THE UNIVERSITY OF CALGARY

Automated Approaches to Object Measurement and Feature Extraction from Georeferenced Mobile Mapping Image Sequences

by

Chuang Tao

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

OCTOBER, 1997

ABSTRACT

The rapid increase in the availability of data, contributed by the full-scale development of multi-platform and multi-sensor data acquisition systems, demands a rapid, accurate and highly automated information processing and analysis strategy. This is a great challenge that is being faced in the geomatics industry. Mobile mapping technology represents a new trend towards integrated data acquisition. With the enormous data volume collected by mobile mapping systems, rapid and accurate information extraction becomes a critical issue.

The principal objective of the research endeavors to improve the capability and exploit the potential of effective and efficient information extraction from terrestrial mobile mapping image sequences. Extensive exploration of the combined technology of computational vision and digital photogrammetry forms the basis of the methodology used in this research. The emphasis has been placed on automation as well as accuracy of the approaches to object measurement and feature extraction using mobile image sequences.

The dissertation consists of two parts. In the first part, an overview of mobile mapping technology is provided, visual motion analysis of mobile mapping image sequences is addressed, and quantitative geometric analysis of mobile mapping imaging systems is conducted. In the second part, three automated approaches to information processing using mobile mapping image sequences are developed and evaluated. Three approaches are, object measurement based on multiple image matching, road centerline reconstruction using

"shape from image sequences", and vertical line feature extraction based on feature detection and correspondence. A large number of tests are carried out to evaluate the approaches in terms of reliability, accuracy and efficiency. The proposed approaches have been implemented and used as core modules in ImagExpert, a software package for automatic processing of mobile mapping images.

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my supervisors. Dr. Michael Chapman, my current supervisor, has been a marvelous advisor to me all through my graduate studies. His critical guidance and insightful comments at every stage of the research have been valuable. Dr. Rongxing Li, my previous supervisor, provided me the unique opportunity to pursue this research. His valuable guidance and constructive suggestions have influenced my thesis studies.

Special thanks go to Dr. K. P. Schwarz, who has led the VISAT project from concept to implementation. His valuable advice and discussions are appreciated. Appreciation also goes to my fellow graduate students and friends, L. Qian, B. Chaplin, W. Zou, Y. Xin, A. Alesheikh, H. Ebadi and D. Lichti for fruitful discussions and kind help, and for making my studies at the University of Calgary enjoyable.

Financial support from the Canadian Natural Science and Engineering Research Council (NSERC), the Dean's Special Doctoral Fellowships from the University of Calgary, the SPIE NIKON international scholarship from the International Society for Optical Engineering, and GEOFIT Inc., Laval, Quebec is gratefully acknowledged. The support provided by the Robert E. Altenhofen Memorial Award by the American Society for Photogrammetry and Remote Sensing, ISPRS Best Young Author Prize by the International Society for Photogrammetry and Remote Sensing, Lockheed Martin Best Paper Award by the International Society for Optical Engineering, Best Paper Award by the Chinese

Professional Association of Geographic Information Systems, and Jim Davery Award for Best Doctoral Paper by the Canadian Transportation Research Forum is also greatly acknowledged.

Dedicated

to my wife, Xiaoxue (Sherry), for her persistent understanding,

sacrifices and for taking care of our new baby girl, Luncida Yanran, and

to my parents, Huiying Fang and Benzao Tao,

for their continuous support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

"One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity."

---- *J.W. Gibbs*

1.1 Multi-Platform/Multi-Sensor Integrated Data Acquisition

It has been demonstrated that mapping science has been steadily stepped into a digital mapping era. The core technologies of remote sensing, photogrammetry, geographic information systems (GIS) and spatial positioning are becoming fully integrated, resulting in the tremendous expansion and rapid growth of the mapping markets. Modern digital mapping technology is characterized by the capabilities of ***multi-discipline combination, multi-platform compensation, multi-sensor integration and multi-data fusion.***

The multi-platform, and multi-sensor integrated technology have established a trend towards spatial data acquisition. Multi-sensor systems can be mounted on various platforms, such as satellites, aircrafts or helicopters, land-based vehicles, water-based vessels, and even hand-carried by individual surveyors. As a result, every vehicle or individual surveyor becomes a potential data collector, responsible for globally integrated

data acquisition. The recent development of land-based mobile mapping systems represents a typical application of this integrated technology.

1.2 Background Information on Mobile Mapping Systems

Photo-Logging

In the 1970's, many highway departments had been using the so-called ***photo-logging*** technology to monitor pavement performance, signing, maintenance effectiveness, encroachments, etc. These services are usually required at intervals of about two years. Often 35-mm film cameras were used to capture photos through the windshield of a van-type vehicle. An inertial device (e.g., gyroscopes and accelerometers) and a wheel counter were employed to determine the instantaneous positions of the captured photographs. Consequently, each photo was stamped with time and geographic position information. These photos were stored mainly as a pictorial record of highway performance, and investigated, usually using a back-screen projector at about a 10:1 enlargement. Due to the rather poor accuracy of vehicle positioning and only a single camera configuration in these systems, the functionality of 3-D object measurement was not available. However, the use of consecutive photos to achieve 3-D measurement was possible (Birge, 1985).

Video-Logging

With the advent of the Global Positioning System (GPS) technology in the late 1980's, the vehicle navigation capability has been greatly enhanced. GPS technology, along with video imaging technology, motivated the development of video-logging systems. Video-logging systems offer a fast and low-cost approach to highway inventory (Lapucha, 1990; Schwarz et al., 1990). The collected video images can be georeferenced with respect to a global coordinate system by using continuous GPS navigation and positioning techniques. The turn-around time of data processing is reduced since no film processing is involved. Furthermore, the digitally georeferenced video data allow quick retrieval and effective management of the image data collected. The capability of interpretation of highway video data is also strengthened through the use of image processing technology. Video-logging has become a widely accepted and utilized method in most departments of transportation. Visual recording and feature documenting along road corridors remain the major purpose of these kinds of systems.

Mobile Mapping

The development of mobile mapping systems from video-logging systems was realized by the early efforts of the Center for Mapping, The Ohio State University (GPSVan™ system), U.S.A; GEOSPAN Corp. (GeoVAN™ system), U.S.A; and the joint venture of The University of Calgary and GEOFIT Inc. (VISAT™ system), Canada (Bossler et al., 1991; and Schwarz et al., 1993b). A common feature of these systems is that more than one

camera is mounted on the mobile platform, offering the capability of stereo imaging for 3-D data acquisition. Accurate georeferencing of the digital imagery collected has been accomplished by advanced navigation and positioning techniques. Multiple navigation sensors, GPS, Inertial Navigation System (INS) and dead-reckoning (DR), are fully combined for direct georeferencing. The ground control required for traditional mapping is largely eliminated, thus, leading to the innovative concept of "mobile mapping". Centimeter accuracy of vehicle positioning is achieved, and meter or sub-meter 3-D coordinate accuracy of objects derived from the georeferenced image sequences are available.

New types of cameras with fully digital, high resolution and large sensor capabilities are commercially available and can be applied in mobile mapping systems. Another advantage of mobile mapping systems is that the data link to a geospatial database is easy and straightforward. The collected geometric and attribute information can be directly used to build and update a database. With the development of fast communication and image compression technologies, real-time image data link from a field mobile mapping system to an office GIS can be realized. Furthermore, such data can be disseminated and accessed through a widely distributed internet. The fusion of image, vector feature and base map data could be dealt with in a multi-media GIS. There is no doubt that mobile mapping systems define a new generation of modern digital mapping system family.

Table1.1 provides an overview of video-logging and mobile mapping systems (Bossler et al., 1991; Chapman and Baker, 1996; El-Sheimy and Schwarz, 1993; He, 1996; Hock et al., 1995; Li, 1997; Mekuria, 1995; Pottle, 1995). Most systems integrate multiple navigation

and mapping sensors, such as GPS, INS, DR, gyro, barometer, CCD camera, Laser. Among them, GPS and CCD cameras are currently the principal navigation and mapping sensors, respectively.

Table 1.1 Overview of Video-Logging and Mobile Mapping Systems

System	Developer/Researcher	Navigation sensors	Mapping sensors
ARAN	Roadware Corp., Ontario, Canada	Accelerometers/IMU/ DGPS	1 VHS, 2 CCD, Laser
DGI	Data Chromatics, Inc., USA	GPS	1 video camera
Geografix	Geografix sys., Singapore	GPS/DR	3 video cameras
GeoVAN	GeoSpan Corp., USA	GPS/DR	10 VHS, voice recorder
GIM	NAVSYS Corp., CO, USA	WA-DGPS/IMU	1 CCD
GPSVan	The Ohio State University, Columbus, OH, USA	GPS/Gyro/wheel counter	2 CCD, voice recorder
GPSVision	Lambda Tech. Int. Inc., USA	GPS, INS	2 Color CCD, optical disc
KiSS	Univ. of Bundeswher Munich and GeoDigital, Germany	GPS/IMU/Inclination Odometer/Barometer	1 SVHS, 2 BW CCD, voice recorder
ON-SIGHT	TransMap Corp., OH, USA	GPS/INS	4 color CCD
RGIAS	Rowe Surveying and Engg, Inc., Alabama, USA	GPS	Video, Laser
TruckMAP	John E. Chance and Associate, Inc., LA, USA	GPS/multi-antenna gyro/WA-DGPS	Laser range finder, 1 video camera
VISAT	The Univ. of Calgary and Geofit Inc., Canada	GPS/INS/ABS	8 BW CCD, 1 color SVHS

This research is based on the mobile mapping system, VISAT (**V**ideo imaging system, an **I**nertial navigation system, and the NavStar Global **SAT**ellite System), jointly developed by the University of Calgary and Geofit Inc., Laval, Quebec. A detailed introduction to the VISAT™ system will be given in Chapter 2.

1.3 Motivations for the Research

1.3.1 An Engineering Point of View

In general, there are three fundamental issues towards mobile mapping technology:

1. Accurate and reliable georeferencing of image sequences captured from a mobile mapping vehicle,
2. ***Rapid and accurate information extraction from georeferenced image sequences***, and
3. Complete representation and efficient management of the spatial and attribute information collected.

The development of the advanced navigation capability is a key to unlocking the full potential of mobile mapping systems. It is recognized that accurate and reliable georeferencing relies on integrated navigation techniques. The integration of GPS with

other navigation sensors, such as INS and DR, has defined a trend for mobile mapping georeferencing. The research work has been extensively conducted by Goad (1991), Schwarz et al. (1993a), and Wei and Schwarz (1990).

Although GIS is considered to be a tool for representation, management, retrieval, and storage of the large amount of data collected by mobile mapping systems, an efficient database engine must be designed to handle the sequential stream of information. Object-oriented data representation and network-based data dissemination are very suitable for these types of input (Chen et al., 1995; Kucera and Graves, 1995; Milne et al., 1993). The recent research work on building an object-oriented 3-D GIS using mobile mapping data can be found in Qian (1996).

Since a huge volume of image data has been collected by mobile mapping systems, rapid and accurate extraction of features of interest from image data is highly required during post-mission processing. The low efficiency of manual feature extraction is not compatible with rapid data acquisition by mobile mapping systems. It is also one of the major impediments in the development of on-line mapping or real-time mapping systems. For these reasons, the emphasis of this research is placed on the development of methods for **automatic**, and **accurate** object measurement and feature extraction using mobile mapping image sequences.

1.3.2 A Scientific Point of View

The world is essentially dynamic. The mobility of the human visual system plays a key role in sensing and observing objects in our dynamic 3-D environment. Consequently, the development of visual motion analysis, dynamic vision, image sequence analysis and mobile robot technologies has been pursued in computer vision research. Mobile mapping systems provide a perfect experimental platform to research and application of visual motion theory and methodologies.

Photogrammetry has long been recognized as an efficient large-area surveying technology. The exploration of information from an image/photo sequence captured in a strip or a block form characterizes photogrammetry. Mobility of aerial imaging enables photogrammetry to generate a strip- or block-based terrain model simultaneously, with the use of sophisticated photogrammetric block triangulation techniques.

With improvements in electronic imaging technology, video/digital cameras hare being used by photogrammetrists as a substitute for conventional film-based optical cameras. Applications of video/digital signal based photogrammetry, "***video-grammetry***" (Gruen and Streilein, 1994), are dramatically increasing. A typical example of video-grammetry application is mobile mapping.

Photogrammetry and computer vision are closely related disciplines with different goals. In recent years, they have naturally grown toward each other (Forstner, 1993; Strat, 1994).

Computer vision practitioners seek to develop algorithms for **automatic** information extraction from imagery, while photogrammetrists are concerned with obtaining **accurate** measurements from imagery. A certain amount of duplication is important because a deep understanding of the concepts is best gained through personal involvement. Joint research will provide the cross-fertilization that enables both communities to leverage the fruits of related work. Therefore, research on automatic and accurate information extraction from mobile mapping image sequences will be of considerable interest to both the computer vision and photogrammetric communities.

1.4 Overview of Previous Research Work

The development of mobile mapping systems is at an early stage of commercialization. The main efforts of research and development have been placed on improvement of quality and reliability of image georeferencing. Although automated information extraction from mobile mapping images is of considerable importance, not much work has been done in this area. Only two research groups, one at Ohio State University (OSU) and one at the University of Calgary (UC) have been reported on the development of these techniques. However, the related techniques of automatic information extraction from image sequences have been extensively researched, especially in the areas of mobile robotics, autonomous vehicle navigation and intelligent transportation systems.

1.4.1 Object Measurement

The main task of mobile mapping systems is to map objects from images into a spatial coordinate system. The objects could be footprints of houses, street edges, centerlines, curbs, lane markers, manholes, culverts, fire hydrants, traffic signs, telephone booths, electric poles, etc. In order to calculate the 3-D coordinates of an object from images, at least two conjugate points in the images need to be determined. The OSU group (He and Novak, 1993) applied the image matching technique to automate such an object measurement procedure. Since the orientation parameters of the cameras are known, the corresponding epipolar lines in a stereo image pair can be computed. Once a point in the left image is measured manually, the corresponding point in the right image can be determined by using the epipolar-line based image matching method. The area-based cross-correlation criterion was used in OSU group's image matching method.

The UC group (Xin, 1995) proposed a two-stage matching strategy for object measurement. Firstly, an image matching method similar to OSU's was applied to find the match point at a coarse level, and then the least squares image matching method (Baltsavias, 1991) was used to find the precise match. In order to improve the reliability of image matching, Xin (1995) employed edge features to constrain the matching results. If a point measured in the left image is on an image edge, the corresponding point in the right image should be on an image edge too. Thus, an edge detection algorithm has to be executed before image matching. However, this constraint is sensitive to the reliability and accuracy of the results

of edge detection. This is a typical "chicken and egg" problem, often encountered in computer vision research.

Further improvement of the 3-D coordinate accuracy of object measurement was researched by Li et al. (1996) and Zou (1996). Their work focused on the use of multiple images to perform photogrammetric triangulation. Their results demonstrated that the final 3-D coordinate accuracy can be greatly improved if multiple corresponding points in image sequences can be used. It is required that multinocular point correspondence be established first.

In fact, multiple images covering the same object are available in the VISAT mobile mapping system, that is, multiple corresponding points can be identified in the image sequences. The use of such redundant image information would be valuable not only for enhancing the reliability of image matching, but also for increasing the 3-D coordinate accuracy of photogrammetric triangulation. Thus, multiple-image-based object measurement needs to be investigated. This is one of the major tasks in this thesis research.

1.4.2 Road Centerline Reconstruction

The management of vehicles and infrastructure requires a high-quality and up-to-date road-related spatial information system. Road centerline information has been extensively used to generate road network information systems. Such information can also be used to derive road inspection parameters such as the longitudinal profile and surface deformation, which

are important indicators for road maintenance. The acquisition of up-to-date road centerline data using conventional field survey is fairly difficult due to the cost and logistical reasons.

Since the image features of road centerlines are relatively distinctive, it is expected that automatic reconstruction of road centerlines from mobile mapping images can be realized. The first trial was performed by He and Novak (1992). They applied two methods to automate the reconstruction of road centerlines from a stereo image pair. The first one relies on the definition of the centerline by a certain pattern, e.g., by a bright line. Once the defined patterns are detected, they are approximated by an analytical function. The second one permits the user to define a starting point of a centerline on the screen. A line following algorithm is then employed to trace this centerline from the bottom to the top of the image until it disappears. Xin (1995) implemented a similar method in the VISAT system. He tested three line following algorithms, namely direction-matrix-based line following, least squares line following, and polynomial edge fitting. However, none of them performed superior than the others. It is observed that these methods can only detect and reconstruct the centerlines on the basis of a stereo image pair. The capability of continuous reconstruction of road centerlines from a complete image sequence is limited.

Road based visual navigation was an active research area in mobile robotics and autonomous vehicle navigation. Many algorithms have been developed in this area (Dickmanns and Zapp, 1986; Dickinson and Davis, 1990; Thorpe et al., 1988; Waxman et al., 1987). The goal of this research is to identify clues of roads from images so that the vehicle can be manoeuvred in the right path. This goal is different from ours since we are

concerned with accurate 3-D object reconstruction rather than 2-D object identification. The imaging system configuration is also different from that of mobile mapping systems, in which high-accuracy georeferencing components are used, so that the captured images can be georeferenced and the objects of interest from images can be mapped. However, the results of research on mobile robotics and autonomous vehicle navigation are valuable to our research. As we shall see, an innovative approach based on a state-of-the-art theory, deformable models, is developed to address the issue of automatic and continuous reconstruction of road centerlines from long image sequences.

1.4.3 Feature Extraction and Object Recognition

It is understood that the two processes of feature extraction and object recognition have to be combined in order to identify objects automatically. Without sufficient knowledge about objects, object recognition results can never be reliable. Some research work have been conducted to attempt to identify some specific objects, e.g., road signs, using image sequences. He and Novak (1992) have published their results on the detection of mile makers from mobile mapping images. The focus was placed on the recognition of numbers appearing on mile-maker images. Geiselmann and Hahn (1994) developed a strategy of identification and location of simple objects (i.e. stop signs). Firstly, mathematical morphology-based operations were applied to detect the region of interest. Then affine-invariant features were extracted and employed to identify the simple shapes of objects. Priese et al. (1994) and Sheng et al. (1994) both presented their own methods for traffic sign recognition. Color information was mainly used in their methods for object

classification. It is recognized that automatic object recognition using image sequences has a long way to go. The reliable recognition of even simple objects is still challenging us.

It is found that, during the post-mission processing of the mobile mapping images, it is easy for a human operator to recognize an object from images, but it takes much time on browsing image sequences to find and locate the objects of interest. In this thesis, a semi-automatic approach to extract vertical line features is developed to assist the user to locate the objects of interest from image sequences. This is a novel idea on improving the efficiency of object measurement using feature extraction.

1.5 Objectives of the Research

This research serves as an on-going work to further development of information processing for the VISAT mobile mapping system. Both ***automation*** and ***accuracy*** of the information extraction methods are emphasized. The combination of computational motion vision and digital photogrammetry technologies is the principal methodology used throughout the thesis research. Great efforts have been placed on the development and employment of constraints from the mobile mapping system for design and implementation of the reliable information extraction methods. It is believed that the use of constraints is a key to the success of automation. The following constraints are derived and extensively applied to the proposed and developed approaches:

(1) ***Stereoscopic and sequential imaging geometry constraint***: multinocular vision methods and stereo-motion image analysis techniques can be applied by using this constraint.

(2) ***Image geo-referencing constraint***: rigorous epipolar line information and a direct image-to-scene transformation are available, since all the images have been geo-referenced in a global coordinate system.

(3)***Known vehicle ego-motion constraint***: the viewer's motion trajectory is determined by using GPS/INS navigation technology. This information can be used to develop and optimize a road-network based information collection approach.

Three automated approaches are investigated and developed:

- Multiple-image-based semi-automatic object measurement,
- automatic and continuous reconstruction of road centerlines from long image sequences, and
- semi-automatic extraction of vertical line features from stereo image sequences.

It is excepted that the developed approaches are of practical value to enhance the capability of information processing for mobile mapping applications.

1.6 Organization of the Dissertation

The dissertation basically consists of two parts. In the first part (Chapters 1 to 4), problem identification, an overview of mobile mapping technology, visual motion analysis of image sequences, and geometric analysis of the mobile imaging system are addressed. The major contributions of this dissertation are represented in the second part (Chapters 5 to 8). Three automated approaches to object measurement, road reconstruction, and vertical feature extraction from mobile mapping image sequences are proposed and developed. The outline of each chapter is described below.

Following the introduction in Chapter 1, an overview of VISAT mobile mapping technology is given in Chapter 2. The concept of multi-sensor integrated mapping is discussed. The system configuration of the VISAT system and the characteristics of the VISAT images are described.

In Chapter 3, the general methodology of visual motion analysis of image sequences is addressed. The approaches to short-range and long-range visual motion processing are briefly reviewed. We are essentially dealing with the issue of long-range image sequence analysis with known ego-motion. However, we are more concerned with feature correspondence under known motion parameters, rather than reconstruction of motion parameters of viewers or objects. A further understanding of terrestrial image sequences also proceeds in this chapter. The ground plane constraint and some useful geometric inferences from georeferenced image sequences are derived. It is understood that a

successfully automated visual analysis algorithm is based on the understanding of imaging geometry and the development of valid vision constraints.

Geometric analysis of mobile mapping imaging system is conducted in Chapter 4. A detailed analysis of the errors from the terrestrial stereo imaging system is performed. This work serves as a guideline to the development of accurate and reliable algorithms for object measurement and feature extraction from mobile mapping images. It is indicated from the results of geometric analysis that the use of multiple images for object measurement is necessary, if high-accuracy results of object measurement are required.

In Chapter 5, a semi-automatic approach to object measurement is proposed. Multiple image matching and multiple-baseline-based photogrammetric intersection methods are used to accomplish the measurement process. The approach can be used to measure any object of interest with increased 3-D coordinate accuracy and reliability. The human operator is required to give only one point in one image. The rest of the work on the determination of corresponding points in the multiple images will be automatically carried out by the computer.

In Chapter 6, a robust approach to automatic and continuous reconstruction of road centerlines from long image sequences is proposed. The closely coupled bottom-up and top-down scheme, physically-based 3-D snake theory, and sophisticated feature extraction and image matching techniques are developed. The various tests demonstrate that the approach functions reliably under different road scene and image conditions. The approach is able to

generate a road network of centerlines directly from a long sequence of stereoscopic images.

In Chapter 7, an approach to semi-automatic extraction of vertical line features is developed and investigated. The approach is designed as a tool to find and locate important vertical line features, such as light poles and traffic signs. from image sequences. A cooperative human-computer interaction scheme is developed to implement this tool. The experiments show that the developed approach is able to assist the user in performing vertical object identification and location. The time-consuming work of image browsing and object searching from the large amount of image data can be reduced. It is also worth mentioning that this research forms the basis to the development of vision based vehicle navigation, calibration and quality control.

Conclusions based on the thesis research and recommendations for further work are given in Chapter 8.

CHAPTER 2

OVERVIEW OF VISAT MOBILE MAPPING TECHNOLOGY

2.1 Conceptual Modeling

Mobile mapping systems represent a significant advance in multi-sensor integrated digital mapping technology, which provides an innovative path towards rapid and cost-effective collection of high-quality and up-to-date spatial information (Bossler et al., 1991; Li et al., 1994b; Novak, 1995; and Schwarz et al., 1993b). The development of mobile mapping systems is promoted by the use of multi-sensor integration technology. In general, we classify the sensors into three categories:

(1) Absolute orientation sensors

- Environment-dependent external positioning sensors: GPS, radio navigation system, and Loran-C
- Self-contained inertial positioning sensors: INS, dead-reckoning systems, gyroscopes, accelerators, compasses, odometers, and barometers

(2) Relative orientation sensors

- Passive imaging sensors: Video and digital cameras
- Active imaging sensors: Laser range finders or scanners, and Radar (SAR)

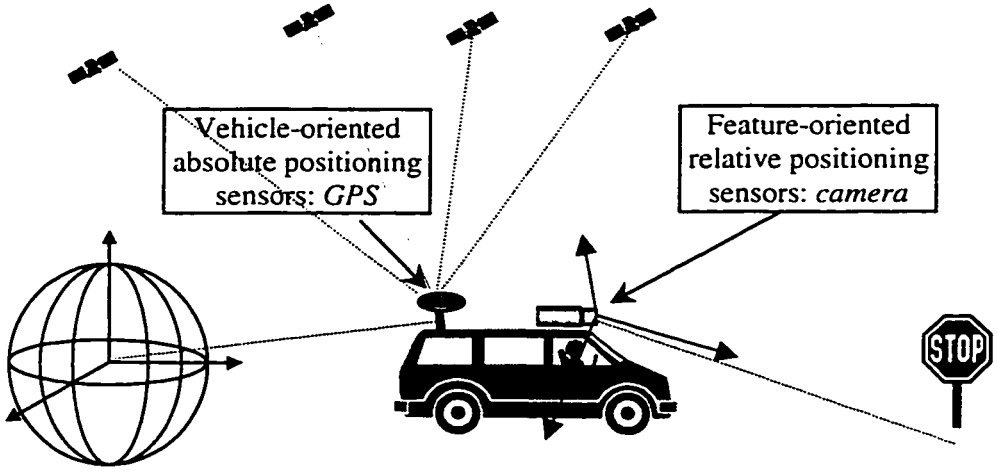
(3) Attribute collection sensors

- Passive imaging sensors: video/digital, BW/Color frame cameras, multi-spectrum/hyper-spectrum push-broom scanners
- Active imaging sensors: SAR, Laser range finders or scanners, and pavement sensors
- Manual recording: voice recording and touch-screen recording
- Others: temperature, air pressure, gravity gauges

Absolute orientation sensors are platform-oriented. They are used to determine the absolute locations of the mobile mapping platform, for instance, a van-type vehicle, with respect to a global coordinate system, e.g., WGS-84. On the other hand, relative orientation sensors provide the positional information of objects relative to the platform in a local coordinate system. Both relative orientation sensors and attribute collection sensors are feature-oriented, and many have capabilities of providing both orientation and attribute information, such as imaging cameras and laser range finders/scanners.

One of the very important concepts of mobile mapping systems is direct-georeferencing. The conceptual layout of direct georeferencing is shown in Figure 2.1. Direct-georeferencing refers to the use of absolute orientation sensors to determine the exterior orientation of the sensors without using ground control points and photogrammetric block triangulation. For example, if a camera sensor is used, any captured image can be "stamped" with the georeferencing parameters, namely three positional parameters and three attitude parameters by using absolute orientation sensors (GPS/INS). As a result, 3-D reconstruction using stereo images becomes straightforward, since the exterior orientation parameters of each image are available. Direct

georeferencing greatly facilitates the mapping procedure. The rapid turn-around time of data processing and reduced cost of ground control surveys are very beneficial.



A global coord. system ⟺ A local coord. system ⟺ Objects of interest

Figure 2.1 The concept of direct georeferencing

2.2 VISAT Mobile Mapping System

The VISAT mobile mapping system has been developed at The University of Calgary in cooperation with Geofit, Inc. The project began November, 1992, and the first demonstration was given in July, 1993. During 1993-1994, the prototype system was developed and the production system was available in 1995. The VISAT has evolved from a 3-camera system to a 5-camera system, and is currently an 8-camera system.

The overall objective of the VISAT project was to develop a precise mobile mapping system for road inventory and general GIS data acquisition, which would be capable of providing an absolute positioning accuracy of 0.3 m and a relative accuracy of 0.1 m for object points within a 30 m corridor, at vehicle speeds of 50-60 km per hour.

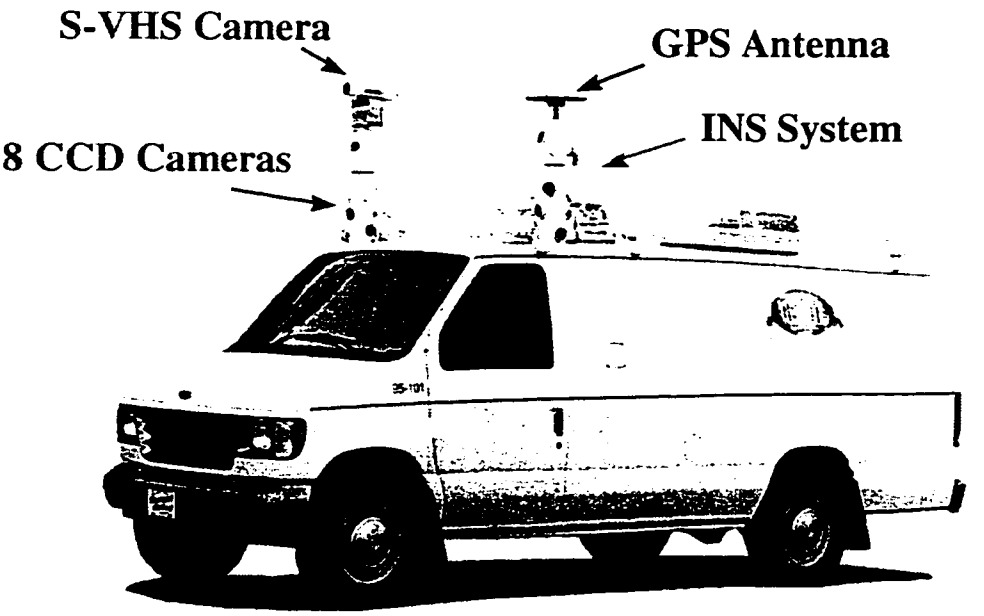


Figure 2.2 VISAT system

The van-based VISAT system is shown in Figure 2.2. This system consists of a strapdown INS system, two L1/L2 GPS receivers, a cluster of B/W CCD cameras, an Antilock-Braking System (ABS), an image control unit, and a color S-VHS camera. The function of each component can be subdivided into primary and secondary functions (Schwarz and El-Sheimy, 1996). In terms of primary functions, the cameras provide 3D positioning with

respect to the VISAT vehicle reference, which in most cases is the perspective center of one of the cameras. The position of this reference with respect to the existing control is determined by differential GPS techniques, while the camera attitude information is given by the INS. The ABS system will trigger the camera at constant distance intervals using the VISAT controller trigger channel. In terms of secondary functions, the cameras provide redundant image information of the objects, i.e., more than two images of the same object are available. The GPS is used to control INS drift errors, while the INS is used to bridge GPS outages, correct GPS cycle slips, and perform precise interpolation between GPS fixes. The ABS data can be used to update the INS data in case of GPS signal blockage.

The road-related spatial data can be acquired very efficiently through the on-line synchronization of GPS, INS, ABS, and image data. The georeferencing parameters of the moving VISAT reference can then be obtained after processing the DGPS/INS inputs. Locations of objects of interest in the road corridor with respect to the VISAT reference are determined during the post-mission processing, using two or more georeferenced images. The information extracted can be directly used to update or generate GIS databases. For more detailed information on the VISAT system, one can refer to (El-Sheimy, 1996b; Li, et al., 1994a and 1994b, Schwarz et al., 1993b; Schwarz and El-Sheimy, 1996).

2.3 VISAT Imagery

The imaging component of the VISAT consists of eight COHU 4980 video cameras, each with a resolution of 640 x 480 pixels. The technical specifications of the camera are given in Table 2.1. The cameras are housed in a pressurized case and mounted inside two fixed-base towers on top of the van. The details of hardware implementation of the imaging component in terms of synchronization, image transfer rate and buffer memory management, etc., can be found in Schwarz, et al., (1994) and El-Sheimy (1996b).

Table 2.1 Technical specifications of the COHU 4980 and 4912 CCD cameras (RS 170 standard)

Parameters	**Specifications**
Image area	6.4 x 4.8 mm
Imager type	IT CCD with on-chip microlenses
Cell Size	8.4 x 9.8 micrometers
Resolution	768 x 494 pixels
Field of view	43° (H) and 37° (V)
Electronic shutter	Eight options: 1/50 ~ 1/10 000 second
Synchronization	Genlock, Crystal Lock, Asynchronous Reset, H&V Drive, Internal clock: 28.6363 MHz

This thesis research is based on the image data collected from the VISAT prototype system. The images were captured by COHU 4912 cameras at a vehicle operating speed of about 50-60 km/h, with the camera shutter speed of 1/1000 second, and image capture rate of 0.4 second. Under these circumstances, most objects in the road corridor appear in at least three image pairs.

It is noted that, in the design of the imaging component, the choice of imaging parameters, such as the baseline B, the focal length f, the overlap degree α, the field of view angle φ, and camera types (pixel spacing δ, sensing area L, and related camera features), has to be taken into account in order to achieve the best performance. The detailed analysis of parameter selection relative to imaging performance and accuracy will be given in Chapter 4.

2.4 Applications of Mobile Mapping Systems

It is very costly that the use of traditional surveys to update GIS databases, to complement missed data, and to add new geometric or attribute data to databases. Mobile mapping is an innovative technology towards low-cost, fast, precise and high-quality data acquisition. Its potential applications appear to be limitless (Bossler et al., 1991). Typical applications include:

- Digital mapping of roads and railways,

- Surveying of pipelines, water and sewer systems (e.g., manholes),
- Surveying and tracking the evolution of distribution networks (e.g., transformers, power lines, telephone lines, street light poles),
- Inventory of structures and infrastructures (e.g., pavements, sidewalks, bridges, buildings),
- Data collection for the generation of a transportation database (e.g., centerlines, road signs, traffic signs, mile markers, pavement conditions), and
- Data collection for creating E-9-1-1 information systems (e.g., phone booths).

It is worth mentioning that Intelligent Transportation Systems (ITS) is a closely related area of research on mobile mapping systems. On one hand, there are a number of common problems encountered in both areas. On the other hand, mobile mapping systems are sometimes considered as a portion of ITS research, since a major application of mobile mapping systems is transportation data acquisition. For more detailed information on ITS, please refer to Appendix C.

CHAPTER 3

VISUAL MOTION ANALYSIS OF MOBILE MAPPING IMAGE SEQUENCES

In recent years, the computer vision community has extensively addressed the computational aspects of visual motion analysis. This chapter examines the general methodology in visual motion studies. We attempt to identify our problem under the framework of visual motion analysis, so that the well-developed methods and the accumulated experience can be utilized for our task. It is also understood that the way to resolving an ill-posed vision problem is to exploit any possible sources of constraints. With a thorough understanding of the properties of terrestrial imaging geometry, we are able to develop application-specific constraints. In this reason, visual inference between image and space is addressed.

3.1 Visual Motion Analysis Methodology

3.1.1 Basic Issues in Visual Motion Analysis

In principle, the study of visual motion analysis, or, motion and structure from image sequences, consists of two basic issues:

1. Determining image optical flow and/or feature correspondences from image sequences, and
2. Estimating motion and structure parameters using the determined optical flow and/or feature correspondences.

In mobile mapping systems, the captured image sequences have been georeferenced by using GPS/INS integrated positioning techniques. The orientation parameters of each camera exposure center are determined with respect to a global coordinate system, i.e., the motion is known. By using techniques of photogrammetric intersection, the computation of 3-D object coordinates is very straightforward. However, conjugate points need to be identified, that is, point (feature) correspondence needs to be established. Therefore, our research emphasis will be placed on the first issue – *determination of correspondences from image sequences.*

In fact, the photogrammetry community has been paid a great amount of attention to the second issue. The research results from computer vision studies not only enhance the understanding of projective geometry and algebraic geometry, but also extend and enrich photogrammetric theories and methods, for example, video-grammetry. Some of the concepts, theories and methods derived by computer vision groups have been investigated by photogrammetrists and applied in photogrammetric applications, such as quality control, camera calibration, sensor navigation and object reconstruction. Reviews of references on

algorithms for estimation of motion/structure parameters from image sequences have been provided by Aggarwal and Nandhakumar (1988), and Huang and Netravali (1994).

3.1.2 Short Range vs. Long Range (Continuous vs. Discrete) Motion Analysis

Generally, there are two complementary classifications of schemes to compute visual motion. The first classifies according to the spatio-temporal range over which methods are applicable, analogous to the human visual system: (1) **short range motion (continuous) process** and **long range motion (discrete) process.** The second classification distinguishes between the fundamentally different processes involved: (2) **optical flow** and **correspondence.** In fact, the optical flow scheme, which uses image gradients to derive image motion, is intrinsically of short range, while correspondence or similarity matching schemes can be of short range or long range.

In terms of short range motion analysis, images are taken at video rate. Thus the emphasis is generally placed on the estimation of the optical flow field between two successive frames, or on the direct use of the spatio-temporal derivatives of the image brightness (Negahdaripour and Horn, 1987; and Subbarao, 1988). These observations must also be combined with a measure of the camera velocity (instead of camera displacement) to determine the 3-D structure of objects.

In long range motion analysis, images are acquired at larger time intervals, and a large camera displacement is observed. Since the image motion of the features is "large"

compared to the temporal sampling rate, the eye has to solve the correspondence problem, i.e., it has to establish which feature at one time instant corresponds to which feature at the next time instant. Therefore, in long range motion analysis, a set of relatively sparse, distinguishable two-dimensional features, such as points, straight lines, curved lines, corners and regions, in the successive images is first extracted (Cox and Hingorani, 1996; Faugeras, 1993; and Zhang et al., 1996). Secondly, feature correspondences are established between consecutive features, and finally, the 3-D structure of the object and its relative motion with respect to the camera can be determined based on the motion of these features. It is worth mentioning that most of the research for long range motion analysis has concentrated on determining motion estimation and feature correspondences over a short image sequence (two to three images).

In general, if the scene has many easily identifiable feature points or lines, the discrete approach based on feature correspondence is suitable. If the surfaces in the scene are smooth and have no texture, then the continuous approach based on intensity derivatives is better. However, robust and accurate computation of feature correspondence and optical flow still remains a difficult problem. The optical flow field is often corrupted by image noise or occlusion, leading to generally poor and unstable results in the 3-D reconstruction. Feature correspondence also easily fails in areas where either the distortion is large, or the occlusion occurs. Hybrid approaches combining both feature correspondence and optical flow would be a way to alleviate the above problems (Baker et al., 1994; Hanna and Okamoto, 1993; and Navab and Zhang,1994).

