Vision-based navigation with reality-based 3D maps

By

Xun Li

A thesis in fulfillment of the requirements for the degree of

Doctor of Philosophy

Surveying and Geospatial Engineering

School of Civil and Environmental Engineering

Faculty of Engineering

The University of New South Wales

August 2013

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Li

First name: Xun

Other name/s:

Abbreviation for degree as given in the University calendar: PhD

School: School of Civil and Environmental Engineering

Faculty: Faculty of Engineering

Title: Vision-based navigation with reality-based 3D maps

Abstract 350 words maximum: (PLEASE TYPE)

This research is focused on developing vision-based navigation system for positioning and navigation in GPS degraded environments. The main research contributions are summarized as follows:

a.A new concept of 3D map, which mainly consists of geo-referenced images, has been introduced. In this research, it provides the map-matching function for vision-based positioning.

b.A method of vision-based positioning with use of photogrammetric methodologies has been proposed. It mainly obtains geometric information of the navigation environment from the 3D map through SIFT based image matching and uses photogrammetric space resection to solve the position in 6 degrees of freedom. The algorithms have been tested in an indoor environment. The accuracy has reached around 10 cm.

c.A multi-level outlier detection scheme for the vision-based navigation system has been developed. It mainly combines RANSAC with data snooping. The former one deals with high percentage of mismatches, while data snooping removes outliers from different sources in the least squares adjustment for both 3D mapping and positioning solution.

d.The deficiency of using RANSAC for outlier detection in image matching and homography estimation has been identified. In this research, a novel method which combines cross correlation with feature based image matching has been proposed. It is able to evaluate the RANSAC homography estimation and improve the image matching performance. The method has been successfully applied to the vision-based navigation solution to find corresponding view from the database and improve the final positioning accuracy.

e.The positioning performance of the system has been evaluated through the analysis of mathematical model and experiments. The focus has been on various image matching conditions/methods and their impact on the system performance. The strength and weaknesses of the system have been revealed and investigated.

f.The vision-based navigation system has been extended from indoor to outdoor with corresponding changes. Besides camera, it also takes advantage of multiple built-in sensors, including GPS receiver and a digital compass to assist visual methods in outdoor environments. Experiments demonstrate that such system can largely improve the position accuracy in areas where stand-alone GPS is affected and can be easily adopted on mobile device.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

Xun Li
Signature

Witness

21 / 11 / 2013
Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

COPYRIGHT STATEMENT

Signed ......Xun Li......

Date ......21/11/2013......

AUTHENTICITY STATEMENT

Signed ......Xun Li......

Date ......21 / 11 / 2013......

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ....Xun Li..........

Date ....21/11/2013..........

ABSTRACT

Positioning and navigation applications have never been so accessible. Some of the most important positioning or localization techniques include satellite-based positioning (e.g. GPS), beacon-based positioning, dead reckoning and so forth. There are however, many unmet needs, especially for urban and indoor environments. Various strategies and sensors have been adopted to fill in this gap, within which vision is regarded to be one of the most promising but challenging approach. This research is focused on developing vision-based navigation system for positioning and navigation in GPS degraded environments. The main research contributions are summarized as follows:

a. A new concept of 3D map has been introduced. The new 3D map mainly consists of geo-referenced images, and provides reality-based visualization in terms of images as well as 3D geo-referenced geometric information. In this research, it provides the map-matching function for vision-based positioning. Its development process and applications have been discussed. Multi-image matching has been introduced into the 3D mapping procedure.
b. A method of vision-based positioning with use of photogrammetric methodologies has been proposed. It mainly obtains geometric information of the navigation environment from the 3D map through SIFT based image matching and uses photogrammetric space resection to solve the position in 6 degrees of freedom. The algorithms have been tested in an indoor environment. The accuracy has reached around 10 cm.
c. Vision sensor is inherently fragile against errors. In this research, a multi-level outlier detection scheme for the vision-based navigation system has been proposed. It mainly combines RANSAC, which deals with high percentage of mismatches, with data snooping, which removes a small number of outliers at the least squares

adjustment for both 3D mapping and positioning solution.

d. The deficiency of using RANSAC for outlier detection in image matching and homography estimation has been identified. In this research, a novel method which combines cross correlation with feature based image matching has been proposed. It is able to evaluate the RANSAC homography estimation, detect poor ones and improve the image matching performance. The method has been successfully applied to the vision-based navigation solution to detect mismatched reference image(s) as well as the image matching for final positioning. Experiment proves such method has improved the system performance effectively.

e. In this research, the positioning performance of the system has been evaluated through the mathematical model and experiments. The focus has been on various image matching conditions/methods and their impact on the system. The characteristics, including both strength and weaknesses of the system have been revealed and investigated.

f. In recent years the low cost built-in sensors on mobile devices (e.g. smartphone), especially high resolution cameras have placed greater demand for a breakthrough in their applications for seamless positioning. In the later stage of research, the vision-based navigation system has been extended from indoor to outdoor with corresponding changes been made to cater for outdoor environments. It mainly uses visual input to match with geo-referenced images for positioning solution, and also takes advantage of multiple sensors onboard, including GPS receiver and a digital compass to assist visual methods. Experiments demonstrate that such system can largely improve the position accuracy in areas where stand-alone GPS is affected and can be easily adopted on mobile devices.

ACKNOWLEDGEMENT

First of all, I would like to gratefully and sincerely thank my supervisor, Associate Professor Jinling Wang, for his guidance, support and encouragement throughout my PhD study. I would also like to thank Dr. Weidong Ding for his assistance and guidance at early stage of my PhD study. Special thanks also go to Professor John Trinder for his valuable advice on this work.

I also would like to express my thanks to all the staff members, especially Professor Chris Rizos, Associate Professor Bill Kearsley and Dr Craig Roberts for their support as well as the introduction to Australian Culture. I would like to express my heartfelt gratitude to Dr. Jinghui Wu for her valuable advice on my research and years of friendship. Special thanks to Professor Ruizhi Chen from Texas A&M University Corpus Christi and my PhD panel Associate Professor Linlin Ge and Dr. Binghao Li for their valuable advice on research study. I am also very grateful to my university friends, especially Ms Ye Shi and Ms Nahid Sultana, as well as friends from our research group: Mr Tao Li, Ms Yiping Jiang, Mr Youlong Wu, Mr Wenyang Liu, who have filled my PhD life with happy memories.

I would like to sincerely acknowledge the Australian Government for awarding me the International Postgraduate Research Scholarship (IPRS) to pursue my PhD studies at the University of New South Wales.

Last but not least, I would like to express my deepest appreciation to my parents for supporting me all the way. And my sincerest love to my husband Lvbing Gong. Without your love and support, I would never have been able to live such a happy and fulfilled life during the PhD study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASIFT	Affine-Sift
ADOP	Attitude DOP
CCD	Charge-Coupled Device
CV	Computer Vision
DOF	Degrees Of Freedom
DoG	Difference Of Gaussians
DOP	Dilution Of Precision
DR	Dead Reckoning
GCP	Ground Control Points
GLONASS	Globalnaya Navigatsionnaya Sputnikovaya Sistema
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
KLT	Kanade-Lucas-Tomasi
LBS	Location-Based Services
LMS	Least Median of Squares
LoG	Laplacian Of Gaussian
LSH	Locality-sensitive Hashing
LSM	Least Squares Matching
MDB	Minimal Detectable Bias

NMEA	National Marine Electronics Association
PDOP	Position DOP
PGCP	Pseudo Ground Control Point
PPS	Pulse Per Second
RMSE	Root Mean Square Error
SAD	Sum Of Absolute Differences
SFM	Structure From Motion/ Shape From Motion
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localization And Mapping
SPP	Single Point Positioning
SSD	Sum Of Squared Differences
RANSAC	Random Sample Consensus
UAV	Unmanned Aerial Vehicle
UWB	Ultra-Wideband

CHAPTER 1
INTRODUCTION

1.1 MAPPING AND NAVIGATION CONCEPTS

Navigation is a technique for the determination of position and velocity of a moving platform with respect to a known reference, but can also include the attitude of the platform (Groves, 2007). In this research, position and attitude have been mostly discussed and sometimes the term positioning is used instead of navigation.

Methodologies used for navigation generally fall into two groups: position fixing and dead reckoning (DR). Position fixing is obtained by making observations with respect to known reference positions, also known as reference-based systems (e.g. GPS). Dead reckoning on the other hand determines the current position based on the previous one, therefore the initial position must be given. In the following, various navigation technologies are introduced with emphasis on the ones been adopted in this research.

The dominating navigation technology has been satellite-based navigation: Global Navigation Satellite System (GNSS), which is a typical reference-based navigation system. GPS is the most popular and the first constellation that been developed (Figure 1.1). In full operation it includes 31 satellites distributed uniformly around 6 circular orbits. These satellites transmit encoded radio frequency signals. Then the receivers on ground can calculate their own positions by the travel time of satellite signals and information from encoded signals. Three satellites are the minimum number to calculate the receiver‘s latitude, longitude and altitude, and four can help correct clock bias. Generally, a good positioning result requires sufficient satellite coverage, and redundancy can help improve the accuracy. Therefore, current trend of satellite-based navigation is moving towards multi-constellation period. The second system is the

GLONASS supported by the Russian Republic. It has 24 satellites in 3 orbital planes. The Galileo system is the third satellite-based navigation system that currently under development. It will have 10 circularly polarized navigation signals in 3 frequency bands. The BeiDou Navigation Satellite System is being developed by China, consisted of 27 medium earth orbit satellites. New constellations, signals and associated frequency diversity will help improve capabilities for calibrating ionospheric propagation delays, robustness against incidental interference and most importantly result in better accuracy (Groves, 2013). However, even with four or more satellite constellations, GNSS still suffer from outages due to signal blockage, interference, multi-path effects or jamming, which means there are places like indoors and urban canyons that unable to receive direct line-of-sight signals from satellites. Satellite-based system alone is unable to full-fill the need for ubiquitous navigation and alternative technologies are needed.

Figure 1. 1 The GPS satellite constellation (Griffin, 2011)

Navigation and mapping are intricately coupled problems. A map is consisted of a coordinate system with a set of information (e.g. location, scale, etc.) about the features in the mapped environment. Today digital maps have commonly been used together with GPS for navigation. It provides the visual information for drivers about the routes, and can also be used to impose constraints on the positioning solution since a vehicle runs within road networks (Skog and Handel, 2009). However, the potential of maps for navigation hasn‘t been fully explored by such applications. As a matter of fact, map itself can be used as a position fix technique. It is called map-based positioning. This method can be adopted when the moving platform is navigating in a mapped environment. In this approach, the moving platform uses its sensors to perceive its local environment, and this perception is then compared to a map that previously stored. If a match is found, the position of this platform can be determined (Aboelmagd et al., 2013). Cameras and laser range finders are typical sensors used in map-based positioning. The map is constructed from prior knowledge of the environment.

Magnetic sensors have been used for ‗navigation‘ for centuries. Today electronic compasses can be easily found on portable low cost equipment and measure at least two orthogonal axes X and Y of the earth magnetic field. In the strict sense, a digital compass alone does not give absolute position information like GPS, thus cannot be used for navigation on its own. However, it uses magnetometers to provide heading measurements relative to the Earth‘s magnetic north and can be applied to assist other systems for position fix.

In some specific environments, active beacons are employed to provide accurate positioning information. Wireless communication is used, such as Wi-Fi and UWB networks. Triangulation and fingerprinting are typical algorithms used in position resolution. However, the problem with active beacons is that it requires high cost installation and maintenance, which limit the application to specific environments.

Despite of various position-fixing methods, DR has also been used either independently or assist position-fixing systems. Typical examples of DR include inertial navigation systems (INS) and odometry. Inertial navigation origins from the Cold War and has been under development for decades. The basic idea is to compute the real-time state of a moving vehicle (object) using motion sensors. More specifically, gyroscopes and accelerometers are motion sensors that measure the rotation rates and forces, which are used to define the translational motion of the vehicle with respect to the inertial reference frame. Then the position, velocity and orientation are calculated as the –state” of the vehicle. Logically, the system requires inertial measurement units (IMUs) for measurement and a system to compute and update the state. With the development of micro-electromechanical system (MEMS), low-cost and light weight IMUs have been enabled. Since inertial navigation depends on the sensed measurements from IMU, the sensor error may accumulate and lead to solution failure. Therefore techniques to compensate for or mitigate the sensor errors are of significant value for research in INS. Odometry on the other hand measures the rotation of the wheel axes and the steer axes. Then wheel rotation is translated into linear displacement. In summary, such systems have the advantage of being self-contained, but subject to cumulative errors. Therefore in navigation systems, DR based technologies are usually used together with position fixed systems, an typical success has been GPS/INS integration.

In order to provide more reliable and comprehensive navigation solution, integrated navigation has attracted growing attention in recent years. It means a combination of two systems or more from the two categories mentioned previously. Typical combination includes GNSS/ inertial navigation, or GNSS, odometry and map-matching combined. The basic role is that the wider the range of technologies deployed, the better the performance will be. Therefore, the corresponding development trend is to develop more components on a mobile platform that uses various positioning methods. For instance, a smartphone today contains a camera, inertial and magnetic sensors, mapping, a GNSS receiver, Wi-Fi transceiver and the phone itself, all of which could potential be used for navigation (Groves, 2013).

1.2 VISION-BASED NAVIGATION

In the research domain of indoor navigation or more generally, navigation in a GPS-denied environment, vision is believed to be the most promising but challenging technology so far. It is an effective method to find the 3D position of a target. Compared with other sensors that may hold the promise to supplement satellite-based positioning, since vision sensors are cheap, ubiquitous, self-contained (no external infrastructure such as beacons, radio stations is required) and works well for both indoor and outdoor environment. In fact, vision-based navigation has begun to attract attention since late last century. Up today countless research contribution has been made. Most vision-based navigation systems depend on the exploitation of one or more cameras, either map-based or map less navigation is adopted.

The main purpose of vision-based navigation is to determine the position (and orientation) of the vision sensor, then mobile vehicle‗s (the platform carrying the vision sensor) motion/trajectory can be recovered. However, it is not without its limitation. Vision sensor can measure relative position with derivative order of 0 but senses only a 2D projection of the 3D world – direct depth information is lost. Several approaches have been made to tackle such a problem. Some people use stereo cameras by which the distance to a landmark can be directly measured. Another way is to use monocular vision with the integration of data from multiple viewpoints (structure from motion) or rely on the prior knowledge of the navigation environment (e.g. map, or models). Sometimes a combined use of these methods is presented in the same system. For example in Christensen et al. (1994), CAD model is used together with stereo vision. Despite the various methods that have been used, generally any vision-based navigation system falls into two categories: the one that depends on prior knowledge of the navigation environment for positioning, and the one that do not need any previous knowledge of the environment but calculate its position as they perceive it. We refer to the former one as map-based approach and the later as mapless approach.

In this research, the map-based approach is used. Therefore the review will be focused on the former group and the later will be briefly discussed in the follow context. Different from previous literature review works made by Guilherme et al. (2002) and Francisco et al. (2008) that map is restricted to more specific forms, in this research map is referred to more general concept: pre-defined knowledge of the navigation environment.

1.2.1 Map-based approaches

The essential idea for map-based visual navigation is to store the information of the landmarks in the navigation environment first, then during navigation, the system use the visual input to match the map, identify the landmarks previously stored and estimate its own position. The four steps (Borenstein et al., 1996) the computation normally involved are: firstly acquire sensor information through camera(s); secondly detect landmarks through image processing; thirdly establish matches between observation and expectation by image/map matching; last step is to calculate its position according to the relationship between observed landmarks and its information stored in the map. The form of the map however is changing with the advance of the technology, from CAD models with varying complexity to simpler models, then models are replaced by appearance based approach, later even images are used as 3D map for navigation. With the rich literature available, here we focused on the development of the maps that relevant to reality based 3D map. Some other forms of the maps, such as occupancy map and topological maps, which attempt to –squeeze 3D into 2D” are not included. More information on those topics can be found in Guilherme and Avinash (2002).

1.2.1.1 3D model as map

Van Driel has realized in as early as 1989 that the advantage of 3D lies in the way we see the information. A 3D display of the environment simulates spatial reality, thus allowing the viewer to recognize and understand quickly. Therefore, a navigation system that contains a realistic map will certainly assist the users better. As Coors et al. (2004) point out: a more natural way to present route instructions is to use pseudo realistic instructions, i.e. three-dimensional maps. There has been some evidence that people recognize landmarks and find route in the cities easier using a 3D model than using a symbolic 2D map and that search and visualization of location-based information of a city becomes more intuitive with life-like 3D **(**Rakkolainen, 2000). Moreover, the high visual correspondence between objects on a 3D map and real world objects increases the 3D map‗s navigational value (CoorK, 2008). Therefore, 3D navigation, or the use of a 3D map for navigation purposes is of great value in the development of navigation systems.

