	0	14.3	201	29	76	35	168
249	1	14.7	145	16	75	24	144
249	2	15.3	144	28	59	6	209
249	3	15.7	167	42	68	38	170
249	4	16.5	228	49	87	37	202
249	5	17.2	206	31	93		
250	0	16.4	168	25	62	26	172
250	1	16.9	185	26	69	19	109
250	2	17.3	85	46	3	10	186
250	3	17.7	132	41	52	39	124
250	4	18.5	198	38	83	40	172
250	5	19.2	158	36	61		
252	1	6.6	23	10	8	1	16
252	2	6.9	19	3	8	12	26
252	3	7.4	43	18	18	25	24
252	4	7.9	13	11	0		
253	1	10.5	56	12	18	7	6
253	2	10.8	31	8	11	18	22
253	3	11.1	33	16	3	20	28
253	4	11.6	62	14	20	15	65
253	5	12.1	82	13	34		
314	0	6.7	57	6	33	6	47
314	1	6.5	34	8	22	4	25
314	2	6.6	24	7	15		
315	0	8.3	54	4	32		
315	2	8.1	32	7	20	12	28
315	3	8.2	17	5	2		
316	0	8.3	5	6	5	8	1
316	1	8.2	12	6	5	6	17
316	2	8.1	21	7	8	8	26
316	3	8.2	40	1	26		
317	0	10.2	88	8	50		
317	1	10.0	79	13	46	5	66
317	2	9.9	60	8	34		
319	0	10.1	70	7	39	7	67

Continued on next page

Table A.1 – continued from previous page

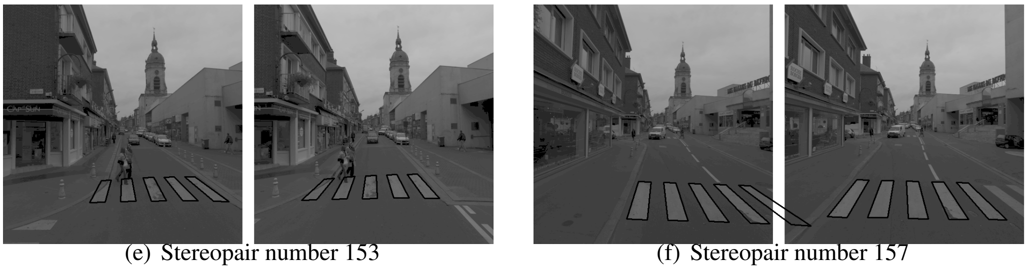
Stereopair's number	Strip's number	Distance in *m*	$\hat{\sigma}_C$	$\hat{\sigma}_T$	$\hat{\sigma}_d$	$\hat{\sigma}_W$	$\hat{\sigma}_S$
			in *mm*				
319	1	9.9	50	11	32	3	40
319	2	9.9	37	9	25	12	31
319	3	9.9	22	6	4		
320	0	10.1	66	7	38	8	64
320	1	9.9	50	10	31	2	53
320	2	9.8	49	20	36	8	29
320	3	9.9	15	7	2		
321	0	10.1	85	9	48	8	79
321	1	9.9	65	12	40	4	53
321	2	9.9	54	8	32	14	51
321	3	9.9	37	7	12	21	61
321	4	10.1	109	31	49		
322	0	10.9	66	10	39	9	58
322	1	10.7	46	11	30	3	39
322	2	10.6	38	9	27	12	29
322	3	10.7	18	5	0		
676	1	6.4	145	20	60	5	128
676	2	6.7	120	16	55	13	127
676	3	7.2	135	8	60	5	126
676	4	7.9	123	9	57		
677	1	10.7	220	26	108	18	232
677	2	10.8	253	20	117	22	230
677	3	11.1	216	19	104	26	227
677	4	11.5	243	18	107	8	229
677	5	12.1	220	15	102	27	220
677	6	12.8	218	26	99	19	222
677	7	13.5	237	34	104		
677	9	15.4	260	29	116		
678	1	15.2	390	28	182	36	357
678	2	15.6	337	32	155	36	354
678	3	16.0	359	21	167	8	322
678	4	16.5	318	26	144	39	316
678	5	17.1	295	26	142	28	293
678	6	17.9	320	30	152		
678	8	19.5	293	17	146	28	272
678	9	20.4	317	71	109		

(a) Stereopair number 71

(b) Stereopair number 253

(c) Stereopair number 115

(d) Stereopair number 117



(e) Stereopair number 153

(f) Stereopair number 157

(g) Stereopair number 158

(h) Stereopair number 164

(i) Stereopair number 165

(j) Stereopair number 166

(k) Stereopair number 252

Figure A.1: Re-projection of reconstructed zebra-crossing in stereo images.

(a) Stereopair number 94 (b) Stereopair number 95

(c) Stereopair number 314 (d) Stereopair number 315

(e) Stereopair number 316 (f) Stereopair number 317

(g) Stereopair number 319 (h) Stereopair number 320

(i) Stereopair number 321 (j) Stereopair number 322

Figure A.2: Re-projection of reconstructed zebra-crossing in stereo images.