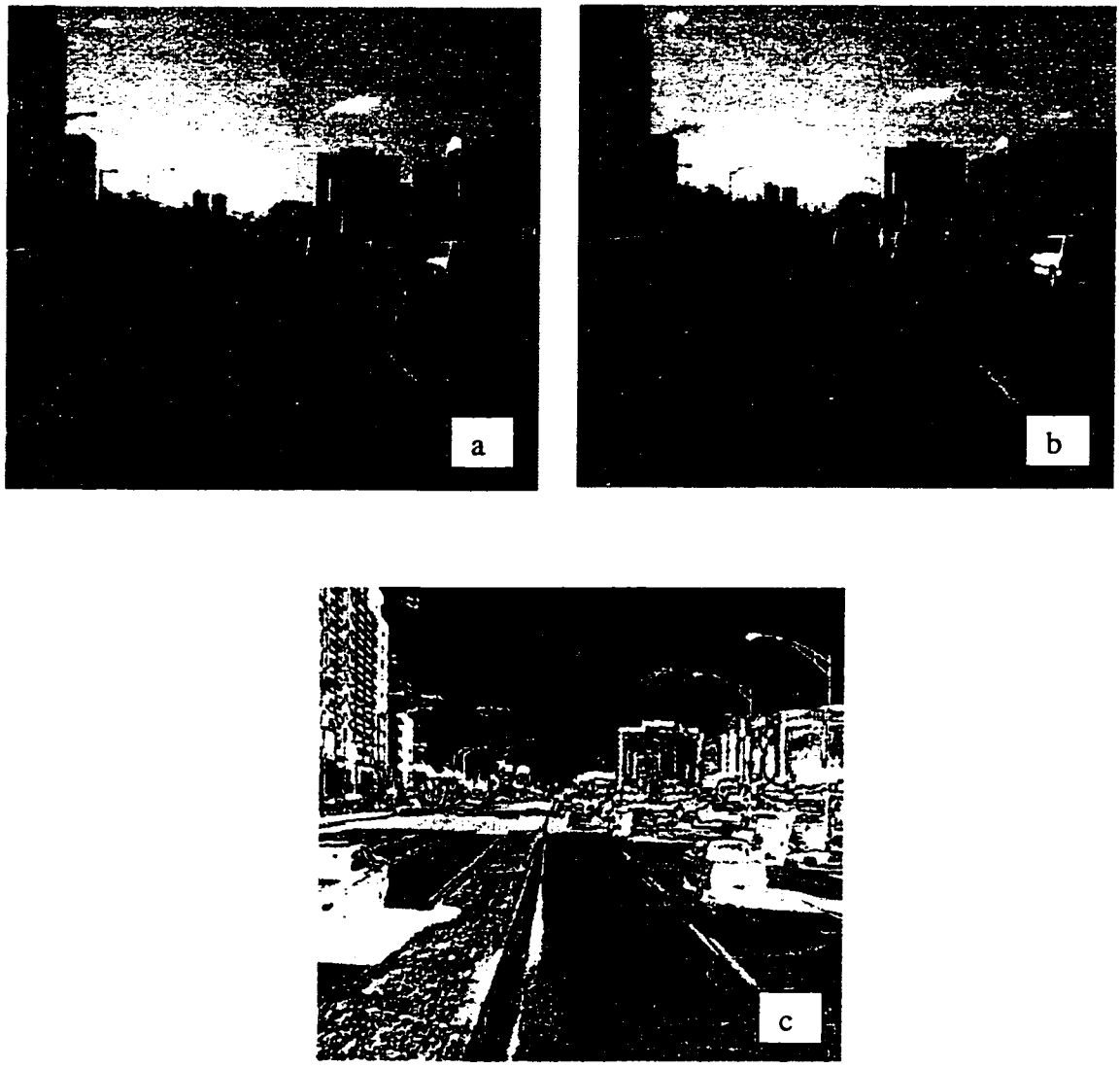


Figure 3.1 Differential image
(a) the image captured at time t (left camera); (b) the image captured at time t-1 (left camera); and (c) the differential image = | image (a) – image (b) |

It is observed that optical flow field based approach is not suitable for the VISAT images, since the image capture interval is about 0.4 second and the camera movement between the imaging intervals is too large of the order of 6-10 meters (see Section 2.3). An example of a differential image computed using two consecutive images is shown in Figure 3.1

(differential image value = | image(a) value – image(b) value |). It can be seen that the displacements of the corresponding features are fairly large, implying that the optical flow algorithm will not be reliable in our case.

Intuitively, our research falls in the category of long range motion analysis. However, compared to the processing of monocular image sequences commonly addressed in the literature, we are dealing with binocular image sequences. This redundant image information allows us to develop more robust algorithms for processing of image sequences. In this research, feature correspondence and image matching techniques are mainly used in the proposed methods. There are a number of good references available with reviews of techniques for feature correspondence and image matching (Baltsavias, 1991; Barnard and Fischler, 1982; Dhond and Aggarwal, 1989; Forstner, 1993; Gruen, 1994; Jones, 1997, and Lemmens, 1988). In Appendix D, a very brief but systematic overview of these techniques is presented.

3.1.3 Visual Motion Analysis with Known Ego-Motion

Vision analysis with known ego-motion refers to motion analysis under known dynamics of the camera (observer). In fact, known ego-motion analysis forms the basis of an active vision system. Under the condition of known ego-motion, the 3-D reconstruction problem can be solved more efficiently. This fact has motivated some investigations in recent studies (Aloimonos et al., 1988; Bajcsy, 1988). On the other hand, accurate geometric constraints,

such as the epipolar line constraint, are also available, resulting in a more robust realization of feature correspondences.

In the VISAT system, the kinematic trajectory of the vehicle can be determined with a high accuracy of 5-15 cm, and the camera dynamics can be examined rigorously by using the GPS/INS georeferencing technique (Schwarz and El-Sheimy, 1996). As a result, visual analysis can be conducted under the constraint of known ego-motion. It will be seen that this constraint is very valuable for automating and optimizing a reliable procedure for object measurement and feature extraction.

3.1.4 Active Vision

A very important advance in the theoretical framework of computer vision is the concept of ***active vision***, proposed by Aloimonos et al. (1988). Active vision represents a behaviorism school, which is directly opposite to Marr's theory of vision, a recovery school (Marr, 1982).

There is a noncontroversial observation that vision is an underconstrained problem. Thus the main goal of vision work is to find and develop constraints. However, rather than focusing on narrow sources of constraints, mostly oversimplified constraints such as smoothness constraints widely used in the recovery school, it is argued that one must exploit constraints from all possible sources and incorporate them systematically. The basic idea of active vision is the introduction of a new source of constraints arising from the

internal architecture of the system itself and the iteration of its components, such as observer-based constraints, e.g., the sensor and/or the computer (Jolion, 1994). Under the constraint that the active observer moves with known motion, a unique solution is available, resulting in a well-posed formulation of the problem. Moreover, the knowledge of these viewpoints of the active observer increases the robustness to noise.

The known motion of the observer can be determined by the use of advanced navigation technology. In this context, an imaging sensor mounted on a mobile mapping system could be treated as an active observer. Accordingly, we may state that mobile mapping systems provide an industry platform for research on active vision. On the other hand, the newly developed theories and methodologies of active vision offer us an invaluable tool to tackle the challenge of automation in mobile mapping systems.

3.1.5 Animate Vision

It is argued that the active observer above defined is not truly active, but only a moving observer. Vision is not perception but a perception-action cycle. This is leading to another vision framework known as ***animate vision*** (Bajcsy, 1988; Ballard, 1991). "We do not see, we look" represents the philosophy of this school. Under this framework, the process of vision is not considered alone, but as part of a global mechanism of an intelligent system, including cognition and motor processing.

Current research work aims at developing active vision systems with great visual abilities, such as control of ocular parameters (e.g., aperture and focus), spatially-varying sensing, and gaze control (Abbott and Ahuja, 1990; Burt, 1988; and Sandini and Tistarelli, 1990). The control of the viewing parameters gives a stable and robust means for visual perception. The control of ocular parameters allows the system to maintain a suitable image quality against the degradations that often occur during the acquisition process. The control of gaze is commonly used in binocular camera heads (Ballard, 1991; Pahlavan and Ekhlund, 1992). This mechanism, called vergence, consists of bringing and maintaining the two camera axes at a specified spatial target position, fixation point. This permits the simplification of the correspondence problem.

Animate vision further facilitates the computational process regarding 2-D correspondence and 3-D reconstruction to a large extent. It is the author's belief that animate vision theory and methodology will make a profound impact on the design and development of a new generation of mobile mapping systems, intelligent data acquisition systems.

3.2 Visual Inference between Image and Space

3.2.1 Ground Plane Constraint (GPC)

Since the vehicle is moving on ground level, and the objects to be surveyed also lie on the ground, the information is very useful for the purpose of a qualitative visual analysis. Generally speaking, the ground plane constraint holds if the following three assumptions can be made: (i) the ground is locally planar over the path which the camera traverses; (ii) the camera is rigidly fixed to a vehicle, so that it undergoes all translations and rotations in a plane parallel to the ground plane; and (iii) the camera is aligned with the ground plane such that the x-axis in the image plane is parallel to the ground plane (see Figure 3.2).

The use of the ground plane constraint can be found in many publications. Gatson and Lazano-Perez (1984), for instance, used the GPC to simplify the interpretation tree search in recognizing objects from range data; Mundy and Heller (1990) applied the known unit normal vector of the ground plane to limit the possible orientations of an object model. Tan (1993) and Kasprzak (1995) used the GPC to simplify pose estimation of moving objects. For a more general camera model of motion under GPC, one can refer to Wiles and Brady (1996). Under the ground plane constraint, the transformation relationship between 2-D image plane to 3-D object space becomes a plane to plane mapping problem.

3.2.1.1 A Direct Image-Space Transformation under GPC

The definition of the coordinate system and its associated notions are shown in Figure 3.2, where object points are restricted to the ground plane and represented as *P(X, Y, 0)*. The camera downward angle is τ, the camera height with respect to the ground plane is *H*, and the focal length is *f*. The transformation between the image coordinates (*x*, *y*) and the space coordinates (*X*, *Y*, *Z*) can be derived as follows:

$$\frac{x}{X} = \frac{\overline{Cp'}}{\overline{CP'}} = \frac{\sqrt{y^2 + f^2}}{H / \cos(90^\circ - \tau + \beta)} . \qquad (3.1)$$

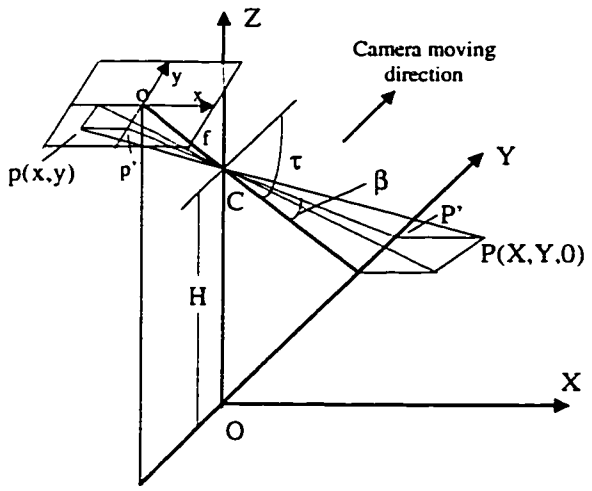


Figure 3.2 Image to ground plane transformation

Therefore,

$$X = \frac{xH / \cos(90° - \tau + \beta)}{f\sqrt{\tan^2 \beta + 1}}$$

$$= xH \cos\beta / f(\sin\tau \cos\beta - \cos\tau \sin\beta) \qquad (3.2)$$

$$= xH/f(\sin\tau - \cos\tau \tan\beta),$$

where, $y/f = tan\beta$. Similarly,

$$Y = Htan(90°-\tau + \beta)$$

$$= H \frac{ctg\tau + y/f}{1 - ctg\tau(y/f)}$$

$$= H \frac{f\cos\tau + y\sin\tau}{f\sin\tau - y\cos\tau} \quad . \qquad (3.3)$$

Considering the image coordinates of the principal points and scale factors of the camera, the transformation relationship can then be represented as

$$(x-x_0)p_x = Xf(\sin\tau - \cos\tau \tan\beta)/H \, .$$

$$(y-y_0)p_y = \frac{f(Y\sin\tau - H\cos\tau)}{H\sin\tau + Y\cos\tau} \, . \qquad (3.4)$$

Once the camera downward angle τ, the camera height H, focal length f, and the image measurements (x,y), are known, the object coordinates (X,Y) on the ground can be determined. Since the parameters involved in the above transformation, τ, H, and f, are

easily obtained, this transformation equation is of great value in many practical applications. Essentially, this principle has been applied in many video-logging systems in which often only one camera is mounted. This transformation allows the one-camera system to have the capability of 3-D measurements when the accuracy requirement is not strict.

3.2.1.2 Ground Plane Equation

As illustrated in Figure 3.3, using a viewer-centered coordinate system, the coordinate values of the camera center C and the nadir point N are, respectively,

$$C: x = 0,\ y = 0,\ z = 0,$$

$$N: x = 0,\ y = -H\cos\tau\ ,\ z = -H\sin\tau\ .$$

The normal vector of the ground plane can then be represented by

$$\boldsymbol{n} = \overline{CN} = (0,\ -H\cos\tau\ ,\ -H\sin\tau\ ). \qquad (3.5)$$

A general plane equation in space is

$$Ax+By+Cz+D=0,\ \boldsymbol{n}=(A,B,C) \qquad (3.6)$$

Substituting Eq. (3.5) into the above equation, we have

$$-yH\cos\tau - zH\sin\tau + D = 0. \qquad (3.7)$$

After solving for D by using the point N, on the ground plane, the final ground plane equation in image coordinates is obtained as

$$y\cos\tau + z\sin\tau - H = 0. \qquad (3.8)$$

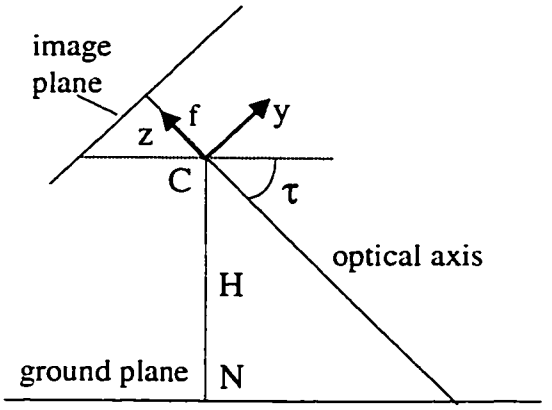


Figure 3.3 Establishment of a ground plane equation

3.2.2 3-D to 2-D Inference

3.2.2.1 Determination of the Vanishing Point (VP)

Considering a 3-D line l represented by its direction cosines, the line consists of the set of points

$$l = (X, Y, Z)^T = a + \lambda b. \qquad (3.9)$$

The line passes through the point $\boldsymbol{a}=(a_1, a_2, a_3)^T$, and has direction cosine vector $\boldsymbol{b}=(b_1, b_2, b_3)^T$, which is assumed to be of unit length, i.e.,

$$b_1^2+b_2^2+b_3^2 = 1. \tag{3.10}$$

VP is the projection of a point which is on the line and is infinitely far away from the center of the lens. Using a simple pinhole model, the VP(x_{vp}, y_{vp}) can be derived as

$$x_{vp} = fX/Z = \lim_{\lambda\to\infty} f\frac{a_1+\lambda b_1}{a_3+\lambda b_3} = f\frac{b_1}{b_3},$$

$$y_{vp} = fY/Z = \lim_{\lambda\to\infty} f\frac{a_2+\lambda b_2}{a_3+\lambda b_3} = f\frac{b_2}{b_3}. \tag{3.11}$$

Therefore, all parallel 3-D lines which have non-zero slope ($b_3\neq 0$) along the optical axis have perspective projections which meet at the same point, called the vanishing point, on the perspective projection image plane.

3.2.2.2 Determination of the Vanishing Line (VL)

Assuming a vector pointing at any VP from the perspective center C is represented as $\boldsymbol{l}(b_1, b_2, b_3)$, and the normal $\boldsymbol{n}$ to the ground plane is known, since these two vectors are perpendicular to each other, we have

$$\boldsymbol{l} \cdot \boldsymbol{n} = \boldsymbol{0} . \qquad (3.12)$$

Substituting Eq. (3.5) into the above equation,

$$b_2(-H\cos\tau) + b_3(-H\sin\tau) = 0 ,$$

$$b_2 / b_3 = -\tan\tau . \qquad (3.13)$$

By examining Eq. (3.11), the VL corresponding to the ground plane is

$$y = y_{vp} = f(b_2/b_3) = -f\tan\tau \qquad (3.14)$$

Therefore, the VL can be determined once f and τ are obtained. The camera downward angle can also be estimated with a known VP or VL.

3.2.3 2-D to 2-D Inference

Given two image lines which are parallel in space, the position of the corresponding VP can be determined.

Given one image line and the known VL, the position of the corresponding VP can also be determined by computation of the intersection point of these two lines in image.

If given two pairs of image lines which are parallel in space, respectively, the corresponding two VPs can be determined, as well the VL equation can be formed by these two VPs.

3.2.4 2-D to 3-D Inference

3.2.4.1 Determination of the Direction of A Group of Parallel Lines in Space

If a VP position is obtained, the direction cosines of the corresponding parallel lines in space whose perspective projections meet at this VP can be inferred. Examining Equations (3.10) and (3.11), three independent equations result in a unique solution of a direction cosine vector with a unit length:

$$(b_1, b_2, b_3)^T = (1/\sqrt{x^2_{vp} + y^2_{vp} + f^2}\ )\ (x_{vp}, y_{vp}, f)^T. \qquad (3.15)$$

3.2.4.2 Determination of the Normal Vector of the Ground Plane

If a VP position is known as $V_1(u_1, v_1, f)$, the vector $\overline{CV_1}$, which is (u_1, v_1, f), is perpendicular to the normal $\boldsymbol{n}(A,B,C)$ of the ground plane. Thus, we have

$$(u_1, v_1, f)\ \boldsymbol{n} = 0\ ,$$

$$u_1A + v_1B + fC=0 \Rightarrow\ \ u_1X + v_1Y + f=0\ . \qquad (3.16)$$

The above equation means that the normal vector is constrained on a line in space. If two VPs are known, the normal of the ground plane is determined by

$$\boldsymbol{n} = (u_1, v_1, f) \times (u_2, v_2, f)$$

$$= \begin{vmatrix} i & j & k \\ u_1 & v_1 & f \\ u_2 & v_2 & f \end{vmatrix} = ((v_1 - v_2)f, (u_2 - u_1)f, (u_1 v_2 - u_2 v_1))$$

$$= (0, (u_2 - u_1)f, (u_1 - u_2)v , \qquad (3.17)$$

where $v_1 = v_2$ in our case. The normalized vector of $\boldsymbol{n}$ is

$$\boldsymbol{n} = (1 / |u_2 - u_1| \sqrt{f^2 + v^2}) (0, (u_2 - u_1)f, (u_1 - u_2)v) . \qquad (3.18)$$

3.2.5 An Example of Qualitative Image Measurements

3.2.5.1 Determination of the Vanishing Point VP

As shown in Figure 3.4 of a pair of stereo images, the corresponding vanishing point to the road centerlines can be determined by using the intersection of road centerlines based on the 2-D to2-D inference. The image coordinates of these two VPs are $V_L (233, 207)$ on the left image and $V_R (215, 178)$ on the right image, respectively. If the VP of a group of spatially parallel lines is available, the VP can be used as a constraint to derive the corresponding image lines. As demonstrated in Figure 3.4, if line 1 and line 2 are used to determine the

VP, one of the end points of line 3 can then be fixed, thus reducing the search space of line 3 to a certain degree.

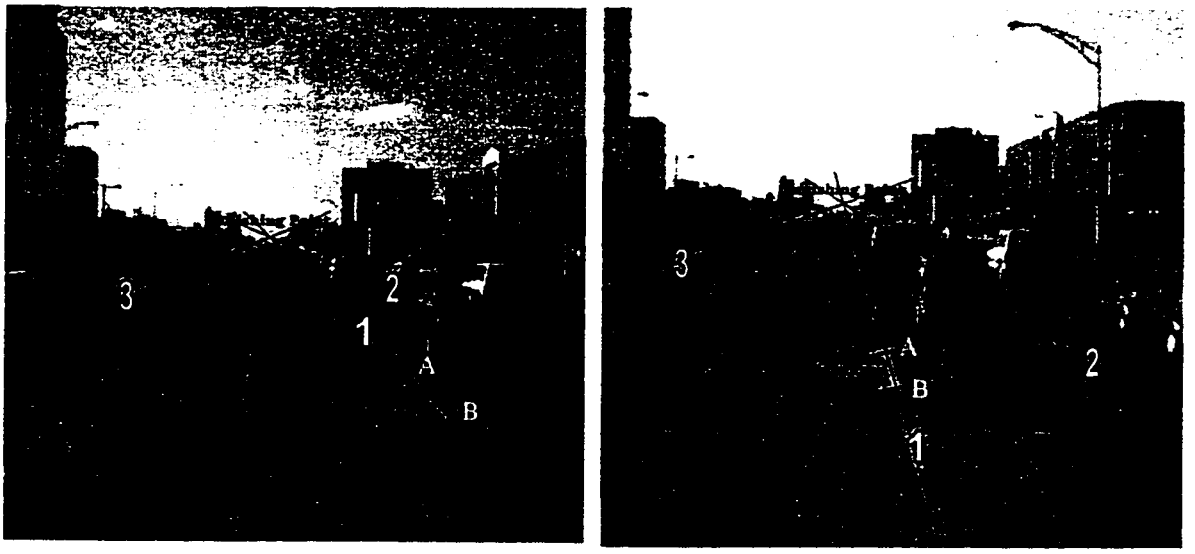


Figure 3.4 Qualitative image measurement

3.2.5.2 Estimation of the Camera Downward Angle τ

By using Eq. (3.14), the camera downward angle can be estimated. Table 3.1 gives the calibrated parameters of this pair of cameras. Inserting the related parameters into Eq. (3.14), we have

$$\tau \ (left) = tan^{-1} (240.8 - 207)/682 = 2.84°$$

$$\tau \ (right) = tan^{-1} (222.5 - 178)/680 = 3.74° \ .$$

The computed angle τ is quite close to the design value of 2.5° (the true angle τ with respect to the ground is not available).

Table 3.1 Qualitative image measurements

Parameters	**Left image**	**Right image**
Principal point coordinates	(247.4, 240.8)	(244.8, 222.5)
Focal length	682 pixels	680 pixels
Camera height	2.15 meters	2.18 meters
Scale factor	Px = 1	Py = -0.82
Image coordinates A	(356, 322)	(252, 291)
Image coordinates B	(385, 351)	(261, 320)

3.2.5.3 Computation of the Coordinates of Ground Plane Points (X, Y)

In order to evaluate the accuracy of results computed using the direct image-space transformation (Equations (3.2) and (3.3)), a comparison with the results computed using the rigorous photogrammetric intersection is performed. Since the local coordinate system is used in the direct transformation relationship, we use the relative differences of the length computation to compare the results. As shown in Figure 3.4, the image coordinates of the two end points A and B of the lane marker are measured and provided in Table 3.1. The length, D = |AB|, of the dashed centerline marker can then be computed using Equations (3.2) and (3.3), considering the scale factors,

- D1 (direct image-space transformation based on the left image measurements): 2.83 m
- D2 (direct image-space transformation based on the right image measurements): 2.85 m.

Using the photogrammetric intersection, the result is

- D3 (photogrammetric intersection based the stereo image measurements): 3.03 m.

Therefore, the relative accuracy is about 6.3%, according to the above results. It shows that the direct transformation expression derived in Section 3.2.1 can be applied as a measurement tool when the rigorous photogrammetric expression cannot be formulated.

The advantage of the use of the direct image-space transformation is that some simple 3-D computations can be conducted using monocular images instead of stereo images. In fact, GPC-based image measurements have been widely used in many robotic vision applications, as well as some video-logging systems.

3.3 Chapter Summary

We are essentially dealing with the issue of long range image sequence analysis with known ego-motion. We are more concerned with feature correspondence under known motion parameters, than reconstruction of motion parameters of viewers or objects. Therefore, feature correspondence and image matching techniques are the main tools to automate our

data processing. The recently proposed concepts of active vision and animate vision are making a profound impact on vision related research, and also are very valuable to the development of mobile mapping systems.

Visual inference between image and space was addressed in the second part of this chapter. A direct image-space transformation equation under GPC was derived. This equation can be applied to compute the geometric shape of objects at ground level. A number of imaging parameters can also be estimated using image and/or space clues. In fact, these qualitative inference methods can be used in video-logging systems as an effective measurement tool for mapping when stereo images are not available. As we shall see, the constraints and parameters derived here can also be employed to improve the reliability of the automatic algorithms developed in this research.

CHAPTER 4

GEOMETRIC ANALYSIS OF MOBILE MAPPING IMAGING SYSTEMS

This chapter presents a quantitative analysis of positioning accuracies by considering different imaging geometry conditions and system calibration errors, independent of the error contribution of the georeferencing component. The work servers as a guideline to develop accurate image measurement and object reconstruction methods.

4.1 Introduction

The errors limiting the accuracy of a mobile mapping system arise from two major sources: the positioning errors of a mobile vehicle coordinate frame with respect to a global coordinate frame, and the positioning errors of object point coordinates with respect to a vehicle coordinate frame. In addition, it is worth mentioning that time synchronization among the system components (e.g., GPS receiver, INS unit and CCD cameras) also introduces an error which has to be considered in the implementation of a mobile mapping system. It is estimated that this error is limited within 1-2 msec, thus yielding a positioning error of less than 4 cm at a vehicle speed of 60 km per hour (Schwarz and El-Sheimy, 1996). For a more detailed discussion about system performance and error analysis of the GPS/INS georeferencing component, the reader is referred to the references (El-Sheimy et

al., 1995; Schwarz and El-Sheimy, 1996). The following issues are addressed in this chapter:

(1) Based on the established error models, the along-track, across-track and total error distributions are addressed. Both the worst-case absolute accuracy and the achievable relative and relative-range accuracy are estimated.

(2) Error contributions of imaging system geometric parameters (focal length, baseline length, field of view, degree of image overlap and camera sensing area) are systematically examined. A method to optimally configure image parameters is proposed and a tradeoff of these parameters is discussed.

For the sake of completeness, quantitative error models of the effects of imaging system calibration errors on the positioning accuracy of object point coordinates are also addressed. The error budget for system calibration is determined thereafter. These research results are attached in Appendix E for interested readers.

4.2 Error Modeling of a Normal Case Stereo Imaging System

A pinhole camera model is used in modeling the error of a normal case stereo imaging system, so that camera lens distortion and other optical non-linearities can be ignored. It assumes the lens to be a point through which all incoming rays of light pass and focus perfectly on all points in the camera's field of view. In the normal case of stereo imaging

(Figure 4.1), the two camera axes are parallel to each other and perpendicular to the baseline. The relationships between the object point coordinates X, Y and Z of an object point and the image coordinates of its corresponding points (x_L, z_L) and (x_R, z_R) can be formulated as:

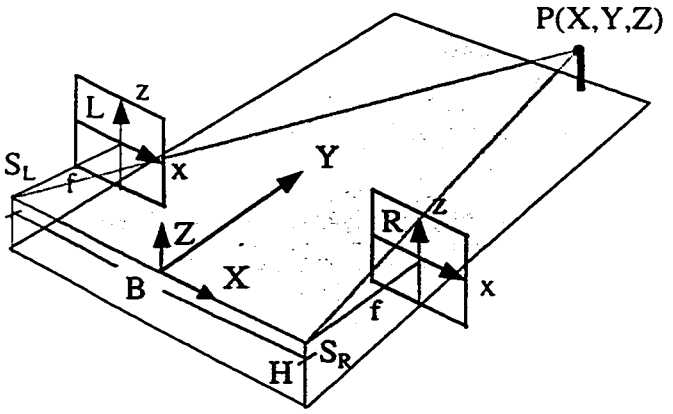


Figure 4.1 A normal case of a stereo imaging system

$$Y = Bf / p$$

$$X = (x_L+x_R)B / 2p = (x_L+x_R)Y / 2f$$

$$Z = (z_L+z_R)B / 2p = (z_L+z_R)Y / 2f$$

$$p = x_L-x_R \ , \qquad (4.1)$$

where p is the x-parallax, and B is the baseline of camera separation. The range and heading angle of the object point are defined to be

$$H = sqrt(X^2+Y^2+Z^2) = sqrt((x_L+x_R)^2+(z_L+z_R)^2+4f^2\ ))\ B / 2p$$

$$\theta \ = tan^{-1}(X/Y) = tan^{-1}(\ (x_L+x_R) / 2f\ ) . \qquad (4.2)$$

The differential relationships of Eq. (4.1) can be represented as:

$$dY = -(Bf/p^2)\, dp = -(Y/B)(Y/f)\, dp$$

$$dX = -(Y/B)(Y/f)((x_L+x_R)/2f)dp + (1/2)(Y/f)dx_L+(1/2)(Y/f)dx_R$$

$$dZ = -(Y/B)(Y/f)((z_L+z_R)/2f)dp+ (1/2)(Y/f)dz_L+(1/2)(Y/f)dz_R$$

$$dp = dx_L - dx_R \ . \qquad (4.3)$$

Substituting $k_1 = Y/B$, $k_2 = Y/f$, and converting the above equations into the relationships of the mean squared errors, we obtain :

$$m_Y = \pm \ k_1 k_2 m_p \qquad (4.4.1)$$

$$m_X = \pm \ sqrt\,(k_1^2 k_2^2((x_L+x_R)/2f)^2\, m_p^2 + k_2^2 m_x^2) \qquad (4.4.2)$$

$$m_Z = \pm \ sqrt\,(k_1^2 k_2^2((z_L+z_R)/2f)^2\, m_p^2 + k_2^2 m_z^2) \qquad (4.4.3)$$

$$m_p = \sqrt{2}\, m_{pxl} \ . \qquad (4.4.4)$$

where m_{pxl} represents the mean squared error of image coordinate measurements and its value depends upon the pixel spacing δ of the cameras. The coefficient k_1 indicates the distance-to-baseline ratio, and k_2 the imaging scale.

Assuming that the camera's field of view is φ, the maximum coordinate range in the x-direction of the image space is

$$max\,|((x_L+x_R)/2f)| \leq \tan(\varphi/2). \qquad (4.5)$$

Since the cameras are mounted on the top of a land vehicle in most terrestrial mobile mapping systems, the height of each camera exposure station is about 2m, i. e., H=2m in Figure 4.1. Therefore, the variation in the Z-direction of object space can be assumed to be restricted to $|Z_{max}| = 2.5$ mm, which results in the range of the image z-coordinates

$$((z_L+z_R)/2f) \leq Z_{max}/Y. \qquad (4.6)$$

By examining Eq. (4.4), the following parameters are considered as major factors influencing imaging system performance and accuracy: **the distance (Y) between camera exposure station and object point, the baseline length (B), the focal length (f), and the mean squared error of image parallax measurements (m_p).**

4.3 Estimated Positioning Accuracy

4.3.1 The Mean Squared Error of Image Coordinate Measurements

Because of the discrete nature of the digital imagery, the maximum error of image measuring coordinates may reach up to half of one pixel spacing ($\delta/2$). Let ε_L and ε_R represent the noise component associated with the true image coordinates, x_L and x_R. The ε_L

and ε_R are assumed to be independent and uniformly distributed between $-\delta/2$ and $\delta/2$. Then, their probability density functions can be written as,

$$f(\varepsilon) = 1/\delta\,, \qquad -\delta/2 < \varepsilon_L = \varepsilon = \varepsilon_R < \delta/2\,. \tag{4.7}$$

The mean squared error of image coordinate measurements can be derived as,

$$M^2_{pxl} = \int_{-\delta/2}^{\delta/2} \varepsilon^2 f(\varepsilon) d\varepsilon = \delta^2/12\,, \qquad m_{pxl} = \pm\ 0.29\,\delta\,. \tag{4.8}$$

The mean squared error of x-parallax measurements is

$$m_p = \sqrt{2}\, m_{pxl} = \pm\ 0.41\ \delta\ . \tag{4.9}$$

The above value indicates the coordinate measurement accuracy using manual pointing, i.e., by a human operator.

4.3.2 Error Distribution

4.3.2.1 Along-Track Error Distribution

Using Eq. (4.4.1), in the case of B=2m, f=8mm, and δ =8.4μm, the magnitude of along-track error m_Y with respect to Y is tabulated in Table 4.1. The along-track error drifts very

fast with an increase of the distance Y. The m_Y reaches up to 30 cm when the distance Y is approximately 35m from the camera exposure station.

Table 4.1. Distribution of along-track errors and upper bounded total errors (unit: cm)

Y (m)	5m	10m	15m	20m	25m	30m	35m	40m	45m	50m
m_Y	0.54	2.15	4.84	8.61	13.46	19.38	**26.38**	34.45	43.60	53.83
m_X	0.27	0.94	2.06	3.62	5.63	8.08	10.98	14.32	18.11	22.35
m_Z	0.33	0.63	0.93	1.24	1.55	1.86	2.17	2.48	2.79	3.09
m_{XY}	0.60	2.35	5.26	9.34	14.59	20.99	**28.57**	37.31	47.21	58.28
m_{XYZ}	0.69	2.43	5.35	9.42	14.67	21.08	**28.65**	37.39	47.30	58.37

4.3.2.2 Across-Track Error Distribution

When the angle φ of the field of view is chosen as 45° (when the imaging area is assumed to be 6.4mm × 4.8mm and f=8mm), the upper bounded across-track errors are shown in Table 4.1, i.e., the values of coordinate $X = X_{max} = Y\tan(\varphi/2)$ is used. This error increases with distance Y as well. When Y is 35m, the upper bounded error of the X component is about 10.98 cm, which is 1/2.4 of the error of the Y component. Comparing the two components m_X and m_Y (Eq. 4.4), we have

$$k_1k_2((x_L+x_R)/2f)\ m_p : k_2m_X = \sqrt{2}\,X : B\,. \qquad (4.10)$$

When $X>>B$,

$$m_X = \pm \; k_1 k_2((x_L+x_R)/2f)\; m_p . \qquad (4.11)$$

Therefore, the following relationship is generally used to estimate the error contributions:

$$m_Y : m_X = Y : X \qquad (4.12)$$

4.2.3 Total Error Distribution

Substituting Eq. (4.5) and Eq. (4.6) into Eq. (4.4), the upper error bounds (m_X, m_Y, m_Z, m_{XY}, and m_{XYZ}) can be obtained. Figure 4.2 shows the distribution of upper error bounds with φ = 45° and $Z=Z_{max}$=2.5m. The error values are given in Table 4.1. It can be seen that the Z-component errors can be neglected. The vertical accuracy m_Z is superior to the horizontal accuracy m_{XY}. The error component m_Y dictates the total accuracy of object coordinates, m_{XYZ}. In other words, the along-track distance Y is a significant factor influencing the total positioning accuracy. In order to meet the accuracy requirement of ± 30 cm, the distance between the positioned object and the camera exposure station should be restricted to within 35m using the assumed set of system parameters mentioned above. Figure 4.3 illustrates the 3-D distribution of horizontal mean squared error m_{XY} . The contour interval is 10 cm. This map provides an overview of positioning errors.

Similarly, the error distribution of the range and heading angle can be derived using Eq. (4.2). The z-image coordinate has a negligible effect on the range and no effect on the heading angle.

A number of practical tests have been conducted by El-Sheimy (1996a) and Szarmes (1994). Their results fit well with the error estimates derived from the error models given above.

4.3.3 Error Analysis

4.3.3.1 Worst Case Error Analysis

Since the m_Y term dominates the total error, only the Y-direction error component will be analyzed in the following sections for the sake of simplicity. Using Eq. (4.1), let the parallax error be $\Delta p = \hat{p} - p$ ($|\Delta p| < \delta$), we have

$$\Delta Y = \hat{Y} - Y = Bf/\hat{p} - Y$$

$$= -Y^2 \Delta p / (Bf + Y\Delta p) . \qquad (4.13)$$

Therefore, the worst case error formula can be written as

$$0 \leq |\Delta Y| \leq Y_{max}^2 \, \delta / (Bf - Y_{max} \delta) . \qquad (4.14)$$

Figure 4.2 Distributions of upper bounded errors

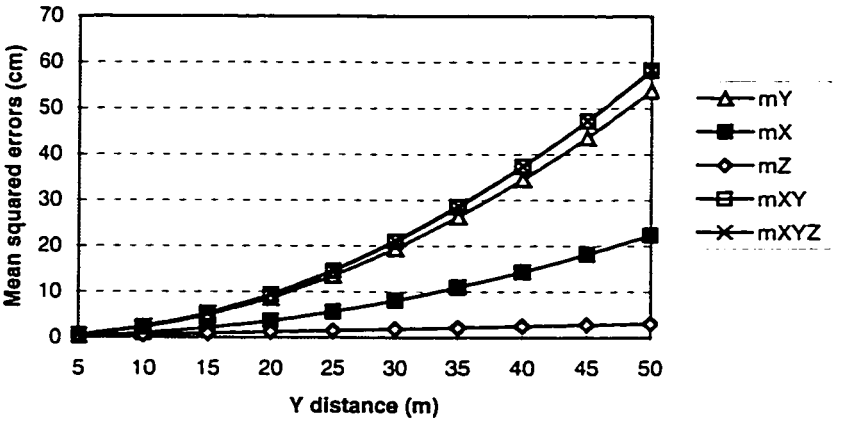
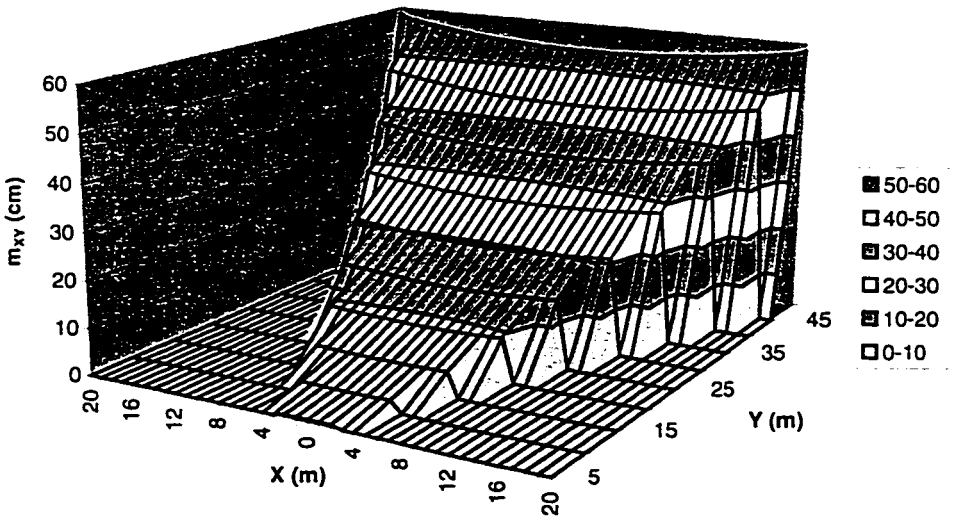


Figure 4.3 3D overview of horizontal error distribution



The formula above facilitates a conservative analysis of the error budget for imaging systems:

(1) Determination of the maximum accessible object distance Y_{max} with a required accuracy: if f=8mm, B=2m, δ =8.4 μm and ΔY is required to be less than 30 cm, then the maximum accessible distance Y_{max} is 24m after solving the quadratic equation above. This result also demonstrates that if a high accuracy level is required, the positioned object should be within a short distance to the camera station.

(2) Determination of the required accuracy of the image coordinate measurement to position an object with a certain distance: if the positioned object is 50m from the camera station, in order to reach 30 cm (ΔY) object accuracy (f=8mm, B=2m), the maximum parallax error Δp should be restricted in 1.9 μm (0.24 pixel). Consequently, the improvement of the image coordinate measurement accuracy will greatly increase the object positioning accuracy.

4.3.3.2 Relative Accuracy

The relative accuracy can be defined as

$$R_Y = |\hat{Y} - Y| / Y = |\Delta p| / (p+\Delta p) = Y|\Delta p| / (Bf+Y\Delta p) . \qquad (4.15)$$

If p or Bf increases, the relative accuracy will be improved. The probability of the relative accuracy R_Y being within a certain tolerance τ has been derived by Blostein and Huang (1987), and can be written as:

$$P(|R_Y|<\tau) = \begin{cases} \frac{1}{\ln(1-p^{-2})}\left(\frac{2\tau(\tau-p^{-1})}{1-\tau^2}+\ln(1-\tau^2)\right), & \tau < 1/p \\ 1, & \tau \geq 1/p \end{cases} \qquad (4.16)$$

The above expression indicates that the larger the parallax p, the greater the percentage of the relative accuracy being within a certain tolerance τ that can be achieved. For instance, if $\tau = 0.01$, the probability $P(|R_Y|<\tau)$ will be 75%, when p is assumed to be 50 pixels

4.3.3.3 Relative-Range Accuracy

The relative-range accuracy is sometimes a more important measure than relative accuracy, and can be defined as

$$RR_Y = |\hat{Y} - Y| / (Y_{max} - Y_{min}) . \qquad (4.17)$$

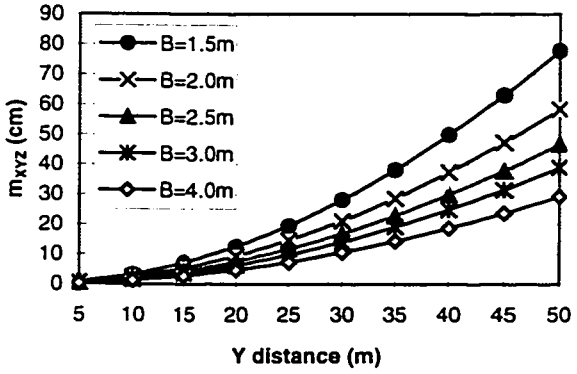
Since the size of an object of interest lies within a narrow range, the above expression can be used to assess a relative accuracy of the size or the length of the measured objects

situated at a certain object distance. It contains more descriptive information in airborne mapping applications. For example, if the flying height of an aerial stereo imaging system is 5000m above the terrain and the objects to be positioned on the terrain vary in a range of 50m, and assuming the absolute range error is 5m, then the relative accuracy R_Y would be 0.1% and the relative range accuracy RR_Y would be 10%.