Most of the early vision-based navigation systems rely on 3D geometric models that contain precise metric measurements of the objects in the environment. Firstly CAD model was used. For example in 1987, Tsubouchi and Yuta proposed a system that used both CAD models and color images for their map-assisted vision system. Later other forms of models are used to represent the geometry of the navigation environment. For instance in FINALE system (Kosaka and Kak, 1992) 3D geometric model of the hallways were built. Self-localization was realized by matching a sequence of image features and landmarks derived from the geometric model of the environment. Fukastu et al. (1998) proposed a manipulation technique to intuitively control the ―bird‗s eye" overview display of an entire large-scale virtual environment in a display system that enables efficient navigation even in enormous and complicated environments using both global and local views. Coors and Schilling (2004) presented routes on mobile devices using 2D and 3D maps, and the problem with this approach is that their 3D maps are found to be slower to use both in initial orientation and route finding compared to 2D maps. According to the users the 3D model should be more detailed and realistic and the target should be highlighted in it. Meijers et al. (2005) propose their method for indoor navigation routing. And their problem is also in that the model still requires refinements.

In the meantime, three-dimensional geo-information has grown to be an important subject within the GIS community. 3D navigation therefore benefits from such an act. Sharkawi et al. (2008) come with a 3D navigation system in virtual 3D (indoor and outdoor) environment by utilizing a freely available 3D game engine couple with GIS elements. But their system hasn‘t yet been able to provide real-time positioning capability. Li et al. (2008) discusses the framework regarding a 3D indoor navigation service, which aims at 3D GIS-based, BIM information-supported and topologic analysis-oriented indoor navigation. It points out the limitation in current research regarding 3D indoor navigation, and the need and importance to develop such systems in various application areas. Wang et al. (2008) use virtual reality technique to develop a 3D navigation system, which can be used for campuses, museums or art galleries, and more importantly, it includes user positioning mechanism using GPS and active RGID.

The above development shows different technologies being used and various methodologies contributing to the research of 3D model based navigation. One relatively common shortcoming with these systems is that the 3D model of the environment is not good enough. The similarity between world model and the real world is not big enough to enable a user to quickly recognize and build up correspondence.

1.2.1.2 Appearance-based approaches

Appearance-based method then emerged to provide an alternative approach for model-based methods. An appearance-based model is created by ―memorizing” the navigation environment using images or templates. By comparing the templates in the model with its current view, a robot can derive control commands to steer itself along a memorized route or to a goal position (Cobzas et al., 2003). The strength of appearance based models lies in their ability to represent the environment through high-level image features, using similarity measures to decide if new information can be added to the map (Zhang et al., 2012). It has attracted growing attention for both indoor and outdoor navigation.

For indoor appearance-based navigation, mobile robot has been one of the major applications. One of the early approaches was developed by Turk and Pentland (1991). Their approach treats the recognition problem as an intrinsically two dimensional recognition problem rather than requiring recovery of the three dimensional geometry. Ohno et al. (1996) used the differences between the currently collected images and the pre-recorded image sequence to continuously estimate the robot‘s position and orientation shifts. While the orientation change can be obtained with relatively high accuracy, position change may not be accurately estimated. Another limitation with this approach is that it is based on the assumption that the correspondence between the current image and the reference image has always been found correctly, leaving mismatches a severe danger jeopardizing the reliability of the whole system. Rivlin et al. (2003) proposed a new algorithm for image-based robot navigation applications. At the core of this idea is to generate the translation and rotation shifts in a robot movement by matching the target image with the images taken in real time. While this idea makes a good point, another contribution of their approach is that RANSAC paradigm is used to deal with outliers caused by mismatches. However, it is not without its limitations. The algorithm is only able to provide three degrees of freedom.

The reason is that it only uses the epipolar geometry to estimate the relative position and orientation between query image and target image. The authors only consider a camera that is rigidly positioned on the robot, which means the difference in the positions of the two cameras is only regarded as motion in the plane parallel to the floor (the X and Z plane) and rotation about the Y axis. In several cases not enough correct matches can be found to compute the position shift. The limitation in the degrees of freedom can also be found in other approaches (e.g. Kitanov et al., 2007).

In the meantime, outdoor image based navigation has also been developed over the years. The traditional approach is to match the real time query image with the reference images in the database. Whenever a match is found, the position information of this reference image is transferred to the query image and used as user position. This is essentially an object-recognition and image retrieval problem. A great variety of work has been done to address the location recognition aspect by using different image matching techniques (e.g. Schaffalitzky, 2002; Goedeme, 2004). A further improvement is to calculate the relative position between the query view and the identified reference view to obtain more accurate position estimation. In 2006 Zhang and Kosecka first used a wide-baseline matching technique based on SIFT features to select the closest views in the database, then the location of the query view was obtained by triangulation. In Robertson (2004) the orientation of the sensor was also estimated since the pose of the query view is obtained from plane-to-plane transformation. Building façade was used as dominant plane.

One common shortcoming for the appearance-based approaches lies in that the pre-stored images can only provide 2D information, which limits the camera pose estimation to certain accuracy with only position information while orientation estimation has been lost.

1.2.1.3 Images used as 3D maps

Images used as 3D maps for navigation purposes is a relatively new approach in the research domain, which has its origin from appearance-based methods and also retains the virtue of 3D model-based approach. Compared with its ancestor –appearance-based methods”, it takes advantage of the 3D geometric information from the 3D map for pose estimation. And in the meantime their advantage over traditional model-based 3D navigation lies in that it provides more realistic view of the navigation environment and at the same time, do not require full 3D reconstruction, which is both time consuming and demand much space for storage. Monocular, stereo as well as panoramic images all have been used as 3D map for research on image navigation systems. And such systems have found their application in computer vision community in areas like mobile robot navigation, simultaneous localization and mapping (SLAM) and unmanned aerial vehicle (UAV).

The earliest of such work can date back to 1987, when Harris and Pick (1987) propose a method to construct an explicit three-dimensional representation from feature points extracted from a sequence of images taken by a moving camera. The points are tracked through the sequence, and their 3D locations are accurately determined with Kalman filters. The ego-motion of the camera is also determined. But the limitation of this attempt is that it fails to notice the strong correlations caused by the camera motion. Manessis et al. (2000) proposed to reconstruct 3D structure of the environment through image sequence. The major contribution of the work is that it uses a recursive structure from motion to realize surface recovery. Different from Manessis‘s approach, Kidono and his colleagues (2002) used stereo vision to obtain range data for 3D reconstruction. More specifically, they used a human guided process to record images with a stereo camera and construct the 3D map on-line from these images. Then the robot can repeat the same route. Following this trend, Royer et al. (2005) developed a system that is able to follow a pre-defined path based on a 3D map, which was built off-line using video sequence obtained from pre-training period. We found this work is of significant importance to our research. It proposed the idea that interest points can be reconstructed in 3D and used as landmarks for the localization process. Royer et al. (2007) refined this previous work and evaluated the accuracy and robustness of the system under various environments. From then on, using images to build/as 3D map for navigation has attracted growing attention and more work can be found in recent years, (e.g. Li et al., 2011; Ruiz-Ruiz et al., 2012).

It is worth mentioning that a close field to vision-based navigation: SLAM (Simultaneous Localization and Mapping) has also contributed to the image-based 3D map. Over the last 10 years, autonomous robot navigation community has seen much progress, especially on the topic of SLAM. It has used both monocular and stereo imaging to build 3D maps of the environment for navigation. Davison and Murray (2002) introduced their work as an improvement of previous work (Davison, 1998) by using a visual SLAM system that can operate in real-time. This system was able to build a 3D map of landmarks from images taken from different viewpoints. Jung and Lacroix (2003) presents an approach to build high resolution digital elevation maps from a sequence of unregistered low altitude stereovision image pairs. The limitation of the work is that it relies on a wide baseline fixed stereo rig to obtain depth information. Tardif et al. (2008) present a system for Monocular Simultaneous Localization and Mapping (Mono-SLAM) relying solely on video input. The main methodological contribution in this paper is that given the last image and a current 3D map of landmarks, they decouple the rotation estimation from the translation in order to estimate the pose of a new image. The image-based 3D mapping process will be further discussed in Chapter 2.

1.2.2 Mapless approaches

For vision-based navigation, a mapless approach does not rely on any prior knowledge of the environment. Mainly there are two techniques that are used in mapless visual positioning: optical flow and feature tracking.

A fundamental problem in the processing of image sequence is the measurement of optical flow (or image velocity). The goal is to compute an approximation to the 2D motion field- a projection of the 3D velocities of surface points onto the imaging surface – from spatiotemporal patterns of image intensity. Once computed, the measurements of image velocity can be used for a wide variety of tasks ranging from passive scene interpretation to autonomous, active exploration (Barron and Beauchemin, 1994). For vision-based navigation, the camera movement is perceived as a relative motion of the field of view. Two essential problems to solve are: what kind of image property to track, and how to track it.

The basic idea for feature tracking is to form sets of correspondences between features on every frame of a sequence. Intensity correlation is usually used for feature tracking on a pixel basis. Two widely applied methods are cross correlation and sum of squared differences. By using dense optical flow algorithms, the optical flow between two consecutive frames is usually represented by a vector for every pixel. By comparison, sparse optical flow algorithms only calculate the displacement for certain selected region of pixels. Invariant image features such as SIFT (Lowe, 1999), Harris corners (Harris and Stephens, 1988), canny edges are usually used for the selection of such regions. The sparse optical flow is preferred in many scenarios since it is more robust against noise.

Optical flow has been used for a variety of navigation applications. The major group is for mobile robot. In 1993 Santos-Victor and his colleagues developed an optical-flow based system inspired by the behaviour of bees. Antonis et al. (2004) presented a robotic centering behaviour based on the exploitation of a panoramic camera. Optical flow has also been used in UAV and other aerial based applications (e.g. Stefan et al., 2005). More detailed description on the optical flow techniques can be found in Barron et al. (1994). In this research, greater emphasis has been placed on map-based approach. Therefore in the following context, vision-based navigation normally refers to the methods depending on prior knowledge of the environments.

1.3 IMAGE MATCHING FOR VISION-BASED NAVIGATION SYSTEMS

Image matching techniques have been used in a variety of applications, such as 3D modelling, image stitching, motion tracking, object recognition and vision based localization. Over the past few years, many different methods have been developed, which can be generally classified into two groups: area-based matching (intensity based, like cross-correlation and least-squares matching (Gruen, 1985)) and feature-based matching, e.g. SIFT (Lowe, 2004).

Area-based methods (Trucco and Verri, 1998) directly work on image intensity values, Area-based image matching usually adopts a two-step approach: first a cost function (error metric) is chosen to evaluate the similarity between candidate corresponding areas on two matching images; then a search function is used to find such corresponding scene from two images. The error metric can be intensity differences or the cross-correlation value. The former approach is trying to find the location with minimum intensity difference. Typical metric includes sum of squared differences (SSD) and sum of absolute differences (SAD). And the latter approach is to locate (pixel) positions that reach maximum similarity. After the goal has been set for certain cost function, suitable search techniques need to be used to find the best match. A straightforward way is to exhaustively search every pixel on the reference image. Considering the computation load, a coarse-to-fine strategy based on image pyramids has been proposed to accelerate the search process. Area-based methods thus may be comparatively more accurate because they take into account a whole neighbourhood around the imagine points being analysed to establish correspondences.

Feature based methods on the other hand uses symbolic descriptions of the images that contain certain local image information to establish correspondence. The general scheme usually involves three important steps. The first step is to use interest operators to extract salient features from both images, which can be in the form of points, corners, edges or regions. The first operator algorithm was developed by Hans. P. Moravec in 1977. It is followed by various approaches, among which the most widely adopted algorithms are Forstner operator (Forstner and Gulch, 1987), the Harris operator (Harris and Stephens, 1988), and Lowe‘s SIFT (Lowe, 1999). As a result of applying an interest operator, an unstructured list of image points with associated attributes is generated for each image. More specifically, the second stage is the construction of feature descriptors around the salient points using mechanisms that aim to keep the region‘s characteristics insensitive to viewpoint and illumination changes (Alhwarin et al., 2008). Then the final step is to find correspondences based on these attributes (feature descriptors). While a full search of features from one image against all the features on the matching image is straightforward, a more efficient approach is to use an indexing scheme to find the nearest neighbour in high-dimension space.

No single algorithm, however, has been universally regarded as optimal for all applications since they all have their pros and cons. Since image matching is regarded as the most important component for visual systems, in this research we first briefly explore the performance of various image matching techniques according to the specific needs of vision-based positioning systems, and we discuss the systems according to their choice of image matching methods.

For stereo vision based approaches, stereo matching is employed to create a depth map (i.e. disparity map) for navigation. It belongs to map-based approach. Usually area based algorithms are used to solve the stereo correspondence problem for every single pixel in the image. Therefore, these algorithms result in dense depth maps as the depth is known for each pixel (Kuhl, 2004). Typical methods include Census (Zabidh and Woodfill, 1994), SAD (Sum of Absolute Differences), and SSD (Sum of Squared Differences). The common drawback is that they are computational demanding. To deal with the problem, some efforts have been made. In Nalpantidis et al. (2009) the authors proposed a quad-camera based system which used a custom tailored correspondence algorithm to keep the computation load within reasonable limits. Meanwhile, feature based methods are less error sensitive and require less work load. But the resulting maps will be less detailed as the depth is not calculated for every pixel (Kuhl, 2004). Therefore, how to achieve a disparity map which is both dense, accurate while the system maintains reasonable refresh rate is the cornerstone of its success, and still remains to be an open question.

For monocular vision sensor, the two approaches: 1) ones that use the structure from motion (SFM) method, which is a mapless approach, and 2) map-based monocular visual navigation systems, also differ from each other. For the former group, consecutive frames present a very small parallax and small camera displacement. Given the location of a feature in one frame, a common strategy for SFM is to use feature tracker to find its correspondence in the consecutive image frame. Kanade-Lucas-Tomasi (KLT) tracker (Zhang et al., 2010) is widely used for small baseline matching. The methodology for a feature tracker to track interest points through image sequence is usually based on a combined use of feature-based and area-based image matching methods. First interest points are extracted by operators from the first image, such as (Lowe, 2004; Harris and Stephens, 1988; Forstner, 1986). Then due to the very short baseline, positions of corresponding interest points in the second image are predicted and matched with cross-correlation, which can be further refined using least squares matching. Some approaches perform outlier rejection based either on epipolar geometry (Remondino and Ressl, 2006) or RANSAC (Zhang et al., 2010) for the last step.

For monocular vision map-based navigation systems, one significant nature that differs them from other visual systems is that at positioning (second) stage the real time query image might be taken at substantially different viewpoint, distance or difference illumination conditions from the map images in the database, or using different optical devices. In other words, the two matching images may have a very large baseline, large scale difference and big perspective effects, which lead to a wide range of image transformation. Due to such significant changes, most image corresponding algorithms working well for short baseline (e.g. stereo, or video sequence) images will fail in this case. For area-based approach, cross correlation method can't get a good performance when rotation is greater than 20° or scale difference is greater than 30% (Lang and Forstner, 1995); an iterative search for least squares matching (LSM) will require a good initial guess of the two corresponding locations, which is not applicable in situations where image transformation parameters are unknown. Feature based algorithms on the other hand prove to be more robust against scene movement and potentially faster. Therefore, a feature-based approach suits the monocular vision-based navigation systems better.

The development of image matching of using keypoints can be traced from Moravec‘s work in 1977. However, early matching methods based on corner detectors (Harris and Stephens, 1988) would fail because of the big perspective effects (Remondino and Ressl, 2006). Therefore, more distinctive and invariant features are needed. The first work for invariant feature was by Schmid and Mohr (1997) who used a jet of Gaussian derivatives to form a rotationally invariant descriptor around a Harris corner. Then a significant contribution to the field of feature-based matching is made by Lowe for his method: SIFT (Scale Invariant Feature Transformation). It is first introduced in 1999 ICCV (Lowe et al., 1999) with some information on its application to object recognition. From then on, it has been applied to a wide range of applications, such as object recognition, pose estimation, image retrieval and so forth. Particularly, SIFT has been used in image matching systems (Se et al., 2002; Boris et al., 2008) for accurate location estimation. The major virtue is that SIFT features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise and change in illumination (Lowe, 2004). Mikolajczyk and Schmid (2005) compared the performances of several local descriptors and their experiments revealed that the SIFT yielded the best matching results. Therefore, for map-based navigation systems, SIFT is one of the best choices. We have mainly chosen SIFT for image matching in our system.

The algorithm bundles both a feature detector and a feature descriptor. The major steps to generate SIFT features are as follows:

1) Scale-space extreme detection:

   Firstly, scale space is created to ensure scale invariance. Traditionally, Laplacian of Gaussian (LoG) is used to find interesting feature points. Instead of using Laplacian of Gaussian (LoG) here, the Difference of Gaussians (DoG) is used to help find potential interest feature points for SIFT. The benefit of DoG over LoG is that it retains scale invariance and computationally less intensive. More specifically, the maxima and minima in DoG images are located as candidate feature regions.

2) Keypoint localization:

   In the second step, we obtain subpixel locations for the candidate features to increase the chances of matching and stability. Meanwhile, some candidates are rejected if they don‘t have enough contrast or lie on an edge.