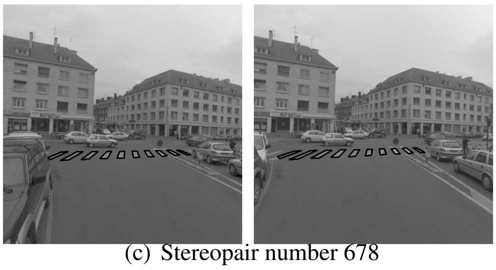
(a) Stereopair number 236
(b) Stereopair number 244

(c) Stereopair number 246
(d) Stereopair number 247

(e) Stereopair number 249
(f) Stereopair number 250

Figure A.3: Re-projection of reconstructed zebra-crossing in stereo images.

(a) Stereopair number 676
(b) Stereopair number 677



(c) Stereopair number 678

Figure A.4: Re-projection of reconstructed zebra-crossing in stereo images.

Bibliography

[Absolute Mapping Solutions Inc.] ABSOLUTE MAPPING SOLUTIONS INC.: *Visat*™. `http://www.amsvisat.com`. – (Accessed 10 August 2007)

[Abuhadrous *et al.* 2004] ABUHADROUS, I. ; AMMOUN, S. ; NASHASHIBI, F. ; GOULETTE, F. ; LAURGEAU, C.: Digitizing and 3D modeling of urban environments and roads using vehicle-borne laser scanner system. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)* Volume 1, Oct 2004, p. 76– 81

[APPLANIX] APPLANIX: *POSLV 200*. `http://www.applanix.com/media/downloads/products/resources/POSLV_200_Datasheet.pdf`. – (Accessed 22 January 2008)

[Baillard and Dissard 2000] BAILLARD, C. ; DISSARD, O.: A Stereo Matching Algorithm for Urban Digital Elevation Models. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 66 (2000), Nr. 9, p. 1119–1128

[Baker and Binford 1981] BAKER, H.H. ; BINFORD, T.O.: Depth from Edge and Intensity-based Stereo. In: *Proceedings of the 7th International Joint Conferences on Artificial Intelligence*, 1981, p. 631–636

[Bauer *et al.* 2002] BAUER, J. ; KLAUS, A. ; KARNER, K. ; ZACH, C. ; SCHINDLER, K.: A semi-automatic city documentation system. In: R.KALLIANY (Editor) ; LEBERL, F. (Editor): *Proc. of the ISPRS Commission III Symposium on Photogrammetric and Computer Vision* Volume 34. Graz, Austria : Institute for Computer Graphics and Vision, Graz University of Technology, sep 2002, p. 22–27

[Bayoud *et al.* 2004] BAYOUD, F.A. ; SKALOUD, J. ; MERMINOD, B.: Photogrammetry-derived navigation parameters for INS Kalman filter updates. In: *Proceedings of the XXth ISPRS Congress* Volume 35 – Part 5B. Istanbul, Turkey, Jul 2004

[Bentrah *et al.* 2004] BENTRAH, O. ; PAPARODITIS, N. ; PIERROT-DESEILLIGNY, M. ; HORAUD, R.: Estimating Sensor Pose from Images of a Stereo Rig. In: *Proceedings of the XXth ISPRS Congress* Volume XXXV part B. Istanbul, Turkey, Jul 2004

[Bentrah 2006] BENTRAH, Ouided: *Reconstruction de la géométrie d'acquisition de séquences d'images acquises par un véhicule en milieu urbain*. Paris, France, Université Paris V, Ph.D. thesis, 2006

[Bertozzi *et al.* 1998a] BERTOZZI, M. ; BROGGI, A. ; FASCIOLI, A.: Stereo Inverse Perspective Mapping: Theory and Applications. In: *Image and Vision Computing* 16 (1998), June, Nr. 8, p. 585–590

[Bertozzi and Broggi 1998] BERTOZZI, Massimo ; BROGGI, Alberto: GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. In: *IEEE Transactions on Image Processing* 7 (1998), Nr. 1, p. 62–81

[Bertozzi *et al.* 1998b] BERTOZZI, Massimo ; BROGGI, Alberto ; FASCIOLI, Alessandra: An extension to the Inverse Perspective Mapping to handle non-flat roads. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*. Stuttgart, Germany, October 1998

[Beucher *et al.* 1990] BEUCHER, S. ; BILODEAU, M. ; YU, X.: Road Segmentation by watershed algorithms. In: *Proceedings of the PROMETHEUS workshop*. Sophia-Antipolis, France, April 1990

[Birge 1985] BIRGE, S. T.: Highway dimension from Photolog. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 51 (1985), Nr. 10, p. 1609–1614

[Bonnifait *et al.* 2006] BONNIFAIT, Philippe ; CHAUMETTE, François ; CHERFAOUI, Véronique ; DHOME, Michel ; JABBOUR, Maged ; LAVEST, Jean-Marc ; LHUILLIER, Maxime ; MALIS, Ezio ; MARTINET, Philippe ; MEIZEL, Dominique ; REMAZEILLES, Anthony ; RENAULT, Stéphane: Le projet BODEGA [2003, 2005]. Paris, April 2006. – URL `http://robea2006.robot.jussieu.fr`. – (Accessed 30 October 2007)