4.4 Imaging Geometry and Positioning Accuracy

4.4.1 Effects of the Length of Baseline

Figure 4.4 Variations of the upper bounded errors relative to B



As B increases (k_I decreases), the positioning accuracy will be improved according to Eq. (4.4.4). The variations of m_{XYZ} relative to B are illustrated in Figure 4.4 (f = 8 mm and δ=8.4μm). If the object coordinate accuracy of 30 cm is required, this figure shows that the object distance can reach up to 50m when B is set to be larger than 4m. However, the extension of the length of the baseline B is restricted practically and physically, since most

stereo imaging systems for mobile mapping applications are mounted on a mobile platform (e.g., on top of or inside a land vehicle). The size of the platform limits the setting of a large baseline. In addition, an increase of B will cause a decrease of the overlap (ΔX) of imaging, thus eventually increasing the system cost for applications (i.e., more survey runs are needed to enlarge the image coverage). It is also worth mentioning that the long baseline will produce many more ambiguities for manual target pointing, automatic image matching and feature extraction from imagery. Consequently, a tradeoff has to be made in the choice of an appropriate baseline. Maintaining a basic degree of overlap of a stereo image pair should be considered first in determining the maximum length of the baseline, which will be discussed in the next section.

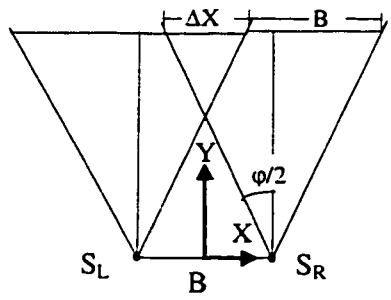


Figure 4.5 Overlap area

4.4.2 Effects of the Overlap Percentage

The horizontal range ΔX of overlap area can be derived from Figure 4.5,

$$\Delta X = 2Y \tan(\varphi/2) - B \,. \qquad (4.18)$$

The degree of overlap, α, of an image is defined as

$$\alpha = \Delta X / 2Y\tan(\varphi/2)$$

$$= 1 - B / 2Y\tan(\varphi/2) . \qquad (4.19)$$

In a terrestrial imaging system, the degree of image overlap is a function of the distance Y. If one wants to maintain a 40-50% overlap degree at Y_{min} = 5m, a baseline length larger than 2.5m cannot be used. In practice, a larger α is always preferable. On the other hand, the overlap degree can be extended using a bigger angle φ of the field of view. However, φ is dependent on the selected CCD camera type (sensing area) and the focal length.

4.4.3 Effects of the Focal Length

The coordinate accuracy of 3-D objects will be improved when using a lens with a larger focal length according to Eq. (4.4). The effect of increasing f on m_Y is equivalent to that of increasing B, examining Eq. (4.4.1). Also, a tradeoff of the length f has to be achieved to maintain a certain angle of the field of view φ .

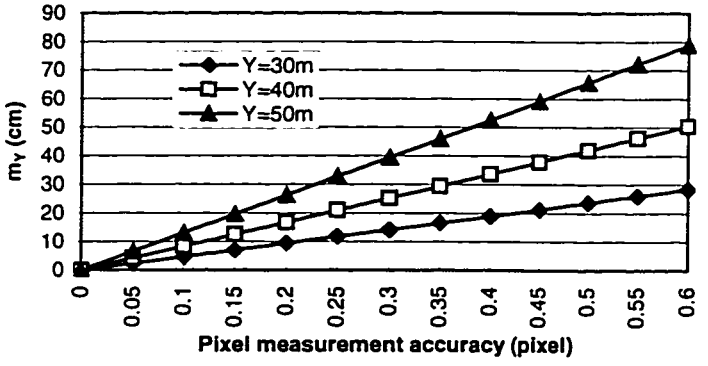
The most commonly used CCD video cameras have two types of sensing area formats, namely, 1/2 inch format (6.4H*4.8V mm) and 2/3 inch format (8.8H*6.6V mm). The small sensing area limits the field of view. It can be seen that only lenses with focal length of 8

mm can provide a field of view angle larger than 45°. However, the use of small focal length will obviously increase effects of errors due to lens distortions.

4.4.4 Effects of the Accuracy of Image Coordinate Measurements

The accuracy of image coordinate measurements is determined by the pixel spacing δ and the pixel measurement accuracy. The pixel spacing depends on the selection of the type of CCD camera. The variations of the major error component m_Y relative to the accuracy of image coordinate measurements using Eq. (4.4.1) are illustrated in Figure 4.6 (for the case of B = 2m, f = 8mm and δ = 8.4μm). The figure shows that the image coordinate measurement accuracy of 0.2 pixel can produce the result of object positioning accuracy of 30 cm with an object distance 50 m (same result as from the worst case accuracy analysis in Section 4.3.1). Therefore, highly accurate image coordinate measurements are of particular importance to improve the total accuracy. The use of sub-pixel image point measurement algorithms or multiple image matching algorithm (Tao et al., 1997) is valuable.

Figure 4.6 Variations of m_Y relative to pixel measurement accuracy



4.4.5 Optimal Configuration of Imaging Parameters

The configuration of geometric imaging parameters is very important in the development of a successful stereo imaging system. The following imaging parameters have to be set or chosen appropriately: the baseline B, the focal length f, the overlap degree α, the field of view angle φ, and camera types (pixel spacing δ, sensing area L, and related camera features). Based upon the previous analysis, the maximum length of B is restricted by the desired overlap degree α; the overlap degree is affected by the field of view angle φ of the camera, while the field of view angle is determined by the focal length f and the camera sensing area L. Although the individual effects of these parameters on the system performance have been analyzed, a tradeoff has to be achieved to realize an optimal, combined system performance. The following constraints can be applied to determine this tradeoff:

$$Bf \rightarrow max \qquad (4.20.1)$$

$$\varphi = 2tan^{-1}(L/2f) \geq T_{\varphi} \qquad (4.20.2)$$

$$\alpha = 1 - B/2Ytan(\varphi/2) = 1 - Bf/YL \geq T_{\alpha}\ , \qquad (4.20.3)$$

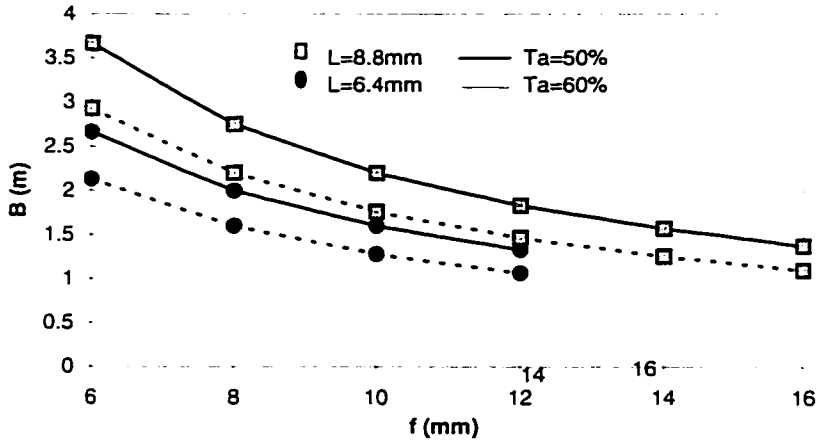
where L represents the maximum dimension of the sensing area, T_{φ} is the required angle of the field of view and T_{α} is the required overlap degree. As a result,

$$f \leq L/2tan(T_{\varphi}/2)$$

$$(Bf)_{max} = YL(1 - T_{\alpha}) . \qquad (4.21)$$

In land-based mobile mapping systems, the objects to be measured lie in the road corridor about 5-50 m from the camera station. If the field of view T_{φ} = 30° at Y_{min}=5m is required, and a standard format camera is selected (either 1/2"or 2/3"), i.e., L=6.4 or L=8.8mm, the value range of combinations of parameters B and f can be calculated (shown in Figure 4.7), by solving the above constraints. It can be seen that the focal length f is in range of 6-12 mm and the corresponding B is limited to 1-3 m, when the ½" format camera is chosen. It is found that the use of 2/3" inch camera allows a more flexible combination of B and f.

Figure 4.7 Optimal configuring parameters B and f



There are a number of combinations of B and f that are available. However, the use of a shorter focal length will increase lens distortion errors. Therefore, the focal length should be chosen as large as possible. In addition, the use of a larger focal length instead of a larger

baseline B will reduce the ambiguities of image point matching and improve the point measurement accuracy.

Furthermore, the tradeoff between the choice of camera parameters of pixel spacing δ, sensing area L, and related camera features is also to be achieved. In practice, the physical features of different camera types should be taken into account first. The choice of camera types in terms of δ, and L can be decided by the quantitative analysis mentioned above.

4.5 Chapter Summary

- The positioning accuracy of object point coordinates derived from a stereo imaging system is determined by four major components: the distance between camera exposure station and object point Y; the baseline length B; the focal length f; and the mean squared error of image parallax measurements m_p.
- The along-track error component m_Y is the predominant error contributing to the total positioning error, while the z-directional error component is relatively small and could be negligible.
- Under the assumed settings of the imaging parameters (B = 2m, f = 8mm and δ =8.4μm), if only the error of image coordinate measurements is considered (m_p =0.41δ), total positioning accuracy of 30 cm can be achieved when the positioned object is 35m from the camera exposure station. If the mean squared error of image coordinate

measurements is improved to be 0.2 pixel (0.2δ), the positioned object can be 50m from the exposure station with the same positioning accuracy of 30cm. Therefore, the development of sub-pixel image measurement methods is particularly important.

- The effects of the imaging system parameters (baseline B, focal length f, overlap degree α, field of view angle φ, and camera sensing area) on object positioning accuracy were systematically examined. The maximum baseline length is restricted by the desired overlap degree. The overlap degree is affected by the field of view angle of the camera. The field of view angle is determined by the focal length and the camera sensing area. A method to determine the optimal imaging parameters was developed to achieve a tradeoff between these parameters.
- It can be concluded that the use of multiple images for object measurement is necessary due to the limited baseline setting in mobile mapping systems.

CHAPTER 5

MULTIPLE-IMAGE-BASED SEMI-AUTOMATIC OBJECT MEASUREMENT

The goal of this research is to develop a robust semi-automatic approach to improve the efficiency and accuracy of object measurement from mobile mapping image sequences. Since high-density image sequences are available in mobile mapping systems, an object along the road corridor is usually visible in a number of consecutive image pairs. By using a multiple-image matching method, once an object point in one image is manually selected, corresponding points in the consecutive image pairs can be determined automatically. Consequently, the whole procedure of object measurement is carried out semi-automatically. Furthermore, the 3-D coordinates of the object points can be calculated using multiple corresponding points, rather than two corresponding points. Due to the multiple-baseline geometry employed in the photogrammetric intersection, the accuracy and the reliability of the 3-D coordinates will be greatly improved.

5.1 Introduction

It has been recognized that the application of fully automatic feature extraction is still restricted and is only possible when tailored to specific objects (Gruen and Streilein, 1994).

We are dealing with the measurement of any type of object of interest in images with no or little knowledge about the objects. It is understood that the cooperative work between the two agents, human operator and computer, will play a key role in intelligent information processing (Tao and Lin, 1994). A successful semi-automatic method will be beneficial not only due to its reliability but also due to its efficiency. Therefore, development of a semi-automatic scheme for feature extraction is the aim of our research. It will be seen that a certain level of human involvement has been applied in the approaches developed in Chapters 5 and 7.

Acquisition of 3-D information through image measurements requires matching of corresponding points. Extensive studies and publications can be found in this area. A brief overview of image matching and feature correspondence methods is given in Appendix D. Since the mobile mapping image sequences are georeferenced, an object in the scene can, usually, appear in a number of stereo image pairs. For example, in the VISAT system, at least 3-4 image pairs covering the same object can be available when the vehicle speed is 50-60 km/h and the image capture rate is one image set per 0.4 second. Such imaging geometry of mobile mapping systems presents a great opportunity for the use of multiple-image matching to achieve robust and accurate object measurement.

In addition, according to the analysis results performed in Chapter 4, the accuracy of 3-D object coordinates degrades dramatically along the depth (vehicle moving) direction, due to the limitation of a short baseline configuration in the VISAT mobile mapping system. Therefore, even in the manual object-measurement scheme, the human operator may be

required to measure more than two corresponding points in consecutive image pairs, so that a multiple-baseline-based photogrammetric intersection using multiple corresponding points can be performed to increase the coordinate accuracy of measured objects. Figure 5.1 shows the current imaging geometry of the VISAT system. The baseline of the current stereo image pair is about 2m, while the stereopsis between the current image and the next image taken in the next time instant has a 6-10 m baseline. The use of multiple points with larger baseline settings for intersection to increase the coordinate accuracy has been widely applied in high-accuracy close-range photogrammetry (Gustafson and Handley, 1992; Fraser, 1992; Fraser et al., 1995). Therefore, if high-accuracy results are required, the human operator will be asked to browse the image sequences and capture the corresponding points visible in a number of images. This procedure is not only time-consuming and error-prone, but also unrealistic for practical operations.

In this research, a semi-automatic object measurement approach is developed to expedite the manual measurement procedure. There are two key techniques involved in this approach: (1) **multiple-image matching**: automatic determination of the corresponding points in consecutive image pairs; (2) **multiple-baseline-based photogrammetric intersection**: optimized calculation of 3-D object coordinates using multiple-baseline geometric configurations. In this thesis, the development of a multiple-image matching method is a goal.

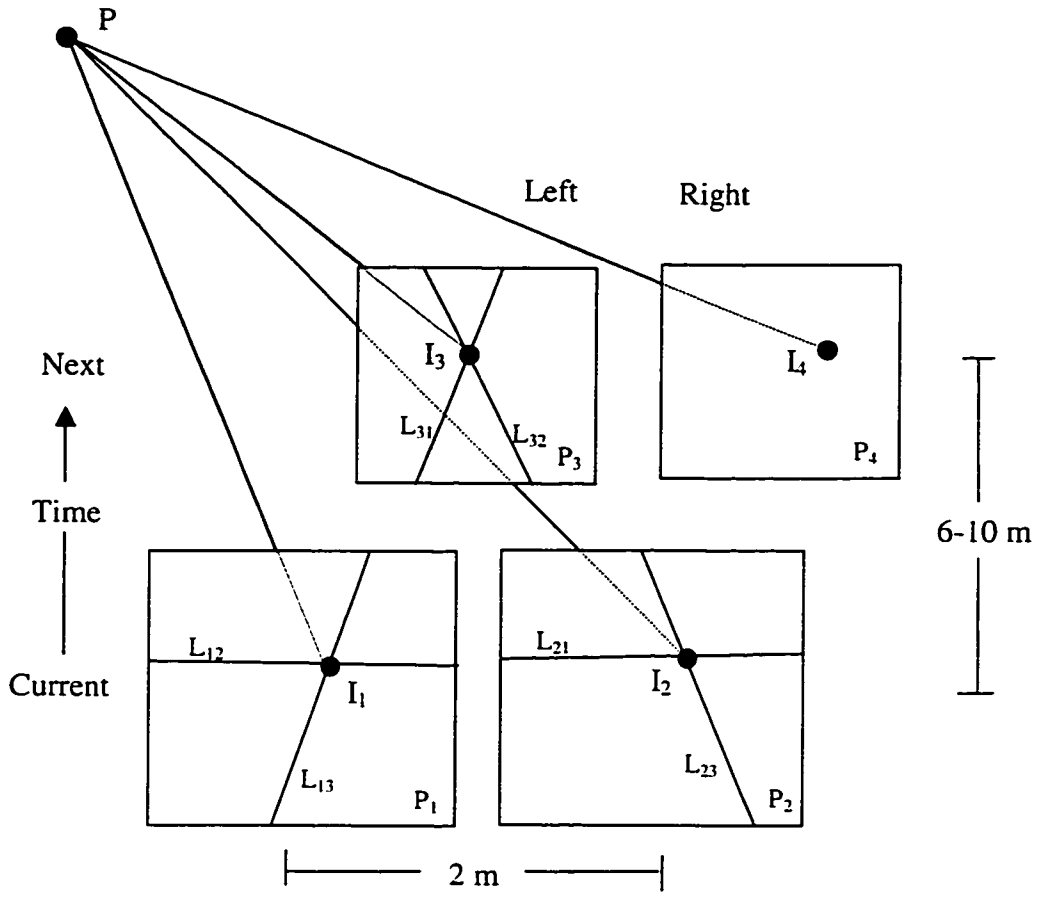


Figure 5.1 Multinocular stereo geometry and multiple epipolar constraint

5.2 Multiple-Baseline-Based Photogrammetric Intersection

Recall from the analysis of Section 4.2, the mean squared error of the 3-D coordinate (Y) along the depth direction can be expressed as

$$m_Y = k_1 k_2 m_p, \qquad k_1 = Y/B, \qquad k_2 = Y/f \ . \tag{5.1}$$

In land-vehicle based mobile mapping systems, the baseline length is usually limited, say 2m. In the case of object distance Y = 30m, f = 800 pixel units, and m_p = 0.4 pixel, the resulting depth coordinate error m_Y is 22.5cm. However, if the baseline B of 8m is applied, the object accuracy of m_Y = 4.5cm will be achievable.

On the other hand, depending upon the vehicle speed and the image capture rate, the along-track baseline (the distance from the current camera station to the next camera station, shown in Figure 5.1) could be 6-10m long, for instance, in the VISAT system. The use of multiple baselines provides a viable means to improve the accuracy and reliability of the 3-D intersection. Due to the unavoidable calibration error of imaging systems, the conjugate light rays originating from different cameras may not intersect at one point in the 3-D object space. Therefore, an estimation algorithm has to be derived to determine an optimal position of the intersection point. When using the maximum likelihood estimation method, the estimator for the $\hat{Y}$ is given by Roberts and Ganapathy (1987) as

$$1/\hat{Y} = \frac{\sum_{i=2}^{N} \sum_{j<i} B_{ij}^{\,2} \frac{1}{Y_{ij}}}{\sum_{i=2}^{N} \sum_{j<i} B_{ij}^{\,2}} \ , \tag{5.2}$$

where B_{ij} is the baseline of the i and j camera pair, and Y_{ij} is the intersection result using the i and j camera pair. The reciprocal of the maximum likelihood estimator $\hat{Y}$ for N cameras turns out to be a weighted sum of the two-camera intersection results. The weight for each stereo image pair is the square of the pair's baseline. The variance of $\hat{Y}$ can be derived as

$$\sigma_Y = \frac{\sum_{i=2}^{N}\sum_{j<i}\frac{B_{ij}^{\,2}}{Y_{ij}^{\,2}}\sigma_{Yij}}{\sum_{i=2}^{N}\sum_{j<i}\frac{B_{ij}^{\,2}}{\hat{Y}^2}}, \qquad (5.3)$$

In fact, the orientation accuracies of camera stations are different from one to another resulting from the results of GPS/INS navigation data. The relative orientation accuracy of the fixed stereo rig (2m baseline) is also different from that of the along-track stereo rig (6-10 m baseline) formed by the current camera station and the subsequent camera station. The fixed camera rig is calibrated by a photogrammetric bundle adjustment, while the relative orientation of the along-track stereo rig is determined by the GPS/INS positioning method. A Kalman filter based method considering both accuracy and reliability of the camera orientation parameters has been discussed by Li, et al. (1996) and Zou (1996). The method is applied to select the best image pair from an image sequence to perform an optimal 3-D intersection. For a more detailed analysis of advantages of the use of multiple-baseline intersection, one can refer to Mason (1995). In our implementation, the bundle adjustment based multiple-point photogrammetric triangulation method is used (Szarmes, 1994).

5.3 Multiple-image Matching

5.3.1 Problem Identification for Matching

Due to the large geometric distortions in terrestrial imagery, image matching becomes very difficult when applied in either a left-right image pair or a forward-backward image pair. Such distortions are caused by the large variations in object distance to a camera and by the large differences in camera view angles. Since no knowledge of the measured object distance is obtained, the distortions cannot be easily removed. If an object is located at ground level, the geometric distortion can be corrected by using a ground plane constraint. The ground plane constraint has been described in Chapter 3 and successfully applied in road centerline reconstruction (Chapter 6).

The second common problem arises from occlusions. Occlusions occur when an object that can be seen from one camera viewpoint cannot be seen from another camera viewpoint. Consequently, an object in one image may not have a match in another image. Figure 5.2 gives a typical example. Due to the large differences in the camera view angles and the partial occlusion, the window background of the object of interest, a traffic sign, in the left image is very much different from that in the right image. A very low rate of successful matching is found in this case, when an area-based matching method is applied to establish the correspondence. Feature-based matching is an alternative to this problem. However, reliable extraction of object features in outdoor scenes is still very complicated.

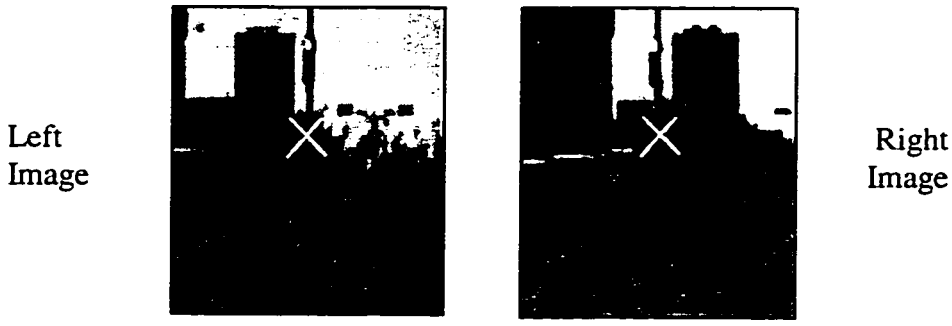


Figure 5.2 Problems for image matching

It is recognized that the constraints from binocular stereo are not sufficient to impose a unique solution to the matching ambiguities caused by the above factors. The exploitation of information from a third image or more images can disambiguate matching results to a large extent:

- The use of more than two images presents a strong geometric constraint, the **multinocular epipolar constraint**. This constraint has been proven to be very effective to reduce matching ambiguities.
- There are a number of combinations of stereo pairs among a stereo image sequence with different baselines. Generally speaking, the ambiguities will be less when matching is performed on an image pair with a short baseline. If a longer baseline is used, a larger disparity range has to be searched to find a match. Thus, a high possibility of a false match may occur (Kanade, et al., 1992). However, the 3-D

intersection may have a more precise result when using a large baseline. Therefore, the use of multiple-images can improve both the reliability of matching and the accuracy of intersection through an appropriate trade-off.

- The availability of multiple views provides the possibility of circumventing the occlusion problem. For instance, if a corresponding point cannot be seen in the right image taken from the current camera station, it may be visible in the next (or previous) left image taken from the next (or previous) camera station.

5.3.2 Multinocular Epipolar Constraint

Multinocular epipolar constraint is the main constraint derived from multinocular stereopsis for disambiguating the matching results. As shown in Figure 5.1, for the case of a four-camera configuration, camera i (i=1, 2, 3, or 4) is represented by its optical center C_i and its image plane P_i. Given a scene point P, its image point I_i by camera i is given by the intersection of the line PC_i with the plane P_i. Points I_1, I_2 and I_3 form a triplet of homologous image points.

Given a pair (i, j) of cameras and a physical point P, the conjugate epipolar lines L_{ij} and L_{ji} can be determined. As one can see, a scene point P produces three pairs of homologous epipolar lines. When the image point (I_i, I_j, I_k) forms a triplet of homologous image points, then I_i is necessarily located at the intersection of the epipolar lines L_{ij} and L_{ik}, respectively, defined by I_j and I_k. Therefore, the search for homologous image points between two images can now be reduced to a simple verification at a precise location in the third image.

For instance, checking that (I_1, I_2) form a pair of homologous image points consists of verifying the presence of I_3 at the intersection of L_{31} and L_{32}. The more the images used, the more the constraints that are available. However, the computational cost must be considered.

5.3.3 Review of Related Work

In the recent past, trinocular and multinocular matching has been of considerable interest in stereo analysis (Ayache and Lustman, 1991; Baltsavias, 1991; Dold and Mass, 1994; Ito and Ishii, 1986; Mass, 1991; Okutomi and Kanade, 1993; Pietikainen and Hardwood, 1987; Shibasaki and Murai, 1988; Yachida et al., 1986). Dhond and Aggarwal (1989) performed a cost-benefit analysis of the use of a third camera for stereo matching. It was concluded that trinocular matching reduces the matching ambiguities by more than half, while the corresponding increase in the computational cost is only about 25%. Mass (1992) conducted a quantitative evaluation of disambiguating performance by using more than two cameras, and applied four cameras to establish image correspondences for particle tracking. In photogrammetry, the use of multiple-image matching techniques to generate 3-D terrain surfaces from a number of overlapping images is increasing. Shibasaki and Murai (1988) used three SPOT image matching to improve mapping accuracy; Baltsavias (1991) developed a multiple-image-based, least-squares matching method. The multinocular epipolar constraint was recently utilized by Mass (1996) to derive matching approximate values. Automatic point transfer for aerial triangulation can be also accomplished by multiple-image matching, the related publications can be found in Agouris (1992) and

Krzystek et al. (1995). In computer vision, multiple-image matching has been widely used for scene reconstruction (Dold and Mass, 1994 and Yachida et al., 1986) and robotics navigation (Ayache and Lustman, 1991; and Okutomi and Kanade, 1993). The recent trend in robotics towards reliable reconstruction is to further integrate correspondence methods with optical flow techniques, using stereoscopic and sequential image data (Baker et al., 1994; Hanna, 1993; Navab and Zhang, 1994). It is demonstrated by many research results that stereo-motion cooperation has a capability to augment stand-alone correspondence methods, image matching or optical flow. However, the establishment of an optical flow field from image sequences requires a very high rate of image capture, say 25Hz or 30Hz. This condition cannot be met in the VISAT mobile mapping system. However, the stereo-motion cooperation approach could be applied to other mobile mapping image data.

In the above mentioned methods, even though the goals, system configurations, and imaging conditions are different from one another, the primary aim of the use of multiple images is to improve the reliability of image matching using the redundant information existing in multiple images, for example, via the multinocular epipolar constraint. It is believed that the success of a multiple-image matching method is based on the design of a matching strategy and the utilization of the redundant information.

5.4 Object Measurement Based Upon Multiple Images

5.4.1 Method Overview

The integrated GPS/INS navigation technology allows the collected image sequences to be georeferenced rigorously in a global coordinate system. Thus, the multinocular epipolar constraint is available from the image sequences with known ego-motion parameters. The proposed multiple-image matching method consists of the following steps (Figure 5.3):

(1) Manually select an object point in the current left image using the mouse pointer.

(2) Automatically snap a critical point close to the input point using a digital snapping technique.

(3) Perform the initial image matching between the current left image and the right image along the corresponding epipolar line to find a set of match candidates. If no match candidate is found, use the forward left image or the backward left image in the image sequence forward-current or backward-current image matching.

(4) Conduct consistency filtering to determine and validate the best match point from the match candidates using the multinocular epipolar constraint.

(5) Calculate 3-D coordinates of the object point using multiple-baseline photogrammetric intersection.

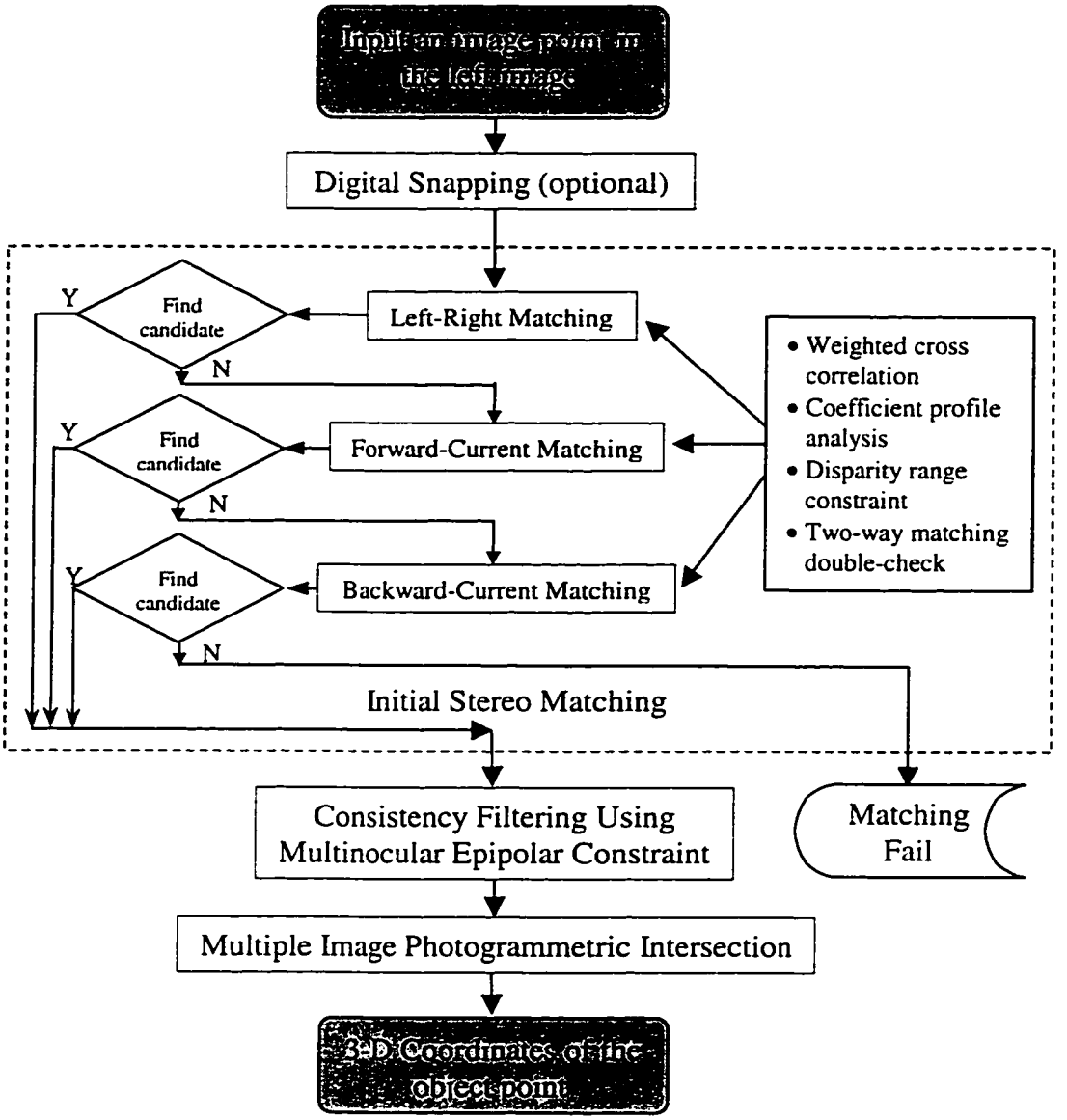


Figure 5.3 Flowchart of the semi-automatic object measurement procedure

5.4.2 Digital Snapping

A digital snapping technique is proposed to increase the positional precision of the image point measured by a human operator. Due to the low resolution of the image (512 x 480), it is not easy for a human operator to point an object at the desired position in the image. The image enhancement and zoom-in tools have been developed to facilitate the measurement process. However, it is our experience that the desired point of the object to be measured usually occurs at the place where gray values vary distinctively. In other words, the desired point is essentially a critical image point in most cases. Since human visual ability is not sufficient to sense such a point, an automated technique called *digital snapping* is designed to snap such a neighboring critical point when a manually clicked point is given. Figure 5.4 demonstrates the digital snapping process. This function is optional depending on the user setting.

Figure 5.4 Digital snapping process

("+" manually selected point, "x" automatically selected point by the snapping process)

In the current implementation, digital snapping is performed by using the Moravec operator (Moravec, 1979). The Moravec operator evaluates the squared differences of gray values in the four main directions, defined by pixel rows and columns and the two diagonals. The squared difference in each direction is computed as the sum of squared gray value differences between neighboring pixels. The Moravec operator is defined as

$$M_1 = \sum_{k=-N/2}^{N/2-1}[g(I+k,J)-g(I+k+1,J)]^2$$

$$M_2 = \sum_{k=-N/2}^{N/2-1}[g(I+k,J+k)-g(I+k+1,J+k+1)]^2$$

$$M_3 = \sum_{k=-N/2}^{N/2-1}[g(I,J+k)-g(I,J+k+1)]^2$$

$$M_4 = \sum_{k=-N/2}^{N/2-1}[g(I+k,J-k)-g(I+k+1,J-k-1)]^2 \ , \qquad (5.5)$$

where $g(I, J)$ represents the gray value of pixel (I, J), and N is the defined window size. If the $M = min(M_i)$, $i=1,2,3,4$, of that point, is the maximum in a local neighborhood and surpasses a certain threshold in the defined window, the point is defined as a critical point to be snapped. If no such point exists, the original input point will be taken as the output.

When using a small window for snapping, we find that this technique is of very practical use to assist the user to locate precisely the desired point in an image. Some examples and their effects are illustrated in Figure 5.5 (the window size for snapping is 9x9 and the window size for Moravec computation is 5x5 in the examples). Both the original images

and the enlarged images are given. The '+' sign represents the clicked point by a human operator and the 'x' sign marks the point selected by the snapping process. It can been seen that digital snapping yields improved pointing precision.

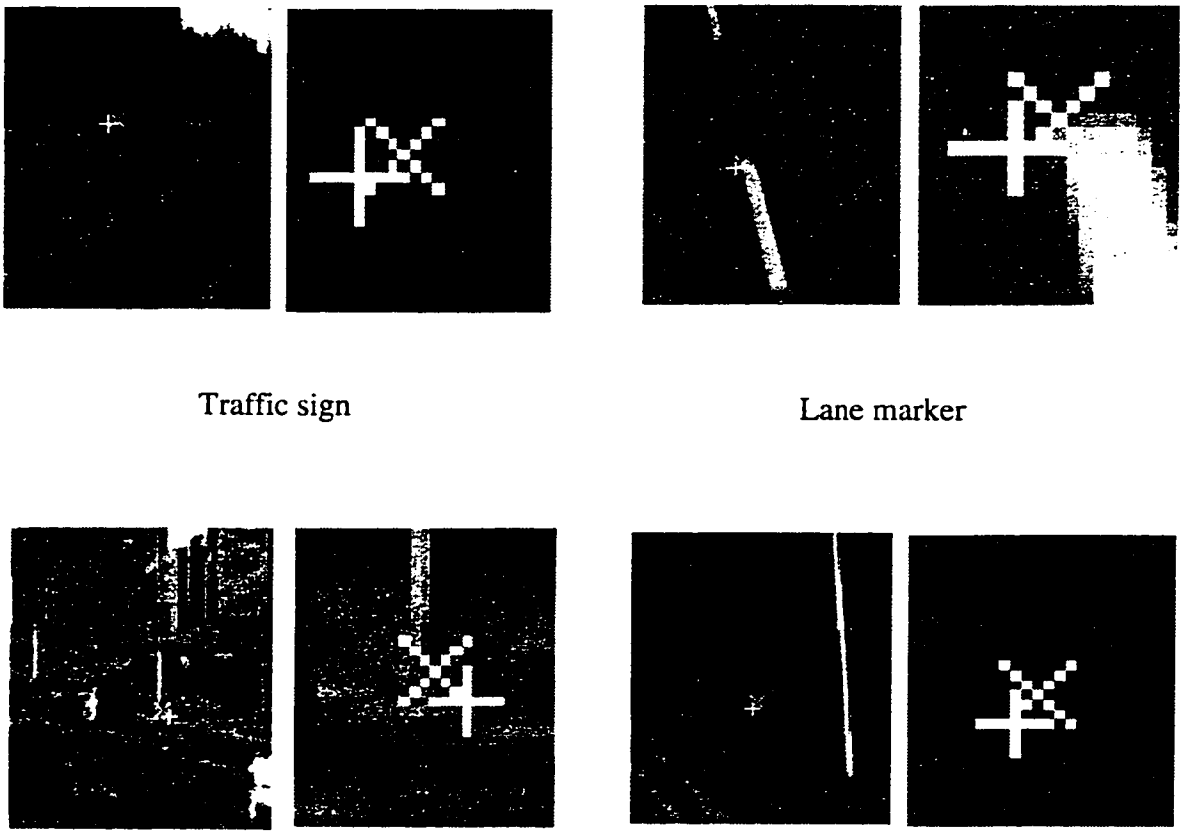


Figure 5.5 Effects of digital snapping

It has also been found that the digital snapping can improve the success-rate of image matching. An example is given in Figure 5.6. In this case, if the critical point is not clicked,

the image matching step will fail. Aided by digital snapping, this problem can be overcome to some extent.

Left image Right image

Figure 5.6 Image matching based on digital snapping

5.4.3 Initial Stereo Matching

In the current implementation, three consecutive image pairs of an image sequence are used for image matching. If a point is measured in the left image, the corresponding five epipolar lines, L_{i1}, i= 2, 3, ...,6, in the five other images can be determined. However, it has been observed that the orientation accuracy of the along-track stereo rig is not adequate to perform matching using 1-D epipolar line search. For example, the rate of successful matching is low between image No.1 and image No.4 if matching is only performed along the 1-D epipolar line. As mentioned in Section 5.2, the relative orientation accuracy of the fixed stereo rig is much higher than that of the along-track stereo rig, and as well, the

baseline is shorter, which is good for limiting the matching search range. Therefore, we use the current stereo pair, i.e., the fixed stereo rig, to derive a set of initial match candidates. If no match candidate is found in this step, the previous left image or the next left image in the image sequence will be used to perform the initial stereoscopic image matching step (see Figure 5.3).

In the initial stereo matching step, a number of techniques are employed, namely a weighted cross-correlation measure, correlation coefficient profile analysis, disparity range constraint and two-way matching double-check. A weighted cross correlation measure is designed to compute a correlation coefficient profile along an epipolar line. An analysis of the obtained correlation coefficient profile is conducted to determine a set of match candidates. The matching search range is restricted to a certain length along the epipolar line by using a disparity range constraint. At this step, in order to avoid complete match failure, i.e., the correct match is suppressed, a very small threshold of correlation coefficient, e.g., 0.4, is applied, and all peak points in the obtained profile of correlation coefficients are chosen as initial match candidates. A correlation window of 15 (H) x 11 (V) size is used in all experimental tests. As a result, a relatively large number of points are available in the set of matching candidates.

The two-way matching double-check is then applied to filter out some inconsistent candidates. We take each candidate as a master point and perform matching backwards in the current left image around the input point. If the matched point is the same as or close to the original input point within 2 pixels, the candidate is acceptable; otherwise it is rejected

from the candidate set. The two-way matching double-check scheme was proposed by Hannah (1989) and has been widely adopted for area-based image matching. Its advantage in consistent matching has been demonstrated by Fua (1991). In our system, this function is treated as an option since it increases the computational cost.