3) Orientation assignment:

In the third step, orientation is assigned to each selected key point to provide the nature of rotation invariance. The basic idea is to collect gradient directions and magnitudes in the small region around each key point. Then the most prominent orientation(s) is assigned to the feature point.

4) Keypoint descriptor:

The final step is to generate a unique signature for each key point to prepare for matching. The signature is a vector (128 dimensions by default) that describesthe local image gradients of the key point region.

In summary, the detector extracts a number of regions from the images, then a descriptor is associated with each of the region, which contains properties that describe the appearance of the region. As shown in Figure 1.2, 1195 keypoints are found in this image. They are displayed as vectors indicating scale, orientation and location.

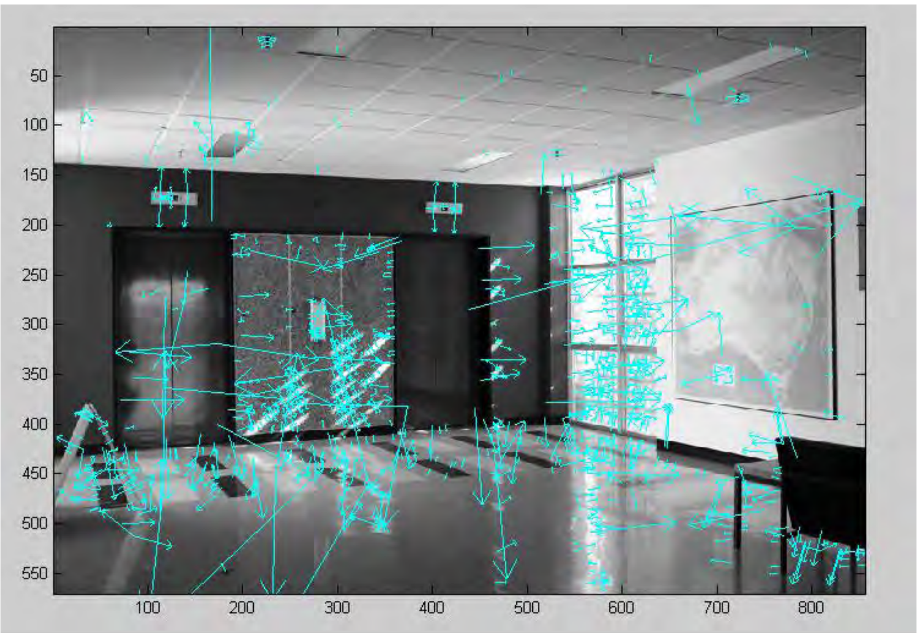


Figure 1. 2 SIFT Feature extraction

In the next step, matching is performed based on the feature extracted. A new image is matched by individually comparing each feature from the new image to the previous image and finding candidate matching features based on Euclidean distance of their feature vectors. In fact, the best candidate is the nearest neighbour from the reference descriptor vector. The determination of whether it is a correct match depends on a probability calculated by taking the ratio of distance from the closest neighbour to the distance of the second closest. Lowe (2004) rejected all matches in which the distance ratio is greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches. Here we set the ratio to 0.6. Figure 1.3 illustrates one of the matching between two images, which have 67 matched keypoints. Then the image coordinates of these matched pairs are stored in the database. It should be noted that a number of mismatches are generated during the process, which need to be taken care at later stage (outlier detection function). The problem will be addressed in later chapters.

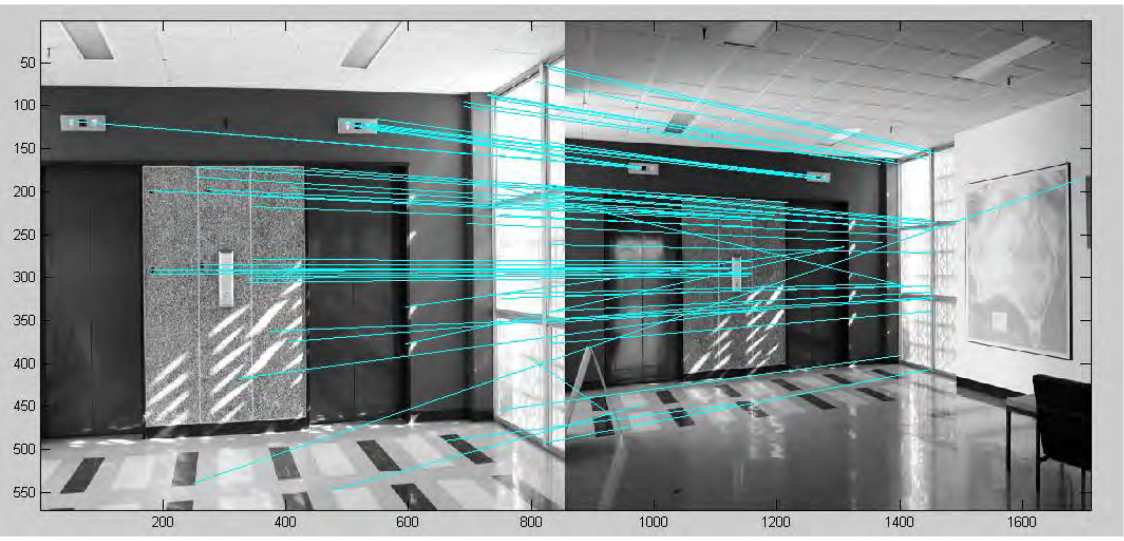


Figure 1. 3 SIFT based image matching

1.4 CHALLENGING ISSUES AND OBJECTIVES

Positioning and navigation applications have never been so accessible. Some of the most important positioning or localization techniques include satellite-based positioning (e.g. GPS), dead reckoning, beacon-based positioning and so forth. There are however, many unmet needs, especially in the area of indoor navigation. Nowadays, most localization services can only be provided for outdoor environments with an adequate GNSS signal availability. In urban and indoor environments the signal may be degraded due to various reasons, such as shadowing, signal attenuation, multipath, intentional denial or deception, and so forth. In order to strengthen and extend the positioning capabilities to provide a more robust navigation solution, alternative positioning techniques and navigation schemes are sought for.

In the research domain of indoor navigation or more generally, navigation in a GPS-degraded environment, vision is regarded to be highly promising because of its ubiquitous and self-contained nature. Given the rich literature available, most existing vision-based navigation systems depend on the exploitation of one or more cameras, either map-based or mapless navigation is adopted. However, available vision-based approaches are still far from mature to supplement GNSS. Major challenging issues versus current limitations are identified in four aspects: mapping, poisoning accuracy, reliability, and coverage.

1) Mapping

   Mapping and navigation are intrinsically coupled question. As has been reviewed in previous section, map can be in various forms, from 2D to 3D, from models to images. However, current digital maps as well as maps developed in the navigation research are still limited to certain applications. The next generation of navigational maps require richer information to be provided and able to support various location-based services. Therefore in this research, one objective is to develop a 3D map, which is capable of providing reality-based view and accurate geometric information of the environment for navigation. Moreover, it can be extended to support other location-based services, such as 3D geometry measurement of the landscapes.

2, 3) Accuracy and reliability

In the navigation community, the accuracy and reliability of the system are regarded as two priorities. For vision sensors, direct depth measurement is lost. The real world geometry information can only be indirectly obtained through photometric effects from images, which means a mathematical model needs to be built. Meanwhile, vision sensors have a very high input data rate and inherently fragile against error. Therefore, both systematic error and gross errors can easily affect the system performance. This will pose great challenge for the vision-based positioning function in terms of accuracy and reliability.

Most of the available research on vision-based navigation came from the mobile robotic field. The techniques adopted are mostly based on algorithms developed by the computer vision (CV) community. CV makes certain contributions to the field, such as object recognition, structure from motion techniques, 3D modelling and so forth. However, the major limitation lies in that it focuses on fast, preferably linear techniques, which is insufficient to provide accurate estimation. In its application to vision-based navigation, much emphasis has been placed on enabling a robot to safely and effectively navigate in an indoor environment with a high level of autonomy. However, as long as the navigation performs without failure (hitting any obstacle or unable to follow the pre-defined path), self-localization process is considered as satisfactory. The accuracy and reliability aspects of positioning have hardly been paid much attention or fully investigated. On the other hand, photogrammetry has been developed to obtain the best possible accuracy with certain imaging networks. It has also developed quality control measures to strengthen the reliability of the outcome. Photogrammetry has been widely applied to applications in mapping, industrial metrology, archaeological surveying. But its application in vision-based navigation has been rarely to be found.

Therefore in this research, we aim to introduce photogrammetric methods into vision-based navigation. The basic idea is to jointly use technologies from the two fields to complement each other. More specifically, image matching algorithms from CV are used for feature recognition, and photogrammetric mapping is used to develop the 3D map, space resection is adopted for positioning. The objective is to improve the positioning accuracy.

Meanwhile, quality control measures from both communities are adopted for the vision-based navigation system. The objective is to develop a dedicated outlier detection mechanism to strengthen the robustness of the system.

4) Coverage

Ubiquitous positioning is considered to be a highly demanding application for today‘s Location-Based Services (LBS). However, the significant difference between outdoor and indoor environments has divided the early stage of the research into two different groups: outdoor and indoor visual navigation. As a matter of fact, vision sensor is able to function in both environments. Therefore, it is high time that a consistent framework of vision-based navigation technology to be developed, which is capable of filling the gap in satellite-based system deficiencies, providing coverage from outdoors, urban canyons to indoor environments.

In summary, the objective of this research is to develop a new system that is able to address these problems and bring the technology forward.

1.5 CONTRIBUTIONS OF THIS RESEARCH

This research is focused on developing a new vision-based navigation system for positioning and navigation in GPS degraded environments. The main research contributions are summarized as follows:

a. A new concept of 3D map has been introduced. The new 3D map mainly consists of geo-referenced images, and features in both reality-based visualization of the environment and 3D geo-referenced geometric information. In this research, it provides the map-matching function for vision-based positioning. Its development process and applications have been discussed.

b. A method of vision-based positioning with use of photogrammetric methodologies and computer vision techniques has been proposed. It mainly obtains geometric information of the navigation environment from the 3D map and uses photogrammetric methods to solve the position. More specifically, a least squares based space resection is used to solve the position and orientation in 6 degree of freedom at high accuracy. Both function model and stochastic model have been built. SIFT features have been used as 3D landmarks during map matching, and served as pseudo ground control points (PGCPs) for positioning resolution. The algorithms have been implemented and tested in an indoor environment. The accuracy has reached around 10 cm.

c. Vision sensor is inherently fragile against errors. Therefore, any vision-based system requires a robust quality control mechanism to ensure good performance. In this research, a multi-level outlier detection scheme for the vision-based navigation system has been proposed. It mainly combines RANSAC, which deals with high percentage of mismatches, with data snooping, which removes a small number of outliers at the final adjustments for both 3D mapping and positioning resolution.

d. The deficiency of using RANSAC for outlier detection in image matching and homography estimation has been identified. In this research, a novel method which combines cross correlation with feature based image matching has been proposed. It is able to effectively evaluate the RANSAC homography estimation, detect poor ones and improve the image matching performance. The method has been successfully applied to the vision-based navigation solution to find corresponding view with the query image from the database and improve the final positioning accuracy.

e. Image matching has been the essential component for visual systems. However, given the rich literature on image matching, there‘s still lack of analysis on image matching in the context of vision-based navigation systems, especially for a map-based approach. In this research, factors that influence the positioning performance of the system have been evaluated through the mathematical model and experiments. The focus has been on various image matching conditions/methods and their impact on the geometry of PGCPs. The characteristics, including both strength and weaknesses of the system, have been revealed and investigated. Multi-image matching has been introduced into the 3D mapping procedure, and ASIFT has been used to deal with dramatic viewpoint changes.

f. In recent years the low cost built-in sensors on mobile devices (e.g. smartphone), especially high resolution cameras have placed greater demand for a breakthrough in their applications for seamless positioning. In the later stage of research, the vision-based navigation system has been extended from indoor to outdoor with corresponding changes been made to cater for outdoor environments. It mainly uses visual input to match with geo-referenced images for positioning resolution, and takes advantage of multiple sensors onboard, including GPS receiver and a digital compass to assist visual methods in various aspects. Experiments demonstrate that such system can largely improve the position accuracy in areas where stand-alone GPS (SPP) is affected and can be easily adopted on mobile devices.

1.6 THESIS OUTLINE

This thesis consists of eight chapters. The contents of each chapter are outlined as follows.

Chapter 1 first gives a general overview of navigation technologies, including basic principles from position fixing and dead reckoning (DR) methods. Then the literature of vision-based navigation is reviewed and divided into two categories in this research: map-based and maples approach. Image matching algorithms are introduced in the context of vision-based navigation system.

Chapter 2 first provides an overview of 3D mapping methodologies with the emphasis on image-based methods. Then the newly defined 3D map is introduced with its development process. Geo-referencing procedure has been the essential step for the mapping process. Experiments focused on geo-referencing are presented.

Chapter 3 introduces the methodology of the vision-based positioning resolution, including both mathematical model and implementation procedures. The algorithms have been implemented with results given in the experiment.

Chapter 4 investigates different outlier detection strategies in the context of vision-based navigation. The multi-level outlier detection scheme proposed in this research is introduced. Experiments have revealed the nature of the outliers in the system and proved the efficiency of the outlier detection scheme.

Chapter 5 introduces the enhanced RANSAC homography estimation method proposed in this research. It integrates the cross-correlation information into feature based RANSAC combined image matching. Experiments prove that it can effectively mitigate for the random nature of RANSAC and identify the poor RANSAC estimation, so as to improve the performance of image matching. It largely improves the positioning accuracy of the system by optimizing the image matching procedure.

Chapter 6 evaluates the performance of the vision-based navigation system through experiments with varying real world conditions and simulations. Factors that influence system performance are investigated. Two major components that determine positioning accuracy, geometry of PGCP and measurement accuracy are identified and discussed.

Chapter 7 presents a comprehensive system that adopted hybrid vision-based method with combined use of onboard sensors (GPS, camera and digital compass) to achieve a seamless positioning from indoor to outdoor environments. It mainly extends the previous approach to outdoor environment. The system adopts the same strategy: geo-referenced images are used as 3D maps for vision-based positioning. Due to the difference between two environments, corresponding changes to the algorithms are introduced.

Chapter 8 summaries the contributions of this research, draw conclusions and makes recommendations for future research.

CHAPTER 2
3D MAP DEVELOPMENT BASED ON GEO-REFERENCED IMAGES

2.1 INTRODUCTION

Three-dimensional geo-information has become an important subject within the GIS community for many years. Research mostly has been concentrated on aspects such as 3D data collection and modelling, data management (e.g. topological, geometrical models), 3D data analysis and visualization (e.g. virtual reality, augmented reality, etc.). The target application has mostly been 3D modelling of landscapes, urban and city models. As three-dimensional geo-information technique becomes so ubiquitous and shows good potential in navigation applications, researchers try to include 3D representations of the environment via different approaches for navigation purposes.

The 3D map for navigation is optimally both realistic and geometrically accurate. Such a character differs the requirement and procedure of its development from other 3D applications that mainly used for visualization, such as ones in the movie industry. Moreover, compared with 3D modelling of specific targets, such as heritage documentation, the 3D map for navigation usually do not require the acquisition of full detailed information of the target but need to cover greater areas.

2.1.1 Range-based and image-based 3D mapping/modelling

Currently, there are two main stream 3D modelling/mapping strategies: range-based and image-based modelling/mapping. The former one normally uses laser/sonar scanner to directly produce 3D point cloud and depth information in aerial and terrestrial mapping. It is able to provide highly detailed and accurate representation of the target object. However, most of the systems only focused on the acquisition of the 3D geometry, providing a monochrome intensity value for each range value. Few systems attached a colour camera to the instrument so that the acquired texture is registered with the geometry (Remondino, 2006). But the difference requirement for imaging and scanning also poses challenge for such an act. The procedure for range-based 3D mapping is shown in Figure 2.1.

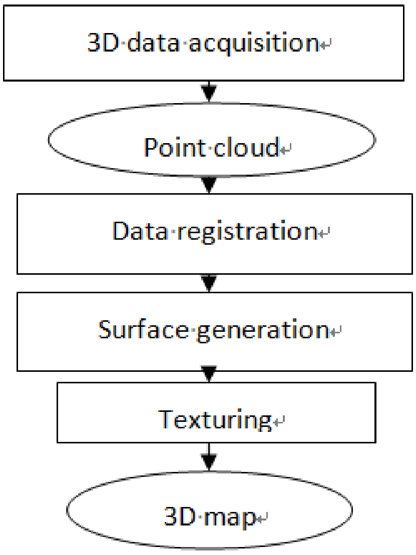


Figure 2. 1 Range-based 3D mapping

For image-based 3D mapping, a mathematical model need to be built in order to derive the object coordinates (3D geometry). Compare the two approaches, the quick and direct solution of a laser scanner may be easily assumed to be superior to image-based methods. As a matter of fact, it remains to be a bulky instrument and suffers from the loss of semantic and colour information. Image-based methods, on the other hand, can provide an economical and efficient alternative with context and geometric information (Aguilera and Lahoz, 2006). Therefore in our approach, an image-based mapping method is adopted.