[Borkowski *et al.* 1999] BORKOWSKI, A. ; BURGHARDT, D. ; MEIER, S.: A Fast Snakes Algorithm Using the Tangent Angle Function. In: *Proceedings of the ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*. Munich, Germany, 1999, p. 61–66

[Broggi 1995] BROGGI, A.: Robust real-time lane and road detection in critical shadow conditions. In: *Proceedings of the International Symposium on Computer Vision*. Coral Gables, FL, USA, november 1995, p. 353–358

[Canny 1986] CANNY, J.: A computational approach to edge detection. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (1986), Nov, Nr. 6, p. 679–698

[Chaplin and Chapman 1998] CHAPLIN, B.A. ; CHAPMAN, M.A.: A procedure for 3D motion estimation from stereo image sequences for a mobile mapping system. In: *The International Archives of Photogrammetry and Remote Sensing* Volume 32 (3W1), 1998

[Charbonnier *et al.* 1998] CHARBONNIER, P. ; NICOLLE, P. ; GUILLARD, Y. ; CHARRIER, J.: Road boundaries detection using color saturation. In: *Proceedings of the 9th European Signal Processing Conference*. Rhodes, Greece, Sept 1998, p. 2553–2556

[Chehata *et al.* 2003] CHEHATA, N. ; JUNG, F. ; PIERROT-DESEILLIGNY, M. ; STAMON, G.: A Region-Based Matching Approach for 3D-Roof Reconstruction from High Resolution satellite Stereo Pairs. In: *Proceedings of the Digital Image Computing: Techniques and Applications* Volume II. Sydney, Australia, dec 2003, p. 889–898

[Chen and Medioni 1990] CHEN, J. S. ; MEDIONI, G.: Parallel multiscale stereo matching using adaptive smoothing. In: *Proceedings of the first European Conference on Computer vision (ECCV)*. New York, NY, USA : Springer-Verlag New York, Inc., 1990, p. 99–103

[Cheng *et al.* 2007] CHENG, Wang ; HASSAN, Taher ; EL-SHEIMY, Naser ; LAVIGNE, Martin: ARVEE, Automatic Road Geometry Extraction System for Mobile Mapping. Padua, Italy, May 2007

[Chiu and Lin 2005] CHIU, Kuo-Yu ; LIN, Sheng-Fuu: Lane Detection using Color-Based Segmentation. In: *Proceedings of the Intelligent Vehicles Symposium (IVS)*, 2005, p. 6–8

[Cipolla *et al.* 1999] CIPOLLA, R. ; DRUMMOND, T. ; ROBERTSON, D.: Camera calibration from vanishing points in images of architectural scenes. In: *Proceedings of the The British Machine Vision Conference (BMVC)*, 1999

[Collado *et al.* 2005] COLLADO, J. M. ; HILARIO, C. ; ESCALERA, A. de la ; ARMINGOL, J. M.: Detection and classification of road lanes with a frequency analysis. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, Jun 2005, p. 78–83

[Dahyot *et al.* 2000] DAHYOT, R. ; CHARBONNIER, P. ; HEITZ, F.: Robust Visual Recognition of Colour Images. In: *IEEE proceedings of the conference on Computer Vision and Pattern Recognition* Volume 1. Hilton Head Island, USA, June 2000, p. 685–690

[Deriche 1987] DERICHE, R.: Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. In: *The International Journal of Computer Vision* 1 (1987), may, Nr. 2, p. 167–187

[Deriche *et al.* 1992] DERICHE, R. ; VAILLANT, R. ; FAUGERAS, O.: *From Noisy Edges Points to 3D Reconstruction of a Scene : A Robust Approach and Its Uncertainty Analysis*. Volume 2. p. 71–79. In: *In Theory and Applications of Image Analysis* Volume 2, World Scientific, 1992. – Series in Machine Perception and Artificial Intelligence

[Devernay 1995] DEVERNAY, Frederic: A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy / INRIA. 1995 (RR-2724). – Research Report

[Dietmayer *et al.* 2005] DIETMAYER, K. C. J. ; KAEMPCHEN, N. ; FUERSTENBERG, K. ; KIBBEL, J.: Roadway Detection and Lane Detection using Multilayer Laserscanner. In: *Proceedings of Advanced Microsystems for Automotive Applications*. Berlin, Germany, March 2005

[Douglas and Peucker 1973] DOUGLAS, David H. ; PEUCKER, Thomas K.: Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. In: *The Canadian Cartographer* 10 (1973), Nr. 2, p. 112–122

[El-Sheimy 1996] EL-SHEIMY, N.: A Mobile Multi-Sensor System For GIS Applications In Urban Centers. In: *Proceedings of The International Society for Photogrammetry and Remote Sensing (ISPRS) 1996, Commission II* Volume XXXI, Part B2. Vienna, Austria, July 1996, p. 95–100

[Faugeras *et al.* 1996] FAUGERAS, O. ; HOTZ, B. ; MATHIEU, H. ; VIÉVILLE, T. ; ZHANG, Z. ; FUA, P. ; THÉRON, E. ; MOLL, L. ; BERRY, G. ; VUILLEMIN, J. ; BERTIN, P. ; PROY, C.: Real time correlation based stereo: algorithm implementations and applications. In: *The International Journal of Computer Vision* (1996)