5.4.3.1 Weighted Cross-Correlation and Coefficient Profile Analysis

The weighted cross-correlation (WCC) measure was introduced by Mori (1973) in the early 70's, and has seldom been used. Recently, it has been found that the WCC measure has a good performance in tracking features in image sequences (Zheng and Chellappa, 1993). Since greater weights are placed on pixels that are closer to the window center, WCC favors the center parts at the cost of the surrounding portion, thus the deformation and occlusion ambiguities can be addressed to some extent by using the WCC measure.

The WCC measure is constructed as follows in our implementation:

$$w(I, J) = w_{min}+(D_{max}-D)(1-w_{min})/D_{max}$$
$$= 1 - D(1-w_{min})/D_{max}, \qquad D=max\{|I|,|J|\} \qquad (5.6)$$

where $w(I, J)$ is the weight value for the pixel (I, J). w_{min} is the minimum weight value (e.g., 0.5) defined in the correlation window controlling the contribution of the surrounding pixels. The maximum weight is normalized to 1. D and D_{max} represent, respectively, the distance from the current pixel (I, J) to the window center and the maximum distance value

within the window (i.e., ½ of window size). This is a linear weighting form. Other weighting functions can also be applied, such as a Gussian function or an exponential function. Finally, the WCC coefficient, ρ, for the correlation windows *i* and *j* becomes

$$\rho = \frac{\sum wg_i g_j - (\sum wg_i \sum wg_j)}{\sqrt{(\sum wg_i^2 - \sum wg_i \sum wg_j)(\sum wg_j^2 - \sum wg_j \sum wg_j)}}, \qquad (5.7)$$

where, g_i and g_j represents a gray value in master and slave window, respectively, and is the weight defined in Eq. (5.6). Figure 5.7 shows an example of a comparison of the correlation coefficient profiles along an epipolar line computed using an unweighted correlation coefficient (CC) measure and the WCC measure. The solid green profile was computed by the CC measure and the red profile was computed by the WCC measure. We find that the difference between the CC profile and the WCC profile is not significant in most cases. However, in general, the profile shape of WCC peaks is sharper than that of CC peaks. The averaged WCC peak values are about 15% higher than the CC peak values. This is due to the fact that the WCC measure places more emphasis on the window center part rather than the whole window area.

It can be assumed that the best match point appears at one of the peak points of the WCC profile. Consequently, a search algorithm is developed to detect these local peaks. The peak is defined as a match candidate, if its associated correlation coefficient surpasses a threshold (0.4). This threshold level is represented as a yellow line in Figure 5.7. The obtained candidates are marked by "x" on the epipolar line.

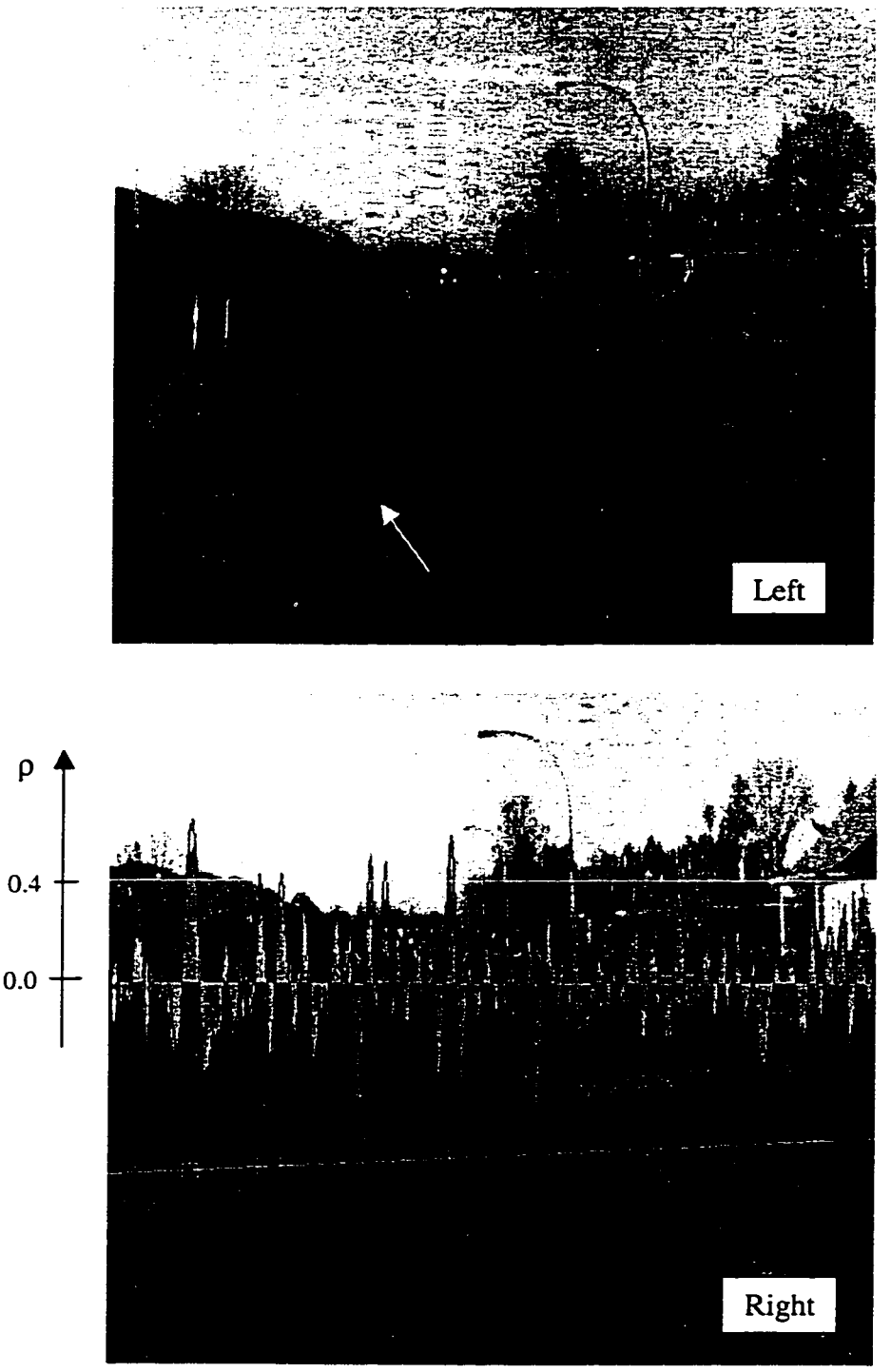


Figure 5.7 Correlation coefficient profiles and match candidates

5.4.3.2 Disparity Range Constraint

As discussed in Chapter 4, the object to be measured should lie in a certain object distance (depth) range in order to meet the accuracy requirement. This depth range controls a disparity range in images and forms a valuable constraint. As a result, the matching search can be restricted to the disparity range along an epipolar line. Considering a normal case stereo, Figure 5.8(a), the following relationships hold:

$$x'_R/f = x_1/Y_{max},$$

$$x''_R/f = x_2/Y_{min},$$

$$x_L/f \cong (B-x_1)/Y_{max} \cong (B-x_2)/Y_{min}. \qquad (5.8)$$

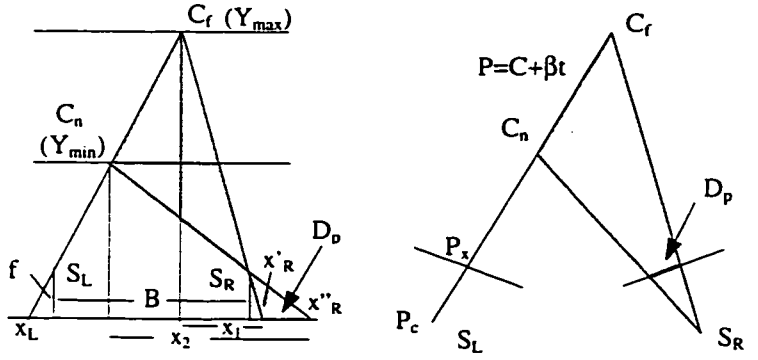


Figure 5.8 Disparity range constraint (a) Normal case (b) General case

The disparity range D_p in an image can be determined by the known maximum and minimum depth values ($Y_{min} \sim Y_{max}$) as

$$D_p = x''_R - x'_R = fB(Y_{max} - Y_{min})/Y_{max}Y_{min} . \quad (5.9)$$

The starting search position x'_R can also be computed as

$$x'_R = (Bf - x_L Y_{max})/Y_{max} . \quad (5.10)$$

When using a rigorous camera model, shown in Figure 5.8(b), an arbitrary 3-D ray originating from the camera center to a 3-D object point can be represented as

$$P = C + \beta t , \quad (5.11)$$

where, P is an arbitrary point on the line, C can be considered as the furthest (or nearest) object point on the line, β is the orientation vector of the line, and t is a parameter representing the distance from the point C to the point P. Using the known coordinates of the measured image point, P_x, and the perspective center point, P_c, along with the furthest (or nearest) object distance value t (e.g, $5 \leq t \leq 60$ m), the furthest (or nearest) object point coordinates C_f (or C_n) can be calculated by

$$P_x = C_f + \beta\, t \qquad or$$

$$P_c = C_f + \beta\,(t+d), \qquad t = 60m,\ d = |P_x - P_c|,\ \beta = (P_x - P_c)/d\,. \qquad (5.12)$$

Thus the disparity range D_p can be determined using a back projection of the points C_f and C_n.

The disparity range constraint is applied to screen the match candidates whose object distance is too far to calculate the object coordinates from the current camera station. Any candidate passing this condition is marked by a white plus '+' sign (see Figure 5.7).

5.4.4 Consistency-Filtering Using Multinocular Epipolar Constraint

After the determination of a set of left-right match candidates, what remains is to filter out the spurious match candidates and validate the best match. Based upon the multinocular epipolar constraint, an algorithm for consistency filtering is designed as follows, illustrated in Figure 5.9:

(a) Calculate the epipolar lines L_{i1} of the input point in the other four images No.3~No.6.

(b) Take a match candidate in the right image (No.2) as a group master, marked by the double signs, red cross 'x' and white plus '+', and then compute the corresponding epipolar lines, L_{i2} (i=3,4,5,6), in the other four images.

(c) Using the WCC measure, search each of the four images for corresponding points, within the window of which the center is determined by the intersection of two epipolar lines L_{i1} and L_{i2}. The size of the search window is set as 11 x 11 considering the camera orientation errors. A point is considered as a group member of the group master, if its correlation coefficient is maximum in the search window and surpasses a threshold (0.6 rather than 0.4 is used in this step for the sake of reliability).

(d) Repeat the steps (a) to (c) and go through all match candidates found in image No.2.

(e) Accumulate the number of members for each group. The best match group is found if its number of group members is a maximum. As shown in Figure 5.9, the best match group points are marked by the white boxes, which indicates a successful match.

In some cases, a maximum as described above may not be found. If there are more than one group having the same maximum number of members, the group in which a member has a maximum weighted correlation coefficient is considered as the best match group. The yellow boxes, indicating "caution", will be used in this case.

Finally, all the points belonging to the best matching group will be used to perform a multiple-baseline photogrammetric intersection and to calculate the 3-D coordinates of the corresponding object point.

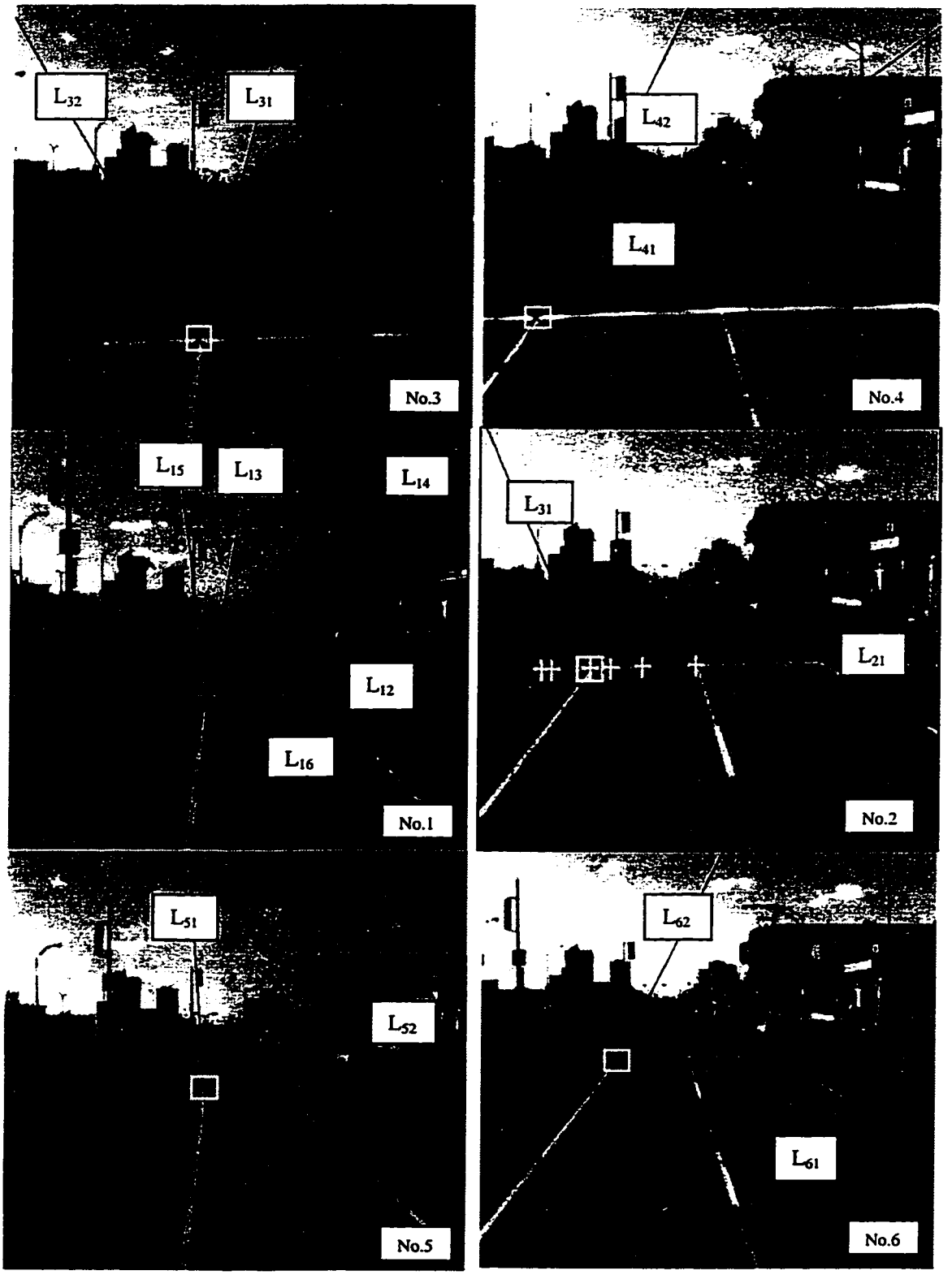


Figure 5.9 Multiple-image matching (The black crosses represent match candidates, and L_{ij} represent the epipolar line of the i and j image pair)

5.5 Test Results and Evaluation

5.5.1 Evaluation of Reliability

Test I

To evaluate the reliability of the method, two types of the tests were designed. The first test was used as a statistical evaluation of the performance of the multiple-image matching method. To this end, a number of stereo image pairs with different road scenes were chosen from the test image sets (see Appendix B for a detailed description). Firstly, the feature points in the left images were extracted by using the Moravec interest operator (described in section 5.4.2). An algorithm for the local non-maxima suppression was applied to determine the local maximum of the interest points (Forstner and Guelch, 1987). The local maxima of the interest values were taken as feature points. In the test, the threshold of interest value was set as 25, so that a certain number of feature points were extracted. Totally, ten test images were used. One of them is shown in Figure 5.10. The detected feature points are marked as black dots overlaid on the original left image. There are 45 feature points detected in this image (Figure 5.11a). As we shall see that some of detected points represent features on moving objects (e.g., cars) or objects which are too far from the camera (e.g., buildings, trees). After screening these points manually, 17 feature points (illustrated in Figure 5.11b) remain in this image. Using the same procedure, a total of 214 feature points was obtained from ten test images.

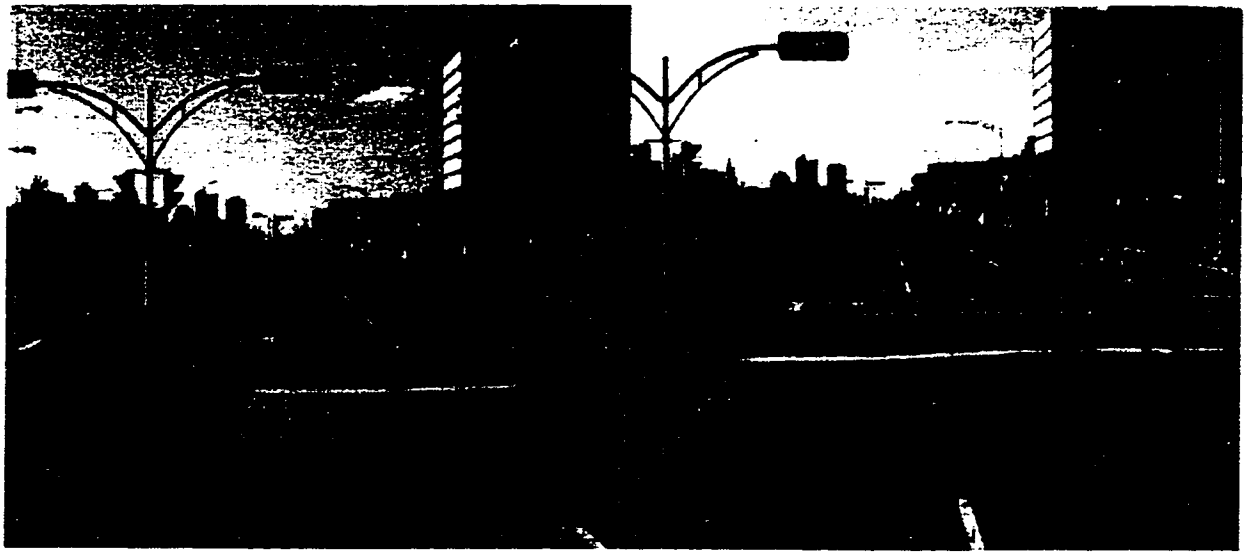


Figure 5.10. Test images for image matching (detected feature points are marked by black dots in the left image)

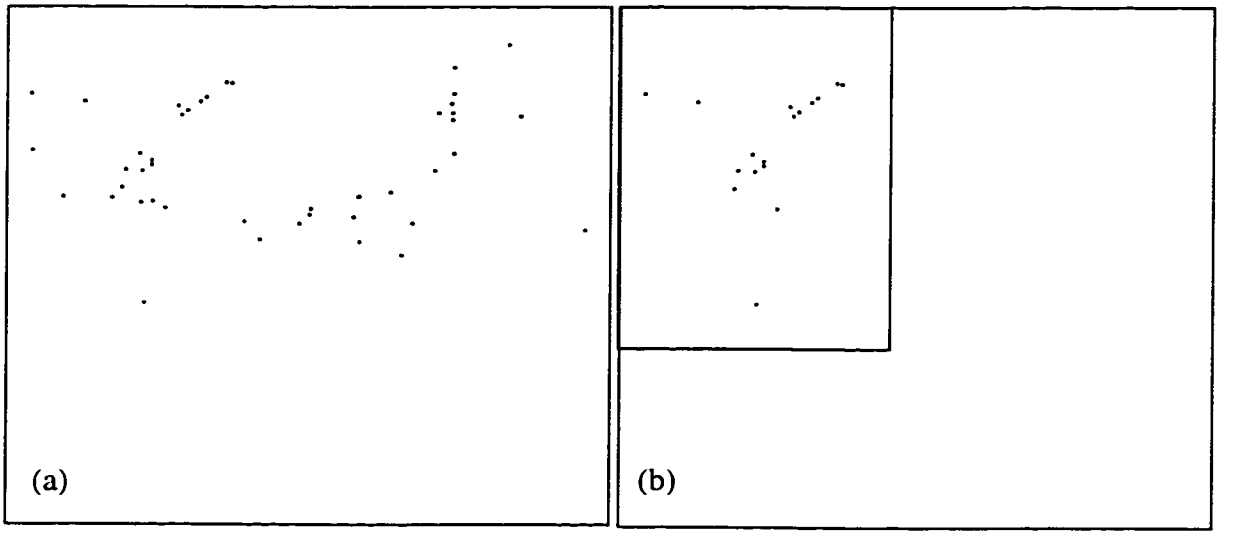


Figure 5.11 Detection of feature points (a) Detected feature points (threshold 25), and (b) Valid feature points for image matching

Secondly, all remaining feature points were taken as input points. The multiple-image matching method was performed using the current, forward and backward consecutive image pairs. Finally, a manual check of matching results was conducted on a point-by-point basis. It was found that 203 matches were successful. This indicates that the successful matching rate is 94.9% for feature point matching when using the multiple-image matching method.

The distribution of the number of determined corresponding points for each feature point is tabulated in Table 5.1. The results demonstrate that, there are about 85% cases in which more than one corresponding points can be found.

Table. 5.1 Distribution of the number of corresponding points

# of corresponding points for each feature points	0	1	2	3	4	5
# of determined corresponding points for total 214 points	11	24	49	59	29	42
Percentage (%) of 214 points	5.1	11.2	22.9	27.6	13.6	19.6

The selection of the input points for image matching is objective in this test, since the feature points are detected by an automatic interest operator, rather than selected by a human operator. Therefore, human factors are avoided in the test. These feature points are distinct in terms of intensity variations, but have no physical meaning with respect to the

objects of interest. Essentially, this test demonstrates that the developed multiple-image matching method can also be used for a general purpose of feature matching.

Test II

The second type of test was to evaluate the performance of the multiple-image matching method for the measurement of objects of interest, including point-, line- and polygon-type objects, such as, electrical poles, traffic signs, manholes, fire hydrants, road lines, and road boundaries. For comparison, stereo image matching using CC and WCC, and the multiple-image matching (MIM) methods were applied on the same data set. Totally 69 object points of interest were manually selected from the image sets. Then, three matching methods were performed to find the best matches. Based upon a manual check, the successful matching rates of three methods were 82.6% (CC), 84.1% (WCC) and 92.8% (MIM), respectively. The unsuccessful matches were 12 points (CC), 11 points (WCC) and 5 points (MIM), respectively. The improvement of the use of the WCC measure against the CC measure was not significant from the test; however, the MIM method demonstrates that great progress is made towards the matching reliability. The reliability of the developed method of multiple-image matching can achieve 90% in both test-I and test-II.

A typical example of multiple-image matching is given in Figure 5.9. The image shows a horizontal feature, a traffic stop line. As can be seen, the matching ambiguities are distinct in this case, because the epipolar line on the right image almost overlaps on the feature of interest. Neither the CC method nor the WCC method was successful in this situation. However, by using the multiple-image matching method, the correct match point was

uniquely determined, because the ambiguities were greatly reduced by the use of the forward and backward images. Actually, using forward image No.3, the corresponding point can be very easily determined, guided by the epipolar line intersection of L_{31} and L_{32}. It was experienced that the disparity range constraint is very useful in reducing the number of match candidates. Figure 5.9 is a good example of the application of this constraint.

Another example is shown in Figure 5.12. There are four consecutive image pairs, namely No. T-1 to No. T-4, containing two traffic signs of interest. One is a "Stop-line Sign" and the other is a "Pedestrian Sign". Here we need to measure the footprint of the "Stop-line Sign". As can be seen, the footprint images of both signs are very similar, and thus matching ambiguities are very significant. The matching methods, CC and WCC, both failed on the image pairs No. T-1, No. T-2 and No. T-3, but were successful on the pair of No. T-4, because the information from only two stereo images is not enough to resolve the ambiguities. The multiple-image matching method worked well in this case, no matter which image was used as a master image, and no matter which sign was to be measured.

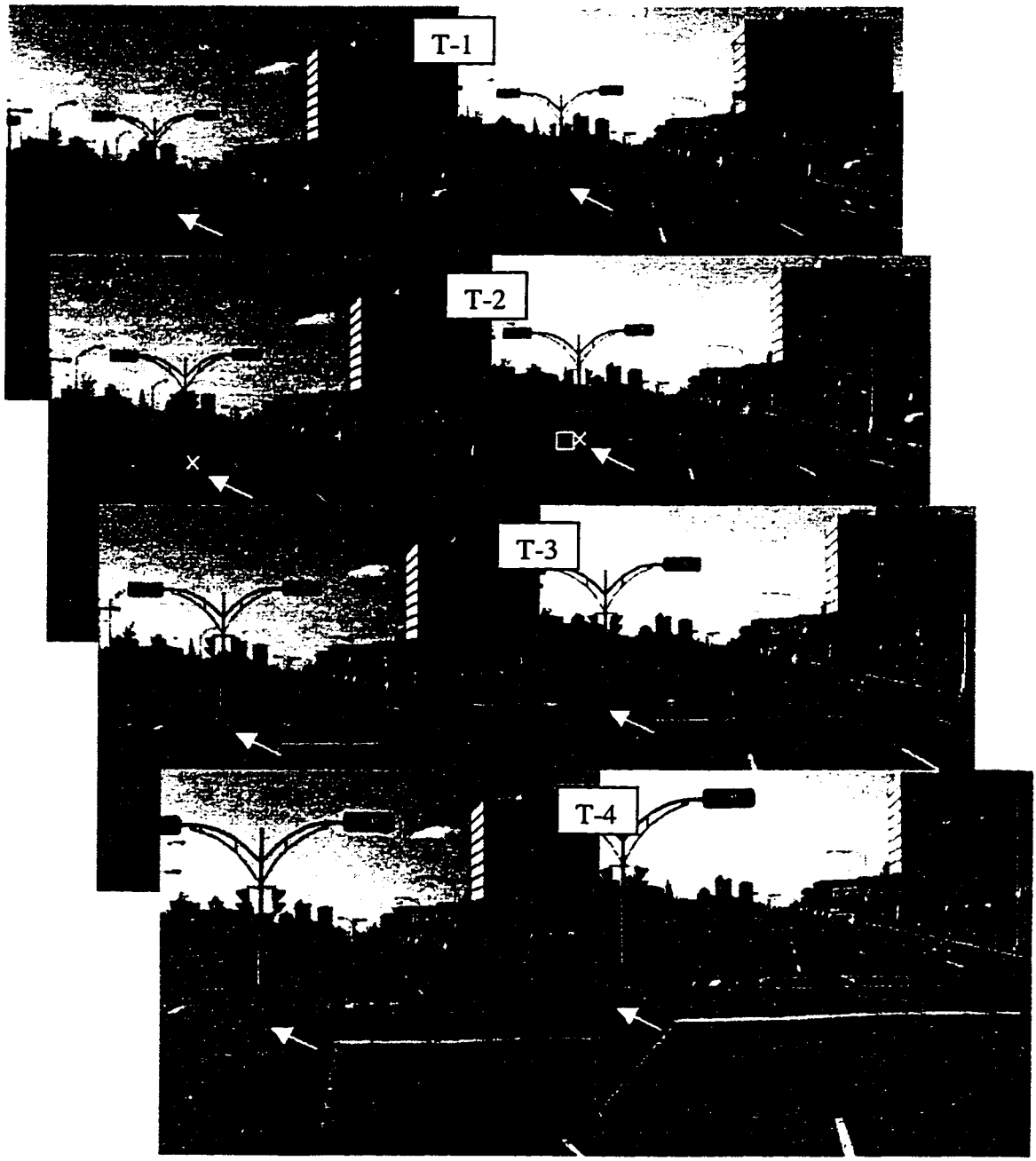


Figure 5.12 A comparison of the two image matching
and the multiple-image matching procedures

5.5.2 Evaluation of Accuracy

By using the multiple-image matching approach, more than two corresponding points can be found and used in the photogrammetric intersection. It is seen from the tests documented in Table 5.1 that, in most cases, more than two corresponding points can be determined from the image sequences. Based on long-term photogrammetric practices and recent research results using the VISAT image data, it can be stated that the intersection results computed from multiple corresponding points with different baseline settings are much better than those from only a pair of corresponding points. The related references have been given in Section 5.2. A number of practical tests have been conducted by Szarmes (1994).

The following example demonstrates the importance of the use of multiple-image matching. Due to large scale variations in terrestrial images, high accuracy sub-pixel image matching is very difficult, and so far there is no good method available to address this problem. The cross-correlation based matching method gives a theoretical accuracy of 0.29 pixel (same as the accuracy of manual measurement described in Section 4.3.1). If the object to be measured is at a distance of 25 m, the resultant error in the object space will be around 15 cm (Table 4.1), in addition to the errors of the georeferencing component (Schwarz and El-Sheimy, 1996). Figure 5.13 (a) and (b) show a stereo image pair. Two correct matching points (marked by "x" in the image) were determined by using the CC method, and their corresponding 3-D point coordinates were also calculated. When this 3-D point was projected onto the third image (a consecutive forward-left image, Figure 5.13c), it is clearly seen that the projected point has a shift error of 8 pixels to the real corresponding point in

this image. Therefore, the use of two images for matching cannot yield accurate reconstruction, while the use of multiple-images with different baseline setting will greatly improve the final accuracy in object space.

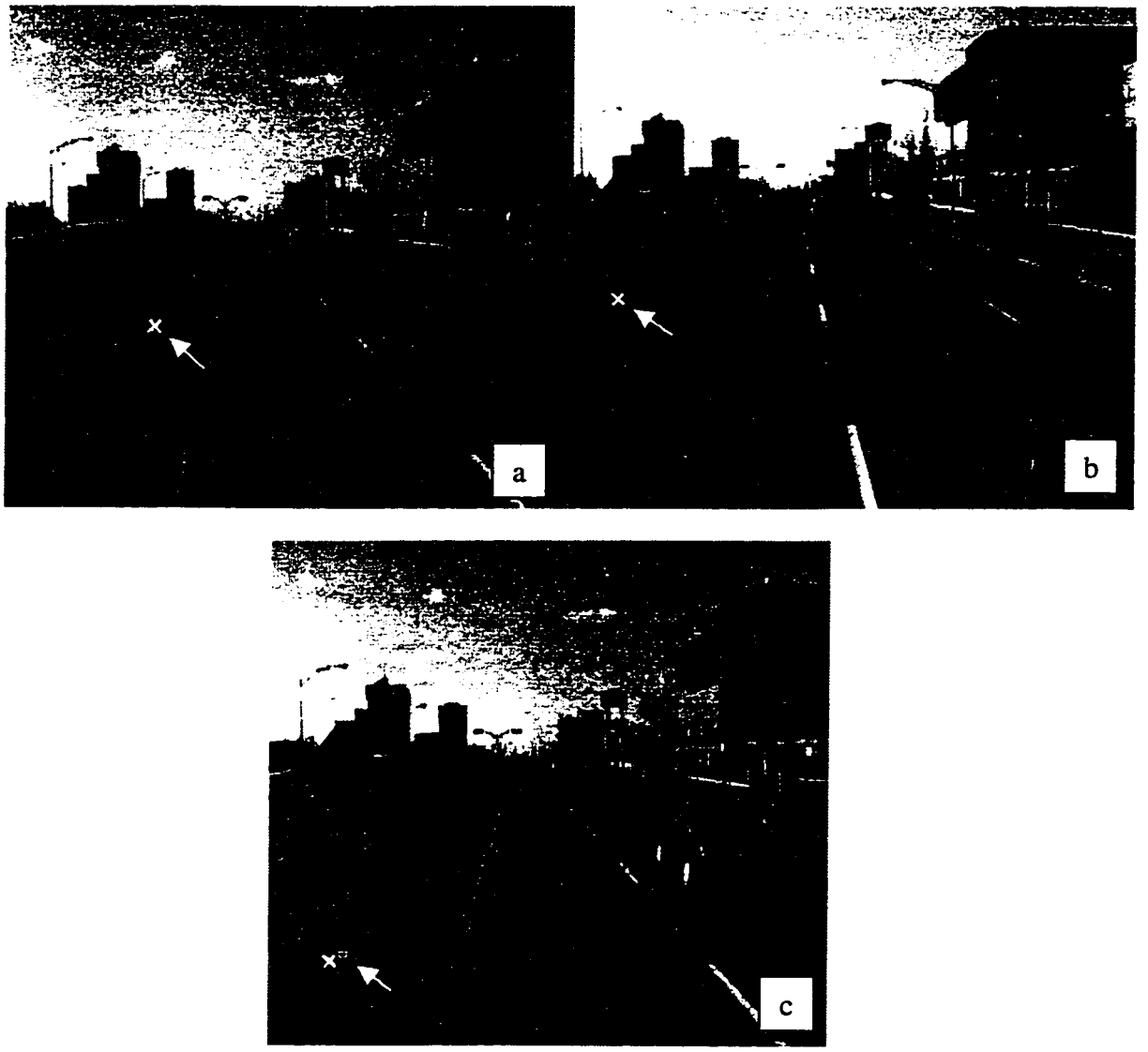


Figure 5.13 Measurement error and mulitple image matching

5.5.3 Evaluation of Efficiency

It is recognized that the identification of a desired object by a human operator is easy, but the accurate measuring procedure undertaken by a human operator takes much time. In this semi-automatic scheme, the human operator is required only to identify an object of interest and locate that point in the left image; the rest of the work can be done automatically. Moreover, by using the digital snapping technique, precise location of the point may not be necessary. Consequently, comparing with the traditional manual measuring procedure, the semi-automatic approach avoids the most time-consuming process, i.e., manual measurement, and takes the advantage of the human capability of object identification.

Although the speed of the multiple-image matching method is not high, the reliability can be greatly enhanced. The reliability of the matching is our top priority. So far, it generally takes about 15-25 seconds to complete the matching using the multiple-image matching method based on a low-end Unix computer, SUN-IPX. A major reason for this is that the two consecutive image pairs used for the matching are not stored in the memory in the current implementation, so that the time cost on reading images is significant. Generally speaking, the multiple-image matching method is two or three times slower than the CC or WCC method (two image-based matching). It is believed that, with enhanced computing hardware and an optimized version of the algorithm, the speed of multiple-image matching will not be a barrier to its practical applications.

5.6 Implementation of Object Measurement in ImagExpert

ImagExpert is a software package developed for intelligent processing of the VISAT image data. Appendix A gives a more detailed description of the software package. The core function in ImagExpert is object measurement, since it is considered a major tool during the data processing. To facilitate the measurement procedures, a number of options for object measurement have been developed. In accordance with the differences of the operation requirements for a human operator, four options are available and listed in Table 5.2.

Table 5.2 Functions of Object Measurement

Options (1~4)	**Operations by operator**		**Support Techniques**
	Left image	**Right image**	
Mode I	Pointing precisely	Pointing precisely	FBVM
Mode II	Pointing precisely	Pointing approximately	LRM+FBVM
Mode III	Pointing precisely	Pointing not required	MIM
Mode IV	Pointing approximately	Pointing not required	DSP+MIM

[Table Notes] FBVM: Forward-backward verification matching
LRM: Left-right image matching
DSP: Digital snapping
MIM: Multiple-image matching

Option I

(OPERATOR) The human operator is required to precisely measure the points in two images (left and right). The basic functions, such as zoom-in, undo, and edit, have been developed to assist in the manual measurement process.

(COMPUTER) Since the two corresponding points are given, the 3-D coordinates of the object point can be calculated. By using a back projection of the 3-D point onto the forward two images (No. 3 and 4) and the backward two images (No. 5 and 6), the approximate positions of the corresponding image points of the original given two points are available. The matching is performed between the current left image and the forward/backward four images around these projected positions to search for the corresponding points. At the end, all the matched points are used to perform a multiple-baseline photogrammetric intersection. The above process is termed as forward-backward verification matching (FBVM). This method increases the accuracy of 3-D object coordinates (Li et al., 1996; Zou, 1996) and has been used a recommended option.

Option II

(OPERATOR) The human operator is required to precisely measure a point in the left image and approximately give a point position in the right image.

(COMPUTER) Since the approximate position of the corresponding point in the right image is given, image matching can be constrained in a small window. The rate of successful matching can be increased to a large extent. Followed by this step (left-right

image matching, LRM), FBVM will be executed to further improve the accuracy of object measurement.

Option III

The human operator only gives a precisely measured point in the left image and the rest of the work will be done automatically. The multiple-image matching method is employed in this case.

Option IV

The human operator gives a roughly measured point in the left image and the rest of the work will be done automatically. This task is fulfilled by using the digital snapping technique along with the multiple-image matching method.

5.7 Chapter Summary

An approach to semi-automatic object measurement using multiple-image matching was proposed and developed. The reliability of the multiple-image matching method comes from the use of georeferenced multinocular image constraints. The matching ambiguities are greatly reduced by using a multinocular epipolar constraint associated with the digital snapping, the weighted cross-correlation, the correlation coefficient analysis, the disparity range constraint, and the two-way matching double-check techniques. Moreover, the final accuracy of 3-D object coordinates is increased. It is believed that this method is of particular importance to assist the user to perform fast, accurate and reliable feature extraction. It is also worthwhile research that if a higher image-capture rate is available, further improvement to robust matching could be achieved by using a cooperation of feature correspondence and optical flow approaches.

CHAPTER 6

AUTOMATIC RECONSTRUCTION OF ROAD CENTERLINES FROM LONG IMAGE SEQUENCES

An automatic approach to road centerline reconstruction using stereo image sequences is proposed based on a shape from sequences scheme. The reconstruction of 3-D road centerlines is formulated as a physical problem of solving an energy-minimizing 3-D B-spline shape model. The reconstruction is described as a process whereby a 3-D road centerline shape model is deformed gradually, driven by internal and external forces. The road centerline model is represented as a deformable and elastic 3-D curve, composed of an abstract material. The external forces, generated from road centerline data extracted from image sequences, control the local shape of the model. The internal smoothness forces, arising from the road shape knowledge, control the global shape of the model, and also take care of the fact that the external forces are noisy or sparse.

6.1 Previous Work and Problem Identification

As discussed in Section 1.4.2, a number of methods have been developed for the extraction of road centerlines from image sequences in the areas of mobile robotics and autonomous

vehicle navigation. However, these methods have primarily aimed at visual navigation rather than object reconstruction (Dickmanns and Zapp, 1986; Thorpe et al., 1988). In general, there are two schemes for road extraction: region-based segmentation and edge-based following.

The region-based segmentation scheme is mainly used when color images are available. Using color classification techniques, the region of road pixels can be segmented from the image background so that the orientation of the right road can be determined. The development of adaptive color models for pixel classification is the main research focus.

The basic principle of the edge-based following scheme is to, firstly, apply an edge detector to detect the edges of centerline markers, and then use a grouping/following algorithm to link the edges. In terms of edge following, three main algorithms have been proposed:

(a) Approximation using an analytic function: when edges are detected, an analytic function, e.g., a polynomial function, is used to approximate the edge points (Dickinson and Davis, 1990; Schneiderman and Nashman, 1994);
(b) Hough transformation: the Hough transformation algorithm is applied to find a straight line from edge points (Davis and Kushner, 1985); and
(c) Least squares matching: the area-based least squares matching method can also be applied to follow the centerline edges (He and Novak, 1992).

We have tested the above three algorithms using the VISAT images. Each algorithm has its own pros and cons, but none of them is very reliable. They are all sensitive to image contents and noise. Long-term experience from the previous research also indicates that automatic extraction of road centerlines from image alone is still restricted.