2.1.2 Image-based 3D mapping procedure

Image-based 3D mapping often involves three major steps: image collection, image matching and derivation of 3D geometric information for 3D reconstruction.

2.1.2.1 Image collection and matching

In order to extract 3D information from images for mapping, corresponding points from different views need to be found. Usually, a stereo vision sensor can capture simultaneously several images of the same scene on a rigid frame at slightly different locations. This is also called passive stereo. It allows the recovery of depth (or range) information based on the triangulation of the matching features in the stereo image pair. The disparities can be used to compute the relative positions of the landmarks in the images through triangulation. When using a stereo vision to produce image data, one of the most essential elements is the length of the baseline. It will exert a big influence on the resolution of depth estimation. As has been mentioned in Chapter 1, Kidono and his colleagues (2002) used stereo vision to obtain range data to generate 3D map and utilized the map and observation to realize safe and efficient navigation. Sabe et al. (2004) developed a humanoid robot named QRIO. Their approach is based on plane extraction from data captured by a stereo-vision system that has been developed for QRIO. Using a 5cm baseline, the error of the depth measurement at a distance of 1.5 meter is over 80mm. The depth estimates of objects with the distances more than 2 meters are omitted. So the depth (range) measurement is limited by the baseline. As a result, many efforts use stereo vision to extract 3D information have been limited to indoor positioning applications.

By combining the information from multiple views, a single camera can be used to generate 3D data. Huang et al. (2005) described a direct method (in the sense it does not use an iterative search) based on vision for localizing a mobile robot in an environment with only two observations along a linear trajectory. Most importantly, it demonstrates that when a robot moving straight, the estimation of a landmark range from a monocular vision can be abstained, and is actually very similar to the technique used in a stereo vision. Such a technique is often named shape from motion (SFM). Generally, it refers to the process of building a 3D map of a static scene by a moving camera. With regard to the SFM methodology itself, Huang and Netravali (1994) present a review of algorithms and their performance for determining three-dimensional (3D) motion and structure of rigid objects when their corresponding features are known at different times or are viewed by different cameras. The idea is basically similar with stereo vision in that the 3D map is built from two images of the same landmark. In both cases, the same landmark is shot in several images, and the disparities of these images are used to compute the 3D information of the landmark. The only difference is that in stereo vision, images are taken simultaneously based on a rigid frame, while in SFM, images are taken at different time steps. It should be noted that SFM overcome the limitation of stereo vision by having a flexible baseline.

An omnidirectional camera can also be used to collect image data. It provides a 360 degree panoramic view of the environment by either using dioptric fish-eye lenses, or catadioptric systems which combine cameras and mirrors. But it has the disadvantage of low image resolution, which makes it unsuitable to produce a 3D map for navigation purposes.

No matter what kind of cameras is used for image collection, image matching is essential for 3D mapping. As has been discussed in Chapter 1, among the rich availability of image matching methods, a feature-based approach suits the monocular vision-based navigation systems better. In two dimensional intensity images, which are the most commonly used in photogrammetry, the automatic extraction and identification of features, such as targets, is one of the initial steps required to determine the three dimensional coordinates of points using a multitude of images. Once features are extracted, corresponding features can be found using image matching techniques. Provided the information of control points and corresponding (tie) points, the 3D object coordinates of these tie points in the overlapping areas of images can be determined.

2.1.2.2 3D mapping/modelling

Both the field of photogrammetry and computer vision have contributed to 3D mapping/modelling. In the field of computer vision, the relation between 3D objects and 2D images is always expressed with central projection model, but a linear representation is used and achieved by means of projective geometry (Jazayeri, 2010). Most of the research has been focused on 3D object reconstruction and robotic applications. Unlike photogrammetry, the accuracy of the 3D geometry obtained is not the priority. In robotic navigation for instance, much emphasis has been placed on enabling a robot to safely and effectively navigate in an indoor environment with a high level of autonomy. For our research, however, the positioning accuracy is considered as a very important aspect, thus the geometric accuracy of the map is of significant importance.

In the field of photogrammetry, one of the main tasks is to reconstruct precisely certain object or region in 3D given a set of images. The theory behind is photogrammetric geo-referencing. It is used to establish relationship between images and object coordinate systems. Direct geo-referencing and indirect geo-referencing approaches are both adopted for various mapping systems. The former approach collects data from multi-sensors platform which includes signal-synchronously integrated GPS, IMU and vision sensor. It is an efficient and direct process if the coupled sensors could work together cooperatively. The latter one needs ground control points to compute the exterior parameters of camera, which can yield high quality geo-referenced data but time consuming. In the following context, we briefly introduce both efforts with emphasis on the one we adopted for current research.

In outdoor environment, especially in applications like UAV and mobile mapping, direct geo-referencing is used. Normally GPS and IMU data is used to obtain the position and orientation information of the platform and then space intersection is used to calculate the object coordinates. The flowchart for direct geo-referencing procedure is shown in Figure 2.2. Camera integrated with GPS is an important step for direct geo-referencing, which means the position of camera perspective centre can be directly collected from GPS. However, the time reference between vision sensor and GPS is quite different. To integrate GPS and camera, possible solution is to add a time stamp on each frame of images from CCD cameras, and an external trigger is implemented for synchronizing the camera with computer through National Marine Electronics Association (NMEA) GPS messages, or image time stamp synchronized to receiver via pulse per second (PPS).

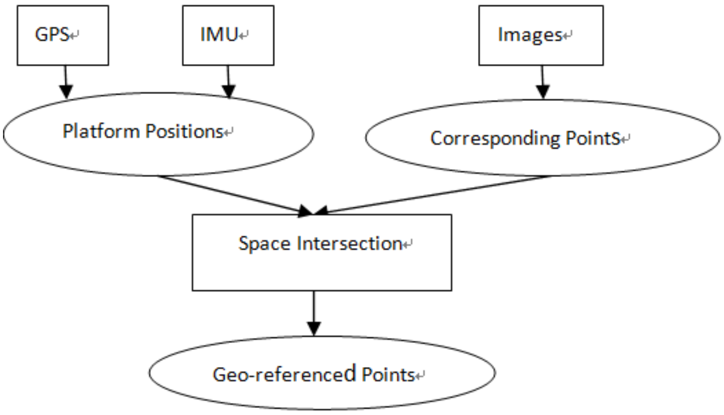


Figure 2. 2 Flowchart for 3D map direct geo-referencing

Under circumstances that GPS signal, or GPS offset and INS drift angles are not available, direct geo-referencing cannot be applied. In case of imaging by the conventional metric/non-metric cameras located on the terrestrial stable stands, indirect geo-referencing of data is usually executed in the post-processing stage (Bujakiewicz et al., 2011).

For indirect geo-referencing, the main idea is to locate an object point in three dimensions through photogrammetric methods, which uses central projection imaging as fundamental mathematical model. In aerial photogrammetry, it makes satellite and aerial as well as terrestrial imagery useful for mapping. The main function for indirect geo-referencing is called bundle triangulation, or bundle block adjustment. It means the simultaneous least squares adjustment of all bundles from all exposure stations, which implicitly includes the simultaneous recovery of the exterior orientation elements of all photographs and the positions of the object points (Faig, 1985). The technique was developed in the very early stage of the field. Tewinkel published his work on ―future of analytical aerial triangulation" in 1958. Then the method of bundle adjustment was introduced to close range applications (Brown, 1976). However, the two have certain differences in terms of camera networks, structure of normal system of equations, camera type and so forth. Today in close-range photogrammetry, bundle adjustment enables the production of accurate as-built measurements and 3D reconstruction. Examples of 3D geometry recovery can be found mostly in building documentation (e.g. Lisowska, 2007), traffic accident reconstruction, engineering measurement and so forth. For mapping and navigation, such an approach has rarely been found. The disconnection of two fields: mobile robotics and photogrammetry, has result in the fact that photogrammetric approach has rarely been introduced to 3D mapping for navigation purposes. However in this research, we adopt photogrammetric methods for both 3D mapping and navigation to increase the accuracy and reliability of the system. The flowchart of the bundle adjustment process we used is shown in Figure 2.3.

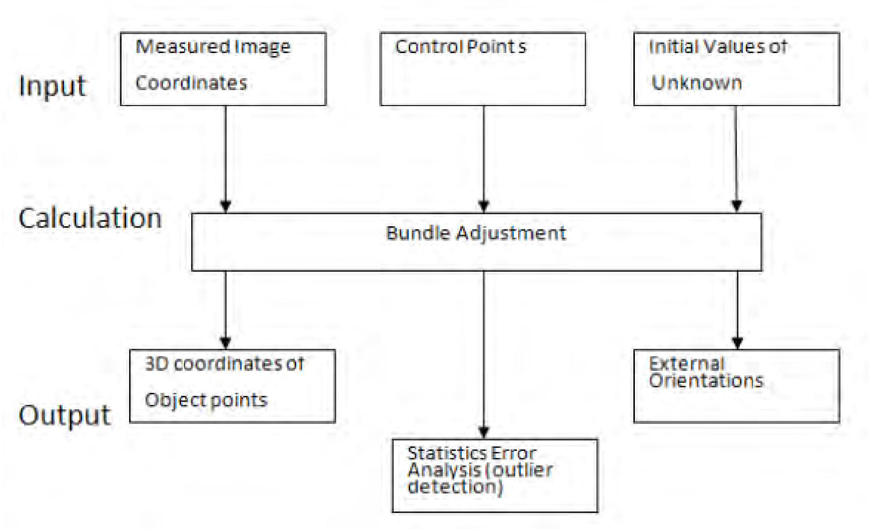

Input
Measured Image Coordinates
Control Points
Initial Values of Unknown
Calculation
Bundle Adjustment
Output
3D coordinates of Object points
External Orientations
Statistics Error Analysis (outlier detection)

Figure 2. 3 Flowchart of bundle adjustment process for 3D map indirect geo-referencing

2.2 NEWLY PROPOSED REALITY-BASED 3D MAP

Most 3D maps we see today are more emphasized on the aspect that the map can be visualized in 3D, disregard its accuracy in terms of geometry. The newly proposed 3D map in this research on the other hand pays greater attention to the 3D information contained by the map. More specifically, feature points are extracted from the map images and their 3D spatial information are obtained in a geo-referenced coordinate system. Therefore the major function of the map depends on the geo-referenced 3D point cloud. Since the map is originated and produced both in the form of images, it also has a photo-realistic nature.

The reality-based 3D map is defined as a sum of geo-referenced points with three dimensional (3D) local or global coordinates that are overlapped on images of the environment. Users of the 3D map will have the benefits of geo-referencing with 3D coordinates as well as realistic visualization (Figure 2.4). Since it contains 3D information and uses images as maps, we give it the name ~~-r~~eality-based 3D map". One basic function of the 3D map is for positioning and navigation. Whenever a new image is taken, it can be matched with the images stored in the 3D map database and therefore enables the user to locate its position. The main difference between our approach to the available image-based navigation methods lies in the fact that the map images are geo-referenced, which means they themselves can give absolute position information (local or global) in 3D, functioning like a sensor (eg.GPS), and at the same time can be used as a map for location-based services.

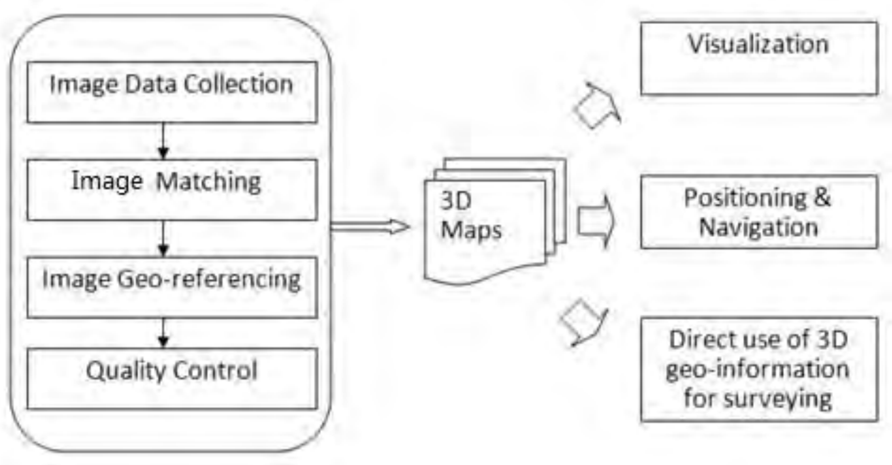


Figure 2. 4 Newly defined reality-based 3D map

The main purpose of photogrammetry is to achieve the 3D world coordinate from flat 2D images and reconstruct objects in 3D in digital form or graphical form (images, maps, etc). Image geo-referencing has been one of the major processes, which matches features in the image to real world coordinates on the ground. In this research, an essential part of the mapping process is image geo-referencing. We produce the geo-referenced images by obtaining their feature information in 3D and overlapping the 3D point cloud to the original images.And an indirect geo-referencing method has been adopted. The 3D mapping process mainly consists of 3 steps: firstly, images of the navigational environment are collected and ground control points are set in the navigation environment; secondly, image matching between images with overlapped areas is performed to extract feature points; thirdly, these feature points on the map images are geo-referenced through photogrammetric bundle adjustment (indirect geo-referencing). The quality of the map depends on the accuracy of geo-referencing. Detailed procedure is shown with flowchart in Figure 2.5.

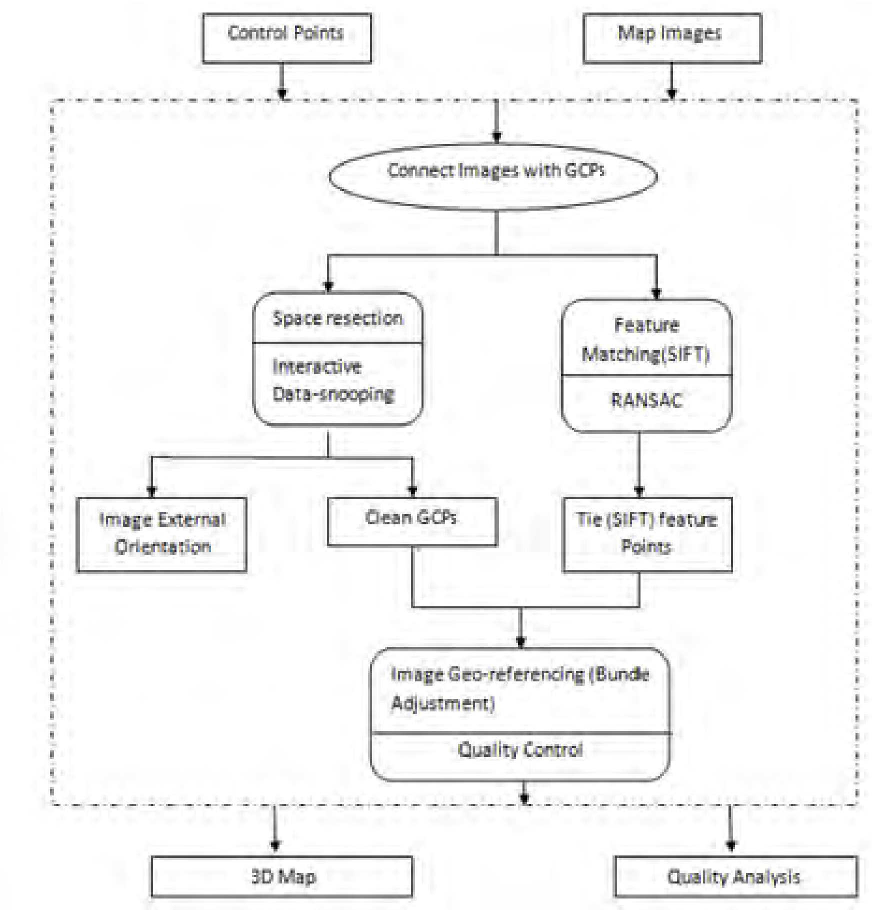


Figure 2. 5 Flowchart of 3D mapping procedure

The final 3D map is presented in the form of geo-referenced images, including images and 3D feature point cloud overlapped on the map images. For better illustration, Wu's VisualSfM software (Wu, 2011) is used to visualize the features (in Figure 2.6 and Figure 2.7), but this software has not been used in data processing. Although the resulting map is not in the form of 3D models, which can be visualized in 3D, it more emphasized on the fact that 3D geometric information is contained by the map. And it is the very 3D information from the map enables the vision-based positioning function to proceed.