[Flamanc *et al.* 2003] FLAMANC, D. ; MAILLET, G. ; JIBRINI, H.: 3-D City Models: an Operational Approach using Aerial Images and Cadastral Maps. In: H. EBNER AND C. HEIPKE AND H. MAYER AND K. PAKZAD (Editor): *ISPRS Conference Photogrammetric Image Analysis (PIA)* Volume XXXIV, Part 3/W834. Munich, Germany : Institute for Photogrammetry and GeoInformation University of Hannover, Germany, Sep 2003, p. 53–58

[Frueh and Zakhor 2001] FRUEH, C. ; ZAKHOR, A.: 3D Model Generation for Cities Using Aerial Photographs and Ground Level Laser Scans. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, p. II:31–38

[Früh and Zakhor 2005] FRÜH, Christian ; ZAKHOR, Avideh: Data Processing Algorithms for Generating Textured 3D Building Faccade Meshes From Laser Scans and Camera Images. In: *International Journal of Computer Vision* 61 (2005), Nr. 2, p. 159–184

[Gagalowicz 1994] GAGALOWICZ, A.: *Data Visualisation.* Chap. Modeling Complex indoor scenes using an analysis/synthesis framework, Academic Press, 1994

[GeoAutomation] GEOAUTOMATION: *website of GeoAutomation company.* `http://homes.esat.kuleuven.be/~vergauwe/geoautomation/index.html`. – (Accessed 12 February 2008)

[Giraudon *et al.* 1992] GIRAUDON, G. ; GARNESSON, P. ; MONTESIONOS, P.: MESSIE: un système multi spécialistes en vision. Application à l'interprétation en imagerie aérienne = MESSIE: a multi expert system in computer vision. Application for aerial imagery interpretation. In: *Traitement du signal* 9 (1992), Nr. 5, p. 403–419

[Gontran *et al.* 2006] GONTRAN, H. ; SKALOUD, J. ; JANVIER, N.: Open-source software-operated CMOS camera for real-time mapping. In: *ISPRS Commission I Symposium, From sensors to Imagery.* Paris, France, 2006

[Gontran *et al.* 2003] GONTRAN, Hervé ; SKALOUD, Jan ; GILLIÉRON, Pierre-Yves: A mobile mapping system for road data capture via a single camera. In: *4th International symposium on mobile mapping technology.* Kunming China, August 2003

[Gonzalez and Ozguner 2000] GONZALEZ, J.P. ; OZGUNER, U.: Lane Detection Using Histogram-Based Segmentation and Decision Trees. In: *Proceedings of the IEEE Intelligent Transportation Systems Conference.* Dearborn, MI, USA, 2000, p. 346–351

[Goulette *et al.* 2006] GOULETTE, F. ; NASHASHIBI, F. ; ABUHADROUS, I. ; AMMOUN, S. ; LAURGEAU, C.: An Integrated on-board laser range sensing system for on-the-way city and road modelling. In: *Proceedings of the ISPRS Commission I Symposium, From sensors to Imagery.* Paris, France, 2006

[Habib *et al.* 1999] HABIB, A.F. ; UEBBING, R. ; NOVAK, K.: Automatic Extraction of Road Signs from Terrestrial Color Imagery. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensin* 65 (1999), May, Nr. 5, p. 597–601

[Habib 2000] HABIB, Aymen F.: Matching road edges in stereo-image sequences using data association techniques. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 66 (2000), January, Nr. 1, p. 91–98

[Han and Park 2000] HAN, Joon H. ; PARK, Jong-Seung: Contour Matching Using Epipolar Geometry. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), Nr. 4, p. 358–370

[Hannah 1989] HANNAH, M-J.: A system for stereo image matching. In: *ASPRS journal of Photogrammetric Engineering and Remote Sensing* 55 (1989), Nr. 12, p. 1765–1770

[Harris and Stephens 1988] HARRIS, C. ; STEPHENS, M.: A combined corner and edge detector. In: *Proceedings of the 4th Alvey Vision Conference*, 1988, p. 147–152

[He 2002] HE, G.: Design and Application of the GPSVision Mobile Mapping System. In: *Proceedings of The International Society for Photogrammetry and Remote Sensing (ISPRS) , Commission II, Working Group 1* Volume XXXXIV, Part 2. Vienna, Austria, July 2002, p. 163–168

[He and Novak 1992] HE, G.P. ; NOVAK, K.: Automated Analysis of Highway Features from Digital Stereo-Images. In: *Proceedings of The International Society for Photogrammetry and Remote Sensing (ISPRS) , Commission II, Working Group 1* Volume XXIX, Part B3. Washington, DC, USA, Aug 1992, p. 119–124

[Heipke *et al.* 1997] HEIPKE, C. ; MAYR, H. ; WIEDEMANN, C. ; JAMET, O.: Evaluation of Automatic Road Extraction. In: *ISPRS* Volume 32, 1997

[Heipke *et al.* 1995] HEIPKE, C. ; STEGER, C. ; MULTHAMMER, R.: A Hierarchical Approach to Automatic Road Extraction from Aerial Imagery. In: *SPIE* 2486 (1995), p. 222–231