The real problematic situations affecting the extraction method can be summarized as follows (Tao et al., 1996): centerlines with missing, intermittent and faded markings; a variety of lane marking types, noisy markings with respect to the centerline markings (stop lines, junction lines or direction lines); lighting conditions (sunshine, rain or cloud) and shadows (trees or buildings); road types (highway, city street or rural road) and material (concrete or asphalt); road shape (straight, smooth or sharp curve, or intersection); road coverage (snow, water, leaves or stones); and road quality (featureless, cracked, or pot-holed). In addition, some characteristics of terrestrial imagery makes the problem more complicated, namely, large scale variations within an image, large geometric and perspective distortions, and occlusions (e.g., due to moving vehicles). Several typical examples of road images captured by the VISAT mobile mapping system are shown in Figure 6.1 (6.1a: noisy markings, gaps and road surface cracks; 6.1b: shadows and occlusions (car blockage); 6.1c: large gap in a T-intersection; 6.1d: curved and dashed markings).

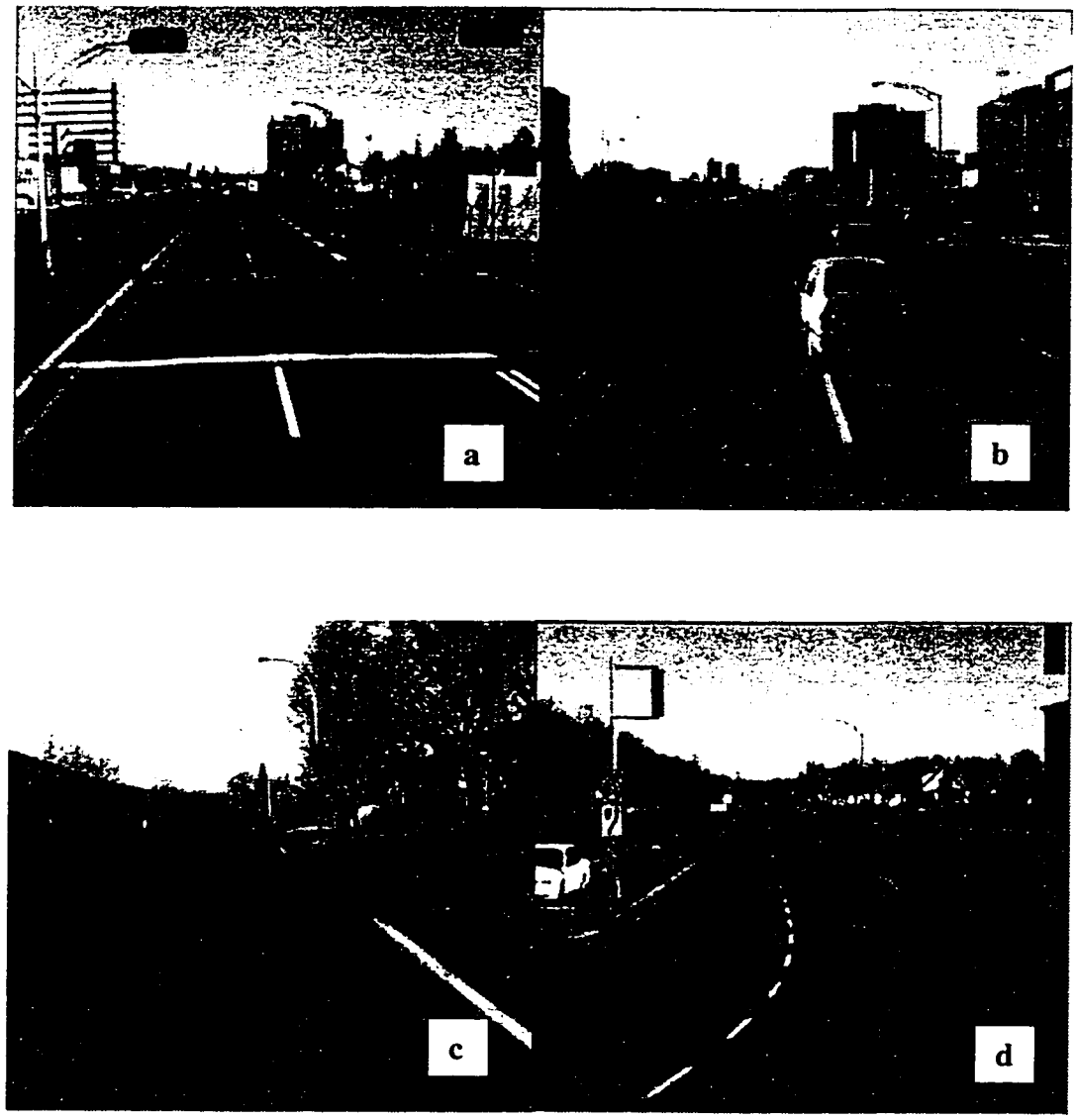


Figure 6.1 Examples of road images

It is observed that the above mentioned methods rely on image information alone and only use a single image. Due to the fact that many scenarios and road conditions have to be taken into account, an innovative approach has to be developed in order to realize an automatic and continuous reconstruction of road centerlines from long image sequences. To this end, a new reconstruction scheme was designed, and external information and constraints must

be explored. As discussed in Section 1.5, three valid constraints are available in the mobile mapping systems, namely, (1) *stereoscopic and sequential imaging geometry*; (2) *image georeferencing constraint;* and (3) *known vehicle ego-motion constraints.*

It is also recognized that the model-driven (top-down) scheme is much more controllable and reliable than the image-based (bottom-up) scheme if the model information can be obtained. Deformable models have been of considerable interest in current computer vision research. It is one of new model-driven methods, which has been used in edge detection, contour extraction, stereo matching and object modeling (Fua, 1991; Kass et al., 1988; Menet et al., 1990; Terzopoulos et al., 1988). Recently, deformable models have also been recognized and utilized by photogrammetric practitioners (Gruen and Li, 1995, 1996; Trinder and Li, 1995; Tao and Lin, 1994).

The main advantage of deformable models is that both geometric and physical constraints can be incorporated into a mathematical model, whereas in the conventional methods, only geometric constraints are considered. The multiple constraints coming from different sources can be synthesized into a mathematical model through an energy combination. For this reason, an approach based on the idea of shape from sequences is proposed, in which a physically-based 3-D deformable curve model is employed to accommodate the combinations of multiple constraints available in the mobile mapping system for 3-D reconstruction of road centerlines. For detailed information on deformable models, one can refer to Kass et al. (1988) and Terzopoulos et al. (1988).

6.2 Overview of the Reconstruction Approach

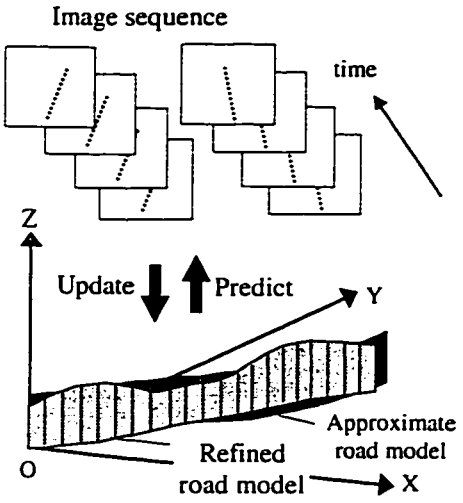


Figure 6.2 The concept of shape from image sequences

The reconstruction of road centerlines from image sequences is considered as a problem of "*shape from image sequences.*" The problem is to use road centerline information available from successive images in a sequence and road shape knowledge to update and refine a 3-D shape model of road centerlines (see Figure 6.2). Thus optimized road centerlines are obtained through a dynamic and global energy minimization procedure. Based on the shape from sequences scheme, a systematic combination of multiple constraints from a mobile mapping system can be implemented (Tao, 1996b). The schematic flowchart of the approach is illustrated in Figure 6.3.

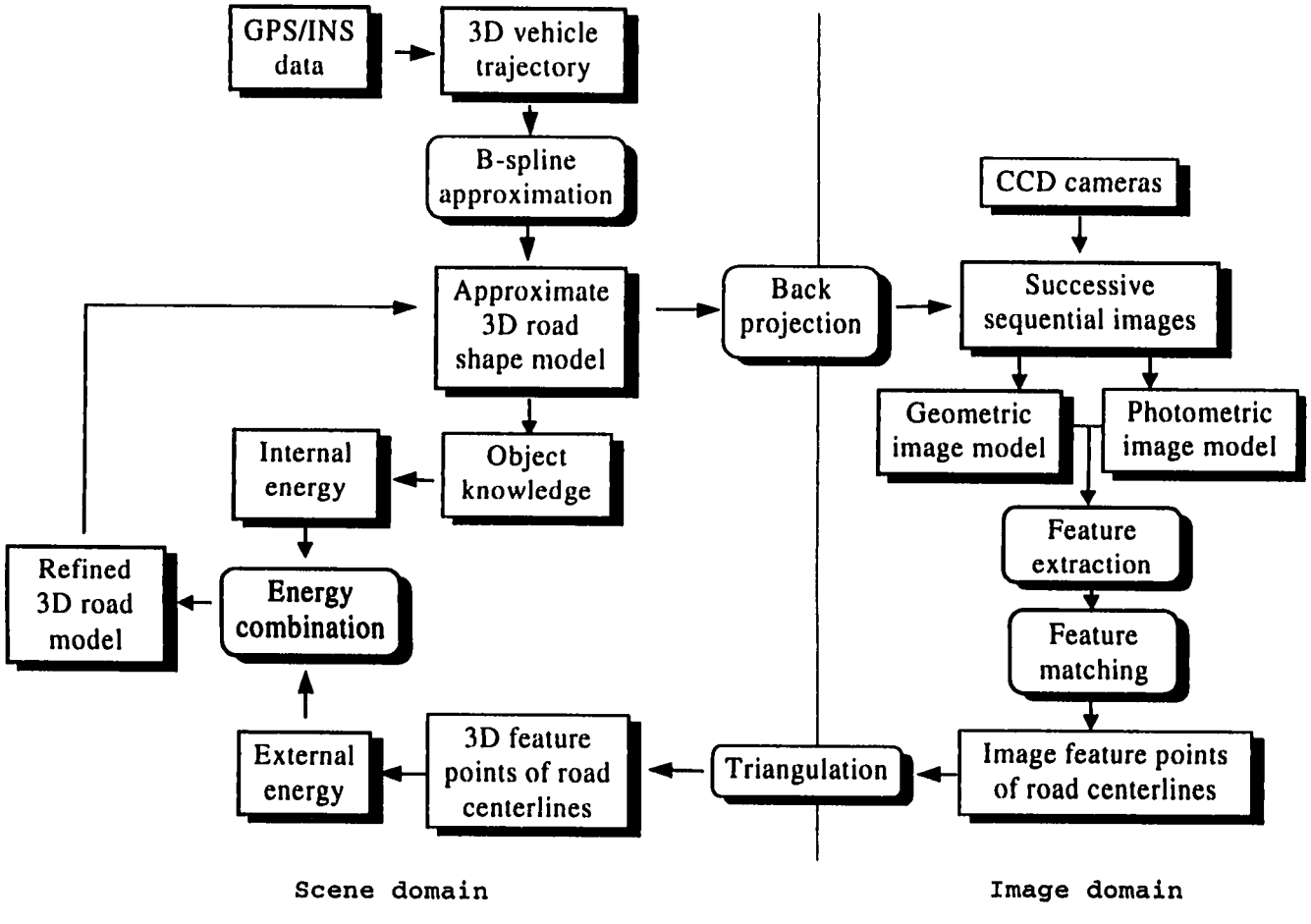


Figure 6.3 A schematic flowchart of the reconstruction approach

This is a closely coupled bottom-up and top-down scheme. Firstly, an approximate 3-D road centerline model is set up by means of a B-spline approximation of the kinematic vehicle trajectory data. Secondly, the road model can be back-projected onto the image sequences using the known orientation parameters of the cameras. Guided by an established geometric and photometric image model of road centerlines, model-driven extraction of road centerline feature points is performed, constrained by the projected road centerline. Furthermore, a stereo-motion matching algorithm based on the stereo and motion constraints is applied to find a set of matched feature point pairs. After using photogrammetric triangulation of feature point pairs, a set of 3-D feature points of road centerlines in the object space is obtained. Thirdly, in order to refine the approximate road model using the information extracted from the image sequences, the model is defined as an active and deformable 3-D curve, the 3-D snake (Kass et al., 1988). A physically-based deformation mechanism is incorporated such that the model can be progressively deformed driven by the internal and external energies. Internal energy arises from smoothness constraints representing the natural characteristic of the shape of road centerlines. It maintains the a priori knowledge about the shape of the road model. The 3-D feature points extracted from image sequences act as the external energy which enforces the model to deform towards its desired position. Under a combination of the actions of internal and external forces, the model will be deformed incrementally towards the final state at which forces from different sources are balanced. The newly deformed and refined model will then be used to update the approximate model, and the new information available from successive new images will be applied to further refine this model. The above process is

executed repeatedly until the entire image sequence is processed. The finally obtained deformation curve can be treated as an optimized reconstruction result of 3-D road centerlines, because such a result is derived by using globally combined constraints.

In the following sections, three key methods to the implemention of this approach will be addressed:

(1) generation of an approximate 3-D shape model of road centerlines;

(2) extraction of reliable information of road centerlines from image sequences; and

(3) dynamic refinement of a physically-based 3-D road centerline model.

6.3 Generation of an Approximate Road Shape Model

The approximate 3-D shape model of road centerlines forms the basis of the implementation of the shape from sequences scheme. To this end, the dynamics of the vehicle motion (ego-motion constraint) are utilized. The 3-D vehicle trajectory data have been used to generate an approximate shape model of road centerlines. In VISAT, the integrated GPS/INS component has been employed to determine the vehicle trajectory (Schwarz et al., 1993b). After a post-mission Kalman filtering, the trajectory is obtained with high accuracy, and represented in a global coordinate system.

A 3-D cubic B-spline is introduced in defining the 3-D road shape model:

$$Q(u_j) = (x(u_j), y(u_j), z(u_j)) = (\sum_{i=0}^{m-1} V_i B_i(u_j)), \qquad j=0,1,...,n-1, \qquad (6.1)$$

where B_i are the basis functions of B-splines, $V_i=(X_i, Y_i, Z_i)$ are a set of control vertices of the B-spline curve, and n and m are the numbers of sampling points on the curve and control vertices, respectively. The reasons for choosing B-splines to represent the shape model are as follows: (a) compact representation: the curve can be parameterized very compactly by its control polygon V_i ; (b) local control: only a small part of the curve needs to change if a certain control vertex is modified; (c) feature compatibility: corners or straight lines can be imposed on the curve if multiple control vertices are imposed; (d) invariant characteristics: the control vertices of B-splines are invariant under affine and projective transformation (Cohen et al., 1995); (e) numeric advantages: the number of unknowns in Eq. (6.1) is m, the number of control vertices, instead of n, the number of sampling points on the curve ($m<n$). Moreover, the B-spline representation can be extended to define a deformable curve model (B snake) which is a key to our approach for road centerline reconstruction. By means of a B-spline approximation of the vehicle trajectory data, the approximate shape model is generated. This approximation is accomplished by a least-squares method. Detailed mathematical description of B-spline approximation of a group of discrete data can be found in Bartels et al. (1987).

6.4 Extraction of Road Centerline Features

The extraction of reliable road centerline information from image sequences is performed by two-stage processing: feature point extraction of road centerlines and feature point matching from an image sequence.

6.4.1 Model-Driven Feature Extraction

By the use of the approximate road model, the approximate positions of road centerlines on the images can be predicted through the model projection (scene-to-image transformation). A constrained processing window for feature extraction can be determined for providing the algorithm an area of focus, thus reducing computations. As shown in Figure 6.4, the height of the window can be determined by using the vanishing line condition (Section 5.2), while the width of window is a function of the lane width in the scene domain (Tao et al., 1996). The geometric and photometric model of road centerlines for terrestrial imagery can be represented as a group of rules (Table 6.1). These rules guide the implementation of a reliable feature extraction process.

6.4.1.1 Oriented Edge Detection

An oriented edge detection algorithm is proposed to focus on the utilization of edge direction clues due to the following two reasons: (1) the major clue of the lane marking feature in images lies in the directions of edges rather than the magnitudes of edges. The

magnitudes of edges cannot provide meaningful information for centerline features because they are disturbed by noise, such as from shadows, road coverage, and noisy markings. However, the edge direction along an image centerline is almost perpendicular to the predicted centerline (see Figure 6.4) no matter how big the corresponding magnitude is; (2) as illustrated in Figure 6.5, if the image centerline is oriented at 45 degrees, the nonoriented edge detector will find an intensity slope with gradually varying values. However, if the same edge detector is oriented at 45 degrees, the sharp changes of the intensity values can be detected, and thus the edge can be located.

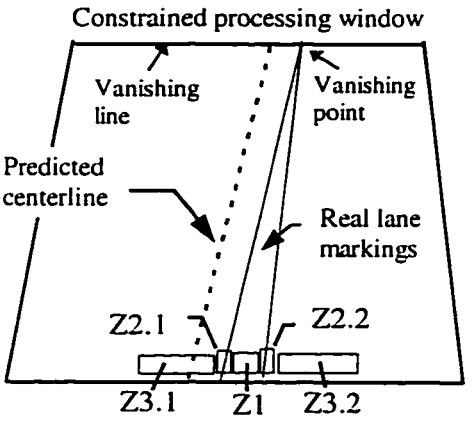


Figure 6.4 Image model of road centerlines (refer to Table 6.1 for details)

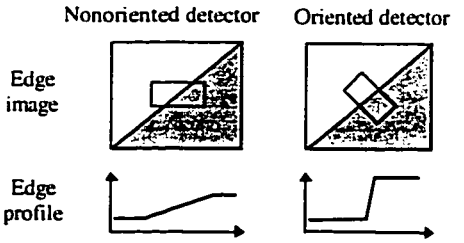


Figure 6.5 Oriented edge detector

Table 6.1 Rules for feature extraction (refer to Figure 6.4)

Rule ID	Rule Name	Rule Description
Rule 1	Magnitude	The magnitude of the edge should surpass a certain threshold.
Rule 2	Direction1	The direction of the desired edge should be perpendicular to the predicted centerline with an acceptable tolerance of ±20°.
Rule 3	Direction2	The gradient directions of edge z2.1 and z2.2 must be in opposite sense with a given tolerance of ±15°.
Rule 4	Distance	The distance between edge z2.1 and z2.2 should be within a few pixels (2-5 pixels).
Rule 5	Zone gray	The average gray value of the zone z1 must surpass a threshold and must be brighter than zones z3.1 and z3.2.
Rule 6	Continuity	If the distance between a dual-edge and its nearest dual-edge is beyond a certain range, the dual-edge will be removed.
Rule 7	Smoothness	The position of the center point of a dual-edge should be located along a smooth line.

The oriented edge detection algorithm is implemented in three steps: coarse edge detection, edge thinning and precise edge detection. Firstly, to avoid the time-consuming computation of the gradient of each point, four directional *Prewitt* masks, coded 0 to 3, are used to convolve the image. Then the direction code (0-3) is assigned to each convolved point at which the associated directional mask gives the highest magnitude. After using rule (1) in Table 1, edge thinning (Nevatia and Babu, 1980) is conducted. On the remaining edges, the *Sobel* gradient operator is oriented according to the direction code of the edge, and the edge gradient direction is computed. By applying rule (2), the final edges are obtained and recorded as a set of candidate feature edges of road centerlines.

6.4.1.2 Dual-Edge Detection

Road centerlines are painted as white or yellow markings with a certain width. In accordance with this knowledge, a dual-edge detection algorithm is designed to refine the set of candidate feature edges. A dual-edge is defined as a pair of edges located on the cross-sectional boundary of the image centerline, namely z2.1 and z2.2 in Figure 6.4. If an edge selected from the previous candidates corresponds to a real edge of the centerlines, there should exist a "dual" edge, whose gradient direction is opposite to the selected edge (rule 3), and the distance between them must satisfy rule 4. Furthermore, the averaged gray value of the zone z1 should be surpass a threshold and be brighter than that of surrounding areas, namely zone z3.1 and z3.2 (rule 5). The above dual-edge detection algorithm greatly reduces the amount of undesired edge features. The result of this stage is shown in Figure 6.6. The dual-edges are represented as the short line-segments.

6.4.1.3 Feature Filtering By Majority Voting

The edges of image centerlines should be located along a smooth line, and may not have a large difference from the predicted line. This knowledge is represented by the connectivity rule (6) and the smoothness rule (7). As shown in Figure 6.6, a dual-edge which is far from the others is treated as a blunder (marked by a circle) and removed according to rule 6. Rule 7 is implemented using a proposed majority voting method. This method is able to reliably find out the distribution of the majority of the dual-edges along a smooth line and thus the blunders can be eliminated. Similar to the commonly used Hough transformation method, a transformation parameter space has to be set up firstly. In our 1-D parameter space, the x-axis with a unit of 0.5 pixel represents the distance value between the center point of a dual-edge and the predicted centerline, while the y-axis represents the accumulated number of distance values of dual-edges. This distance value is calculated for each extracted dual-edge and then projected onto the x-axis (see Figure 6.6). The accumulated number appearing most frequently in a particular location of the x-axis is essentially the distance value of the majority of the dual-edges. By searching for the peak point in the parameter space, the peak point D_0 of the distance value can be determined and also the distance range of the majority of the dual-edges along the centerline is obtained (defined as $D_0 \pm 1.5$ pixels). The dual-edges whose distance values are out of this range will be filtered out as blunders. The output of this stage is a set of feature points defined by the center points of dual-edges.

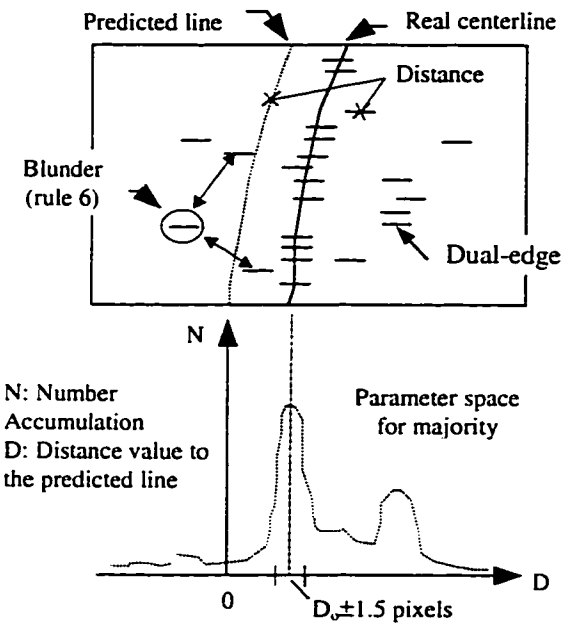


Figure 6.6 Majority voting method for identification of centerline feature points

6.4.2 Constrained Image Matching

This stage is used to (a) further verify the feature points obtained by the previous stage and (b) generate a set of matched feature point pairs among the image sequence.

6.4.2.1 Constrained Matching Range

A useful assumption employed here is that the ground of the road is a flat plane. Since the approximate camera height with respect to the ground and orientation parameters of each camera station are available, the corresponding point position on image B of a given point on image A can be directly estimated using the camera(left)-ground-camera(right)

transformation (Tao et al., 1996). Along with the epipolar line constraint, the searching range for matching a point on image B can be restricted to a relatively small area. In Figure 6.7, the constrained matching range is illustrated as a box and the dotted straight line depicts the epipolar line. This constraint greatly improves the reliability of matching, and reduces the searching space.

6.4.2.2 Ground Plane Based Matching

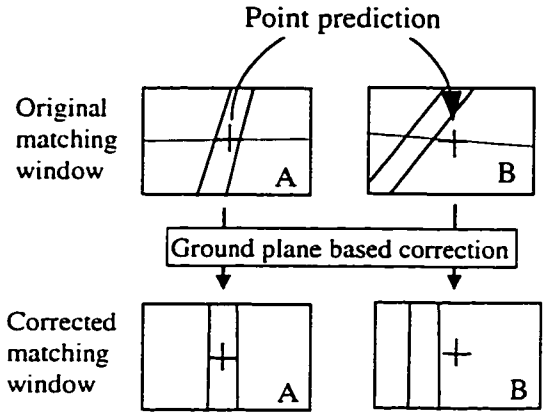


Figure 6.7 Ground-plane-based distortion correction and matching

Although the searching range has been constrained, geometric discrepancies mainly caused by perspective distortion exist between the two windows to be matched (see Figure 6.7). If the two matching windows can be projected onto a plane which is parallel to the ground plane, the perspective distortion problem could be alleviated. Therefore, before the matching step, each matching window on both images is resampled according to a plane-to-

plane transformation using a bilinear sampling technique. Finally, the matching is conducted on the two windows which are "free" of perspective distortions (Figure 6.7).

6.4.2.3 Multiple-Image Matching for Consistency Check

A multiple-image matching method, which is similar to the method developed in Chapter 5, is used as a consistency check. As can be seen in Figure 6.8, a multiple epipolar line constraint is available, in which three epipolar lines corresponding to three other images pass through a single point P_0 on the current left image. Making use of this constraint, an algorithm for consistency matching is designed as follows (refer to Figure 6.8, assuming that the current left image is the master image): (1) for each point from a set of feature points on the master image, perform matching on the three other images constrained by the estimated matching ranges. After applying the ground plane based corrections for matching windows, use Kanade's SSD (Sum of Squared Difference) correlation criteria (Kanade et al., 1992) for matching; (2) for the matched point on the next left image and the current right image, locate the best matches on the next right image in their constrained matching ranges (the arrow indicates its own constrained matching range in Figure 6.8); (3) if two of three matched points on the next right image are within 1.5 pixels, the matched point pair is determined. In the case illustrated by Figure 6.8, P_0 and P_3 are considered as a matched point pair because the distance between the point "Δ" and the point "$\bullet$" is within 1.5 pixels on the next right image. If no matched point pair is found, P_0 will be rejected.

Each point from the feature point set can be taken as the master point and its corresponding image is taken as the master image. The same process described above is executed until a new set of matched feature point pairs is generated from the current images. At the end, the 3-D coordinates of feature points of road centerlines will be calculated by the intersection of the matched feature point pairs. These 3-D feature points will be used to update and refine the 3-D shape model of road centerlines.

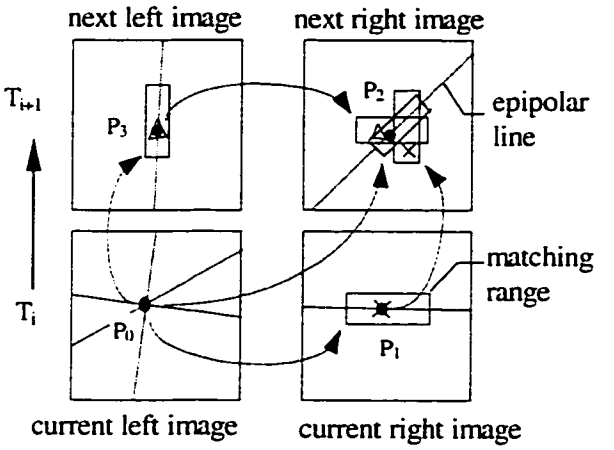


Figure 6.8 Multiple image matching for consistency check

6.5 Dynamic Refinement of a Physically-Based 3-D Road Centerline Model

6.5.1 Physically-Based Deformation Mechanism

The 3-D parametric curve represented by a B-spline (Eq. 6.1) can be treated as a deformable and elastic model (*3-D B-snake*), composed of an abstract material. The model can evolve from its initial position towards its desired position driven by the externally applied forces and internally generated forces. In accordance with the least action principle (Hamilton's principle, Courant and Hilbert, 1953), the dynamics of a deformation model can be described as a motion equation:

$$E = e_1 E_{int}(Q(u)) + e_2 E_{ext}(Q(u)) \rightarrow min\,, \qquad (6.2)$$

where E_{ext} represents the external energy acting on the model, and E_{int} is the internal energy resisting deformation away from the natural state; e_1 and e_2 are constants for weighting these two terms. The minimization of the total energy of the motion equation governs the behavior of the deformation model.

6.5.1.1 Internal Energy

E_{int} is described as the smoothness energy consisting of two terms (Kass et al., 1988):

$$E_{int} = \alpha \left(\int_{\Omega} |Q'(u)|^2 du \right) + \beta \left(\int_{\Omega} |Q''(u)|^2 du \right), \qquad (6.3)$$

in which the first derivative $Q'(u)$ is a measure of the distance discontinuities (stretching effectiveness), while the second derivative $Q''(u)$ is a measure of the orientation discontinuities (bending effectiveness). Adjusting the weights α and β controls the relative balance of the stretching and bending force. E_{int} maintains the local continuity and connectivity of the shape model. It ensures that the curve model will not tear apart, fold onto itself or exhibit high curvature points. The other significant role is that smoothness is constrained on the curve when the external information (external force) is noisy or sparse. Therefore, the internal forces control the global shape of the model and make it less sensitive to the external noisy input.

6.5.1.2 External Energy

The external forces play a key role in pushing the model into the desired real position. They control the local shape of the model. As shown in Figure 6.9, the model is deformed subject to the external forces between the extracted 3-D feature points P and the curve points $Q(u)$. In order to quantify this kind of force, a gravity-like field is used to describe the external forces. The reason for using such a field is that the closer the distance between the feature point and the curve, the greater the force. To avoid the singularity when the distance is approaching zero, the following function in quantifying the external energy is employed:

$$E_{ext} = f(x), \quad f(x) = \begin{cases} x^2/2, & D_p < r_1 \\ 3/2 - x^{-1}, & r_1 \le D_p \le r_2 \\ 0, & D_p > r_2 \end{cases} \tag{6.4}$$

where $x = Dp / r_1$, $D_p = \|Q(u) - P\|$, the distance between $Q(u)$ and P; and r_1 and r_2 are the coefficients representing the influence range of the feature point to the curve. $r_1=0.1$(meter) and $r_2/r_1=4$ are chosen according to the system accuracy. Thus $f(x)$ acts like a spring when the point is close to the curve ($D_p<r_1$), and has no influence when $D_p>r_2$. The effects of blunders are minimized, since they are distant from the curve. The corresponding external force can be derived as:

$$F_{ext} = \nabla E_{ext} = f'(x) = \begin{cases} x, & D_p < r_1 \\ x^{-2}, & r_1 \le D_p \le r_2 \\ 0, & D > r_2 \end{cases}. \tag{6.5}$$

Considering the "historic" effect of external forces, the total F_{ext} becomes

$$F_{ext}=\omega_1 F_{ext}(T_{i-1}) + \omega_2 F_{ext}(T_i) + \omega_3 F_{ext}(T_{i+1}) \tag{6.6}$$

ω_1, ω_2 and ω_3 are weighting coefficients. During the process of iterative solution of Eq. (6.6), the F_{ext} of each point $Q(u)$ along the model should be calculated repeatedly until the motion equations reach a convergent state.

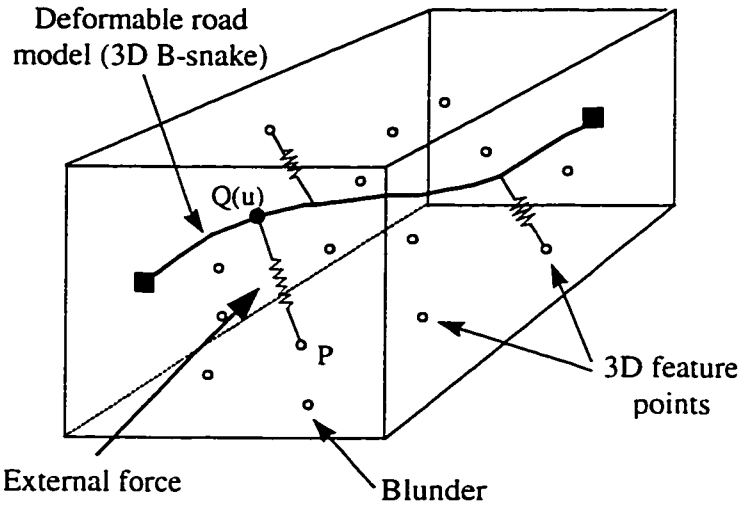


Figure 6.9 Deformation behavior of the model driven by external energy

6.5.2 Solution to Energy Minimization Problem

The combination of the internal energy and the external energy is accomplished by solving for energy minimization of the motion equations. Substituting Equations (6.1) and (6.3) into (6.2), the discrete version of the motion equation can be written as

$$E = \sum_{j=0}^{n-1} \{e_1[\alpha(u_j)(\sum_{i=0}^{m-1} V_i B_i'(u_j))^2 + \beta(u_j)(\sum_{i=0}^{m-1} V_i B_i''(u_j))^2] + e_2 E_{ext}[Q(u_j)]\} . \qquad (6.7)$$

Since this equation is quadratic, the minimum of E can be obtained

$$\partial E / \partial V_k = 0 \quad k=0,1,2, \ldots, m. \tag{6.8}$$

The resultant equations can be derived

$$\sum_{i=0}^{m-1} V_i [\sum_{j=0}^{n-1} \alpha(u_j) B_k'(u_j) B_i'(u_j) + \beta(u_j) B_k''(u_j) B_i''(u_j)]$$

$$= -(e_2 / e_1) \sum_{j=0}^{n-1} B_k(u_j) F_{ext}(Q(u_j)), \qquad k=0,1,2, \ldots, m. \tag{6.9}$$

The matrix form of the above equations is represented as

$$AV = -F_{ext}(V) \ . \tag{6.10}$$

The unknowns of the motion equations are the coordinates of the control vertices V_i. Assuming the curve is made of a massless material and that it is placed in a viscous medium, the above equations can be solved using Lagrangian dynamic motion (Courant and Hilbert, 1953; Terzopoulos et al., 1988), namely,

$$\gamma (\partial V_k / \partial t) + (\partial E / \partial V_k) = 0, \quad \text{for } k = 0,1,2, \ldots, m, \tag{6.11}$$

where γ is an Euler step size and t represents the time of each deformation. After discretization of Eq. (6.11), we have

$$V_t = (A+\gamma\ I)^{-1}(\gamma\ V_{t-1} - F_{ext}(V_{t-1})) \qquad (6.12)$$

where A is a banded matrix ($m \times m$). This matrix can be solved by LU decomposition in linear time $O(m)$.

As can be seen, the total energy E is a function of the set of control vertices, which uniquely control the shape of the model. Minimizing the energy of the curve results in a final control polygon of the model. Therefore, the final shape of the road centerline model can be determined using such a deformation process.

6.6 Implementation and Evaluation of the Approach

6.6.1 Computational Implementation

Several parameters must be set to apply the motion equation. A desirable trade-off between the internal energy and the external energy has to be made to control the local and global shape of the deformation model. In the current implementation, $e_1=0.3$ ($e_2=1-e_1$) is applied. The smoothness coefficients $\alpha=0.7$ and $\beta=0.5$ are determined empirically to maintain the road shape. $\omega_1=0.5$, $\omega_2=1.0$ and $\omega_3=0.8$ are selected for representing historical effects of the external energy.

The major time-consuming parts of the approach involve (a) feature point extraction and matching among sequential images; and (b) the computation of the inversion of the matrices in Eq. (6.12). Since the constrained processing window for feature extraction and the constrained matching range for image matching are available, the time-consuming computation can be greatly reduced. In order to reach a fast convergence of motion equations, the motion of the entire model is made by the composition of the partial deformation of piecewise segmented models (Tao, 1996a). At first, the internal and external energy for the curve segment S_i to S_{i+3} (the distance of three exposure stations) are computed and such curve segment will be deformed using the information from the images on which this segment is visible. Then the above process is repeated on the next segment S_{i+1} to S_{i+4}. As a result, the model will be deformed incrementally, driven by the successive images, and thus a fast convergent state of the deformation can be reached. In the current implementation, the control vertices are seeded every 2.0-2.5 meters along the model. The dimension (m) of the inversion is 10-20 in Eq. (6.11). Without using B-spline curve representation, the dimension (n) would be 80-160 (assuming that the sampling interval on the curve is 0.3 meters). This method eliminates the huge computations of matrix inversion.

6.6.2 Evaluation of Reliability

Before starting the automatic reconstruction program, it is required that two initial image points are given indicating the direction and approximate location of the road centerline to be extracted. Then, the reconstruction can be executed in a batch mode. The reconstructed segments of road centerlines can be back-projected onto the image and displayed

simultaneously on an image interactive interface, MultiView, and an AutoCAD vector interface, OverView (Appendix A). This implementation permits the users to monitor the processed results in both image and scene domains. It also allows the user to interrupt the processing and correct the processed results on-line or off-line. All the available image data has been used for the test (Appendix B). The longest data set used is a highway image sequence of Quebec city, consisting of 1665 images. In order to evaluate the robustness of the approach, the raw image data sets without any preprocessing were used. It was demonstrated that the approach is able to reliably reconstruct the complete road centerlines from the image sequences without any human assistance. It was found that the final reconstructed centerlines are very smooth, and the approach functions reliably on bridging gaps (Tao, 1996a).

A number of special tests have also been conducted to evaluate the approach with images under different scenarios, such as various marking appearances and patterns, different road types, shapes, and conditions, and so on. Typical examples are given in Figure 6.10 (the image projections of the road centerlines obtained finally are represented by dotted black lines): case 1 (Figure 6.10a), the pedestrian lines crossing the road caused a gap and noisy markings, and the distinctive cracks can be found on the road; case 2 (Figure 6.10b), the shadow caused a contrast on the images and the moving car blocked a part of the centerline markings; case 3 (Figure 6.10c), a big gap is visible since the vehicle was approaching a T-intersection; case 4 (Figure 6.10d), the curved and dashed markings were encountered when the vehicle was making a turn.

6.6.3 Evaluation of Accuracy and Efficiency

The accuracy of the reconstruction result relies on the accuracy of the camera orientation parameters, the quality of the determined vehicle trajectory data, and the performance of the reconstruction approach. The evaluation of the resulting accuracy is conducted by two methods: (a) the reconstructed centerlines in the scene domain are back-projected onto images to visually check the projections matching the original image edges of the corresponding centerlines. If the procedure does not introduce any errors, differences between the projected and original image edges should be caused by the errors in the orientation parameters only. This provides a measure of relative accuracy. No significant difference (within a tolerance of the georeferencing error, namely < 5 pixels) was found on checking the centerline projections in the images; (b) A check of the absolute accuracy is carried out by overlaying the reconstructed centerlines over the existing road maps of the same area. The result of the reconstructed centerlines overlaid over a Laval city digital map (1:1000) is given Figure 6.11. The WGS-84 GPS coordinates of the reconstructed road centerlines have been converted to the map coordinates, UTM coordinate system, before performing the overlay. It is seen that the reconstructed road centerlines are very smooth and fit quite well with the map.

This approach functions in a fully automatic manner, except that the approximate road direction has to be given before starting the program. Under the current implementation and hardware configuration, it takes about 3-5 seconds to complete the processing of an image

pair (equivalently, 6-10 meter survey of road centerlines). As a result, the efficiency is very impressive.

6.7 Chapter Summary

An integrated approach has been proposed to achieve a globally optimized automatic reconstruction of road centerlines, in which the reconstruction has been considered as a *shape from sequences* problem. This new approach leads to a successful solution of the integration of multiple constraints in mobile mapping systems. Based on the tests, the approach has the following characteristics:

- *Closely coupled bottom-up and top-down reconstruction:* The whole reconstruction process is designed based on a 3-D shape model and image sequences. This design leads to a robust result, because the local and global shape of the model is controlled under combinations of the actions of the internal energy (model) and the external energy (image). The method is insensitive to sparse, spurious and noisy data from imagery. Accurate model shape can be obtained through a dynamic refinement using reliable image information.
- *Physically-based mechanism for constraint combination*: The synthesis of multiple constraints has been implemented by the least action principle. The advantages of the method are that constraints from both model assumptions and image sequences can be incorporated into a mathematical model, and the constraint combination can be solved by a Lagrangian motion equation.