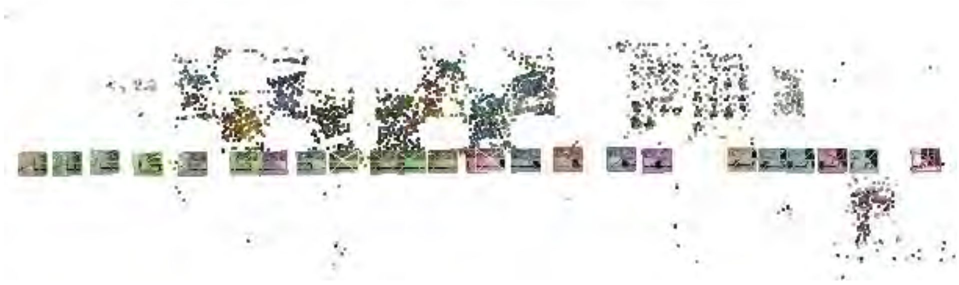


Figure 2. 6 Visualization of SIFT features for geo-referencing (point cloud) produced by Wu‘s VisualSfM software from the reference images, which are shown by square patches.

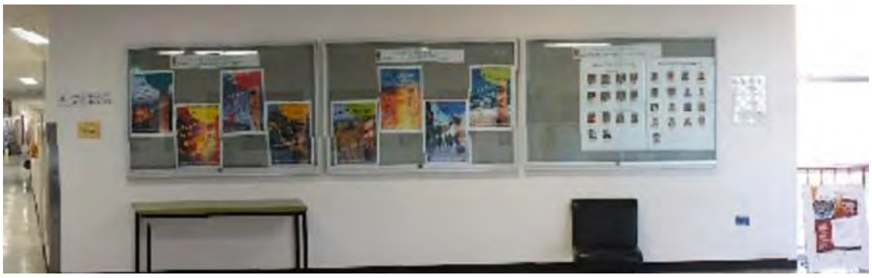


Figure 2. 7 Panorama view of the mapped area in Figure 2.6

2.3 3D MAP DEVELOPMENT

2.3.1 Image collection

At the first stage of mapping, images of the navigational environment are collected using a calibrated camera with fixed focal length. In this research, single camera is used to take images with high percentage of overlapping areas. This is to make sure that the 3D information can be calculated for most of the feature points extracted at later stage. Meanwhile, object information of the environment is provided under a local/global coordinate system through the set up of ground control points (GCPs). It is important that these ground control points are widely and uniformly distributed over the area covered by the images, for the performance of geo-referencing can be improved with better geometry. Then these points can be related to images either manually or automatically. Currently ―Photomodeller" software is utilized (Figure 2.8). Up till now, the raw datasets are ready to be processed. The raw datasets mainly consist of image measurements of ground control points, their surveyed 3D coordinates and images. Here follows an outlier detection process based on iterative data snooping, which will be further explained in Chapter 4. It is mainly used to remove gross errors from control point coordinates. Then these ground control points can be introduced as error-free reference points later in the bundle adjustment process (geo-referencing). Otherwise these errors will be interpreted as errors in observations at bundle adjustment and are difficult to be detected and removed.

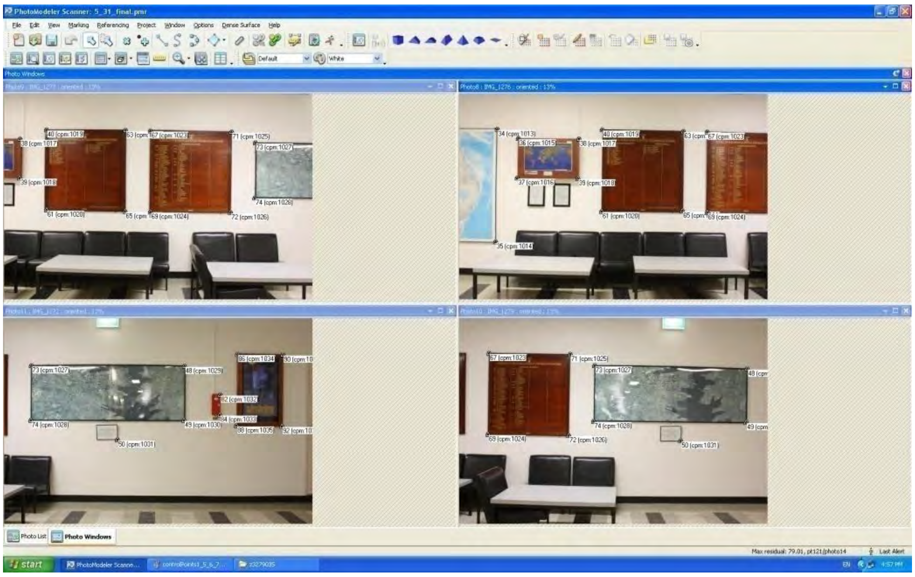


Figure 2. 8 Pre-processing

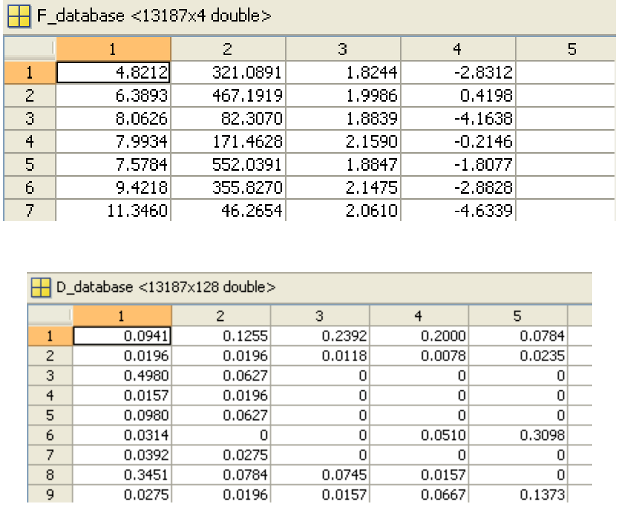
2.3.2 Multi-image matching for 3D mapping

It is noted that for this system, the image matching algorithms used for the 3D mapping and positioning process need to be kept in consistent, otherwise geo-referenced information cannot be transferred from map to real time query images for positioning. Therefore, we mainly adopt methods from feature based group for our system in both 3D mapping and positioning, more specifically SIFT is used in indoor environments.

Traditionally people carry out image matching between every pair of images that have overlapping areas and import the information from each pair to a bundle adjustment for geo-referencing. The major drawback is that the image geometry is weak especially when only neighbouring images are considered. Therefore here we introduce multiple image-matching based on the SIFT algorithm into the system. By definition, it means correspondences are located and matched over multiple images. One obvious benefit is that multi-view constraints are stronger than pair-wise constraints. This allows for more accurate solution for the image geometry and more incorrect matches to be rejected (Brown, 2005). Such an approach has been mostly applied for image stitching to produce panoramic mosaics (e.g. Brown, 2005) but to the authors' knowledge, it has not yet been used for vision-based navigation applications. This research represents a good example of improvement over pair-wise matching by introducing multi-image matching to the mapping process.

The whole process takes 5 steps and here mapping process from one experiment is used as sample data for illustration. A total of 8 map images (No.5 -No.12) are used.

First a feature database is generated. It is a collection of all SIFT features extracted from map images. In the sample data (Figure 2.9), a total of 13187 SIFT features are found. The database in fact consists of two sub-databases: keypoint database (F_database) and descriptor database (D_database). For the keypoint database, each extracted keypoint creates a 4 parameters record, indicating its 2D location relative to the training image, the scale and orientation. Meanwhile, each feature (keypoint) is associated with a descriptor (128-dimension vector) and together to form a descriptor database.



F_database <13187x4 double>

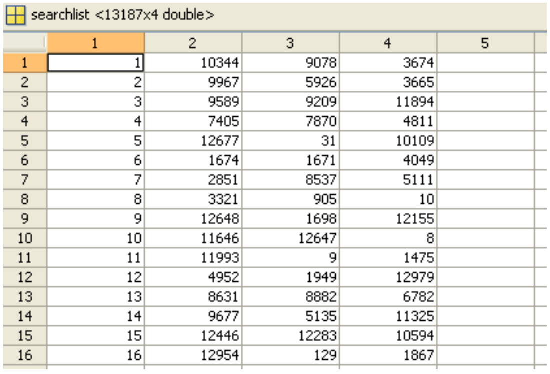
	1	2	3	4	5
1	4.8212	321.0891	1.8244	-2.8312	
2	6.3893	467.1919	1.9986	0.4198	
3	8.0626	82.3070	1.8839	-4.1638	
4	7.9934	171.4628	2.1590	-0.2146	
5	7.5784	552.0391	1.8847	-1.8077	
6	9.4218	355.8270	2.1475	-2.8828	
7	11.3460	46.2654	2.0610	-4.6339	

D_database <13187x128 double>

	1	2	3	4	5
1	0.0941	0.1255	0.2392	0.2000	0.0784
2	0.0196	0.0196	0.0118	0.0078	0.0235
3	0.4980	0.0627	0	0	0
4	0.0157	0.0196	0	0	0
5	0.0980	0.0627	0	0	0
6	0.0314	0	0	0.0510	0.3098
7	0.0392	0.0275	0	0	0
8	0.3451	0.0784	0.0745	0.0157	0
9	0.0275	0.0196	0.0157	0.0667	0.1373

Figure 2. 9 Feature database

After the features have been extracted, multi-image matching is performed. Since multiple images may overlap a single ray, each feature may have multiple corresponding matches. SIFT features are described by 128 dimension vectors, therefore Euclidean distance is used to measure the similarity between features. More specifically, a K-NN search function is used to find nearest neighbours for each feature in the database. An exhaustive search is used for K-NN search since it works well on high dimensions. Other algorithm may also be considered such as LSH (Indyk and Motwani, 1998). It is performed on the 128 dimension feature descriptors in the D_database. The main difference from pair-wise comparison is that single feature may have several matches among the dataset. Therefore, in this approach, a 4-NN search is performed to find potential matches for each feature. It is noted that for the feature points that have no correspondence in the database, still some random points will be found as potential candidates. As shown in Figure 2.10, for every feature in the database, its 4 nearest neighbours (including itself) have been found and listed in a row.

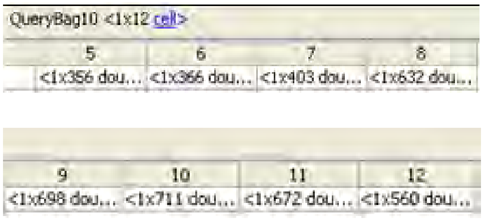


searchlist <13187x4 double>

	1	2	3	4	5
1	1	10344	9078	3674	
2	2	9967	5926	3665	
3	3	9589	9209	11894	
4	4	7405	7870	4811	
5	5	12677	31	10109	
6	6	1674	1671	4049	
7	7	2851	8537	5111	
8	8	3321	905	10	
9	9	12648	1698	12155	
10	10	11646	12647	8	
11	11	11993	9	1475	
12	12	4952	1949	12979	
13	13	8631	8882	6782	
14	14	9677	5135	11325	
15	15	12446	12283	10594	
16	16	12954	129	1867	

Figure 2. 10 A 4-NN search is performed to find potential matches for each feature

The next step is to find images with big overlapped areas that form an image group. In other words, we identify the images that have a large number of matches between each other. A voting strategy is used. First each map image is selected as current query image for once, the features that belong to it are grouped together. Four nearest neighbours found previously associated with each of these features are placed in a ―bag”. Within each bag, each candidate match votes for the image it comes from, therefore the one with biggest number of votes is the image that has biggest number of matching features with current query image. We find 3 images with the highest rank for each current query image. And it can be expected that the map image itself comes first. It is also noted that points with no corresponding feature points or false matches are still involved in the vote, since they are randomly distributed in the images, when the database has large number of features, the final rank will not be affected. In the example (Figure 2.11) map image No.10 is selected as current query image. It can be seen that except itself, map images No. 9 and No. 11 are the images with the biggest number of votes (matches), which are the ones with the largest overlapping area. Map images No. 9 and No.11 together with No. 10 are grouped together as a unit. Using the sample data, a total 8 groups are generated with a few redundant groups (e.g. <5,6,7> and <6,5,7>), which are deleted afterwards.



QueryBag10 <1x12 cell>

5	6	7	8
<1x356 dou...	<1x366 dou...	<1x403 dou...	<1x632 dou...

9	10	11	12
<1x698 dou...	<1x711 dou...	<1x672 dou...	<1x560 dou...

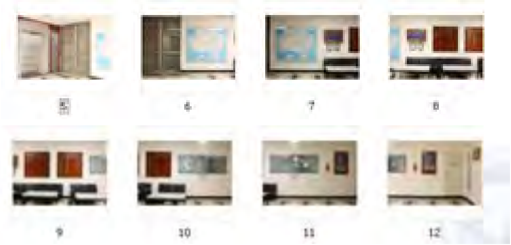


Figure 2. 11 Using voting strategy to find images that overlapped with map image No.10

After image groups have been generated, geometric constraint is used to remove mismatches among the dataset. Within each image group (consisted of 3 images with overlapping areas), RANSAC (Fischler and Bolles, 1981) is used pair-wise. It is a robust estimation method that use feature correspondence to compute homography between a pair of images and remove the matches that do not agree to the geometric model.

Now from each image group we got a bunch of SIFT feature points, and each SIFT feature point actually has at least one counterpart (corresponding point) on another image and their image coordinates are known. Next, we use these SIFT feature points as tie points and put them into bundle adjustment for geo-referencing.

Finally, bundle adjustment is used to geo-reference common feature points of the map images. Different from the previous attempt which imports only two images into a bundle adjustment, we use a multi-image approach. Bundle adjustment is performed on each image group.

2.3.3 Image geo-referencing

After setting up ground control points and extracting tie points from image matching, bundle adjustment is performed to estimate 3D object coordinates, image orientation parameters together with related statistical information about accuracy and reliability (as shown in Figure 2.3).

The mathematical model of the bundle adjustment is based on the collinearity equations:

$$x - x_0 = -f\frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \tag{2.1}$$

$$y - y_0 = -f\frac{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)}$$

These two equations describe the transformation of object coordinates (X, Y, Z) into corresponding image coordinates (x, y) as functions of the interior parameters $(\mathrm{x}_0, \mathrm{y}_0, f)$, which gives the principle points and focal length, and exterior orientation parameters $(\mathrm{X}_\mathrm{s}, \mathrm{Y}_\mathrm{s}, \mathrm{Z}_{\mathrm{s},}\omega, \upsilon, \kappa)$ of one image, which gives the camera position and orientation.

The collinearity equations, linearised at approximate values, can be used as observation equations for a least squares adjustment according to the Gauss-Markov model. It is the basic principle for bundle adjustment. The approximate values served as initial values for the unknown in bundle adjustment are generated using combined intersection and resection.

For the function model of the adjustment, the original model is used, which means coordinates of GCPs are introduced as error-free reference points and the camera external orientations are not surveyed by other device. Additional information about the object or additional non-photogrammetric measurements are not considered. The reason for it is as follows:

1) In the mapping environment, ground control points can be set easily with a good distribution (widely and uniformly distributed) and stability. The coordinates of the ground control points are surveyed by a total station with high accuracy.
2) The geo-referencing process is off-line post processing. Therefore the camera(s) used is pre-calibrated with known interior parameters.
3) Outlier detection has been applied at every step of the process. Before bundle adjustment, the two groups of input: 3D coordinates of GCPs and image coordinates of tie points have gone through outlier detection processes at space resection and image matching respectively. More details can be found in Chapter 4. Therefore, the coordinates of GCPs are treated as fixed values.

The least squares models are listed as Eq.2.2 and Eq.2.3.

$$\boldsymbol{A_c t} - \boldsymbol{L_1} = \boldsymbol{V_1} \quad , \qquad \boldsymbol{L_1} \sim \left(\boldsymbol{0}, \boldsymbol{\sigma_0^2 P_1}^{-\boldsymbol{1}}\right) \tag{2.2}$$

$$\boldsymbol{A_u t} + \boldsymbol{B_u X_u} - \boldsymbol{L_2} = \boldsymbol{V_2} \quad , \qquad \boldsymbol{L_2} \sim (\boldsymbol{0}, \boldsymbol{\sigma_0^2 P_2}^{-\boldsymbol{1}}) \tag{2.3}$$

in which

•$\boldsymbol{A_c}$ is a $2\text{n(number of control points on all images)} * 6$ matrix containing partial derivatives with respect to the exterior orientation parameters, and $\boldsymbol{t}$ contains the incremental changes to the initial values of external orientation parameters;

• $\boldsymbol{A_u}$ is a $2\text{m(number of tie points on all images)} * 6$ matrix containing partial derivatives with respect to the exterior orientation parameters, and $\boldsymbol{t}$ contains the incremental changes to the initial values of external orientation parameters;

•$\boldsymbol{B_u}$ is a $2\text{m} * 3\text{p}$ (number of tie points) matrix containing the partial derivatives with respect to the three coordinates of the tie points, and $\boldsymbol{X_u}$ contains the incremental changes to the initial values of ground coordinates of tie points;

- $\boldsymbol{L_1}$ is a $2\mathrm{n}(\text{number of control points on all images }) * 1$ matrix, which denotes the first group of observations, image measurement of control points in this case. P_1 is its corresponding weight;

- $\boldsymbol{L_2}$ is a $2\mathrm{m}(\text{number of tie points on all images}) * 1$ matrix, which denotes the second group of observations, image measurement of tie points in this case. P_2 is its corresponding weight;

- $\boldsymbol{V_1}$ and $\boldsymbol{V_2}$ denotes the residual.