[Horaud and Skordas 1989] HORAUD, R. ; SKORDAS, T.: Stereo Correspondence Through Feature Grouping and Maximal Cliques. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989), Nr. 11, p. 1168–1180

[Hu *et al.* 2004] HU, Minghao ; YANG, Wenjie ; REN, Mingwu ; YANG, Jingyu: A vision based road detection algorithm. In: *Proceedings of the IEEE Conference on Robotics, Automation and Mechatronics* Volume 2. Computer Science Department, Nanjing University of Science & Technology, China, Dec 2004, p. 846– 850

[Hunt and Nelson 1996] HUNT, Galen C. ; NELSON, Randal C.: Lineal Feature Extraction by Parallel Stick Growing. In: *Workshop on Parallel Algorithms for Irregularly Structured Problems*, 1996, p. 171–182

[Ieng *et al.* 2004] IENG, Sio-Song ; TAREL, Jean-Philippe ; CHARBONNIER, Pierre: Estimation robuste pour la détection et le suivi par caméra. In: *Traitement du Signal* 21 (2004), Nr. 3, p. 205–226

[Ieng *et al.* 2003] IENG, Sio-Song ; TAREL, Jean-Philippe ; LABAYRADE, Raphael: On the design of a single lane-markings detectors regardless the on-board camera's position. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, June 2003, p. 564 – 569

[Jang and Hong 2002] JANG, Jeong-Hun ; HONG, Ki-Sang: Fast line segment grouping method for finding globally more favorable line segments. In: *Pattern Recognition* 35 (2002), Nr. 10, p. 2235–2247

[Jeong and Nedevschi 2005] JEONG, Pangyu ; NEDEVSCHI, Sergiu: Efficient and Robust Classification Method Using Combined Feature Vector for Lane Detection. In: *IEEE Transactions on Circuits and Systems. Video Technology* 15 (2005), Nr. 4, p. 528–537

[Jung and Kelber 2004] JUNG, Claudio R. ; KELBER, Christian R.: A Robust Linear-Parabolic Model for Lane Following. In: *Proceedings of the Computer Graphics and Image Processing, XVII Brazilian Symposium (SIBGRAPI).* Washington, DC, USA : IEEE Computer Society, 2004, p. 72–79

[Kim *et al.* 2006] KIM, Jin G. ; HAN, Dong Y. ; YU, Ki Y. ; KIM, Yong I. ; RHEE, Sung M.: Efficient extraction of road information for car navigation applications using road pavement markings obtained from aerial images. In: *Canadian Journal of Civil Engineering* 33 (2006), October, Nr. 10, p. 1320–1331(12)

[Kluge 1994] KLUGE, K.: Extracting road curvature and orientation from image edge points without perceptual grouping into features. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, Oct 1994, p. 109 – 114

[Kluge and Johnson 1995] KLUGE, K. ; JOHNSON, G.: Statistical characterization of the visual characteristics of painted lane markings. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, 1995, p. 488–493

[Kluge and Lakshmanan 1995] KLUGE, K. C. ; LAKSHMANAN, S.: A deformable template approach to lane detection. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS).* Detroit, MI, USA, 1995, p. 54–59

[Kreucher and Lakshmanan 1999] KREUCHER, C. ; LAKSHMANAN, S.: LANA: a lane extraction algorithm that uses frequency domain features. In: *IEEE Transactions on Robotics and Automation* 15 (1999), Apr, Nr. 2, p. 343 – 350

[Labayrade *et al.* 2002] LABAYRADE, R. ; AUBERT, D. ; TAREL, J. P.: Real time obstacle detection in stereovision on non flat road geometry through *v-disparity* representation. In: *Proceedings of the IEEE Intelligent Vehicle Symposium* Volume 2, Jun 2002, p. 646–651

[Lai and Yung 2000] LAI, A.H.S. ; YUNG, N.H.C.: Lane Detection by Orientation and Length Discrimination. In: *IEEE Transaction on Systems, Man, and Cybernetics Part B* 30 (2000), August, Nr. 4, p. 539–548

[Lapucha *et al.* 1990] LAPUCHA, D. ; SCHWARZ, K.P. ; CANNON, M.E. ; MARTELL, H.: The use of INS/GPS in a highway survey system. In: *Proceedings of the IEEE, Position Location and Navigation Symposium (PLANS).* Las-Vegas, USA, Mar 1990, p. 413–420

[Laurent *et al.* 1997] LAURENT, J. ; TALBOT, M. ; DOUCET, M.: Road Surface Inspection using Laser Scanners Adapted for the High Precision Measurements of Large Flat Surfaces. In: *Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM)*. Washington, DC, USA : IEEE Computer Society, 1997, p. 303

[Li *et al.* 1996] LI, R. ; CHAPMAN, M.A. ; QIAN, L. ; XIN, Y. ; TAO, C.: Mobile mapping for 3D GIS data acquisition. In: *ISPRS* Volume XXXI, Part B2. Vienna, Austria, July 1996, p. 232–237

[Luong *et al.* 1995] LUONG, Q.-T. ; WEBER, J. ; KOLLER, D. ; MALIK, J.: An integrated stereo-based approach to automatic vehicle guidance. In: *Proceedings of the Fifth International Conference on Computer Vision (ICCV)*. Washington, DC, USA : IEEE Computer Society, 1995, p. 52–57