- *Model generation using sensor integration*: Accurate vehicle trajectory determined by a multi-sensor (GPS/INS) integration technique was used for generating an approximate 3-D shape model of road centerlines.

- *Model driven feature extraction:* The geometric and photometric model of the image was described as a group of rules. Guided by the rules, algorithms for oriented edge detection, dual-edge detection and edge filtering by a new majority voting algorithm were developed.

- *Stereo-motion matching*: Effective stereo and motion constraints were exploited from stereoscopic and sequential images. The multiple epipolar line constraint, ground plane constraint and spatio-temporal consistency constraint were applied to implement a robust matching procedure.

- *B-Spline based 3-D B-snake representation*: The use of B-Splines for the representation of the snake model has demonstrated its numeric advantages on the computations of the matrix inversion.

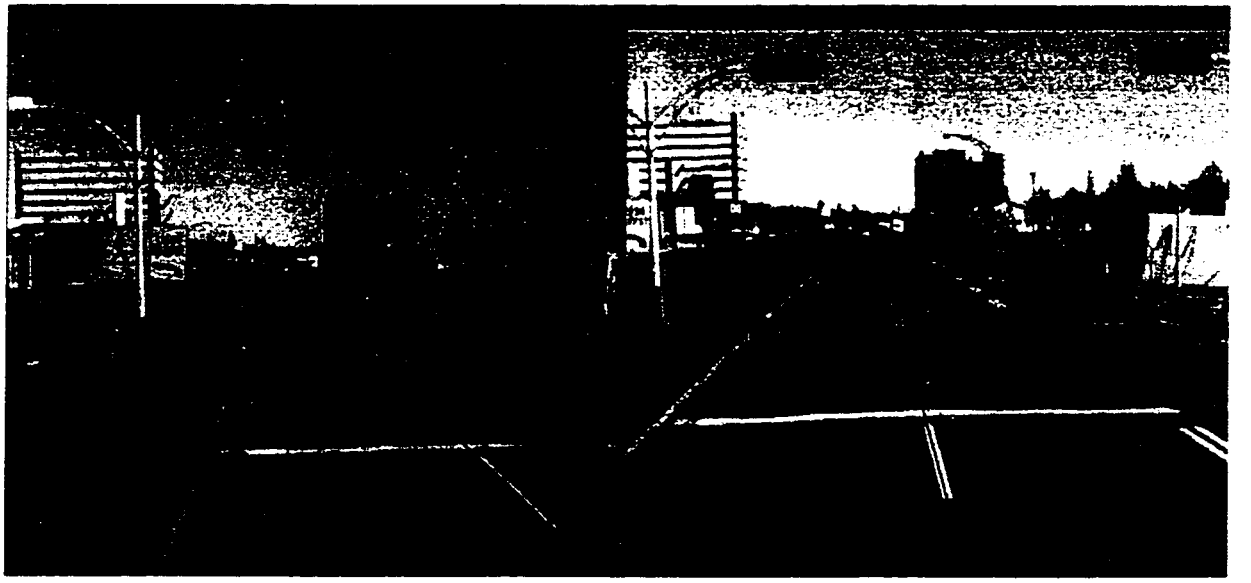


Figure 6.10a Results of road centerline reconstruction

Figure 6.10b Results of road centerline reconstruction

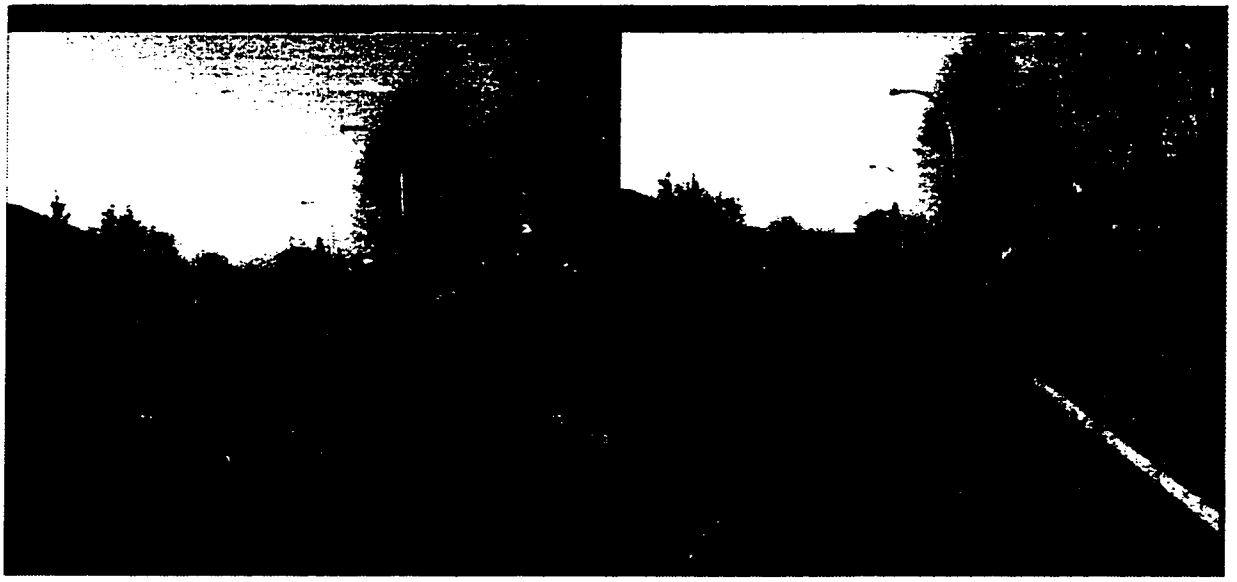


Figure 6.10c Results of road centerline reconstruction

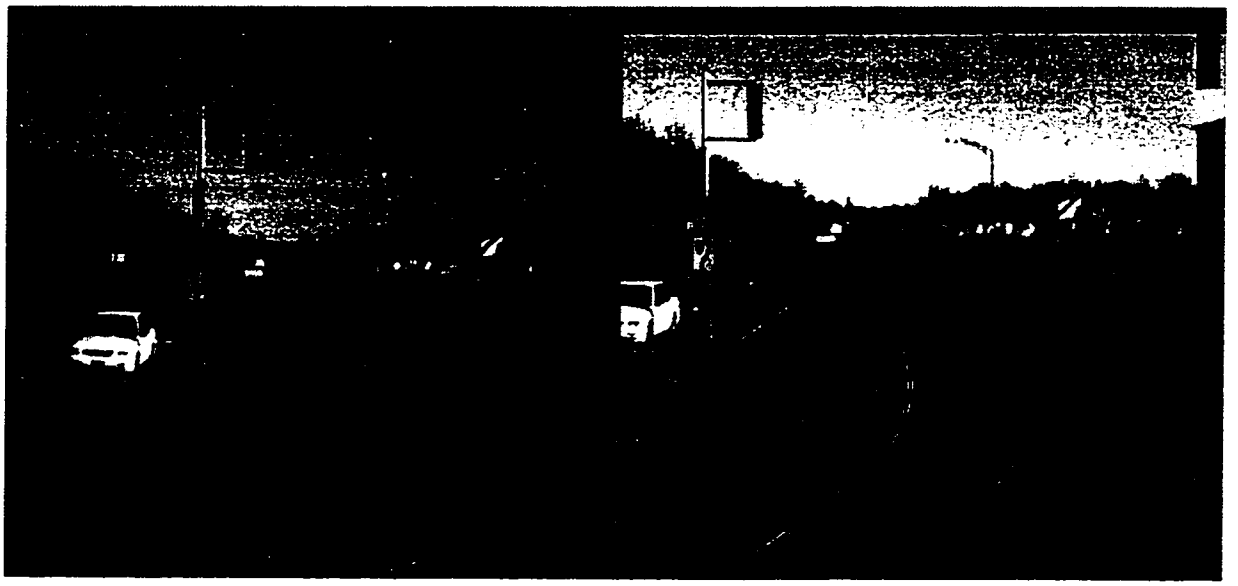


Figure 6.10d Results of road centerline reconstruction

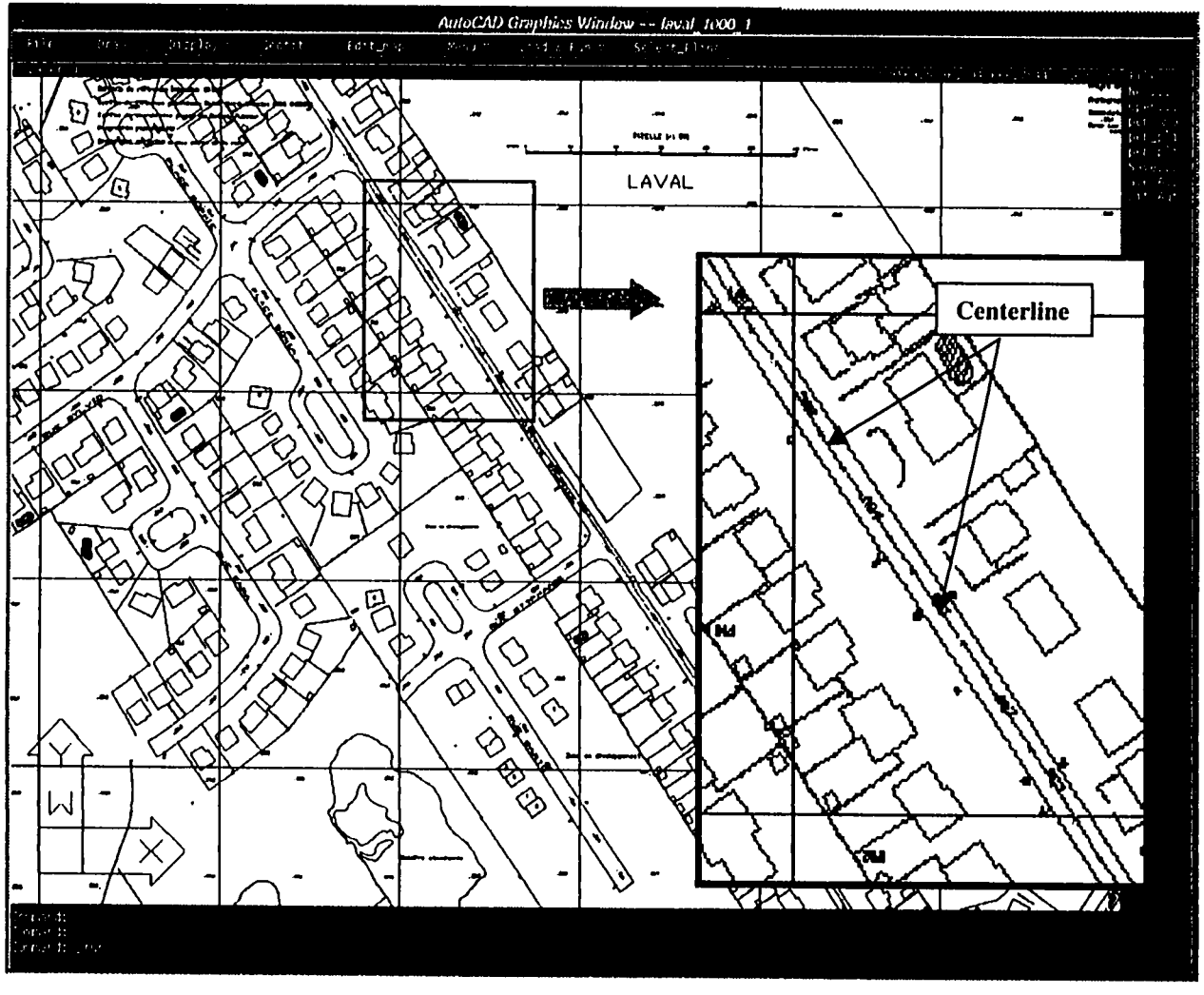


Figure 6.11 Reconstructed road centerlines (red lines) overlaid on the digital map (1:1000).

CHAPTER 7

EXTRACTION OF VERTICAL LINE FEATURES FROM STEREO IMAGE SEQUENCES

A semi-automatic approach to extract vertical line features is proposed and developed in this chapter. The purpose of this work is to assist the user to find and locate the important vertical objects from a large set of image sequences. A cooperative human-computer interaction scheme is developed to implement this technique. After batch processing of automatic feature extraction, a human operator is required to verify the extracted line features. Then, multiple line correspondences will be established by using a multinocular line reconstruction method.

7.1 Introduction

There is a long list of objects to be measured from images collected by mobile mapping systems. The objects of interest can be classified mainly into three categories: utility, transportation and public safety. Among these objects, the vertical type of objects constitutes a large group. These vertical objects include street light poles, power-line poles, traffic signs, telephone booths, and fire hydrants. The common feature of these objects is that they have line features that are vertical with respect to the ground.

The automatic extraction of these objects poses a great challenge. It is understood that two processes, extraction and recognition, have to be incorporated in order to automatically reconstruct the objects. The ability of automatic object recognition is still limited (Forstner, 1993). In this research, the technique to identify and locate objects using line feature extraction is investigated. The focus is placed on extraction rather than recognition. The advantage of this approach is that the time-consuming task of image browsing and object search from a large set of image sequences can be reduced.

It is also recognized that this work forms the basis of object recognition. With the use of color or other sensory information, object recognition technology for specific objects could be successfully developed (Geiselmann and Hahn, 1994; Hock et al., 1996). It is also worth mentioning that this work provides an alternative to perform camera calibration, quality control and vision based navigation for mobile mapping.

The literature regarding extraction of line features is very extensive (Chen and Huang, 1990; Engelbrecht and Wahl, 1988; Faugeras, 1993; Liu et al., 1990; Nevatia and Babu, 1980). It has been difficult to evaluate the various methods due to the differences of the tasks, imaging conditions and system environments. It is believed that a successful method is based on effective cooperation of all individual components. An optimization should be carried out so that the systematic effect will be better than the sum of each individual effect.

7.2 Overview of the Approach

Since the mounting of the stereo cameras in the VISAT system follows the way that the baseline is constrained approximately to be parallel to the ground plane and the camera roll angle is very small. As a result, the projections of vertical 3-D objects are also approximately vertical in images. This constraint greatly simplifies the process of line extraction. The systematic overview of the proposed approach is illustrated in Figure 7.1. It consists of five major modules:

1. Detection of vertical edges
2. Formation of vertical line segments
3. Stereo correspondence of line segments
4. Verification of extracted features
5. Multinocular line reconstruction

The constraints available in the mobile mapping system and the information about the vertical objects are utilized in the design of this approach. The ultimate goal of this approach is to locate the vertical objects from image sequences using the cues of vertical line features. As we shall see, many line segments will be detected after edge detection, line formation and line correspondence. It is expected that the desired lines will appear in the first five records of most likely lines. At this step, the computer shows the user five likely lines, and the user is required to verify the results and confirm the correct lines. Finally, a multinocular line reconstruction algorithm is applied to establish multiple line correspondences and generate 3-D line coordinates.

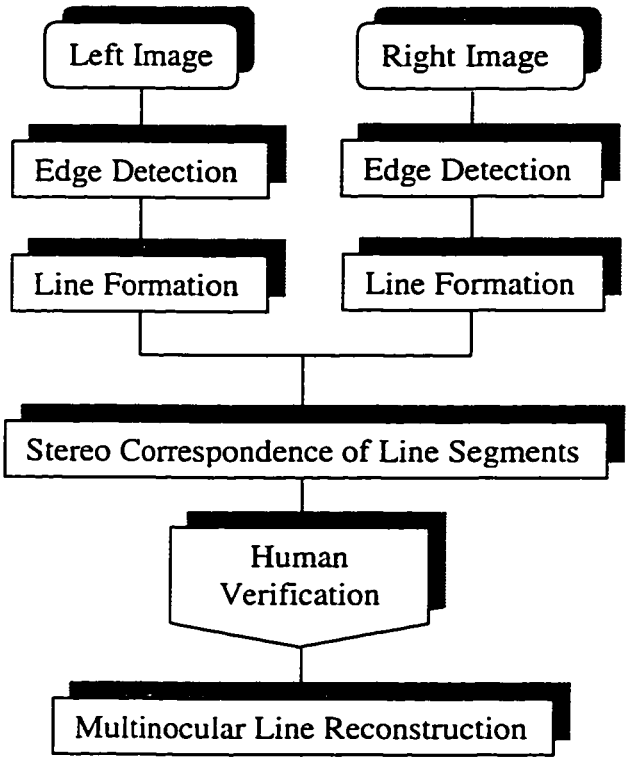


Figure 7.1 Systematic flowchart of extraction of vertical line features

7.3 Detection of Vertical Edges

Edge magnitude and edge direction are two main components of an image edge element (Canny, 1986). We firstly recognize from experiments that information related to edge direction is very useful since the edge magnitude is easily corrupted by noise and physical factors, such as shadows. Therefore, the computed edge direction is used as the main evidence for inference. In the following, a three-step algorithm to determine vertical edges is described:

1. Determination of the existing region of possible edges;
2. Determination of the existing region of thinned edges; and
3. Determination of the existing region of vertically oriented edges.

7.3.1 Determination of the Existing Region of Possible Edges

We start with convolving an input image using the six directional masks. The configuration of the masks is illustrated in Figure 7.2.

$$\begin{vmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{vmatrix} \qquad \begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{vmatrix}$$

0 - 0° 1 - 45° 2 - 135° 3 - 180°

$$\begin{vmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{vmatrix} \qquad \begin{vmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}$$

4 - 225° 5 - 315°

Figure 7.2 Directional masks encoded 0 - 5

After convolving the masks with the original image, the magnitude of the convolved output and the direction of the mask giving the highest output at each pixel are recorded as edge data. A

very small threshold is set in order that all possible edges are detected in the edge existing region. The direction value is encoded as 0 - 5 corresponding to the six directions previously defined. An example of the obtained edge existing region is shown in Figure 7.9(b). The original image pair is given in Figure 7.9(a).

7.3.2 Determination of the Existing Region of Thinned Edges

The simplest technique for thinning is to retain only those edges whose magnitude is a local maximum (Nevatia and Babu, 1980). In the algorithm, the obtained direction values of edges are used to constrain the thinning procedure. Edge thinning is described as: an edge element exists if the following two conditions are satisfied

- the edge directions of the two horizontally neighboring pixels are within one unit (45°) of that of the center pixel; and
- the output edge magnitude at the pixel is larger than the edge magnitudes of its two neighbors in a direction normal to the direction of this edge.

If the above two conditions are met, the two neighboring pixels are disqualified from being candidates for edges. An example of edge thinning is given in Figure 7.9(c).

7.3.3 Determination of the Existing Region of Vertically Oriented Edges

A further refinement is conducted using the accurate values of edge direction. In this step, we use the Sobel gradient operator to compute the direction values of the existing edges. Since a large amount of the undesired edges have been screened out through the previous two steps, the computation of edge gradients is only performed on a small set of edges, and the savings on the computation time are significant. After the computation of gradients, a rigorous tolerance of direction value, 15°, is applied to filter out all unqualified edges. The processing result after this step is illustrated in Figure 7.9(d).

7.4 Formation of Vertical Line Segments

7.4.1 Vertical Line Grouping

Generally speaking, line grouping is a complicated process since no prior knowledge is available to link isolated edges. However, since the lines to be detected are vertical, the grouping algorithm becomes straightforward. The main concern for grouping is how to achieve a trade-off on gap bridging. A strong linking process may produce many false line segments, while a weak linking process may lose many important line segments.

It is observed that a vertical object of interest, such as a pole, has two vertical boundary lines. As shown in Figure 7.3, an ideal edge model of a vertical object, the edges located along the two

boundary lines of the same object should have opposite edge directions and be within a certain distance. This presents a strong constraint for extraction of reliable vertical line features. Based upon the above analysis, the line grouping algorithm is composed of two phases: (1) grouping of primary line segments (to group the distinct line segments), and (2) grouping of secondary line segments (if the detected primary line is located on one of the object boundary lines, the line

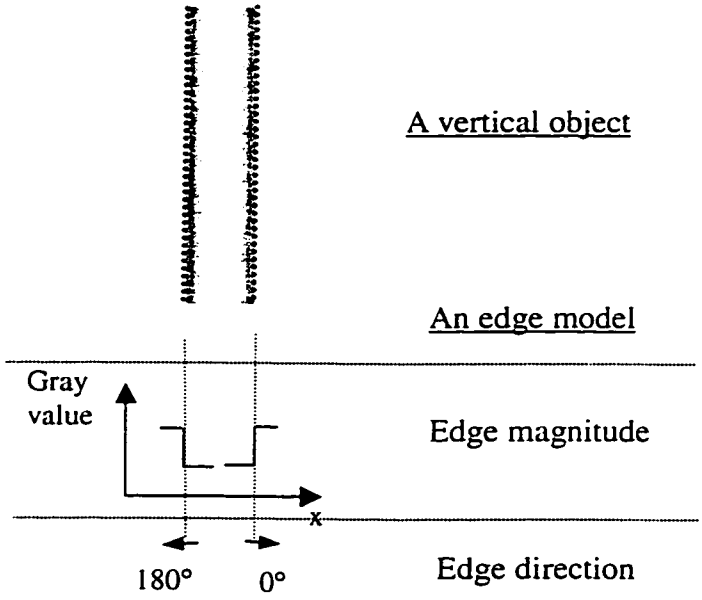


Figure 7.3 An ideal edge model of a vertical object

segments parallel to this primary line are to be grouped).

Phase 1: Grouping of primary line segments

(1.a) Scan the edge image from the left-top to the right-bottom, and find an unprocessed edge;

(1.b) Take this edge point as a starting edge point, and construct a vertical "bridge"

downwards, with a width of 3 pixels (shown in Figure 7.4). The length of the bridge is defined by the parameter "length of bridging gap". In the first phase, only the distinct line segments are to be grouped, so that the length of bridging gap of **15 pixels** is chosen.

(1.c) If an edge is located in this bridge and its direction is the "same" (tolerance is ±15°) as that of the starting edge, this edge is recorded as a "compatible edge", otherwise as an "incompatible edge".

- (case 1) If the distance between a compatible and the starting point is less than half of the bridge length (8 pixels), link this edge to the starting point, take this edge as a new starting point, and go to (1.b).
- (case 2) If no compatible edge exists, but the number of compatible edges in this bridge is larger than the number of incompatible edges, link the closest compatible edge to the starting point, and take this edge as a new starting point. Go to (1.b).
- (case3) If the either of the two cases above is not found, go to step (1.a) and continue scanning.

(1.d) Finally, if the length of the line segment composed of linked edges surpasses a threshold (**50 pixels**), the line segment is considered as a primary line segment and recorded into a line file.

Phase 2: Grouping of secondary line segments

The second phase focuses on the grouping of the line segments parallel to the primary line segments. This work is used to determine the conformance degree of the detected primary line segments. If parallel line segments are found, it is assumed that the primary line segment is

likely located on the boundaries of a vertical object. According to the model of Figure 7.3, an associated line segment should have an opposite direction and be within a certain distance to the primary line segment. The grouping scheme is similar to the first phase except that the length of bridging gap is chosen differently.

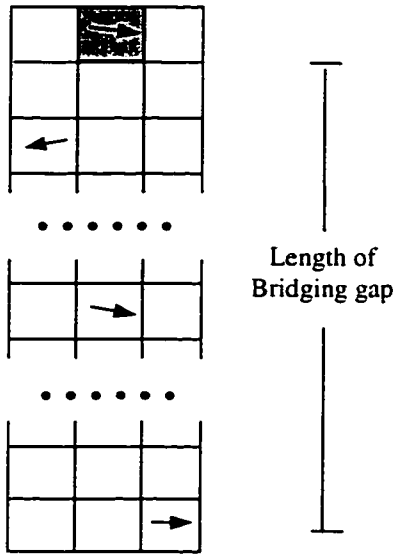


Figure 7.4 Bridge grouping model

(2.a) Scan the surrounding area of a detected primary line segment within a range of ± 10 pixels, and find an edge whose direction is opposite to the direction of the primary line segment. The average direction value of the compatible edges composing the line segment is defined as the line direction.

(2.b) Take this edge as a starting point, and construct a vertical "bridge" downwards, with a width of 3 pixels, and a length of **30 pixels**.

(2.c) Same as step (1.c).

(2.d) If the detected line segment is longer than 1/3 of the length of its associated primary line

segment, the parameter "parallel" in the line file will be set as "True", otherwise "False". The file structure of the primary lines will be described later.

This grouping algorithm represents a trade-off. The first phase is used to extract reliable and distinct line segments from the images, thus the bridging is relatively strict. While the second phase is considered as an evidence collection, so that a large length of bridging gap is applied to collect evidences from noisy data. The results of the line grouping are illustrated in Figure 7.9(e)~(g).

7.4.2 Representation of Line Segments

Table 7.1 Line representation

ID No.	Xs, Ys, Xe, Ye	Length	Number of edges	Direction	Stereo	Parallel
No. 5	.....	78	37	3°	True	True
No. 6	.....	52	18	186°	False	False

The detected line segments have to be represented symbolically in order to facilitate subsequent processing. A line description file is generated after line grouping. The structure of this file is as shown in Table 7.1. All the parameters of line features are computed and recorded. The starting and ending point positions, and the length of the line are two basic feature parameters. The number of edges refers to the number of compatible edges in the line. Line direction is defined as the average direction of all the compatible edges in the line. "Stereo True" means that the corresponding line segment in the right image is found. The establishment of the stereo correspondences will be described in the next section. "Parallel True/False" indicates the state whether or not a secondary line segment of this line is found.

7.5 Feature Correspondence of Line Segments

The method above described is applied on both left and right images (see Figure 7.1). At this step, a matching method is developed to establish stereo correspondence. Since the number of vertical objects of interest in a given image is usually small, the focus of attention of the processing step is placed on the distinctive lines, which are most likely located on the objects of interest. Therefore, ten candidate lines are chosen from the line file of the left image for matching. According to the parametric representation, a sorting algorithm is used to select these lines. The sorting order is "Parallel True" -> "Line length" -> "Number of edges". High priority is given to the line segments whose parallel line segment has been found. Then the line length is considered, the longer the better. Lastly, the number of compatible edges is taken into account.

In the matching method, three constraints are used. The first constraint is the direction compatibility. The corresponding line in the right image exists only if its line direction is the same as the line direction in the left image within a tolerance ±10°.

The second constraint comes from the scene knowledge. First of all, the disparity range constraint described in Chapter 5 is applied. Besides the disparity range constraint, a road corridor condition is also used to eliminate the undesired line features. As shown in Figure 7.5, it is impossible that vertical objects of interest are located right on the road. Therefore, the distance from the object to the vehicle trajectory should be more than 1 m (the camera baseline is 2 m). This constraint is useful to screen out the detected features which are associated with the vehicles on the road.

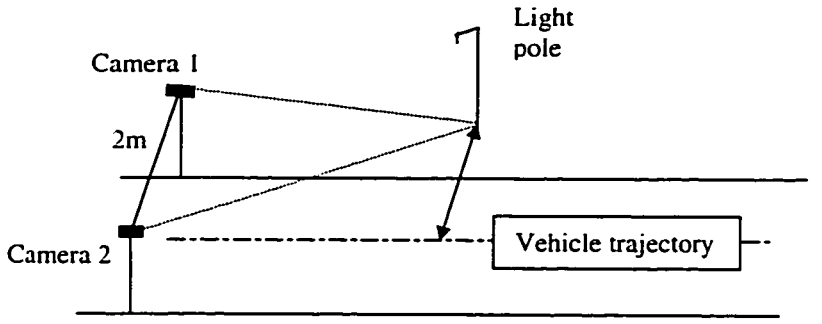


Figure 7.5 Road corridor condition for vertical feature extraction

The third constraint is the similarity of line segments. We use the weighted cross-correlation to perform similarity matching. As illustrated in Figure 7.6 (a part of Figure 7.9(e)), the matching procedure is described as follows:

(a) pick a detected line segment from the left image, and then generate a set of sampling points along the line segment with an interval of 3 pixels;

(b) each sampling point is treated as the center of the master matching window (size of 15H x 11V);

(c) generate the corresponding epipolar line in the right image and determine the intersection of the epipolar line and the vertical detected line segments (or the extrapolation of the line segments) in the right image (see Figure 7.6, the extrapolated line is represented as a dotted line). These vertical line segments must satisfy the previous two conditions, namely, line direction and disparity range constraints;

(d) take the intersection point as the center of the search window (size of 21H x 11V);

(e) if the computed correlation coefficient surpasses the threshold (0.7), record this match point;

(f) once all the sampling points are used, the corresponding line in the right image exists, if the number of match points on that line is more than half of the total number of sampling points used;

(g) finally, update the parameter "Stereo True/False" in the associated line file.

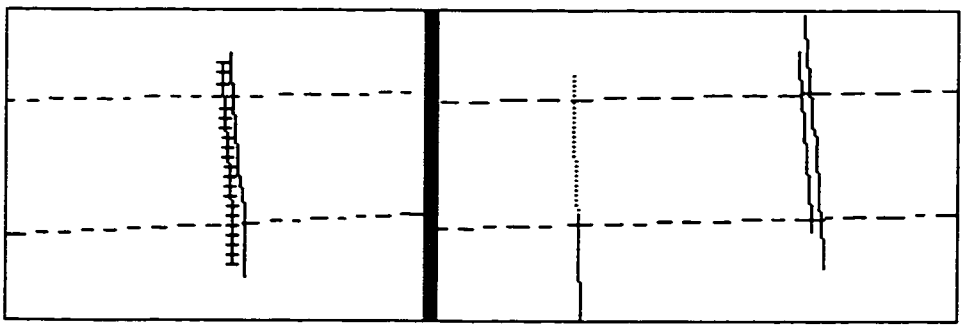


Figure 7.6 Stereo correspondence of line segments

7.6 Verification of Extracted Features

In this step, the user is involved and required to verify the results. When and where the user should be involved is a key to the design of a semi-automatic approach. In order to improve the efficiency of human involvement, the line features are again sorted by a credibility criterion. The sorting is in this order: "Stereo True" -> "Parallel True" -> "Line length" -> "Number of edges". For example, the order of line No. 5 is higher than line No.6 in Table 7.1.

The first five line segments with the higher orders are displayed for verification. Since the detected line segments may be located on the same object, the user should reject duplicated line segments. Basically, three aspects are to be verified by a human operator: (a) whether or not the detected line segment is desired for the application ? (b) whether or not the stereo correspondence result is correct (if available) ? (c) which line segment(s) should be rejected if multiple line segments of the same object are detected ?

7.7 Multinocular Line Reconstruction

Multinocular epipolar geometry is again applied here to improve the reliability of line reconstruction. After human verification, the 3-D coordinates of the sampling points along the line (in the left image) can be calculated using their stereo correspondences. These 3-D points can be back projected in the third image, say the forward left image. The same image matching method described in Section 7.4 is applied to find the correspondence between the forward left image and the current left image. Since the approximate position of the corresponding line in the third image can be predicted, the image matching results are reliable. The search window is set as 25H x 17V pixels, considering the georeferencing errors. The correlation coefficient threshold is set as 0.7.

As illustrated in Figure 7.7 (the match points have been marked by '-'), if the number of match points is more than half of the total points applied, the corresponding line segment exists. Since the match points may not be distributed along a perfect line, a line approximation algorithm using the least median square criterion (Rousseeuw, 1987) is used to filter out the spurious points and determine the best line.

The above method is also applied to the other consecutive images (forward and backward images) to find correspondences. Finally, a straight-line constrained photogrammetric triangulation can be employed to obtain a more reliable and accurate result of 3-D line

reconstruction. This feature triangulation algorithm has been described in a number of publications (Mikhail, 1994; Wilkin, 1996).

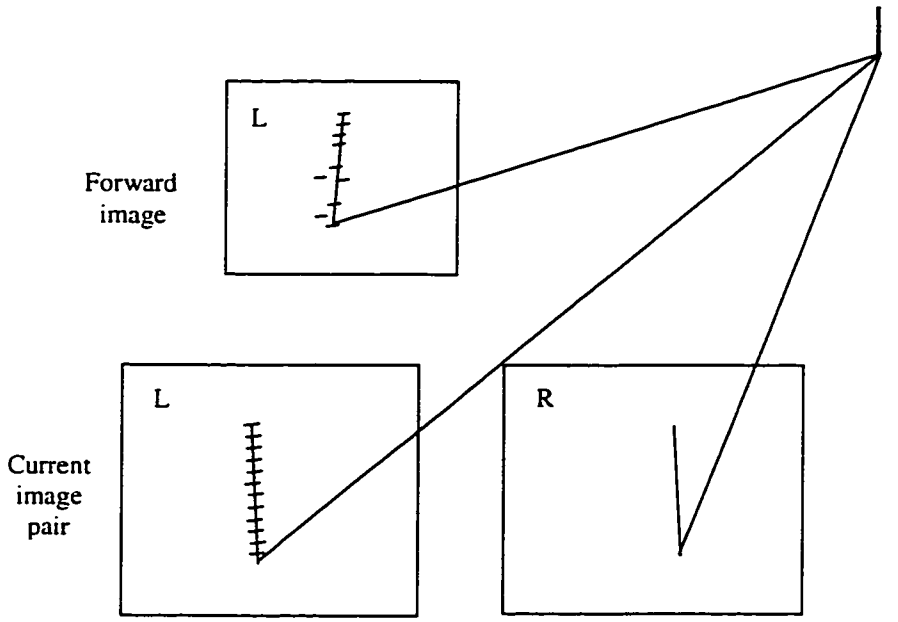


Figure 7.7 Multinocular line matching

7.8 Test Results and Evaluation

7.8.1 Evaluation of Reliability

Figure 7.8(a) shows an image collected by the side-looking camera on the VISAT system. It was used as a test image to evaluate our feature detection algorithm, since there are many vertical features in this image. As we can see, the boundaries of these vertical features are not very clear due to the low resolution of the image. In addition, these features are not

perfectly vertical. The result of the line formation is shown in Figure 7.8(b). The algorithm is able to detect most boundaries of the vertical features. The line segments longer than 20 pixels are displayed.

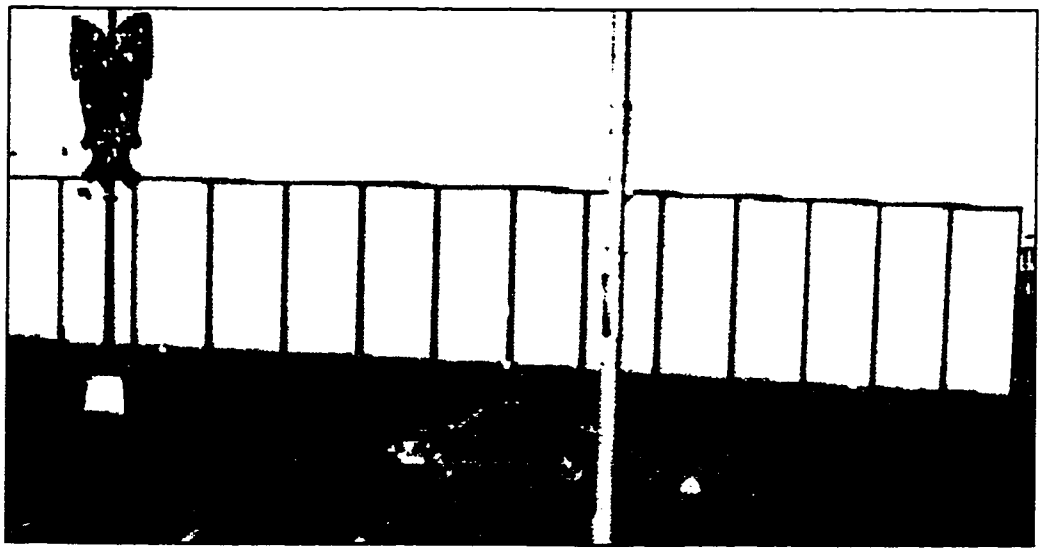


Figure 7.8(a) A test image for vertical feature detection

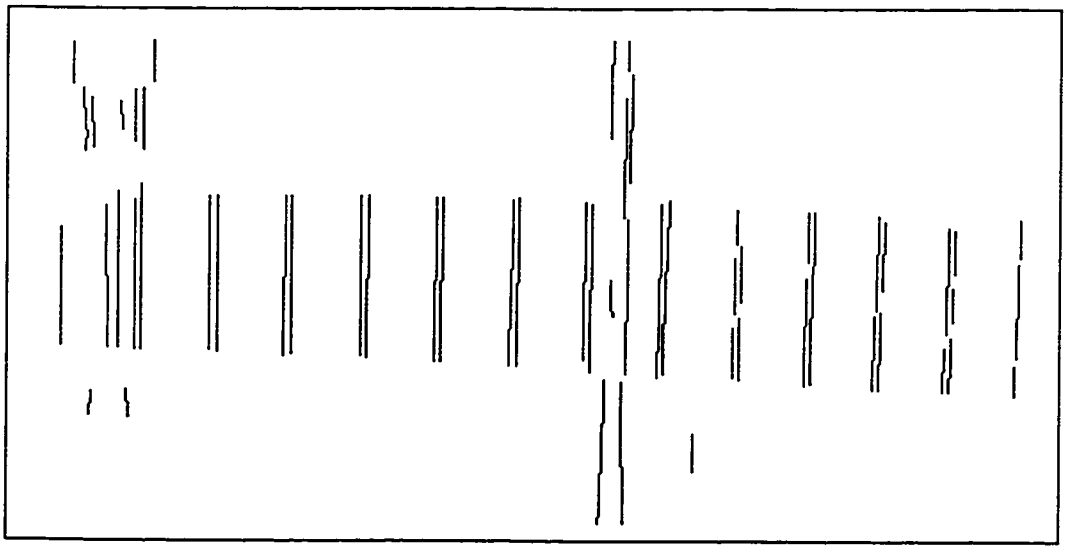


Figure 7.8(b) The result of vertical feature detection

(the length of line segments >20 pixels)

An example of the whole process is given in Figure 7.9 (a) - (g). The Figure 7.9(a) is a stereo image pair. There are four vertical objects of interest (marked on the image). Objects No. 3 and No.4 are far from this view, while object No.2 is too blur to be identified. Object No.1 is definitely important from this view and should be extracted by the developed method. Figure 7.9(b) is the output of edge detection using the six directional masks. All the possible vertical edges have been detected. Figure 7.9(c) shows the effect of edge thinning. It can be seen that a very large number of undesired edges have been filtered out, while the vertical edges have been kept. A refining process is conducted using the Sobel gradient operator. More accurate direction values of existing edges are calculated and only those edges whose directions are vertical with a tolerance of 15° remain. As we can see, the edge points of road boundary in the left image are mostly deleted since these edges are not vertically oriented. The effects of line formation are shown in Figure 7.9(e), (f) and (g). For a comparison, the results of line formation using different thresholds of 20 pixels and 50 pixels, respectively, are both given. Since we are concentrating on the distinct features, only the line segments longer than 50 pixels are recorded in the line file. In this example, ten line segments in the left image and five line segments in the right image are recorded for further processing. After feature correspondence, it is found that four out of the ten line segments in the left image have their corresponding line segments. These four line segments are marked by the small arrows in Figure 7.9(g). These four line segments are located on the same object. In the step of human verification, only one of these four line segments remains. In this example, the left-bottom line segment is used for multinocular

line reconstruction. The corresponding line segments on the forward and backward four images are finally determined in this case.