During the adjustment process, quality control measures have also been taken to improve the quality and reliability of the 3D map. Greater details can be found in Chapter 4. The output of bundle adjustment geo-referencing process are: 3D coordinates of sparse SIFT feature points, camera orientations of the mapping sites, and the geo-referencing accuracy in terms of the mean standard deviation for the 3D coordinates in each direction (X, Y and Z).

2.3.3.1 DOP values for camera 6DOF

The least squares models are listed as Eq.2.4 and Eq.2.5.

$$\boldsymbol{l} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} \tag{2.4}$$

$$\boldsymbol{D} = \boldsymbol{\sigma}_{\boldsymbol{0}}^{\boldsymbol{2}}\boldsymbol{Q} \tag{2.5}$$

in which Eq. 2.4 denotes the function model, Eq.2.5 the stochastic model and σ_0 the a priori standard deviation of measurements. In Eq. 2.4 $\boldsymbol{l}$ denotes the observation; $\boldsymbol{A}$ denotes the design matrix; $\boldsymbol{x}$ denotes the unknowns; $\boldsymbol{v}$ denotes the residual. In Eq. 2.5 $\boldsymbol{D}$ denotes the variance covariance matrix of observations; $\sigma_0$denotes a priori standard deviation and $\boldsymbol{Q}$ the cofactor matrix. Using this model, the covariance matrix for the estimated unknown parameters $(\boldsymbol{C_x})$can be obtained using Eq.2.6, in which $\boldsymbol{P}$ represents the weight matrix. It is listed as:

$$\boldsymbol{C_x} = \boldsymbol{\sigma}_{\mathbf{0}}^{\mathbf{2}}\left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{A}\right)^{-\mathbf{1}} \tag{2.6}$$

In the GPS community, DOP values are used to represent the effect of satellite geometric distribution on the accuracy of a navigation solution. To evaluate the impact of geometry, the covariance of $\boldsymbol{x}$ will be simplified to:

$$\boldsymbol{C_x} = \boldsymbol{\sigma}_{\mathbf{0}}^{\mathbf{2}}\left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\right)^{-\mathbf{1}} \tag{2.7}$$

In fact, the elements in the trace of the matrix $(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-\mathbf{1}}$ are functions of the geometry only.

Here DOP values are used to evaluate the geometric strength for camera 6DOF. For the function model in space resection, which is used for positioning calculation (Section 3.2.3), the diagonal of the matrix $(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-\mathbf{1}}$ is calculated as:

$$\left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\right)^{-\mathbf{1}} = \begin{pmatrix} G_x^2 & & & & & \\ & G_y^2 & & & & \\ & & G_z^2 & & & \\ & & & G_\omega^2 & & \\ & & & & G_\varphi^2 & \\ & & & & & G_\kappa^2 \end{pmatrix} \tag{2.8}$$

Then we give DOP values for 6DOF, which are calculated as follows:

$$XDOP = G_x \qquad YDOP = G_y \qquad ZDOP = G_z \tag{2.9}$$

$$PDOP = \sqrt{G_x^2 + G_y^2 + G_z^2} \tag{2.10}$$

$$\omega DOP = G_\omega \qquad \varphi DOP = G_\varphi \qquad \kappa DOP = G_\kappa \tag{2.11}$$

$$ADOP = \sqrt{G_\omega^2 + G_\varphi^2 + G_\kappa^2} \tag{2.12}$$

in which the PDOP represents the Position DOP, while the ADOP represents Attitude (Orientation) DOP.

It is noted that for the bundle adjustment, the unknown parameters not only include camera external parameters, but also 3D coordinates of tie points. Therefore, DOP values are obtained from part of the matrix $(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-\mathbf{1}}$.

2.4 EXPERIMENTS

In the experiments, a 3D map was produced using a high resolution camera (Cannon EOS4500) with a focal length of 24.7 mm and image resolution of 4272×2848 pixels. A total of 21 images were collected to include the school hallway as mapping area (80m^2), with approximately 60% overlap on neighboring images. Their SIFT feature points were also extracted and geo-referenced for each map image. Both pair-wise and multi-image matching were used during the mapping process. This 3D map was used for navigation at later stage.

2.4.1 Comparison study of multi-image matching and pair-wise matching for 3D mapping

First, a comparison study was carried out between pair-wise based mapping and multi-image based mapping. Then the two maps produced were used for positioning respectively. Here 8 out of 21 map images (the same 8 map images used as sample data in Section 2.3), which cover one side wall of the school hallway, are used for better illustration.

The summary of the two sets of bundle adjustment results are shown in Table 2.1 and Table 2.2. It can be easily observed that a bigger number of image features are extracted and geo-referenced by the multi-image approach, offering better conditions for vision-based positioning. Meanwhile, it is also worth noticing that more outliers are detected when a bigger number of images are involved in the adjustment. One reason is more tie points are calculated, and another is that multi-view constraints are stronger than pair-wise constraints.

Table 2. 1 Summary of the geo-referencing results based on pair-wise image matching

Number of images	2						
Image ID	5,6	6,7	7,8	8,9	9,10	10,11	11,12
Observations	56	690	262	108	326	56	100
Unknowns	39	516	180	69	237	39	66
Number of detected outliers	0	0	0	0	0	0	0
Number of Geo-referenced Feature Points	9	168	56	19	75	9	18

Table 2. 2 Summary of the geo-referencing results based on multi-image matching

Number of images	3					
Image ID	5,6,7	6,7,8	7,8,9	8,9,10	9,10,11	10,11,12
Observations	854	1130	406	644	412	162
Unknowns	633	822	291	459	300	123
Number of detected outliers	0	1	3	0	1	19
Number of Geo-referenced Feature Points	205	268	91	147	94	35

Secondly, imaging geometry was evaluated and compared by DOP values in 6 degrees of freedom. Since each image can participate in more than one adjustment, they are shown in Figure 2.12 and Figure 2.13 at the same 'image ID' with separate icons. It can be observed that multi-image based approach provides generally smaller DOP values compared with their counterparts, which shows a better imaging geometry.

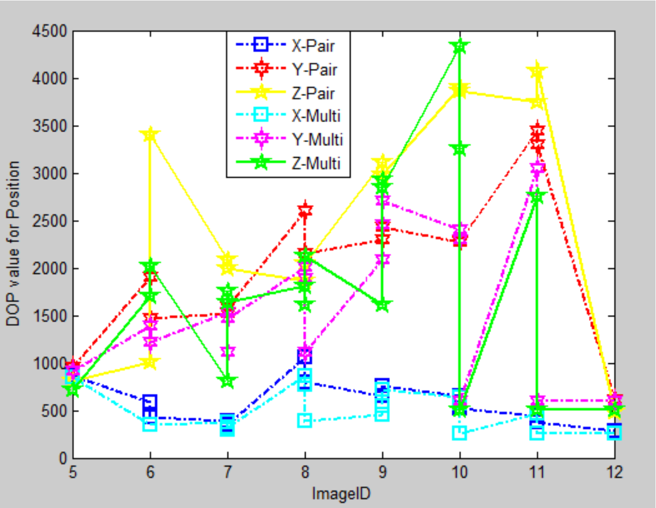


Figure 2. 12 Comparison of position DOP values

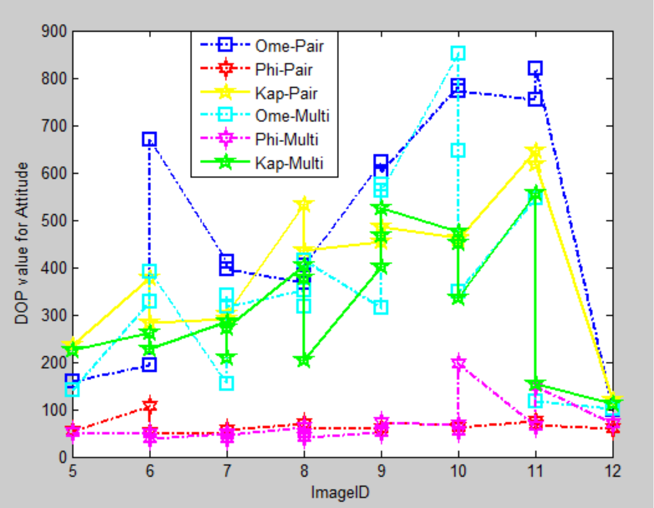


Figure 2. 13 Comparison of attitude DOP values

Finally the geo-referencing accuracy is compared. For every adjustment, first the mean standard deviation for the 3D coordinates in each direction (X, Y and Z) is calculated. For better comparison, then based on each single map image, the mean standard deviations for all its feature points were calculated and shown in Table 2.3. The result is not as expected, as the two groups produce close level of accuracy. Through further study, it is observed that most tie points produced by multi-image matching still come from two images rather than three. The average number of image rays per feature (object) point is 2.03. However if a better accuracy is expected, the number of image rays per object point need to be largely increased.

Table 2. 3 Comparison of geo-referencing accuracy using the mean standard deviation of feature points

Pair-wise								
Map Image ID	5	6	7	8	9	10	11	12
Average sigma_X	0.019	0.022	0.023	0.033	0.032	0.032	0.041	0.024
Average sigma_Y	0.008	0.005	0.005	0.007	0.006	0.006	0.007	0.005
Average sigma_Z	0.006	0.004	0.003	0.004	0.008	0.004	0.003	0.004
Multi-Image								
Average sigma_X	0.026	0.023	0.024	0.025	0.024	0.022	0.017	0.006
Average sigma_Y	0.007	0.006	0.006	0.006	0.005	0.006	0.012	0.015
Average sigma_Z	0.007	0.005	0.005	0.005	0.005	0.006	0.010	0.013

2.4.2 Controlled experiment for 3D mapping

The second experiment was focused on the function of the 3D map used for navigation. It aims to evaluate the geo-referencing accuracy for mapping, both theoretically and against reality. The resulting positioning accuracy using the geo-referenced 3D map has also been analysed. An indoor controlled experiment with coded target was conducted, and the school hallway was used as the testing field. After coded targets have been attached to the wall, map images are collected covering those targets (e.g. Figure 2.14). Since coded targets have very distinctive variation on coded dots against background, there is big chance a feature point is extracted and geo-referenced on those dots (e.g. Figure 2.15). The position of those target dots were surveyed by a total station and the data were used as true values.

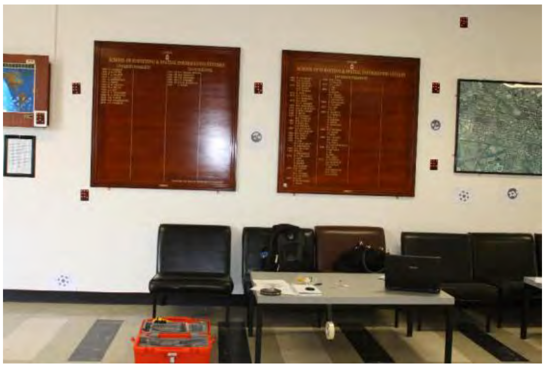


Figure 2. 14 Map image No. 5 with coded targets

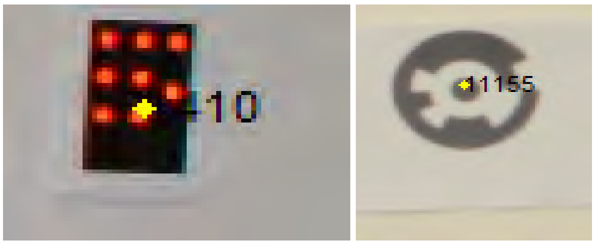


Figure 2. 15 SIFT feature shown with yellow dots are extracted on coded dots: tie point No. 410 and No. 11155

Mapping was proceeded with SIFT feature extraction, matching and indirect geo-referencing based on the images collected. After the bundle adjustment process, the 3D coordinates for each feature have been calculated, and the average standard deviations of the 3D coordinates in each axis from feature points on each map image has also been calculated as shown in Table 2.4. The result indicated that the viewing direction X had the lowest geo-referencing precision. The reason behind is that the geometry of feature distribution has a large influence on geo-referencing accuracy. In our experiment, the depth of the features has the least variation since they are distributed on a plane wall. It can also be deduced that theoretically the geo-referencing precisions on Y and Z axis are at centimetre level, and the X axis is around 5 centimetres.

Table 2. 4 The average standard deviation of 3D coordinates of the feature points on X, Y, Z axis from each map image (m).

IMID	3	4	5	6	7	8	9	10
sigma X	0.037	0.036	0.034	0.051	0.064	0.066	0.057	0.040
sigma Y	0.009	0.009	0.008	0.012	0.015	0.016	0.013	0.008
Sigma Z	0.006	0.006	0.005	0.007	0.009	0.009	0.009	0.006

To further evaluate the absolute accuracy of geo-referencing, and the impact of geo-referencing on final positioning accuracy, a controlled comparison experiment was conducted. First, the mapping was conducted twice to get two sets of results with different geo-referencing accuracy. More specifically, the geo-referenced coded dots produced from the two sets are identified and evaluated against their true values, as shown in Figure 2.15. The root mean square error (RMSE) was calculated. It is noted that different coded dots may be geo-referenced. The result is shown in Table 2.5. It can be observed that the accuracy trend confirmed the theoretical analysis in that X axis has the lowest accuracy. And the absolute accuracy on all the axis are centimetre level. From the results it is deduced that two 3D maps have different geo-referencing accuracy, with the one from Test A superior to that of Test B.

Table 2. 5RMSE of 3D coordinates of the feature points on X, Y, Z axis

RMSE(m)	X	Y	Z	Number of Points
Test A	0.043	0.037	0.008	13
Test B	0.094	0.048	0.011	15

Finally, vision-based positioning was carried out based on the geo-referenced 3D maps produced in Test A and B. The positions of the camera stations are both surveyed using a total station and calculated by the system. Therefore, the positioning accuracy produced by the visual system using the 3D map can be evaluated. The difference between calculated camera positions and their true positions are obtained and compared in Table 2. 6. It can be observed that the position accuracy of the method varies from sub-centimetre to more than 10 centimetres. Comparing the two sets, we can observe that generally the positioning result from the 3D map with better geo-referencing accuracy is more accurate. Therefore, it is deduced that the geo-referencing accuracy exerts certain influence on the final positioning. Further study on such aspect can be found in Chapter 6.

Table 2. 6 The difference between calculated camera positions and their surveyed positions

Difference	Test A			Test B		
ST ID	X0(m)	Y0(m)	Z0(m)	X0(m)	Y0(m)	Z0(m)
2	0.002	-0.020	-0.072	-0.125	0.012	-0.015
3	-0.003	-0.062	-0.079	-0.076	0.022	0.133
4	-0.023	0.038	-0.009	-0.104	-0.001	0.009
5	0.011	0.006	0.075	-0.131	0.055	-0.051

2.5 SUMMARY

This Chapter first gave a brief review on the literature of 3D mapping. Two main methodologies, range-based and image-based methods, are both discussed with the focus on image-based approach. Then the newly defined 3D map in this research was introduced. It features in both reality-based visualization of the environment, since it mainly consists of images, and also contains 3D geometric information. The development process was also described in great details. The main contribution lies in that geo-referenced images used as 3D map for navigation is a novel approach in the research domain. It enables the vision-based positioning at later stage to be resolved in

6 degrees of freedom with high accuracy. Meanwhile, multi-image matching was introduced into the system to improve the imaging geometry and geo-referencing accuracy. Experiments evaluated both the theoretical precision and absolute accuracy of geo-referencing, as been at centimetre level. Experiment has also revealed the impact of geo-referencing accuracy on final positioning accuracy.

Currently, this research mainly uses the 3D map to support vision-based navigation. As a matter of fact, the geo-referenced 3D map aims to provide various location-based services with a realistic view. Another important function, for instance, is to support 3D geometric measurement. While current image-based 3D maps like Google Street View only provide virtual experience in terms of photos, details concerning the topographic and terrain attributes are not available. Next generation of location based services will require much richer information to be provided, and geo-referenced street view is a promising approach. It is believed that such reality based 3D map, which provides both realistic visualization and accurate 3D geometric information of the view, will be of significant importance in next generation of location-based services. Further research to extend the 3D map's function in terms of better geometric accuracy and visualization can be conducted.

CHAPTER 3 VISION-BASED NAVIGATION WITH THE USE OF SINGLE CAMERA AND 3D MAPS

3.1 INTRODUCTION

One key issue in solving a vision-based navigation problem is self-localization. SIFT (the Scale Invariant Feature Transformation) has been successful applied to a variety of vision-related problems based on image matching, such as object recognition, pose estimation, image retrieval and so forth. In Berretti et al. (2010) authors used SIFT for 3D facial expression recognition. In recent years, growing number of researchers choose to use SIFT features as 3D natural landmarks for vision-based navigation applications.