[Ma and Thonnat 1993] MA, Ruihua ; THONNAT, Monique: A Robust and efficient stereo matching algorithm / INRIA - Sophia Antipolis. France, Feb 1993 (1860). – Research Report

[Marr and Poggio 1977] MARR, D. ; POGGIO, T.: A Theory of Human Stereo Vision. Cambridge, MA, USA : Massachusetts Institute of Technology, 1977. – Research Report

[Mayer *et al.* 2006] MAYER, H. ; HINZ, S. ; BACHER, U. ; BALTSAVIAS, E.: A Test of Automatic Road Extraction Approaches. In: *Proceedings of the The International Society for Photogrammetry and Remote Sensing (ISPRS) Commission III Photogrammetric Computer Vision (PCV)*, 2006, p. 209–214

[Mayer *et al.* 1997] MAYER, H. ; LAPTEV, I. ; BAUMGARTNER, A. ; STEGER, C.: Automatic Road Extraction Based on Multiscale Modeling, context and snakes. In: *International Archives of Photogrammetry and Remote Sensing* Volume XXXII - 3-2W3, 1997, p. 47–56

[Mayhew and Frisby 1981] MAYHEW, John E. W. ; FRISBY, John P.: Psychophysical and Computational Studies Towards a Theory of Human Stereopsis. In: *Artificial Intelligence* 17 (1981), Nr. 1-3, p. 349–385

[McCall and Trivedi 2004] MCCALL, J.C. ; TRIVEDI, M.M.: An integrated, robust approach to lane marking detection and lane tracking. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, 2004, p. 533–537

[McKeown *et al.* 2000] MCKEOWN, D. ; BULWINKLE, T. ; COCHRAN, S. ; HARVEY, W. ; MCGLONE, C. ; SHUFELT, J.A.: Performance evaluation for automatic feature extraction. In: SCHENK, T. (Editor) ; VOSSELMAN, G. (Editor): *Proc. of the XIXth ISPRS Congress* Volume 33. Amsterdam, The Netherlands, Jul 2000, p. 379–394

[Moffitt and Mikhail 1980] MOFFITT, F. ; MIKHAIL, E.: *Photogrammetry 3rd Edition*. Chap. 7, Harper and Row New York, 1980

[Mohammed *et al.* 1979] MOHAMMED, J.L. ; LECLERC, Y.G. ; ZUCKER, S.W.: Continuous Relaxation and Local Maxima Selection: Conditions for Equivalence. In: *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, 1979, p. 1014–1016

[Monti *et al.* 2005] MONTI, Gianni ; LINDNER, Christoph ; PUENTE LEÓN, Fernando ; KOCH, Alexander W.: Consumer graphics cards for fast image processing based on the Pixel Shader 3.0 standard. In: *Proceedings of the EOS Conference on Industrial Imaging and Machine Vision*, 13-15 June 2005, p. 15–21

[Moravec 1977] MORAVEC, H.: Towards Automatic Visual Obstacle Avoidance. In: *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, 1977, p. 584

[Nedevschi *et al.* 2006] NEDEVSCHI, S. ; ONIGA, F. ; DANESCU, R. ; GRAF, T. ; SCHMIDT, R.: Increased Accuracy Stereo Approach for 3D Lane Detection. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS).* Tokyo, Japan, 2006, p. 42–49

[Nedevschi *et al.* 2004] NEDEVSCHI, S. ; R..SCHMIDT ; GRAF, T. ; DANESCU, R. ; FRENTIU, D. ; MARITA, T. ; ONIGA, F. ; POCOL, C.: 3D Lane Detection System Based on Stereovision. In: *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC).* Washington, USA, 2004, p. 161–166

[Nicolas Simond] NICOLAS SIMOND: *website of INRIA*. `http://www-sop.inria.fr/icare/personnel/nsimond/index.html`. – (Accessed 13 August 2007)

[Novak and Bossler 1995] NOVAK, K. ; BOSSLER, J.D.: Development and Application of the Highway Mapping System of Ohio State University. In: *The Photogrammetric Record* 15 (1995), April, Nr. 85, p. 123–134

[Okutomi *et al.* 2002] OKUTOMI, M. ; NAKANO, K. ; MARUYAMA, J. ; HARA, T.: Robust estimation of planar regions for visual navigation using sequential stereo images. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002, p. 3321–3327

[Panton 1978] PANTON, D.J.: A Flexible Approach to Digital Stereo Mapping. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 44 (1978), December, Nr. 12, p. 1499–1512

[Paparoditis *et al.* 2005] PAPARODITIS, N. ; BENTRAH, O. ; PENARD, L. ; TOURNAIRE, O. ; SOHEILIAN, B. ; DEVEAU, M.: Automatic 3D Recording and Modeling of Large Scale Cities: the Archipolis Project. In: *Modelling and Visualization of Cultural Heritage*, 2005, p. 227–235

[Poggio *et al.* 1985] POGGIO, Tomaso ; TORRE, Vincent ; KOCH, Christof: Computational vision and regularization theory. In: *Nature* 317 (1985), Nr. 26, p. 314–319

[Pollard *et al.* 1985] POLLARD, S.B. ; MAYHEW, J.E.W. ; FRISBY, J.P.: PMF: A stereo correspondence algorithm using a disparity gradient limit. In: *Perception* 14 (1985), p. 449–470

[Risack *et al.* 1998] RISACK, R. ; KLAUSMANN, P. ; KRÜGER, W. ; ENKELMANN, W.: Robust lane recognition embedded in a real-time driver assistance system. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*. Stuttgart, Germany, October 1998, p. 35–40.