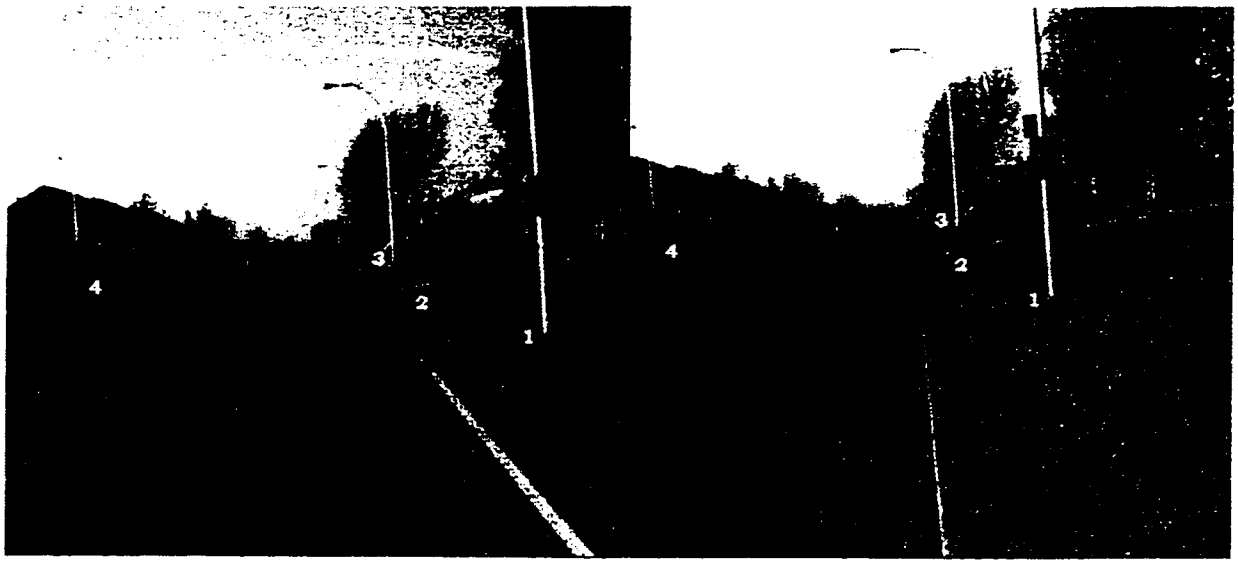


Figure 7.9(a) An original image pair

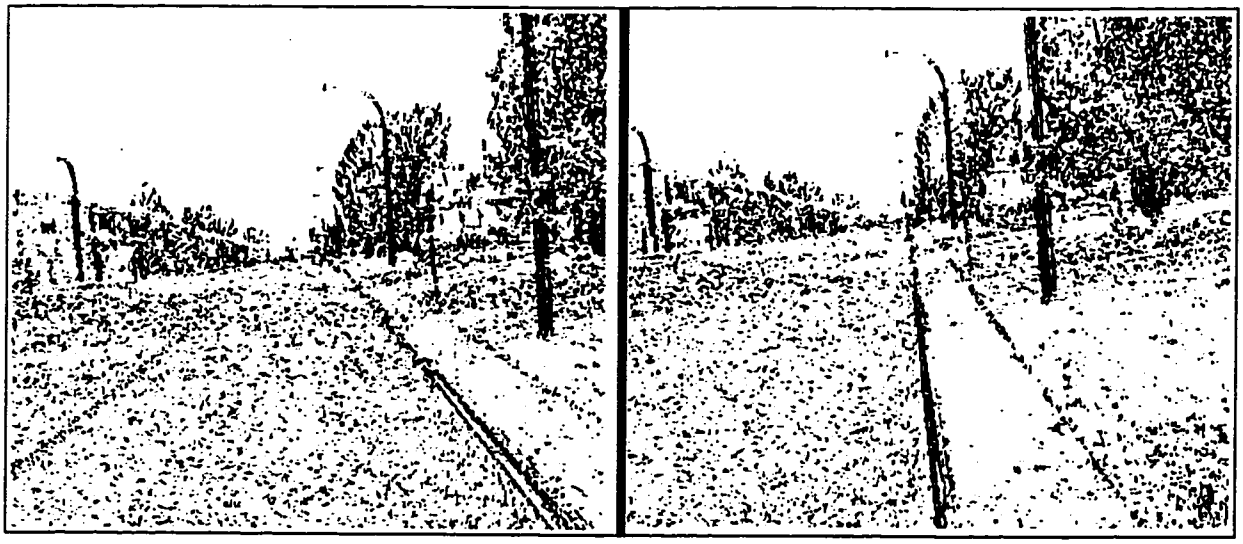


Figure 7.9(b) Edge detection using the directional masks

Figure 7.9(c) Edge thinning

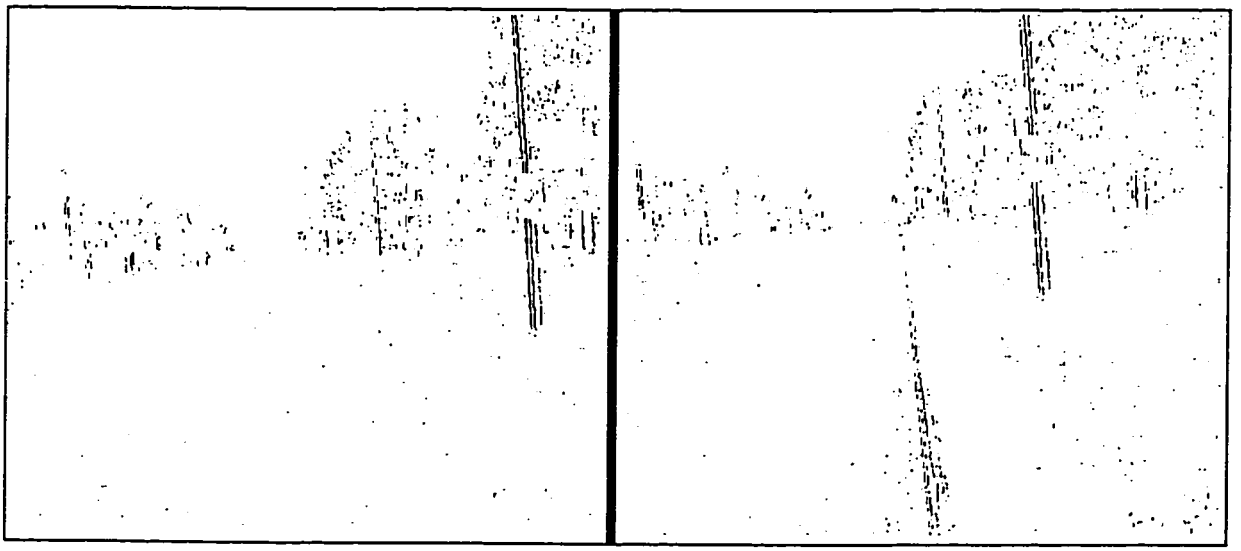


Figure 7.9(d) Detection of vertically oriented edges

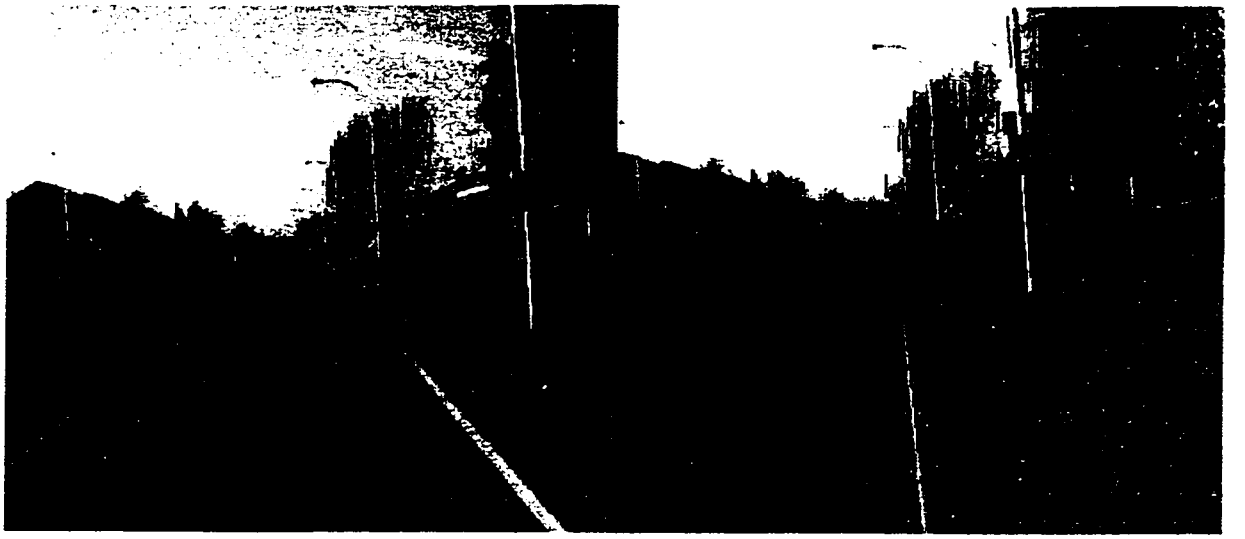


Figure 7.9(e) The detected line segments overlaid on the original images

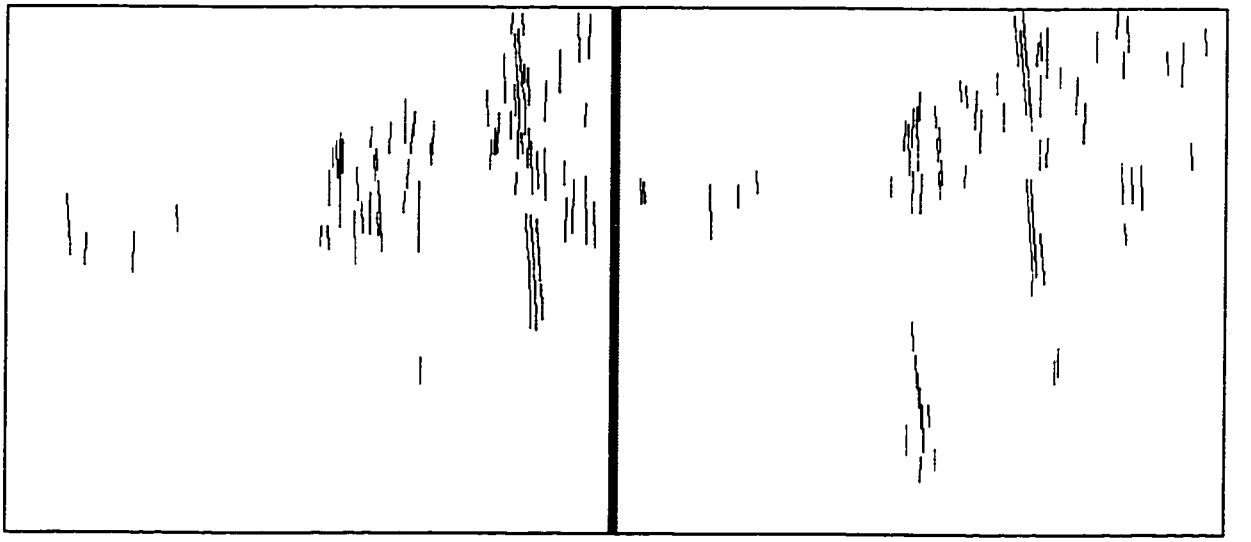


Figure 7.9(f) Line formation (length of the line segments >20 pixels)

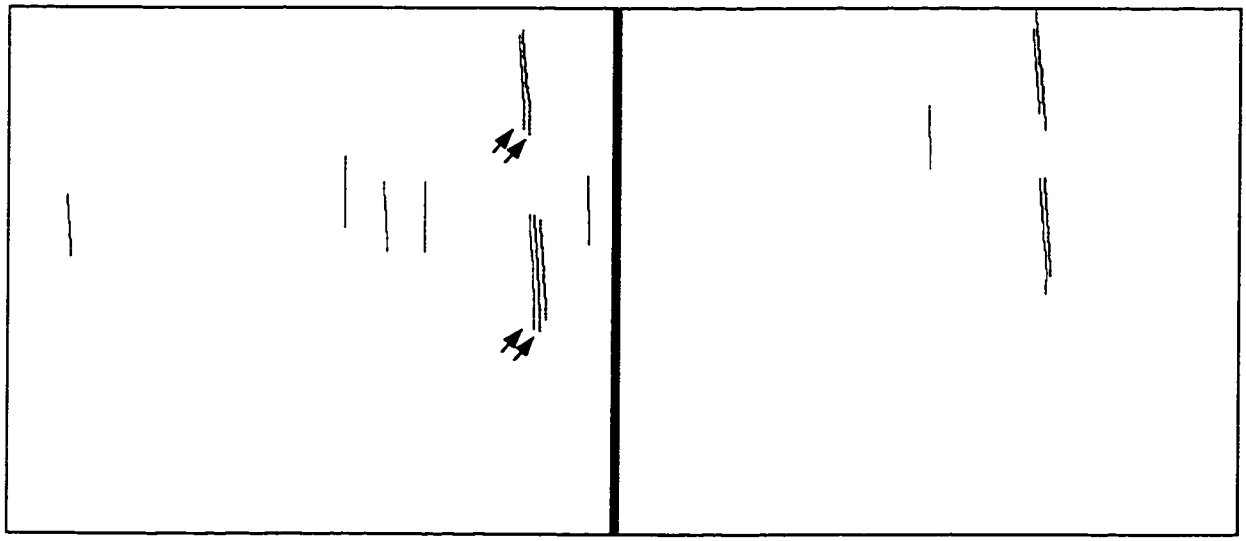


Figure 7.9(g) Line formation (length of the line segments >50 pixels)

The statistical evaluation of the reliability of the method is not easy. The following procedure is designed to evaluate the performance of the approach in a statistical sense. A batch processing step, consisting of edge detection, line formation and line correspondence, was applied to 263 images (left) from the VISAT test set (Appendix B). When the step was completed, five line segments with high ranking were recorded. It was expected that the desired features would appear in the records of the five line segments. If so, a human operator's job was to determine the desired features from the records, thus, the time-consuming part of image browsing and object search can be reduced. The desired feature is defined as that the feature should be extracted in the current image, or say, from the current view. After a careful manual check, the statistical result was obtained. The total number of the desired vertical features was 56, mainly including light poles and traffic signs. The number of the desired features appearing in the records was 48. It meant that the probability

of finding the desired features from the records was 85.7%. The number of the desired features appearing in the records of first three likely line segments is 43, representing a probability of 76.8%. This test demonstrates that the developed method has the capability of finding and locating the vertical line features from image sequences at a probability of 76-86%.

Another example is given in Figure 7.10. In this example, the desired feature (a traffic sign) is detected and appears in the records of the first five likely lines, but not in the records of the first three likely lines. There are a number of reasons that the desired features may not be able to enter into the records: (a) a lot of undesired vertical features exist, such as trees or boundaries of buildings. For instance (see Figure 7.10a), the boundaries of the building have perfect vertical line features; (b) the stereo correspondence may not be established so that the detected line segments cannot get into the records; (c) the image quality is also an important factor influencing the results.

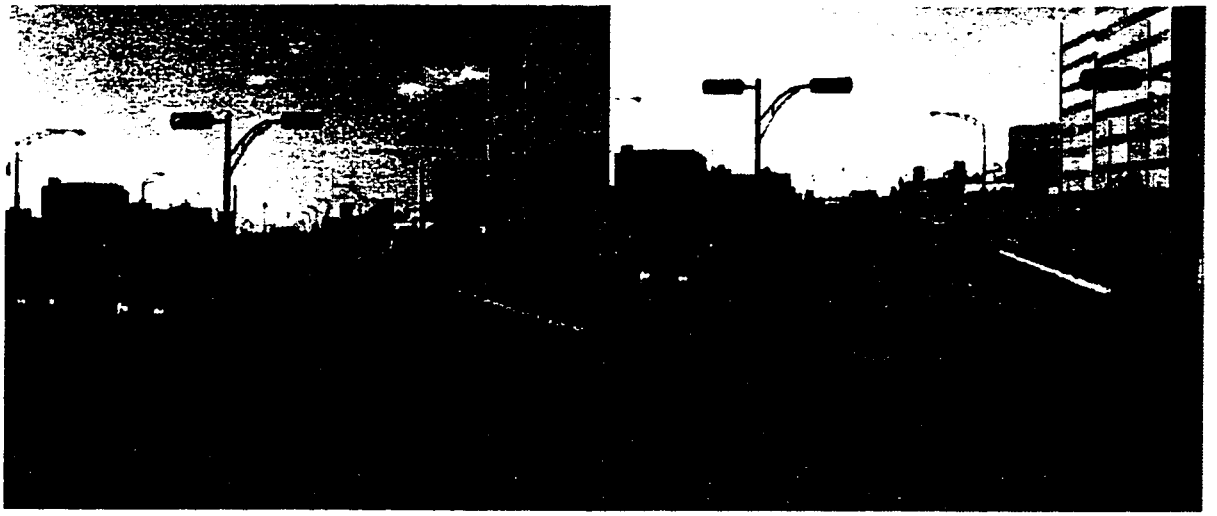


Figure 7.10(a) Detected line segments overlaid on the images

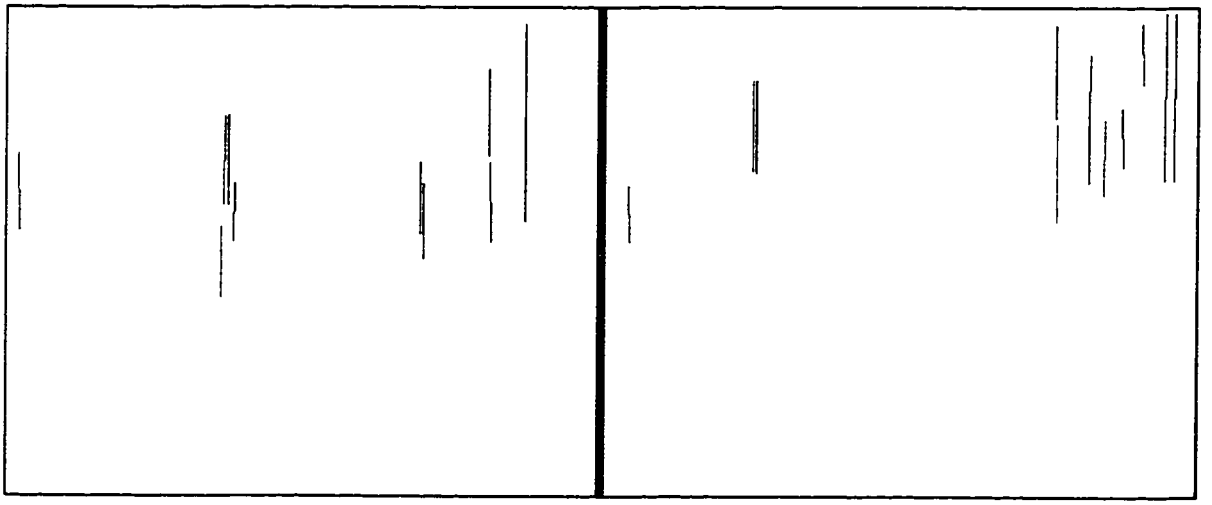


Figure 7.10(b) Detected line segments (length >50 pixels)

7.8.2 Evaluation of Accuracy and Efficiency

The algorithm of the straight-line constrained photogrammetric triangulation has not been implemented, since it is not a focus of this research. However, it was proven that multiple-line based triangulation will present a more reliable and accurate result. This is one of the reasons that the "feature-based photogrammetry" has been gained a considerable attention in the photogrammetric community. The related references can be found in Mikhail (1994) and Barakat et al. (1996).

It is observed that a vertical feature can appear in 3 to even 12 successive images. During the manual extraction procedure, a human operator has to browse the image sequences and located the features image by image. In addition, without having an overview of the images, a human operator does not know which image (view) should be selected for triangulation. It is our experience that sometimes one has to measure the feature again where an image presenting a better quality and/or geometry is found. Therefore, the developed approach is useful to assist the operator to find the likely features from a large set of images. In this approach, the operator acts as a decision-maker and the job is to determine the desired features from the candidate lines prepared by the computer.

7.9 Chapter Summary

In this chapter, an approach to semi-automatic extraction of vertical line features is described. In cooperation with human involvement, the approach is able to find and locate vertical objects of interest using the cues of vertical line features. Moreover, 3-D reconstruction can be carried out automatically by using the multinocular line reconstruction method. It is recognized that human verification is necessary when the knowledge or constraints are not adequate to solve the problem automatically. When and where a human operator should be involved is a key to the success of a semi-automatic approach. It is also clear that the top-down scheme would be appropriate to the design of an automatic approach if the knowledge of the system is available, for example, if a database of the vertical objects, say traffic signs, is available. The developed approach can be extended to detect new objects and update old objects guided by the known object positions along the road corridor.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

The principal objective of the research endeavors to improve the capability and potential for effective and efficient information extraction from mobile mapping image sequences. Based on the understanding of the mobile mapping imaging geometry, the investigation of the state-of-the-art digital photogrammetry and computational vision technologies, and the development of the valid vision constraints from mobile mapping systems, three automated approaches to object measurement, road centerline reconstruction and vertical line feature extraction are proposed and developed. The following conclusions can be drawn from the thesis research.

Methodology

The combination of computational vision and digital photogrammetry technologies is the principal methodology used throughout the thesis research. It is a key to the development and implementation of the three automated approaches. The thorough understanding of the imaging geometry and the extensive development of valid constraints constitute very important parts of the thesis work. A number of state-of-the-art computer vision methods,

namely, model-driven feature extraction, multiple-image matching and shape from image sequences, are developed and extended in this thesis.

Another methodology extensively applied is the human-computer collaborative semi-automatic scheme. An interesting aspect of the history of mankind is the relationship between human and computer. Although the great achievements have been made in recent 10 years, the realization of fully automatic feature extraction is still restricted especially in photogrammetry. Since we acknowledge that it is a long time before the automated feature extraction algorithm can be available as a production tool, and there is no much probability to separate the operator from the system, why should we not take the operator as a component of the system at the beginning of the system design? (Tao and Lin, 1994). A combination of the computational power of the computer and the interpretative ability and knowledge of the human brain has been implemented in the approaches of object measurement and extraction of vertical line features. It has been demonstrated that the use of this methodology makes the system work more reliably and efficiently.

Imaging Geometry

The positioning accuracy of object point coordinates derived from a stereo imaging system is determined by four major components: the distance between camera exposure station and object point Y; the baseline length B; the focal length f; and the mean squared error of image parallax measurements m_p. The along-track error component m_Y is the predominant

error contributing to the total positioning error, while the z-directional (height) error component is relatively small and can be negligible.

The development of high-accuracy image measurement and multiple-point reconstruction methods is particularly important, due to the limited baseline setting in mobile mapping systems. A method to determine the optimal set of imaging parameters has been developed to achieve best tradeoff between these parameters in terms of object positioning accuracy.

Vision Constraints

The only way to resolve an ill-posed vision problem is to exploit any possible constraint sources from the imaging systems. Three effective constraints are developed, (1) stereoscopic and sequential imaging geometry constraint; (2) image georeferencing constraint; and (3) known vehicle ego-motion constraint. The first two constraints are applied in all three proposed approaches, and the third one is employed in the continuous reconstruction of the road centerlines from long image sequences. As demonstrated in the real image tests, these constraints play a key role in enhancing the reliability of the automated approaches.

A direct image-to-space transformation under the ground plane constraint is derived. It forms the basis to the realization of 3-D measurement functionality of the single camera based video-logging systems. This constraint has also been employed in the approaches of

object measurement and road centerline reconstruction to constrain the search space of the algorithms.

Feature Extraction

Feature extraction techniques have been widely developed and used throughout the three approaches. A digital snapping technique is proposed and investigated in the approach of object measurement. It is demonstrated that digital snapping can not only improve the positional precision of point measurement, but also accelerate the measurement procedure.

The model-driven feature extraction methods have been proposed in the extraction of road centerline features and vertical line features. The geometric and photometric models of the specific features are established firstly (see Sections 6.4 and 7.4). Then, the algorithms tailored to these models are designed. For example, in the extraction of centerline features, the algorithms of oriented edge detection, dual-edge detection and edge filtering by a new majority voting techniques are developed. In the extraction of vertical line features, the algorithms of initial edge detection, edge thinning, vertically oriented edge detection, primary line grouping and secondary line grouping are implemented.

Multiple-image Matching

One of the major contributions in this thesis is the use of multiple-image matching for object reconstruction from image sequences. The multinocular epipolar constraint has been successfully used to design an effective strategy of multiple-image matching. The matching

ambiguities can be greatly reduced by using multiple-image matching associated with the weighted cross-correlation, the correlation coefficient analysis, the disparity range constraint, and the two-way matching double-check techniques. The method and its similar versions have been used in all three proposed approaches. It is seen that the method can be applied to not only point-feature correspondence, but also line-feature correspondence. More importantly, the accuracy and reliability of the 3-D reconstruction results can also be greatly improved using the multiple corresponding points or lines derived by the multiple-image matching method.

Shape From Image Sequences

Another contribution made in this thesis is the development of the concept of shape from image sequences. This is, essentially, a closely coupled bottom-up and top-down reconstruction strategy. The whole reconstruction process is designed based on an interaction between a 3-D deformable model and updating information derived from the stereoscopic image sequences. This design leads to a robust result because the local and global shape of the model is controlled under the combinations of the actions of the internal energy (model) and the external energy (image).

The synthesis of multiple constraints has been implemented by the least action principle, a physically-based mechanism for constraint combination. The advantages of the method are that constraints both from model assumptions and image sequences can be incorporated into a mathematical model, and the constraint combination can be solved by a Lagrangian

motion equation. This approach represents a successful accommodation of multiple constraints from a mobile mapping system.

Test Results and Implementation

Various tests using real images have been conducted to evaluate the three approaches in terms of reliability, accuracy and efficiency. The evaluation results can be found in Sections 5.5, 6.6 and 7.8, respectively. According to the test results, the reliability of multiple-image matching for object measurement can achieve a success rate of more than 90%; the approach to extracting vertical line features is able to detect the desired features at a probability of about 85%. The reliability of the approach to the reconstruction of road centerlines is very impressive. The road centerlines can be automatically and continuously reconstructed from the whole image sequences without any human intervention, based on the test data sets. The approach is also insensitive to sparse, spurious and noisy data from imagery.

8.2 Recommendations

Image Quality

Eight CCD video cameras have been used in the VISAT system. It has been found that the image quality from these cameras is not stable. The low resolution affects the measurement

accuracy of objects. It is highly recommended that an investigation into the use of digital cameras and an intelligent dynamic range control for imaging needs be conducted.

Image Capture Rate

The stereo-motion cooperation based method has demonstrated great advantages on the correspondence issue. If a high image capture rate is available, an improvement to a more robust matching can be achieved by using a fusion of feature correspondence and optical flow methods.

Image Compression

The fast image compression and communication techniques should be developed, due to the huge volume of image data generated. Data link from the mobile vehicle to an office GIS and data dissemination through internet/intranet should also be considered.

Multi-Sensor Information Fusion

Without a good input, we may end up with a "garbage in, garbage out" situation. It is believed that the automatic sensing ability will be greatly enhanced by using multi-sensor information, such as color, laser, and sonar. For example, if a laser ranging device is used and angled downward/forward the road, the elevation data of the road surface can be collected. This information can be further incorporated to improve the road centerline

reconstruction method. However, some issues have to be addressed, such as multiple information fusion, registration, management and computational complexity.

Knowledge-Based Feature Extraction and Object Recognition

In many applications, mobile mapping systems are used to update and extend an existing database. Road and/or object knowledge can be derived from databases and used to guide the implementation of feature extraction and even object recognition. For example, If initial knowledge of locations of traffic signs is available, the approach in Chapter 7 can be used to verify the targets and detect new targets. Object recognition is highly complicated, and cannot be performed without the use of knowledge of the road scene.

Vision-Based Navigation, Calibration and Quality Control

Vision-based vehicle navigation is not new and has been addressed very extensively since the 1980's. However, in most robotic systems, few or no navigation sensors such as GPS or INS were used. We believe that the integrated method of navigation and imaging sensors can be used to improve the capability of autonomous vehicle navigation. On the other hand, redundantly overlapped sequential image information can also be used to perform quality control of georeferencing data, and furthermore, to bridge gaps where the absence of the navigation data occurs.

REFERENCES

Abbott, L., and N. Ahuja, 1990. Active Surface Reconstruction by Integrating Focus, Vergence, Stereo and Camera Calibration, Proc. of 3th Int. Conf. on Computer Vision, pp. 489-492.

Ackermann, F., 1996. The Status and Accuracy Performance of GPS Photogrammetry, Digital Photogrammetry, An Addendum to the Manual of Photogrammetry, ASPRS, pp. 108-115.

Aggarwal, J. K., and N. Nandhakumar, 1988. On the computation of motion from sequences of images - A review, Proceedings of the IEEE, 76(8), pp. 917-935.

Agouris, P., 1992. Multiple Image Multipoint Matching for Automatic Aerotriangulation, Ph.D. dissertation, The Ohio State University, 124 Pages.

Aloimonos, Y., Weiss, I., and A. Bandopadhay, 1988. Active Vision, Int. J. Computer Vision, 2(1): 333-356.

Ayache, N., and F. Lustman, 1991. Trinocular Stereo Vision for Robotics, IEEE Transactionss on Pattern Analysis and Machine Intelligence, 13(1): 73-85.

Bajcsy, R., 1988. Active Perception, Proceedings of the IEEE, 76(8): 996-1005.

Baker, H., Bolles, R., and J. Woodfill, 1994. Real Stereo and Motion Integration For Navigation, Int. Arch. of Photogrammetry and Remote Sensing Com. III Symposium, Munich, SPIE-2357:17-24.

Ballard, D. H., 1991. Animate Vision, Artificial Intelligence, 48(1): 57-86.

Baltsavias, E. P., 1991. Multiphoto Geometrically Constrained Matching, Ph.D. Dissertation, Institute of Geodesy and Photogrammetry, Switzerland.

Barakat, H., Doucette, P., and E. M. Mikhail, 1994. Photogrammetric Analysis of Image Invariance, Proc. of Photogrammetry and Remote Sensing Com. III Symposium, Munich, Germany, pp. 25-34.

Barnard, S. T., and M. A. Fischler, 1982. Computational Stereo, Computer Surveys, 14(4): 553-572.

Bartels, R., Beatty, J., and B. Barsky, 1987. An Introduction to Splines for Use in Computer Graphics and Geometric Modeling, Morgan Kaufmann.

Beyer, H. A., 1992. Advances in Characteristics and Calibration of Digital Imaging Systems, Int. Arch. of Photogrammetry and Remote Sensing'92, Washington D.C, Com. V, Part B5, pp. 545-555.

Birge, S. T., Highway Dimensions from Photolog, Photogrammetric Engineering and Remote Sensing, 51(10): 1609-1614.

Blostein, S. D., and T. S. Huang, 1987. Error Analysis in Stereo Determination of 3-D Point Positions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(6): 752-765.

Bossler, J., Goad, C., Johnson, P., and K. Novak, 1991. GPS and GIS Map the Nations Highways, GeoInfo Systems Magazine, March issue, pp. 26-37.

Bossler, J., and C. Toth, 1996. Feature Positioning Accuracy in Mobile Mapping: Results Obtained By The GPSVan™, Int. Arch. of Photogrammetry and Remote Sensing'96, Vienna, Part B4, pp.139-143.

Burt, P. J., 1988. Smart Sensing within A Pyramid Vision Machine, Proceedings of the IEEE, 76(8): 1006-1015.

Canny, J. F., 1986. A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(8): 769-798.

Chapman, M. and L. Baker, 1996. High-Speed Video Road Survey in Singapore, GIS Asia Pacific, April, pp. 30-32.

Chen, H. H., and T. S. Huang, 1990. Matching 3-D Line Segments with Applications to Multiple Object Motion Estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(9): 1002-1008.

Chen, P. M., Tseng, S. Y., Hou, Y. C., and B. B. Loah, 1995. An Object-Oriented Geographic Information System Shell, Proc. of 9th Symposium on GIS, Vancouver, pp. 413-421.

Cohen, F. S., Huang, Z., and Z. Yang, 1995. Invariant Matching and Identification of Curves Using B-Splines Curve Representation, IEEE Transactions on Image Processing, 4(1): 1-10.

Courant, R., and D. Hilbert, 1953. Methods of Mathematical Physics, Vol. I, London: Interscience.

Cox, I. J., and S. L. Hingorani, 1996. An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, (18)2: 138-150.

Davis, L., and T. R. Kushner, 1985. Road Boundary Detection for Autonomous Vehicle Navigation, Proc. of Intelligent Robots and Computer Vision, SPIE-579, pp. 363-365.

Dhond, U. R., and J. K. Aggarwal, 1989, Structure from Stereo - A Review, IEEE Transactions on Systems, Man, and Cybernetics, 19(6): 1489-1510.

Dickmanns E. D., and A. Zapp, 1986. A Curvature-Based Scheme for Improving Road Vehicle Guidance by Computer Vision, Proc. of Mobile Robots. Cambridge, MA, SPIE 727:161-168.

Dickinson, S., and L. Davis, 1990. A Flexible Tool for Prototyping AVL Road Following Algorithms, IEEE Transactions on Robotics and Automation, 6(2):232-242.

Dold, J. and H.-G. Mass, 1994. An application of epipolar line intersection in a hybrid close range photogrammetric system, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 30, Part V., pp. 65-70.

Ebner, H., Heipke, C., and M. Holm, 1993. Global Image Matching and Surface Reconstruction in Object Space using Aerial Images, Proc. of Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision, SPIE-1944, pp.44-57.

El-Sheimy, N., 1996a. A Mobile Multi-Sensor System for GIS Applications in Urban Centers, Int. Arch. of Photogrammetry and Remote Sensing Com II, Vol. 35, Part B2, pp. 95-100.

El-Sheimy, N., 1996b. The Development of VISAT- A Mobile Survey System for GIS Application, Ph.D. Dissertation, Department of Geomatics Engineering, The University of Calgary, Calgary, Canada, 179 pages.

El-Sheimy, N., and K. P. Schwarz, 1993. Kinematic Positioning in Three Dimensions Using CCD Technology, Vehicle Navigation and Information System'93 Conf., Ottawa.

Engelbrecht, J. R., and F. M., Wahl, 1988. Polyhedral Object Recognition Using Hough-Space Features, Pattern Recognition, 21(6): 158-168.

Faugeras, O., 1993. Three-Dimensional Computer Vision: A Geometric Viewpoint. The MIT Press, 663 pages.

Forstner, W., and E. Guelch, 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features, Proc. of Inter-commission Conf. on Fast Processing of Photogrammetric Data, Switzerland, pp. 281-305.

Forstner, W., 1993. Feature Extraction in Digital Photogrammetry, Photogrammetric Record, 14 (82): 595-611.

Fraser, C. S., 1992. Photogrammetric Measurement to One Part in A Million, Photogrammetric Engineering and Remote Sensing, 53(3): 305-310.

Fraser, C. S., Shortis, M. R. and G. Ganci, 1995. Multi-Sensor System Self-Calibration, Proc. of SPIE-2598, pp. 2-18.

French, R. L., and K. Sakamoto, 1994. International Competitiveness in IVHS: Europe, Japan, and The United States, Proc. of Vehicle Navigation and Information System, 1994, pp.525-560.

Fua, P., 1991. Combining Stereo and Monocular Information to Compute Dense Depth Maps That Preserve Depth Discontinuities, Proc. of the 21th Int. Joint Conf. on Artificial Intelligence, pp. 1292-1298.

Gaston, P. C., and T. Lozano-Perez, 1984. Tactile Recognition and Localization Using Object Model: The Case of Polyhedra on A Plane, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(2): 257-265.

Geiselmann, C. and M. Hahn, 1994. Identification and Location of Simple Objects for Real-Time Mapping, Proc. of Photogrammetry and Remote Sensing Com. II Symposium, Ottawa, pp. 459-469.

Goad, C. C., 1991. The Ohio State University Highway Mapping System: The Positioning Component, Proc. of the Institute of Navigation Conference, Williamsburg, VA, pp. 117-120.

Gruen, A., 1994. Digital Close-Range Photogrammetry: Progress Through Automation, Proc. of Close Range Techniques and Machine Vision, Com. V Symposium, Melbourne Australia.

Gruen, A., 1996. Digital Photogrammetric Systems Revisited, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 31, Part 2, pp. 66-76, Austria.

Gruen, A., and A. Streilein, 1994. Videogrammetry, CAAD and Semi-Automated Extraction of Architectural Features, Proc. of the First Int. Colloquium "Photogrammetry for Restoration and History", Italy, Nov. 10-12.

Gruen, A., and H. Li, 1996. Linear Feature Extraction with LSB-Snakes from Multiple Images, Proc. of Photogrammetry and Remote Sensing Com. III, Part B3, pp. 1-7.

Gruen, A., and H. Li, 1995. Road Extraction from Aerial and Satellite Images by Dynamic Programming, ISPRS Journal of Photogrammetry and Remote Sensing, 50(4): 11-20.

Gulch, E., 1996. Deformable Models As A Photogrammetric Measurement Tool – Potential and Problems, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 31, Part B3, pp. 279-284.

Gustafson P. C., and H. B. Handley, 1992. A Video-Based Industrial Measurement System, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 29, Part B5, pp. 501-506.

Hannah, K. J., and N. E. Okamoto, 1993. Combining Stereo and Motion Analysis for Direct Estimation of Scene Structure, Proc. of 4th Int. Conf. on Computer Vision. pp. 357-365.

He, G., 1996. Design of a Mobile Mapping System for GIS Data Collection, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 31, Part B2, Vienna, Austria, pp. 154-159.

He, G., and K. Novak, 1992. Automatic Analysis of Highway Features from Digital Stereo Images, Int. Arch. of Photogrammetry and Remote Sensing Com. III, part B3, pp.119-124.

He, G, and K. Novak, 1993. On Line Data Capture by Image Matching on the Mobile Mapping System, Videometrics, SPIE-1820, pp.50-56.

Hock, C., Behringer, R., and F. Thomanek, 1996. Intelligent Navigation for a Seeing Road Vehicle using Landmark Recognition, Proc. of Photogrammetry and Remote Sensing, Com. V, Part 5, pp. 183-190.

Hock, C., Caspray, W., Heister, H., Klemm, J., and H. Sternberg, 1995. Architecture and Design of the Kinematic Survey System KiSS, Proc. of the 3rd Int. Workshop on High Precision Navigation, Stuttgart, Germany, April, pp. 569-576.

Huang, T. S., and A. N. Netravali, 1994. Motion and Structure from Feature Correspondences: A Review, Proceedings of the IEEE, 82(2): 252-267.

Ito, M., and A. Ishii, 1986. Three View Stereo Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(4): 524-532.

Jolion, J. M., 1994. Computer Vision Methodologies, CVGIP: Image Understanding, 59(1): 53-71.

Jones, G. A., 1997, Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features, CVGIP: Image Understanding, 65(1): 57-78.

Kanade, T., Okutomi, M., and T. Nakahara, 1992. A Multiple-Baseline Stereo Method, Proc. of DARPA Image Understanding Workshop, San Diego, CA, pp. 409-426.

Kasprzak, W., 1995. Ground Plane Object Tracking Under Egomotion, Proc. of Photogrammetry and Remote Sensing Inter-commission "From Pixels to Sequences", Zurich, Switzerland, March, pp. 208-213.

Kass, M., Witkin A., and D. Terzopoulos, 1988. Snakes: Active Contour Models, Int. J. of Computer Vision, pp. 321-331.

Krzystek, P., Heuchek, T., Hirt, U. and F. Petran, 1995. A New Concept for Automatic Digital Aerial Triangulation, Proc. of Photogrammetric Week, Germany, pp. 215-223.

Kucera, H. A., and R. B. Graves, 1995. Open Systems Mean Open Understanding, Proc. of 9th Symposium on GIS, Vancouver, Canada, pp. 918-927.

Lapucha, D., 1990. Precise GPS/INS Positioning for Highway Inventory System, M.Sc Thesis, Department of Geomatics Engineering, The University of Calgary, Calgary, Canada.