Two years after SIFT was proposed, Se et al. (2001) designed a vision-based localization and mapping algorithm by tracking SIFT natural landmarks and building a 3D map simultaneously on their mobile robot, which was equipped with a trinocular stereo system. They improved this approach one year later (Se et al., 2002), in which sparse distinctive visual landmarks (SIFT landmarks) are used for 3D mapping together with an efficient map alignment algorithm. In 2003, Hofman-Wellenhof et al. developed a system that the robot ego-motion is estimated by matching with SIFT 3D landmarks. Experiments show that these features are robustly matched between views. In 2007, some authors (Gil et al., 2007) improve the data association among landmarks so as to improve the quality of the estimated path.

The advantage of using SIFT features as 3D landmarks for vision-based navigation lies in that it is highly distinctive and invariant against many changes, such as scaling, rotation, and partially invariant against illumination and viewpoint changes. Such a nature can provide reliable and stable image matching under different circumstances, making the self-localization process more effective. However, when previous researches have made good efforts, they are not without their limitations. When the robots‘ trajectories have been recovered, the orientation information is usually lost, and some challenging issues like mismatches have yet been properly addressed.

3.2 VISION-BASED POSITIONING USING PHOTOGRAMMETRIC 6DOF POSE ESTIMATION

3.2.1 Methodology

After the 3D map has been constructed with geo-referenced SIFT feature list developed for each map image, a photogrammetric approach of vision-based positioning and navigation can be carried out.

The main function for positioning (self-localization) is based on photogrammetric space resection. It is an algorithm that computes the exterior orientation of a single image based on collinearity equations, which gives 6 parameters that describe the spatial position and orientation of the camera coordinate system with respect to the global object coordinate system (Luhmann, 2009). Knowing the camera position and attitude, the vehicle position can therefore be determined. The requirement of space resection is that at least 3 control points with known object coordinates are measured on the image, in other words, a minimum of three points with image and object coordinates should be provided.

In this research, a vision sensor will take either images or record a video during the navigation time, namely as query images. For each query image, first a SIFT based voting strategy is used to localize the image search space. Then we match the query image with the geo-referenced map images in the localised search space. The aim is to locate SIFT feature points on the query image that have their corresponding feature points on the map images (Figure 3.1). It is noted that an outlier detection mechanism is used to first remove reference images that do not share common view with the query image, meanwhile mismatched feature points are removed. Further details are introduced in Chapter 4 and Chapter 5. When any of the SIFT feature points from the geo-referenced image(s) are found to correspond with the ones on the query image, the geo-information it carried can be transferred to its counterpart on the query image. Therefore, matched SIFT features on the query image obtain both image coordinates from matching process and 3D coordinates from the geo-referenced map images. Then these SIFT feature points can be used as control points for space resection. They are named as pseudo ground control points (PGCPs) in this research, an example is shown in Figure 3.2. By obtaining 3D points and their 2D positions on the query image, camera position and orientation of the query image can be determined through space resection. Thus the platform position can be obtained. The data flow of positioning process is shown in Figure 3.3.

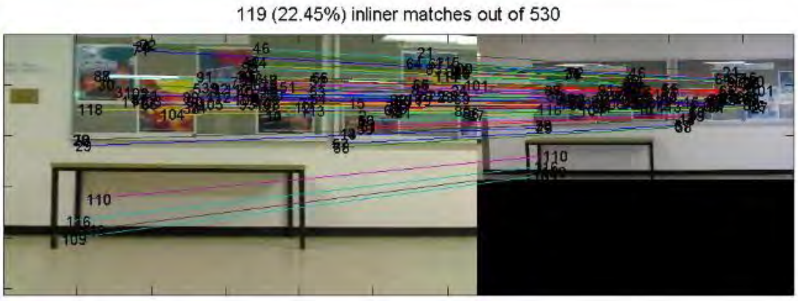


Figure 3. 1 Matching between real time query image with geo-referenced map image

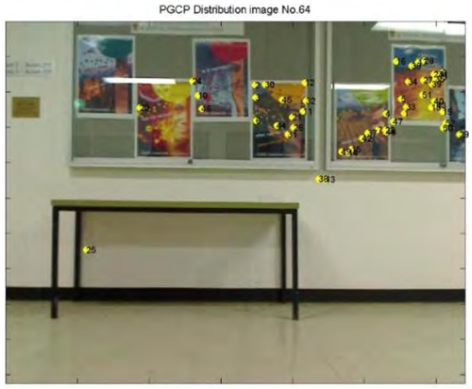


Figure 3. 2 PGCPs (yellow dots) on the query image

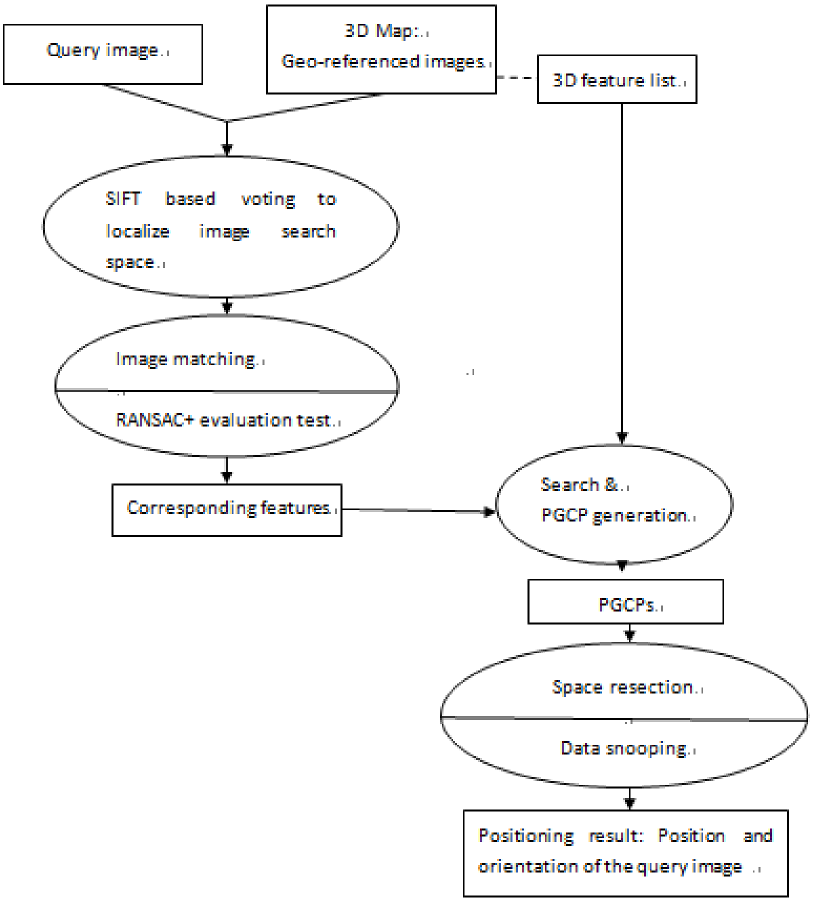


Figure 3. 3 Vision-based positioning

3.2.2 SIFT based voting strategy

The first step is to use a SIFT based voting strategy to localize the search space of geo-referenced images. The mechanism is similar to multi-image matching. A feature database is generated. It is a collection of all SIFT features extracted from geo-referenced map images. Meanwhile SIFT features are extracted from the query image and matched with the feature database. Then a K-NN search function is used to find nearest neighbours in the feature database for each of the features on the query image. The next step is to identify the map images that have a large number of matches with the query image. A voting strategy is used. Each neighbour feature votes for the map image it belongs. The map image that has the largest number of votes therefore has the greatest chance / biggest overlapping area with the query image. In this research, the top 3 images are selected as candidate reference images. Since mismatches may affect the selection of candidates to include false reference images, the matching between the query image and candidate reference images is then evaluated using a newly proposed method, which will be introduced in Chapter 5. Not only bad matching performance can be avoided, reference map images that do not share common view with the query image are removed from the candidate list. An example is shown in Figure 3.4, which is the first epoch of the query image sequence (image No.1). By using the voting strategy, the top 3 ranked candidate images are localised from the database containing 24 map images: reference image No.12, No.11 and No.6 (Figure 3.5). Using the proposed method (Figure 3.6, Figure 3.7 and Figure 3.8), false candidate image No. 6 has been identified, retaining only the correct ones: map image No. 11 and No. 12.

One limitation for the approach is that if it votes on the basis of the whole database for every query image, the algorithm is less efficient. Therefore when the query image is from an image sequence, or the system performs self-localization with a context, the voting is based on a sub-feature database that has been narrowed down by the previous epoch. Only the initial epoch takes longer time. In this example, epoch No.2 is voted from the sub-database that consisted of features from map image No.7-12.

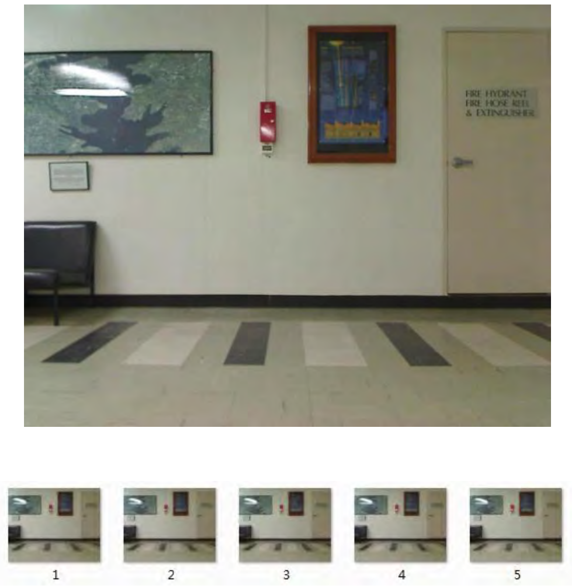


Figure 3. 4 Image No.1 from query image sequence

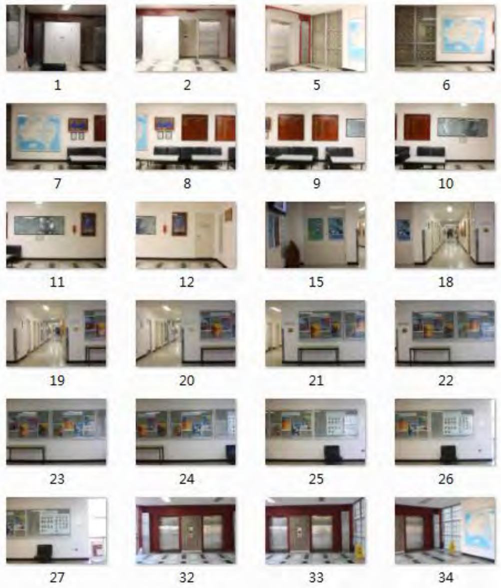


Figure 3. 5 Geo-referenced image database, containing 24 images

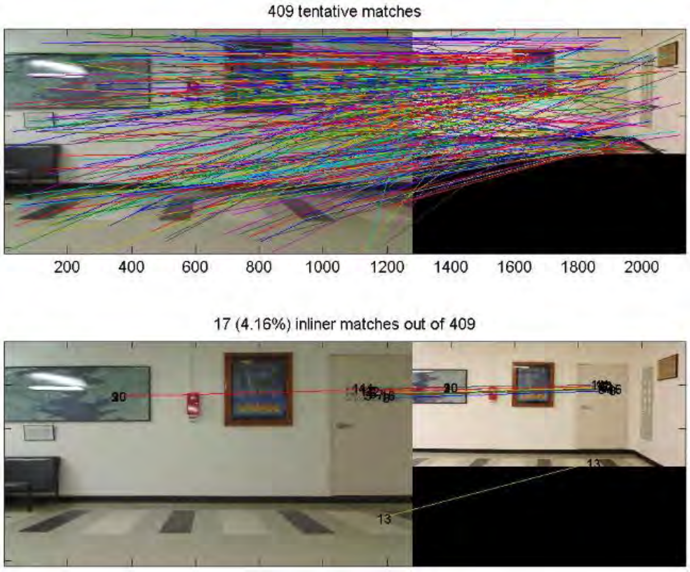


Figure 3. 6 The query image No. 1 matching with map image No. 12, evaluation test passed with pass rate at 94% (the threshold to pass the test is 0.8)

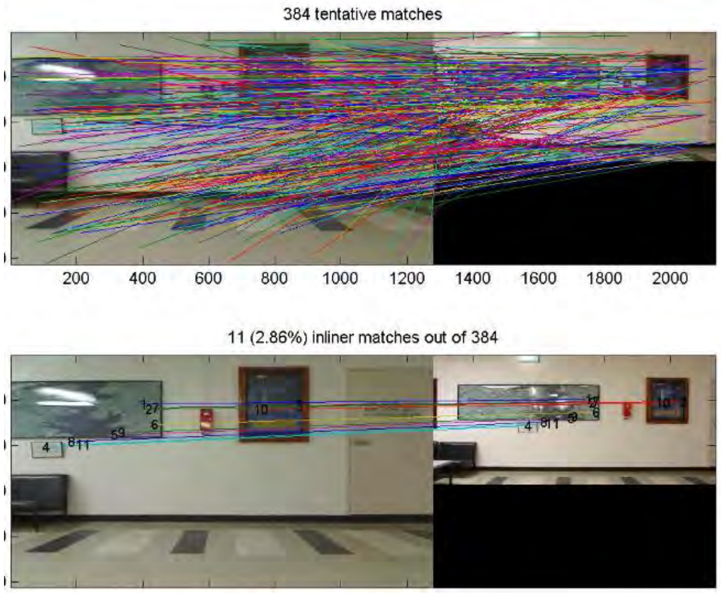


Figure 3. 7 The query image No. 1 matching with map image No. 11, evaluation test passed with pass rate at 80% (the threshold to pass the test is 0.8)

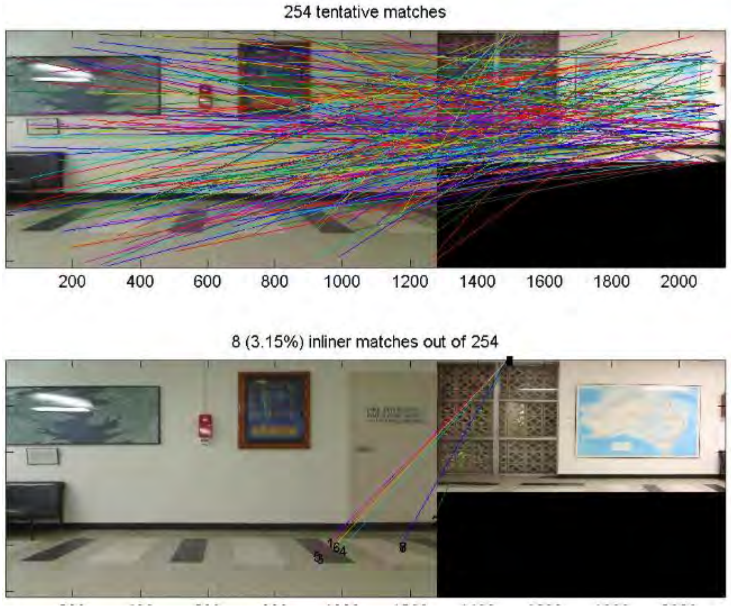


Figure 3. 8 Query image No. 1 matching with map image No. 6, evaluation test failed.

3.2.3 Generation of pseudo ground control points

The key component for positioning is to generate pseudo ground control points for positioning. Therefore in this section we use a simple example from our experiment to show the process of PGCP generation. A low resolution video camera is used for navigation and query images are extracted from the navigation video (image frame No. 457 is used as an example) matched with geo-referenced 3D map images to generate PGCPs. Three best candidate geo-referenced images (No. 10, No. 11 and No 12) from a collection of 24 images of the navigation environment are located in previous step to match with query image No. 457, which have the same landmarks. Map image No.11 is used as an example. First, matching between query image No. 457 and geo-referenced image No.11 is performed. As illustrated in Figure 3.9, SIFT based matching finds 37 candidate matches and RANSAC is used to remove outliers, as a result 14 pairs are retained as inliers. It is illustrated in Figure 3.9 and a detailed result is given in Table 3.1.