[Roadware Group, Inc.] ROADWARE GROUP, INC.: *Automatic Road Analyzer (ARAN©)*. `http://www.roadware.com`. – (Accessed 5 April 2007)

[Robles-Kelly and Hancock 2000] ROBLES-KELLY, Antonio ; HANCOCK, Edwin R.: Grouping Line-segments using Eigenclustering. In: *Proceedings of the The British Machine Vision Conference (BMVC)*, 2000

[Roncella and Forlani 2006] RONCELLA, R. ; FORLANI, G.: Automatic lane parameters extraction in mobile mapping sequences. In: *Proceedings of the The International Society for Photogrammetry and Remote Sensing (ISPRS) Commission V Symposium, Image Engineering and Vision Metrology*. Dresden, Germany, Sep 2006

[Roy and Cox 1998] ROY, Sébastien ; COX, Ingemar J.: A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem. In: *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA : IEEE Computer Society, 1998, p. 492. – ISBN 81-7319-221-9

[Royer *et al.* 2005] ROYER, E. ; BOM, J. ; DHOME, M. ; THUILOT, B. ; LHUILLIER, M. ; MARMOITON, F.: Outdoor autonomous navigation using monocular vision. In: *International Conference on Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ*, August 2005, p. 1253–1258

[Ruskoné 1996] RUSKONÉ, R.: *Extraction automatique du réseau routier par interprétation locale du contexte: application à la production de données cartographiques*, Université de Marne-la-Vallée, Ph.D. thesis, 1996

[Sahoo *et al.* 1992] SAHOO, P.K. ; FARAG, A.A. ; YEAP, Y.-P.: Threshold selection based on histogram modeling. In: *IEEE International Conference on Systems, Man and Cybernetics* Volume 1, 18-21 Oct 1992, p. 351 – 356

[Sappa *et al.* 2007] SAPPA, A. ; DORNAIKA, F. ; GERÓNIMO, D. ; LÓPEZ, A.: Efficient On-Board Stereo Vision Pose Estimation. In: *Computer Aided Systems Theory* Volume 4739, Springer Verlag, September 2007, p. 1183–1190. – (selected paper from EUROCAST 2007)

[Sawano and Okada 2004] SAWANO, H. ; OKADA, M.: Road extraction by snake with inertia and differential features. In: *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)* Volume 4, august 2004, p. 380–383

[Scharstein and Szeliski 2002] SCHARSTEIN, D. ; SZELISKI, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondance Algorithms. In: *International Journal of Computer Vision* 47 (2002), april, Nr. 1-3, p. 7–42

[Schmid and Zisserman 1997] SCHMID, C. ; ZISSERMAN, A.: Automatic line matching across views. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR). San Juan, Puerto Rico : IEEE Computer Society, Jun 1997, p. 666–671

[Schreiber *et al.* 2005] SCHREIBER, D. ; ALEFS, B. ; CLABIAN, M.: Single camera lane detection and tracking. In: *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC)*. Vienna, Austria, 2005, p. 114–119

[Schuster and Weidner 2003] SCHUSTER, H-F. ; WEIDNER, U.: A new Approach Towards Quantitative Quality Evaluation of 3D Building Models. In: SCHIEWE, J. (Editor) ; HAHN, L. (Editor) ; MADDEN, M. (Editor) ; SESTER, M. (Editor): *Proceeding of the ISPRS Commission IV Joint Workshop Challenges in Geospatial Analysis, Integration and Visualization II'*. Stuttgart, Germany, Sep 2003, p. 156–163

[Schwarz *et al.* 1990] SCHWARZ, K.P. ; LAPUCHA, D. ; CANNON, M.E. ; MARTELL, H.: The Use of GPS/INS in a Highway Inventory System. In: *FIG XIX International Meeting* Volume 5. Helsinki, Finland, 1990, p. 237–249

[Schwarz *et al.* 1993] SCHWARZ, K.P. ; MARTELL, H.E. ; EL-SHEIMY, N. ; R. LI, M.A. C. ; COSANDIER, D.: VIASAT - A mobile highway survey system of high accuracy. In: *Proceedings of the IEEE-IEE Vehicle Navigation and Information Systems Conference*, Oct 1993, p. 476 – 481

[Se and Brady 2003] SE, Stephen ; BRADY, Michael: Road feature detection and estimation. In: *Machine Vision and Applications (MVA)* 14 (2003), Nr. 3, p. 157–165

[Serra and Berthod 1994] SERRA, Bruno ; BERTHOD, Marc: Subpixel contour matching using continuous dynamic programming. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* . Seattle : IEEE Computer Society, 1994, p. 202–207

[Simond and Rives 2004a] SIMOND, Nicolas ; RIVES, Patrick: Détection robuste du plan de la route en milieu urbain. In: *RFIA*, 2004