Lemmens, M., 1988. A Survey on Stereo Matching Techniques, Int. Arch. of 16th Photogrammetry and Remote Sensing, Vol. 27, B8, pp. 11-23.

Li, R., 1997. Mobile Mapping – An Emerging Technology for Spatial Data Acquisition, Photogrammetric Engineering and Remote Sensing, 63(9): 1085-1092.

Li, R., M. A. Chapman, L. Qian and Y. Xin, 1994a. Prototype Development of GeoStation System and The Pilot Project, Project Report for Geofit Inc., Department of Geomatics Engineering, The University of Calgary, Calgary, Canada, 49 pages.

Li, R., K. P. Schwarz, M. A. Chapman, and M. Gravel, 1994b. Integrated GPS and Related Technologies for Rapid Data Acquisition, GIS World, 7(4): 41-43.

Li, R., M. A. Chapman, and W. Zou, 1996. Optimal Acquisition of 3-D Object Coordinates From Stereoscopic Image Sequences, Int. Arch. of Photogrammetry and Remote Sensing, Vienna, Austria, Vol. 31, Part 2, pp. 449-453.

Li, R., 1996. Design and Implementation of a Photogrammetric Geo-Calculator in a Windows Environment, Photogrammetric Engineering and Remote Sensing, 62(1): 85-88.

Liu, Y. C., Huang, T. S. and O. D. Faugeras, 1990. Determination of Camera Location from 2-D to 3-D Line and Point Correspondences, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1): 28-37.

Marr, D., 1982. Vision, Freeman, New York.

Mason, S. O., 1995. Conceptual Model of the Convergent Multistation Network Configuration Task, Photogrammetric Record, 15(86): 277-299.

Mass, H. G., 1991. Digital Photogrammetry for Determination of Tracer Particle Coordinates in Turbulent Flow Research, Photogrammetric Engineering and Remote Sensing, 57(12): 1593-1597.

Mass, H. G., 1992. Complexity Analysis for the Establishment of Image Correspondences of Dense Spatial Target Fields, Int. Arc. of Photogrammetry and Remote Sensing, Vol. 29, B5, pp. 102-107.

Mass, H. G., 1996. Automatic DEM Generation by Multi-Image Feature Based Matching, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 31, B3, pp. 484-489.

Mekuria, M., 1995. Application of GIS/GPS to Solve ISTEA Management In A State Highway Department, Proc. of GIS/LIS'95, pp. 730-735.

Menet, S., P. Saint-Marc, and G. Medioni, 1990. Active contour models: overview, implementation and applications, Int. Conf. Syst. Man Cybernet. Los Angeles, pp. 194-199.

Mikhail, E. M., 1994. Linear Features for Photogrammetric Restitution and Object Completion, SPIE-1944, Orlando, FL, April, pp. 16-30.

Milne, A., Milton, S., and J. L. Smith, 1993. Geographical Object-Oriented Database – A Case Study, Int. Journal on GIS, 7(1): 39-55.

Moravec, H. P., 1979. Visual Mapping by a Robot Rover, Proc. of 6th IJCAI, pp. 598-600.

Mori, K., Kidodi, M., and Asada, H., 1973. An Iterative Prediction and Correction Method for Automatic Stereo Comparison, Computer Graphics and Image Processing, Vol. 2, pp. 393-401.

Mundy, J. L. and A. J. Heller, 1990. The Evolution and Testing of A Model-Based Object Recognition System, Proc. of Int. Conference on Computer Vision'90. Osaka, Japan, pp. 268-282.

Navab, N., and Z. Zhang, 1994. Fusion of Visual Data Through Dynamic Stereo-Motion Cooperation. Proc. of Photogrammetry and Remote Sensing Com III symposium, pp. 932-939.

Negahdaripour, S., and B. Horn, 1987. Direct Passive Navigation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(1): 168-176.

Nevatia, R., 1996. Matching in 2-D and 3-D, Int. Arc. of Photogrammetry and Remote Sensing, Vol. 31, B3, pp. 567-574.

Nevatia, R., and K. R., Babu, 1980. Linear Feature Extraction and Description, Computer Graphics and Image Processing, No. 13, pp. 257-269.

Novak, K., 1995. Mobile mapping technology for GIS data collection, Photogrammetric Engineering and Remote Sensing, 61(5): 493-501.

Okutomi, M., and T. Kanade, 1993. A Multiple Baseline Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(2): 353-362.

Ozawa, S., 1994. Image Sensors in Traffic and Vehicle Control, Proc. of Vehicle Navigation and Information Systems, pp. 27-34.

Pahlavan, K., and J. O. Ekhlund, 1992. A Head-Eye System: Analysis and Design, CVGIP: Image Understanding, 56(1): 41-56.

Pentland, A., and S. Sclaroff, 1991. Closed-form Solution for Physically based Shape Modeling and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(7): 715-729.

Pietikainen, M., and D. Harwood, 1987. Progress in Trinocular Stereo, Proc. NATO Advanced Workshop on Real-Time Object and Environment Measurement and Classification, Italy, Aug. 31-Sept. 3.

Pottle, D., 1995. New Mobile Mapping System Speeds Data Acquisition, Geo-Info Magazine, 9(9): 51-53.

Priese, L., et al, 1994. New Results on Traffic Sign Recognition, Proc. of Int. Symposium on Intelligent Vehicles, Paris, France, October.

Qian, L., 1996. Building 3-D GIS by Object Orientation, M.Sc. Thesis, The Univ. of Calgary, Calgary, Canada, 84 pages.

Roberts, K. S., and S. Ganapathy, 1987. Stereo Triangulation Techniques, IEEE Workshop on Computer Vision, Florida, pp. 336-338.

Rousseeuw, L., 1987. Robust Regression and Outlier Detection, John Wiley & Sons, New York.

Sandini, G., and M. Tistarelli, 1990. Active Tracking Strategy for Monocular Depth Inference Over Multiple Frames, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1): 13-27.

Schneiderman, H., and M. Nashman, 1994. A Discriminating Feature Tracker for Vision-Based Autonomous Driving, IEEE Transactions on Robotics and Automation, 10(6): 769-775.

Schwarz, K. P., Lapucha, D., Cannon, M. E., and H. Martell, 1990. The Use of GPS/INS In A Highway Inventory System, Proc. of FIG XIX Congress, Helsinki, Finland, May 10-19, Vol. 5, pp. 238-249.

Schwarz, K. P., Chapman, M. A., Cannon, M. E., and P., Gong, 1993a. An Integrated INS/GPS Approach to the Georeferencing of Remotely Sensed Data, Photogrammetric Engineering and Remote Sensing, 59(11): 1667-1674.

Schwarz, K. P., H. Martell, N. El-Sheimy, R. Li, M. Chapman, and D. Cosandier, 1993b. VISAT - A Mobile Highway Survey System of High Accuracy, Vehicle Navigation and Information System Conference'93, Ottawa, Canada, October 12-15.

Schwarz, K. P., N. El-Sheimy, and Z. Liu, 1994. Prototype Development and Testing of the VISAT System, VISAT Project Report, Univ. of Calgary, Calgary, Canada, 58 pages.

Schwarz, K. P., and N. El-Sheimy, 1996. Kinematic Multi-Sensor Systems For Close Range Digital Imaging, Int. Arch. of Photogrammetry and Remote Sensing, Vienna, Austria, Vol. 31, Part B3, pp. 774-785.

Sheng, Y. J., et al, 1994. An Adaptive System for Traffic Sign Recognition, Proc. of Int. Symposium on Intelligent Vehicles, Paris, October.

Shibasaki, R. and S. Murai, 1988. Improvement of Mapping Accuracy by Applying Triplet Matching to SPOT Imagery, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 27, B10(IV), pp. 264-273.

Smith, C. E., Richards, C. A., Brandt, S. A., and N. P. Papanikolopoulos, 1996. Visual Tracking for Intelligent Vehicle-Highway Systems, IEEE Transactions on Vehicular Technology. 45(2): 744-760.

Strat, T. M., 1994. Photogrammetry and Knowledge Representation in Computer Vision, Proc. of Photogrammetry and Remote Sensing Com. III Symposium, Munich, Germany, pp. 784-792.

Subbarao, M., 1988. Interpretation of Visual Motion: A Computational Study, Morgan Kaufmann Publishers, Inc., California. 139 pages.

Szarmes, M. C., 1994. Error and Accuracy Analysis of VISAT, Technical Report, The University of Calgary, Calgary, Canada.

Tan, T. N., Sullivan, G. D., and K. D. Baker, 1994. Recognizing Objects on the Ground-Plane, Image and Vision Computing, 12(3): 164-172.

Tao, C., and Z. J. Lin, 1994. A New Approach to the Design of a System for Feature Extraction Based on Human-Computer Collaborative Tactic, Proc. of Photogrammetry and Remote Sensing Com. III Symposium, Munich, Germany, pp. 812-821.

Tao, C., 1996a. An Integrated Approach to Road Centerline Reconstruction Using Stereo Image Sequences From A Mobile Mapping System, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 31, Part B3, pp.857-862.

Tao, C., 1996b. Road reconstruction from sequential images based on shape from sequences, SPIE Proceedings on Visual Communication and Image Processing, Orlando, FL, SPIE-2727, pp. 1251-1262.

Tao, C., Li, R., and M. A. Chapman. 1996. Model Driven Extraction of Road Line Feature Using Stereo-Motion Constraints From A Mobile Mapping System, Proceedings of ASPRS/ACSM Convention, April, Baltimore, MD. pp. 135-144.

Tao, C., Chapman, M. A., and R. Li, 1997. Semi-Automatic Object Pointing Using Multi-Image Matching for Mobile Mapping, Proc. of ASPRS/ACSM Convention, Seattle, WA, pp. 265-274.

Terzopoulos, D., Witkin, A., and M. Kass, 1988. Constrains on Deformable Models: Recovering 3-D Shape and Nonrigid Motion, Artificial Intelligence, 36:91-123.

Thorpe, C., Hebert, M., Kanade, T., and Shafer, S., 1988. Vision and Navigation for the Canegie-Mellon Navlab, IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(3): 362-372.

Trinder, J., and H. Li, 1995. Semi-Automatic Feature Extraction by Snakes, Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhaeuser Verlag, pp. 95-104.

Wang, Z., 1990. Principles of Photogrammetry, Publishing House of Surveying and Mapping, Beijing, China.

Waxman, A., LeMoigne, L., Davis, L., Srinivasan, B., Kushner, T., Liang, E., and T. Siddalingaiah, 1987. A Visual Navigation System for Autonomous Land Vehicles, IEEE Transactions on Robotics and Automation, RA-3, pp. 124-141.

Wei, M., and, K. P. Schwarz, 1990. Testing a Decentralized Filter for GPS/INS Integration, Proc. of IEEE PLANS, Las Vegas, USA, pp. 429-435.

Wiles, C., and M. Brady, 1996. Ground Plane Motion Camera Models, Proc. of European Conf. on Computer Vision, pp.238-247.

Wilkin, A., 1996. Robust 3-D Object Representation By Linear Features, Proc. of Photogrammetry and Remote Sensing Congress, Vol. 31, Part B3, pp. 659-666.

Xin, Y., 1995. Automating Procedures on a photogrammetric software system, M.Sc. Thesis, Department of Geomatics Engineering, The University of Calgary, 112 pages.

Yachida, M., Kitamura, Y., and M. Kimachi, 1986. Trinocular Vision: New Approach for Correspondence Problem. Proc. of 8th Int. Conf. on Pattern Recognition, pp. 1041-1044.

Zhang, Z., Luong, Q. T., and O. Faugeras, 1996. Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction, IEEE Transactions Robotics and Automation, 12(1): pp. 103-113.

Zheng, Q., and R. Chellappa, 1993. Automatic Feature Point Extraction and Tracking in Image Sequences for Unknown Camera Motion. Proc. of the 4th Int. Conf. on Computer Vision, pp. 335-339.

Zou, W., 1996. Optimization of 3-D Object Acquisition from Digital Image Sequences. M.Sc. Thesis, Dept. of Geomatics Engineering, The Univ. of Calgary, Calgary, Canada, 74 pages.

APPENDIX A

Implementation of the software ImagExpert

ImagExpert is a software package developed for automating object measurement and feature extraction procedures using georeferenced image sequences. This research is funded by a Natural Sciences and Engineering Research Council (NSERC) grant. Digital photogrammetry, computer vision, artificial intelligence and image processing theories and methods are applied to automate information extraction from images. This software has been developed through cooperated efforts of professors, graduate students and visiting researchers.

ImagExpert is designed as an intelligent tool to facilitate image measurement procedures. This package is now implemented on a SUN SPARC workstation with an X-Window and MOTIF environment. A high-level object-oriented development toolkit WNDX was used to design the Graphic User Interface, MultiView. The package has been also interfaced with AutoCAD. The built-in communications between our package and the AutoCAD core were implemented. The measurement results can be displayed in the AutoCAD environment, OverView, simultaneously. Accordingly, it is easy for us to check and edit the results in an on-line mode.

ImagExpert can be invoked independently by a group of special icons in GeoStation, a data processing and GIS generation component of the VISAT system. For a more detailed

description of the GeoStation, the interested readers are referred to Li et al. (1994a and 1996).

The following functions of ImagExpert are implemented:

- Image enhancement and preprocessing
- Edge image analysis
- Interest point detection
- Measurement optimization
- Semi-automatic object measurement
- Automatic and semi-automatic road-line reconstruction
- Vertical object extraction

The functions and the associated algorithms of image enhancement and preprocessing have been described in Li et al. (1994a) and Xin (1995). Three modules, namely external preprocessor, global enhancement and local enhancement were developed to improve the image quality for the subsequent measurement procedure. Edge image analysis and interest point detection serve as an image analysis tool to help a human operator locate features more precisely (see Qian (1995), and Chapter 5 "digital snapping"). Measurement optimization has been extensively researched by Li, et al. (1996) and Zou (1996). The efforts of the thesis work have been placed on the development of object measurement, road-line reconstruction and vertical object extraction. The description of these approaches can be found in Chapter 5 to Chapter 7.

APPENDIX B

Description of the Test Image Data Sets

There are two data sets captured in the city of Laval and the Quebec City, respectively. Each data set consists of several consecutive image sequences. The Laval data set was collected after 4:00 p.m. with relatively poor lighting conditions. The road scenes of the images are mainly city streets. Most images have low brightness and low contrast. This causes loss of detail in many images. Furthermore, the low quality of the images affects the accuracy and efficiency of the photogrammetric measurements. Therefore, a number of image enhancement methods were developed to preprocess the data for visual interpretation and manual measurement of images. However, our tests were conducted on the raw data sets without any preprocessing in order to evaluate the robustness of the methods.

The data set of Quebec City was collected appropriately. The quality of the forward-looking images is satisfactory. However, the side-looking images were far too bright and had low contrast. The longest image sequence in this data set consists of 1665 images. The road scenes are city highways. In all of the above cases, the vehicle speed was 50-60 km per hour and the image capture rate was about one image per 0.4 second.

APPENDIX C

Intelligent Transportation Systems (ITS)

C.1 Overview of ITS

The congestion of transportation systems has considerably increased during the past two decades. As a direct result of this situation, the overall safety of these systems has been reduced. Traffic experts believe that adding and enlarging highway systems (physical infrastructure) would not ultimately solve the problem. Accordingly, new approaches to relieving the pressure upon the system must be explored. ITS is considered as information infrastructure producing a more efficient transportation network without requiring the construction of additional roadway capacity.

The 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) established the ITS national program, and set forth its goals (French and Sakamoto, 1994). These goals, as stated in the ITS National Program Plan, include:

- the enhancement of the capacity, efficiency, and safety of the highway system, serving as an alternative to additional physical capacity;
- the enhancement of efforts to attain air quality goals established by the Clean Air Act;
- the reduction of societal, economic, and environmental costs associated with traffic congestion.

The world-wide progresses on the development on ITS have been made due to the consistent support by a number of famous programs, such as the DRIVE and PROMETHEUS programs (Europe), the CACS program (Japan) and the ITS program (US). From a technical point of view, ITS provide a platform allowing the use of information processing, positioning, communication, and control technologies to alleviate congestion, improve safety and reduce the environmental consequences of road traffic (such as pollution and energy wastage).

C.2 Automatic Processing of Image Data in ITS

ITS is often defined as the application of information technology to the conventional transportation system. Therefore, the benefits of ITS are derived from the collection, dissemination, and effective use of information. The creation of high-quality and up-to-data spatial data infrastructure is a key fundamental element to the development of ITS. Such tasks have been assigned as top priority in ITS. Since the emphasis is placed on highway applications, information infrastructure is also called Intelligent Vehicle Highway System (IVHS). Mobile mapping systems are considered a very rapid and cost-effective means to collect road-related data for the generation of ITS databases.

In addition, automatic and rapid processing of sensory information about traffic conditions and environment plays a key role in the development of ITS. A variety of sensors have been investigated and developed in the application of traffic control and vehicle control,

such as, GPS, speedometer, accelerator, gyroscope, radio beacon, ultrasonic, radar, laser, loop detector, and visual imaging sensor.

Of the above mentioned sensors, visual imaging sensors are able to offer richer and more complete information of relatively large regions. They are also capable of being used to meet most of the ITS requirements. The applications of the visual imaging sensors can be categorized as follows:

Traffic control segment

- Detecting and counting vehicles at a fixed road point (e.g., intersection, merging section, or tunnel)
- Detecting parked vehicles (parking lot management and finding illegally parked vehicles)
- Detecting abnormally moving vehicles
- Detecting traffic accidents
- Calculating vehicle speed
- Recognizing the type of vehicles and plate numbers

Vehicle control segment

- Finding driving lane
- Recognizing road surface conditions (such as roughness, bump, and icy surface)
- Recognizing traffic signs
- Detecting obstacles in the road

- Measuring distance to leading/following vehicles
- Observing environment around the vehicle (pedestrians, passing vehicles)
- Recognizing view condition (such as fog, smoke, and rain)
- Observing driver's behavior
- Estimating travel time
- Navigating vehicles

The problem with the use of imaging sensors is that image quality depends heavily on the imaging conditions (weather, color, electronic noise, and camera sensitivity and dynamic range). Since the imaging conditions are subjected to the natural environment, the results derived from image processing may cause ambiguities. This is a great challenge when visual imaging sensors are used in ITS.

It is recognized that ITS research results, especially of the vehicle control segment, are of great value to this thesis work (Ozawa, 1994; Smith et al., 1996), since we are dealing with similar input data, image sequences captured from a mobile vehicle. However, qualitative but real-time processing is generally required in ITS applications while we are more concerned with precise measurement and reconstruction. Nevertheless, intelligent understanding and automatic processing of image sequence data is our common approach. The research results coming from either side will be beneficial to the other.

APPENDIX D

A Brief Overview of Image Matching and Feature Correspondence

In any imaging vision system, three components have to be taken into account when dealing with the development of an image matching or feature correspondence method:

1. Light source
    - Passive vs. active illumination

2. The nature of object to be reconstructed
    - Reflectivity
    - Surface geometry

3. Imaging system
    - Static or dynamic camera configuration
    - Number of cameras
    - Imaging geometry (baseline, focus length, sensor resolution, etc.)

Based on different configurations of the above components, the following schemes of system stereopsis are available:

- Passive vs. active vision
- Static vs. motion (dynamic) vision

- Monocular, binocular and multinocular vision
- Monocular vision: shape from shading, texture, contour, focusing, etc
- Motion vision: correspondence vs. optical flow

A major portion of the common research efforts of both the photogrammetry and computer vision communities has been directed towards the study of object reconstruction using image matching methods. The reader who is interested in this topic is referred to these review papers by Barnard and Fischler (1982), Dhond and Aggarwal (1989), Gruen (1994), Jones (1997), and Lemmens (1988).

There are a number of types of classification for image matching. The most common one, especially widely employed in the photogrammetry community, is

- area-based (gray-level matching or signal-based matching),
- feature-based, and
- structural matching (relational matching).

The above classification is essentially based on the matching primitives (or features, tokens, descriptors) used in matching methods. More generally, a matching paradigm can be constructed depending upon the following three aspects:

- matching primitives used,
- constraints imposed, and

- matching criteria and optimization methods developed.

D.1 Matching Primitives

The commonly used matching primitives (entities, features, tokens and descriptors) are represented by three levels: (1) **areas**: gray levels, functions of gray levels (the first or second derivatives of the gray levels), image transformation (frequency domain of the original image); (2) **features**: point features (edges, points of interest, corners, junctions, zero-crossings, end points or intersections of linear features), linear features (curved and straight lines, contours, polygonal approximations, regional features (blobs, circles, ellipses); and (3) **structures**: relationships of the features and the associated attributes. Accordingly, the corresponding three levels of matching scheme are defined as area-based, feature-based and structural matching.

The attributes of features which can be used for feature-based and relational matching have been summarized by Baltsavias (1991) and Forstner (1993).

It is recognized that the appropriate choice of levels of primitives to be matched depends on three components of a vision system mentioned earlier and the matching task. General observations about the choice are: (1) The higher the level of primitives used, the more complex the matching algorithms required. Simple area correlation cannot perform structural matching, a graph-matching scheme is usually employed. (2) Computational requirements at high-level matching may be less due to the lower number of items to be matched; however, the extraction and representation of high-level matching primitives is a

time-consuming procedure. (3) It is likely that high-level matching is more reliable, but more ambiguities may arise from the process of constructing high-level primitives from the original image data (Nevatia, 1996).

D.2 Constraints

The use of constraints is a key factor in reducing the matching candidates. Since light sources are not controllable in most passive stereo systems, any possible constraints from objects and imaging systems involved should be fully derived and explored,

1. geometric constraints arising from the objects being looked at and physical constraints arising from the models of which the object interact with the illumination, and
2. geometric constraints imposed by the imaging system.

The commonly used constraints accordingly are: (1) (derived from objects) disparity range and gradient constraint, continuity constraint, figural constraint, controlled continuity constraint, ordering constraint, ground plane constraint. (2) (derived from imaging system) epipolar line constraint, collinearity constraint, multi-image geometric constraint (zero-order, first-order and second-order constraint), multi-epipolar intersection constraint, camera kinematic constraint.

D.3 Matching Criteria and Optimization

Two strategies of image matching are often used: (1) **local** vs. **global** strategy, and (2) one scale vs. **multi-scale** (**hierarchy**) strategy. In general, local matching can be very precise, but ambiguous and the various local matches may not be globally consistent with one another. Global matching is more robust but not necessarily accurate in local areas. Therefore, a trade-off must be recognized, and the matching scheme should be chosen carefully, e.g., local-to-global, global-to-local, or combination thereof.

Regarding the surface reconstruction, the global consistency check and hierarchical matching strategy has been predominately used. The major differences among the different surface reconstruction algorithms are the way they handle the global and hierarchical consistency of the matches (Ebner et al., 1993).

It should be noted that the above mentioned constraints developed must be embodied with the optimization method to disambiguate the matching candidates. Two major components of the optimization methods are (1) **matching criteria** and (2) **search algorithms**.

Matching criteria can be classified into area-based gray level similarity or difference, feature-based attribute matching, and energy (cost function) merits. The search algorithms for optimization in the literature are dynamic programming, simulated annealing, gradient descent techniques, graph isomorphism, heuristic searching, relaxation, cooperation, and neural networks.

APPENDIX E

Calibration Errors and Positioning Accuracy

E.1 Errors of the Interior Orientation Parameters

E.1.1 The Error of the Principal Distance

An error of the camera principal distance f leads to a proportional change in the image coordinates x and z. Therefore, the following expression can be obtained:

$$dx = (x/f)df,$$

$$dz = (z/f)df,$$

$$dp = dx_L - dx_R = (p/f)df. \qquad (E.1)$$

When p=50 pixels and f = 800 pixels (1 pixel =10 μm), if $m_p = \pm 0.4$ pixel is required, then $m_f = \pm$ 6.4 pixels ($\approx \pm$ 0.064 mm) or the relative accuracy of 1/125 (6.4:800), has to be achieved from Eq. (E.1). At this error level, the effect of the calibration error of f on the system positioning accuracy is at the same order of the image measurement accuracy. It should be noticed that the pixel scale factor has to be introduced in the calibration of the focal length so as to resolve the inconsistency of the two directional principal distance components, f_x and f_y, due to the height-to-horizon aspect ratio of pixels in most CCD camera and frame grabber assemblies.

E.1.2 The Errors of the Principal Point Coordinates

The effect of the errors of the principal point coordinates (x_p, z_p) can be considered as the errors of image coordinate measurements, i.e., $dx = dx_p$ and $dz = dz_p$. Therefore, if an image coordinate measurement accuracy of $\pm$ 0.3 pixel is required, the error in the principal point coordinate determination should be less than the level of 0.3 pixel. Since the z-directional image error is not a major source to object positioning accuracy, the accuracy requirement for image z-coordinate measurement could not be that strict.

In the object space, the errors influence the positioning accuracy of object point coordinates in a systematic way. It should be pointed out that, due to the imperfect alignment of CCD optical axis, the principal point coordinate may have a large deviation from the center of the imaging sensor. In addition, since the focal length is highly correlated with the lens distortion parameters, and the principal point coordinates are correlated to the exterior orientation parameters of cameras, it is recommended that a multi-station self-calibration method with convergent imaging geometry be used to perform an accurate calibration. The principal point coordinates, principal focal length and the lens distortion parameters should be adjusted in a combined case.

E.2 Errors of the Calibrated Length of Baseline

In mobile mapping systems, B is usually determined using a control point field-based calibration method. The contribution of the error in the calibrated baseline B to the object

positioning accuracy can be assessed from the following expressions. Differentiating Eq. (4.1) yields,

$$m_Y = \pm\ (Y/B)\, m_B$$

$$m_X = \pm\ (X/B)\, m_B$$

$$m_Z = \pm\ (Z/B)\, m_B\ . \qquad (E.2)$$

It can be seen that the effect of the error m_B on the object point coordinates changes linearly with the values X, Y and Z, respectively. Its behavior is equivalent to a constant scale change. If the error m_B is 3mm, then the effect on the total positioning accuracy will be 5.25 cm when the object is at a distance of 35m.

E.3 Errors of the Exterior Orientation Parameters

E.3.1 Errors in the Determination of Rotation Parameters

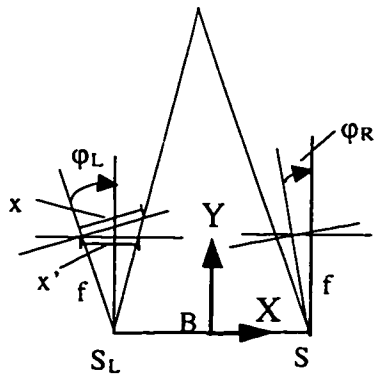


Figure E.1 Rotation parameter errors

As shown in Figure E.1, the correction to the errors from the rotation parameters can be treated as image rectification. Thus, image rectification formulae can be used to correct the image coordinates (x, y) corrupted by the errors of the rotation parameters to the correct ones (x', y') (Wang, 1990),

$$x' = f\frac{a_1x + a_2z + a_3f}{c_1x + c_2z + c_3f} = x + dx\,,$$

$$z' = f\frac{b_1x + b_2z + b_3f}{c_1x + c_2z + c_3f} = z + dz\,. \qquad (E.3)$$

Using small values of the rotation angles, we have,

$$x' = f\frac{x - z\kappa - f\varphi}{x\varphi + z\omega + f} \qquad \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} 1 & -\kappa & -\varphi \\ \kappa & 1 & -\omega \\ \varphi & \omega & 1 \end{bmatrix}$$

$$z' = f\frac{x\kappa + z - f\omega}{x\varphi + z\omega + f} \qquad (E.4)$$

where the rotation angular system of (ω, κ, φ) corresponding to rotations about the (X, Y, Z) axes is chosen. Differentiating the above equations, the error expressions in terms of the image coordinate measurements can be obtained :

$$dx = (x')|_x\, d\varphi + (x')|_x\, d\omega + (x')|_x\, d\kappa = -(f+x^2/f)\, d\varphi - (xz/f)\, d\omega - z\, d\kappa$$

$$dy = (y')|_y\, d\varphi + (y')|_y\, d\omega + (y')|_y\, d\kappa = -(xz/f)\, d\varphi - (f+z^2/f)\, d\omega + x\, d\kappa$$

$$dp = -f\,(d\varphi_L - d\varphi_R) - (x_L^2 d\varphi_L - x_R^2 d\varphi_R)/f -$$

$$(z_L x_L d\omega_L - z_R x_R d\omega_R)/f - (z_L d\kappa_L - z_R d\kappa_R)$$

$$= -f(d\varphi_L - d\varphi_R) - (x^2 d\varphi_L - (x-p)^2 d\varphi_R)/f -$$

$$(z/f)(x\, d\omega_L - (x-p)\, d\omega_R) - z(d\kappa_L - d\kappa_R). \qquad (E.5)$$

As far as dp is concerned, $d\varphi$ is the major error factor. Considering the error of the convergence angle $dv = d\varphi_L - d\varphi_R$, which may be caused by relative orientation errors, then we have (assuming the right imaging station is free from error),

$$dp^v = (f + x^2/f)\, dv \qquad (E.6)$$

After differentiating equations (E.5) and (E.6), the mean square error expression can be obtained. For example, m_p resulting from m_φ, m_ω, and m_κ respectively is,

$$m_p^2 = (2f^2 + x^4/f^2 + (x-p)^4/f^2)\, m_\varphi^2$$

$$m_p^2 = (z/f)^2 (x^2 + (x-p)^2)\, m_\omega^2$$

$$m_p^2 = 2z^2 m_\kappa^2 \quad . \qquad (E.7)$$

If the limits of image coordinate measurement accuracy are given, the accuracy requirements for rotation parameter calibration can be estimated using the above expressions (E.5 - E.7). These requirements are tabulated in Table E.1 (f = 800 pixels is assumed). It is shown that an accuracy 0.02° for the rotation angles φ , *0.09*° for the angle κ, and *0.03*° for the relative orientation angle (φ_L-φ_R) has to be achieved in order to meet the requirement of 0.4 pixel accuracy of image coordinate measurements. Also, the

practical calibration accuracy for ω is not strictly required to be within 0.02° due to the fact that m_z is not the major error source. Consequently, attention should be paid to accurate calibration of the angle φ.

Table E.1 Accuracy requirements for calibrating rotation parameters

Accuracy Requirements	Case1: x=200, z=150, p=50 (pixel)				case2: x=100, z=100, p=50 (pixel)			
	m_φ	m_ω	m_κ	m_v	m_φ	m_ω	m_κ	m_v
m_x =0.3 pixel	±0.02°	±0.46°	±0.11°	---	±0.02°	±1.38°	±0.17°	---
m_z = 0.5 pixel	±0.76°	**±0.03°**	**±0.14°**	---	±2.29°	±0.03°	±0.29°	---
m_p = 0.4 pixel	**±0.02°**	±0.49°	±0.11°	**±0.03°**	±0.02°	±1.64°	±0.16°	±0.03°

E.3.2 Errors in the Determination of Translation Parameters

The effect of the errors (dX_0, dY_0, dZ_0) of the exterior translation parameters of the camera exposure stations on image coordinate measurements can be derived using the following relationships (Figure E.2),

$$x' = fX/Y\,, \qquad x = fX/(Y+dY_0)\,. \tag{E.8}$$

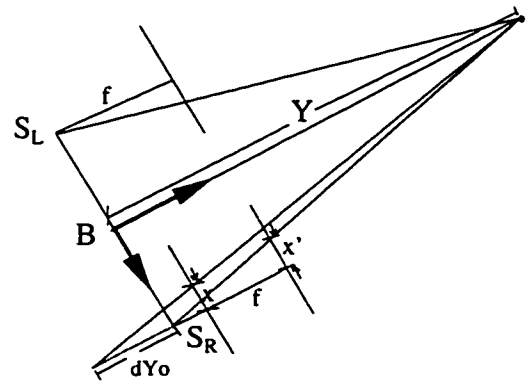


Figure E.2 Translation parameter errors

Thus the differential express is hold as,

$$dx = x' - x = (x/Y)\, dY_0 \,. \qquad (E.9)$$

Similarly, we have

$$dz = z' - z = (z/Y)\, dY_0 \,. \qquad (E.10)$$

Using a similar derivation, it can be found that

$$dx = (f/Y)\, dX_0$$

$$dz = (f/Y)\, dZ_0 \,. \qquad (E.11)$$

The parallax expression can be obtained,

$$dp = (f/Y)(dX_{OL}-dX_{OR}) + (x_L dY_{OL} - (x_L-p)dY_{OR}) / Y . \qquad (E.12)$$

The corresponding mean squared errors can be written as,

$$m_x = \pm \, sqrt\,((f/Y)^2 m_{X_o}^2 + (x/Y)^2 m_{Y_o}^2)$$

$$m_z = \pm \, sqrt\,((f/Y)^2 m_{Z_o}^2 + (z/Y)^2 m_{Y_o}^2)$$

$$m_p = \pm \, sqrt\,(2(f/Y)^2 m_{X_o}^2 + ((x^2+(x-p)^2)/Y^2) m_{Y_o}^2) . \qquad (E.13)$$

By examining Eq. (E.13), the effects of errors in exterior translation parameters on the accuracy of image coordinate measurements decrease as the object distance Y increases. For a comparison, Table E.2 gives the accuracy requirements for the determination of exterior translation parameters when Y=10m and Y=30m (f =800 pixels is assumed). It is revealed that, in the case of Y=10m, the accuracy requirement for m_{X_o} is critical (3.5mm) and, relatively speaking, for m_{Y_o} and m_{Z_o} is not stringent. Since the error contribution of m_{X_o} to the system positioning accuracy behaves as the same as the baseline error m_B, Essentially, only an accurate calibration of the baseline B is crucial.

E.4 Summary

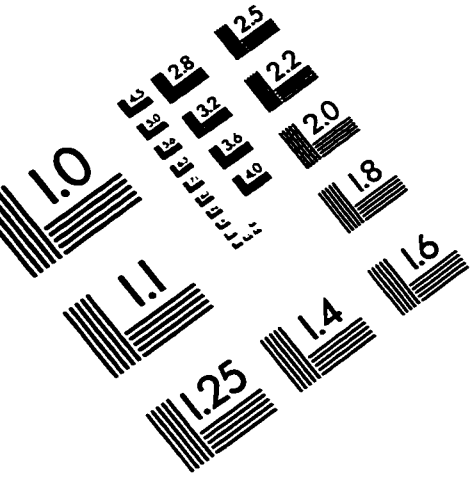
The quantitative error models on the effects of the calibration errors of interior and exterior orientation parameters on the system positioning accuracy were derived. The error budget of the calibration errors has been tabulated in Table E.1 and E.2. A very accurate

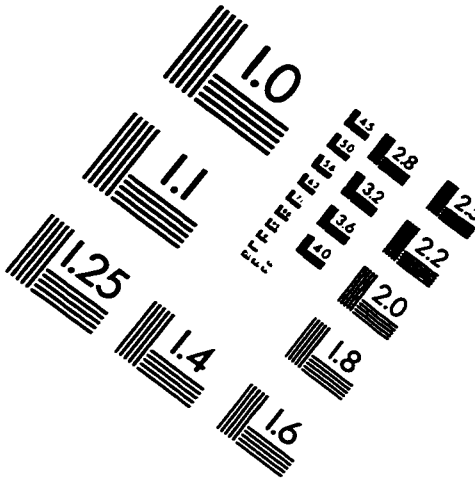
calibration of interior orientation parameters, rotation angles φ (rotation around the axis Z) and the length of baseline is required. The accuracies of 6.4 pixel of m_f, 0.3 pixel of (m_{xp}, m_{yp}), 0.02° for the rotation angles φ , *0.03°* for the relative orientation angle (φ_L-φ_R) and 3.5 mm for the baseline have to be achieved, so that the effects of the calibration errors on the total system positioning accuracy are at the same level as that of errors arising from image coordinate measurements. However, the accuracy requirements for the rotation parameters m_ω and m_κ, translation parameters m_{Yo} and m_{Zo} are not very stringent. The required calibration accuracies are about 0.1° for ω and κ, and 1 cm for Y_0 and Z_0, respectively.

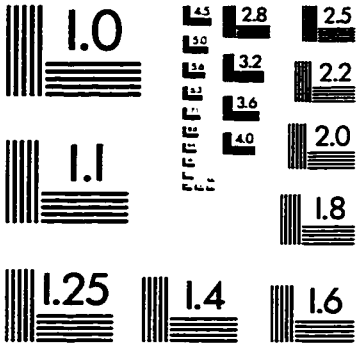
Table E.2 Accuracy requirements for the translation parameters (unit: cm)

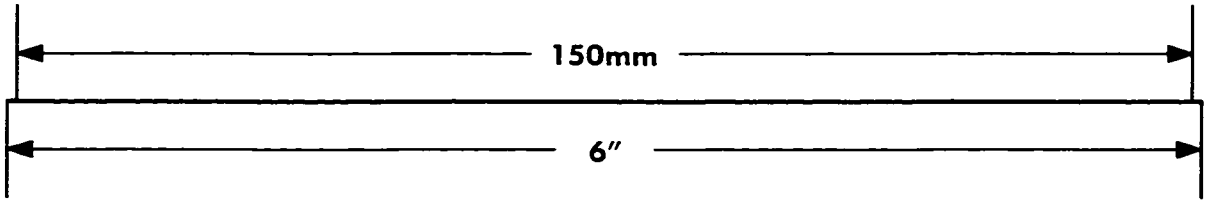
Accuracy Requirements	*Y* (m)	x=200, z=150, *p*=50 (pixel)			x=100, z=100, *p*=50 (pixel)		
		m_{Xo}	m_{Yo}	m_{Zo}	m_{Xo}	m_{Yo}	m_{Zo}
M_x =0.3 pixel	10	±0.38	**±1.5**	---	±0.38	±3.0	---
m_z = 0.5 pixel		---	±3.3	**±0.63**	---	±5.0	±0.63
m_p = 0.4 pixel		**±0.35**	±1.6	---	±0.35	±3.57	---
M_x =0.3 pixel	30	±1.13	±4.5	---	±1.13	±9.0	---
m_z = 0.5 pixel		---	±10.0	±1.89	---	±15.0	±1.89
m_p = 0.4 pixel		±1.06	±4.8	---	±1.06	±10.7	---

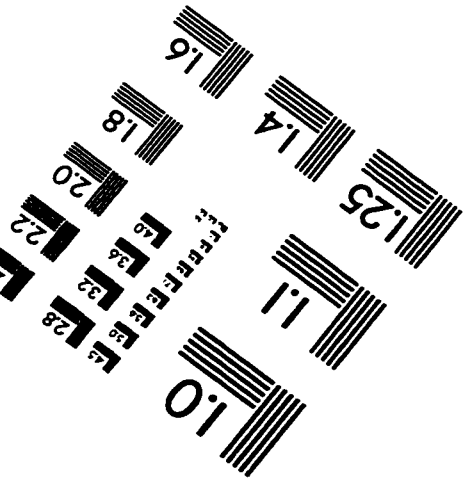
IMAGE EVALUATION
TEST TARGET (QA-3)

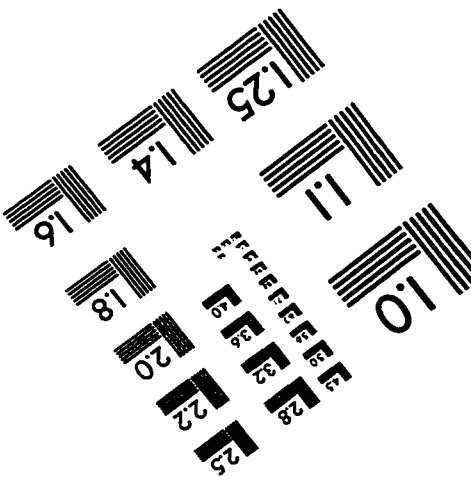












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989