[Simond and Rives 2004b] SIMOND, Nicolas ; RIVES, Patrick: Trajectography of an Uncalibrated Stereo Rig in Urban Environments. In: *IEEE RSJ/International conference on Intelligent Robot and System (IROS'04)*. Sendai, Japan, Sept. 28-Oct. 2 2004, p. 3381–3386

[Sobel and Feldman 1968] SOBEL, I. ; FELDMAN, G.: *A 3×3 isotropic gradient operator for image processing*. 1968. – Never published but presented at a talk at the Stanford Artificial Project

[Stoica *et al.* 2004] STOICA, R. ; DESCOMBES, X. ; ZERUBIA, J.B.: A Gibbs Point Process for Road Extraction from Remotely Sensed Images. In: *International Journal of Computer Vision* 57 (2004), May, Nr. 2, p. 121–136

[Tao *et al.* 1998] TAO, C. ; LI, R.X. ; CHAPMAN, M.A.: Automatic Reconstruction of Road Centerlines from Mobile Mapping Image Sequences. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 64 (1998), July, Nr. 7, p. 709–716

[Tao 1997] TAO, C.V.: *Automatic approaches to object measurement and feature extraction from georeferenced mobile mapping image sequences.* Calgary, Canada, University of Calgary, Ph.D. thesis, 1997

[Tao 2001] TAO, C.V.: Database-Guided Automatic Inspection of Vertically Structured Transportation Objects from Mobile Mapping Image Sequences. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensin* (2001), December, Nr. 12, p. 1401–1409

[Tarel *et al.* 2007] TAREL, J.-P. ; IENG, S.-S. ; CHARBONNIER, P.: Accurate and Robust Image Alignment for Road Profil Reconstruction. In: *Proceedings of IEEE International Conference on Image Processing (ICIP'07)* Volume V. San Antonio, Texas, USA, 2007, p. 365–368. – http://perso.lcpc.fr/tarel.jean-philippe/icip07.html

[Tournaire 2007] TOURNAIRE, Olivier: *Extraction 3D de marquages routiers à partir d'images aériennes multi-vues et quelques applications.* Paris, France, Université Paris-Est, Ph.D. thesis, 2007

[Tournaire *et al.* 2006] TOURNAIRE, Olivier ; PAPARODITIS, Nicolas ; JUNG, Franck ; CERVELLE, Bernard: 3D road-marks reconstruction from multiple calibrated aerial images. In: *Proceedings of the The International Society for Photogrammetry and Remote Sensing (ISPRS) Commission III Photogrammetric Computer Vision (PCV).* Germany, 2006

[Transport Ministry and Interior Ministry 1988] *Instruction interministérielle sur la signalisation routière. Septième partie: Marques sur chaussée.* Paris, France, 1988

[Wang *et al.* 2004] WANG, Yue ; TEOH, Eam K. ; SHEN, Dinggang: Lane detection and tracking using B-Snake. In: *Image and Vision Computing* 22 (2004), April, Nr. 4, p. 269–280

[Wood 1983] WOOD, G. A.: Realities of automatic correlation problem. In: *ASPRS journal of Photogrammetric Engineering & Remote Sensing* 49 (1983), Apr., p. 537–538

[Yu *et al.* 2003] YU, Q. ; ARAÚJO, H. ; WANG, H.: Stereo-Vision Based Real time Obstacle Detection for Urban Environments. In: *Proceedings of the Intternational Conference on Advenced Robotics (ICAR).* Coimbra, Portugal, July 2003, p. 1671–1676

[Yu *et al.* 2007] YU, Si-Jie ; SUKUMAR, Sreenivas R. ; KOSCHAN, Andreas F. ; PAGE, David L. ; ABIDI, Mongi A.: 3D reconstruction of road surfaces using an integrated multi-sensory approach. In: *Optics and Lasers in Engineering* 45 (2007), July, Nr. 7, p. 808–818

[Yu *et al.* 1992] YU, X. ; BEUCHER, S. ; BILODEAU, M.: Road tracking, lane segmentation and obstacle recognition by mathematical morphology. In: *Proceedings of the IEEE Intelligent Vehicles.* Detroit, USA, 1992, p. 166–170

[Zhang 2003] ZHANG, C.: *Updating of cartographic road databases by image analysis*, Institute of Geodesy and Photogrammetry, Zurich, Ph.D. thesis, 2003

[Zhang and H.H. 1994] ZHANG, Jinyou ; H.H., Nagel: Texture-based segmentation of road images. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*. Germany, Oct 1994, p. 260– 265

[Zhao and Shibasaki 2001] ZHAO, H. ; SHIBASAKI, R.: Reconstructing Urban 3D Model Using Vehicle-Borne Laser Range Scanners. In: *Third International Conference on 3-D Digital Imaging and Modeling (3DIM)* 00 (2001), p. 349. ISBN 0-7695-0984-3

[Zhao and Shibasaki 2003] ZHAO, H. ; SHIBASAKI, R.: Reconstructing a textured CAD model of an urban environment using Vehicle-borne Laser Range Scanners and Line Cameras. In: *Machine Vision and Applications (MVA)* 14 (2003), Nr. 1, p. 35–41