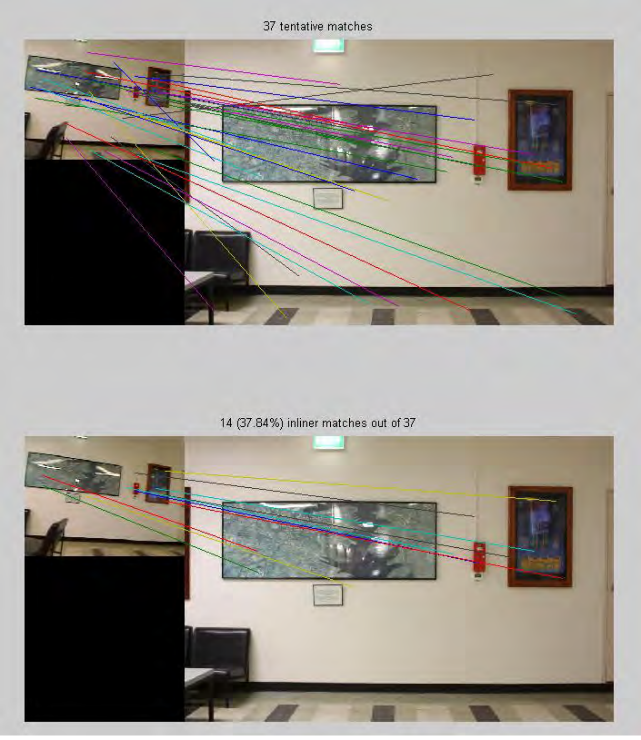


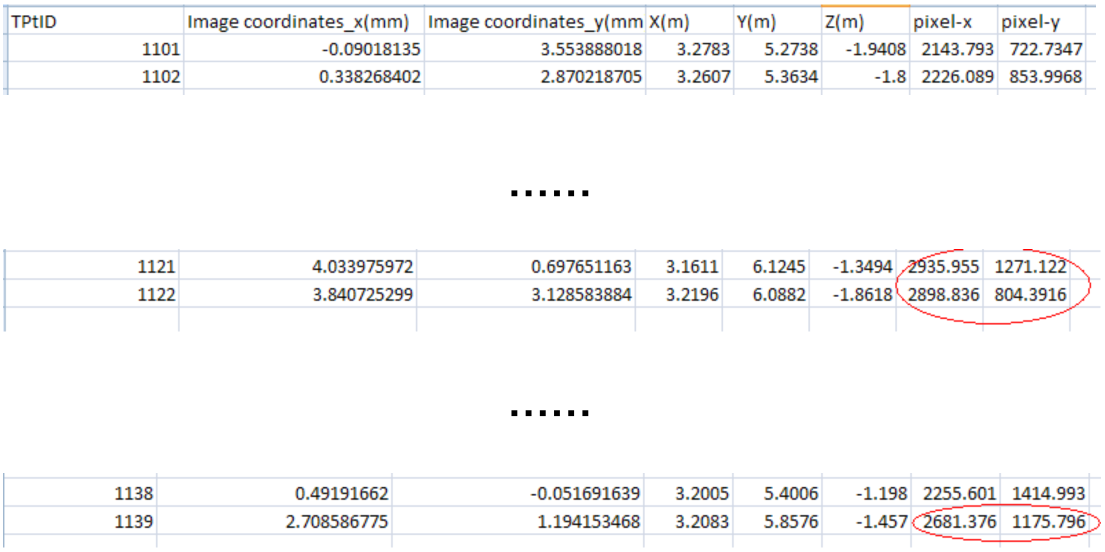
Figure 3. 9 Matching real time image (left) with map image(right)

Table 3. 1 Image coordinates of matching pairs between real time image and map image

x- real time image(pixel)	y- real time image(pixel)	x- reference image(pixel)	y- reference image(pixel)
36.493	98.854	741.414	1366.080
36.493	98.854	741.414	1366.080
36.470	79.206	722.977	1152.396
262.883	105.455	3493.045	1152.387
276.724	117.700	3665.363	1281.670
119.953	119.620	1693.345	1511.543
276.382	117.710	3665.363	1281.670
222.689	109.868	**2935.955**	**1271.122**
282.397	129.787	3771.242	1423.177
282.397	129.787	3771.242	1423.177
202.715	103.427	**2681.376**	**1175.795**
279.949	70.735	3712.411	649.131
279.949	70.735	3712.411	649.131
220.040	74.220	**2898.836**	**804.392**

Then the image coordinates (pixel value) of these 14 features on the reference image (the right two columns) are compared with the image coordinates of the map image feature list. If the same value is found from the list, that feature is identified to be candidate of PGCP. The following table (Table 3.2) is part of the feature list of map image No. 11.

Table 3. 2 Part of feature list of map image No.11



TPtID	Image coordinates_x(mm)	Image coordinates_y(mm	X(m)	Y(m)	Z(m)	pixel-x	pixel-y
1101	-0.09018135	3.553888018	3.2783	5.2738	-1.9408	2143.793	722.7347
1102	0.338268402	2.870218705	3.2607	5.3634	-1.8	2226.089	853.9968

......

1121	4.033975972	0.697651163	3.1611	6.1245	-1.3494	2935.955	1271.122
1122	3.840725299	3.128583884	3.2196	6.0882	-1.8618	2898.836	804.3916

......

1138	0.49191662	-0.051691639	3.2005	5.4006	-1.198	2255.601	1414.993
1139	2.708586775	1.194153468	3.2083	5.8576	-1.457	2681.376	1175.796

It is noticed that the red circled pixel coordinates are the same with the three red signed values appeared in Table 3.1. It means the 3 geo-referenced feature points from map image No.11 are matched by the query image, therefore, the 3D coordinates can be transferred and the 3 corresponding points on the query image can be used as PGCPs. It is noted that 3 PGCPs is not enough for accurate positioning, thus this matching process will be repeated between the query image and other candidate map images in order to generate more PGCPs. The result is shown in Table 3.3-Table 3.5:

Table 3. 3 Matched pairs between real time image No.457 and geo-referenced Image No. 11

x- real time image(pixel)	y- real time image(pixel)	x- reference image(pixel)	y- reference image(pixel)
222.689	109.868	2935.955	1271.122
220.040	74.220	2898.836	804.392
202.715	103.427	2681.376	1175.795

Table 3. 4 Matched feature points found from feature list of Image No. 11

TPtID	Image coordinates_x (mm)	Image coordinates_y (mm)	X(m)	Y(m)	Z(m)	pixel-x	pixel-y
1121	4.034	0.698	3.161	6.125	-1.349	2935.955	1271.122
1122	3.841	3.129	3.220	6.088	-1.862	2898.836	804.392
1139	2.709	1.194	3.208	5.858	-1.457	2681.376	1175.796

Table 3. 5 PGCP generated from this matching process

Image coordinates of real time image		3D object coordinates		
x(Pixel)	y(pixel)	Z(m)	Y(m)	Z(m)
222.689	109.868	3.161	6.125	-1.349
220.040	74.220	3.220	6.088	-1.862
202.715	103.427	3.208	5.858	-1.457

It is noted that after the voting process, the query image has already been matched with the candidate map images to remove misidentified reference map image(s). To save the computation power, during the positioning stage, image matching is performed only once, which both removes the mismatched reference images and generates PGCPs from image pairs that pass the test (introduced in Chapter 5).

3.2.4 Mathematical model for positioning function

Once pseudo ground control points (PGCPs) have been generated, positioning calculation can be carried out. Here we use the classical method for position solution: space resection based on a least squares solution of linearised collinearity equations (Eq.2.1). This method is normally used to compute the exterior orientation of a single image. This procedure requires known coordinates of at least three object points which do not lie on a common straight line. Here pseudo ground control points are used in the same way ground control points are used for traditional space resection. The theory lies in that the bundle of rays through the perspective center from the reference points can fit the corresponding points in the image plane in only one unique (camera) position and orientation (Luhmann et al., 2006). The central projection in space is at the heart of photogrammetric calculations, including space resection as well as bundle adjustment used in Chapter 2.

The least squares models can provide highly accurate results in 6 degrees of freedom with the presence of redundant measurements. Primarily the accuracy of camera external parameters is a function of point distribution and relative positions between the reference objects and the camera (Luhmann, 2009). In this research, the relative positions change during navigation process, therefore the accuracy of positioning largely depends on the geometry of PGCPs.

In the vision-based positioning system, we give DOP values for resolved camera external parameters (position and orientation) to evaluate the precision, which is influenced by PGCP geometry. More details can be found in Section 2.3.3.1.

One major difference between traditional space resection and this approach is that the 3D object space coordinates of PGCPs are transferred from the 3D map, which are photogrammetrically determined by the mapping process. They are not accurate enough to be used as error-free reference. Therefore some modification on the function model of space resection has been made to make it suit the scenario. We gave it the name ―Soft Space Resection" since control point values are not held fixed. The 3D object space coordinates of PGCPs are introduced into the system as observed unknowns (pseudo observations) with a corresponding weight. They receive corrections during the adjustment. The Gauss-Markov Model of the modified space resection for indoor positioning solution is introduced as follows:

$$\boldsymbol{At} + \boldsymbol{BX} - \boldsymbol{l_1} = \boldsymbol{v_1} \quad , \qquad \boldsymbol{l_1} \sim (\boldsymbol{0}, \sigma_0^2 \boldsymbol{P_1}^{-\boldsymbol{1}}) \tag{3.1}$$

in which

•$\boldsymbol{A}$ is a $2\mathrm{n}(\text{number of points}) * 6$ matrix containing partial derivatives with respect to the exterior orientation parameters, and t contains the incremental changes to the initial values of external orientation parameters;

•$\boldsymbol{B}$ is a $2\mathrm{n} * 3\mathrm{n}$ matrix containing the partial derivatives with respect to the three coordinates of the (pseudo) ground control points, and X contains the incremental changes to the initial values of ground coordinates of PGCP;

It should be noted that it is still a multi-solution equation when geo-referencing information from ground coordinates is not available, and design matrix $[\boldsymbol{A} \quad \boldsymbol{B}]$ is rank-deficient. Therefore, absolute orientation information needs to be introduced into the adjustment with stochastic constraints:

$$\boldsymbol{IX} - \boldsymbol{l_2} = \boldsymbol{v_2} \quad , \qquad \boldsymbol{l_2} \sim \left(\boldsymbol{0}, \sigma_0^2 \boldsymbol{P_2}^{-\boldsymbol{1}}\right) \tag{3.2}$$

Combine (3.2) with (3.3):

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{X} \end{bmatrix} - \begin{bmatrix} \boldsymbol{l_1} \\ \boldsymbol{l_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{bmatrix}, \qquad \begin{pmatrix} \boldsymbol{P_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P_2} \end{pmatrix} \tag{3.3}$$

in which $\boldsymbol{l_1}$ denotes the observation, image measurement of PGCP points in this case, $\boldsymbol{P_1}$ is its corresponding weight; $\boldsymbol{l_2}$ denotes the pseudo observation, object space coordinates of PGCPs in this case, $\boldsymbol{P_2}$ is its corresponding weight; $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ denotes the residuals. The modified space resection is named as —Soft Space Resection”.

Such an act will benefit the outlier detection process since outliers in pseudo observations can be detected and removed. More detailed explanation on outlier detection mechanism of the system can be found in Chapter 4.

3.3 EXPERIMENT

An indoor vision-based positioning experiment was carried out in the mapped indoor area of EE building of UNSW. Prior to the experiments, a 3D map has been produced using a high resolution camera (Cannon EOS4500) with a focal length of 24.7 mm and image resolution of 4272×2848 pixels. A total of 24 images are collected to include the school hallway as mapping area. A local orthogonal right-handed coordinate system is used with the Z axis pointing downward.

A calibrated video camera (Logitech Webcam Pro2000) was mounted on a moving vehicle with sampling rate of 1 Hz during navigation. Its relative position to the vehicle was fixed, which means the experiment was partially controlled: camera height Z:-0.725m.

Table 3. 6 Camera property

Focal Length: 5.01 mm
principal point x: 2.92 mm
principal point y: 2.18mm
format width: 6.03mm
format height: 4.50 mm
format width: 1280 pixel
format height: 1024 pixel

We did the positioning by extracting image frames from the video (query image sequence) and match with the geo-referenced images (3D map) frame by frame. Each frame is an epoch; a position in 6 degrees of freedom (6DOF) was calculated. This experiment aims at testing the performance of the navigation system and evaluating the indoor positioning accuracy.

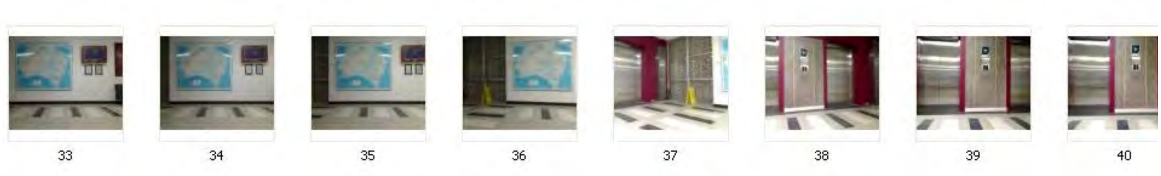


Figure 3. 10 Query image sequence

From the video, a total 83 epochs (frames) were generated & calculated, 20 epochs failed to provide a reasonable result, which is failure rate at 24.1%. As GPS GCPs are not available in indoor areas, commercial software Photomodeller is used to determine camera positions and then use them as references to evaluate the system produced results. Within 10m distance, the software can normally achieve centimeter level accuracy. From Photomodeller 60 reference epochs were generated and the two systems have 55 epochs in common. The trajectory of indoor navigation is shown in Figure 3.11 (horizontal) and Figure 3.12 (vertical). The RMSE of the calculated positions is shown in Table 3.7. It can be observed that the accuracy of this indoor positioning experiment is round 10- 20cm level.

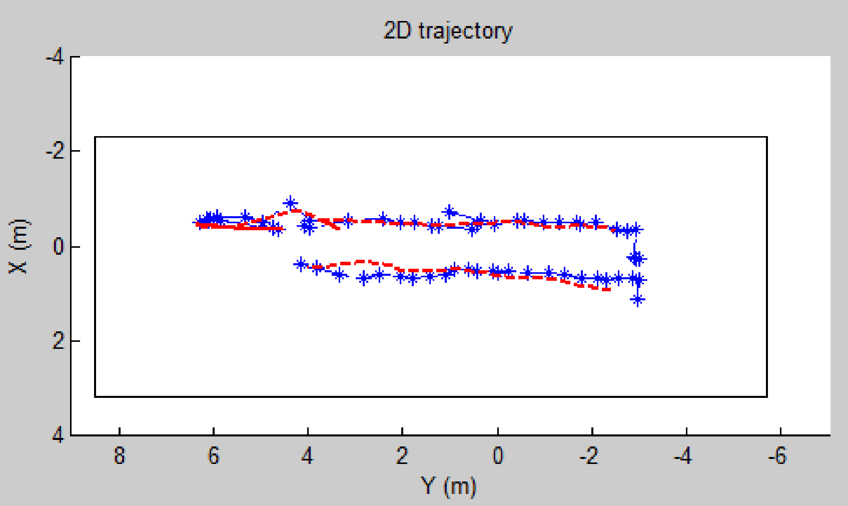


Figure 3. 11 Two dimensional trajectory of indoor navigation recovered by the vision-based system (blue line) with reference to the Photomodeller results (red line)

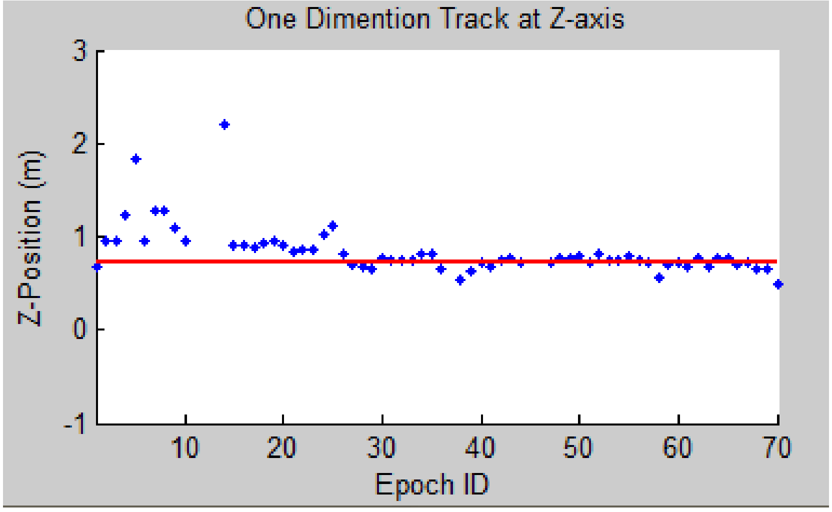


Figure 3. 12 Z positions of indoor navigation calculated by the system (blue dots) with reference to the controlled value (red line).

Table 3. 7 RMSE for indoor positioning

RMSE	X(m)	Y(m)	Z(m)
Calculated	0.126	0.281	0.137

Then we investigate the nature of the positioning results. For better comparison, three typical epochs with adjacent views are chosen for illustration (Figure 3.13-3.15). The PGCPs are shown as yellow dots. The positioning results are shown in Table 3.8 and DOP values in Table 3.9. Since the navigation camera face one side of the wall, 3 camera exterior parameters (Z, ω and ψ) out of 6 are (approximately) controlled.

Table 3. 8 Positioning results in 6DOF for epoch No.13, No.14 and No.15

Epoch ID		13	14	15	Controlled
Positioning result in 6DOF Unit :m & degree	X		-0.890	-0.541	-
	Y		4.366	3.988	-
	Z		-2.191	-0.9	-0.725 (m)
	ω		110.581	91.902	≈90 (°)
	ψ		-2.750	-0.573	≈0 (°)
	κ		-88.350	-91.158	-

Table 3. 9 DOP values in 6DOF for epoch No. 13, No.14 and No.15

Epoch ID		13	14	15
DOP values in 6DOF	X		11974	1069
	Y		36223	2158
	Z		50478	3652
	ω		11612	985
	ψ		5305	152
	κ		8161	516

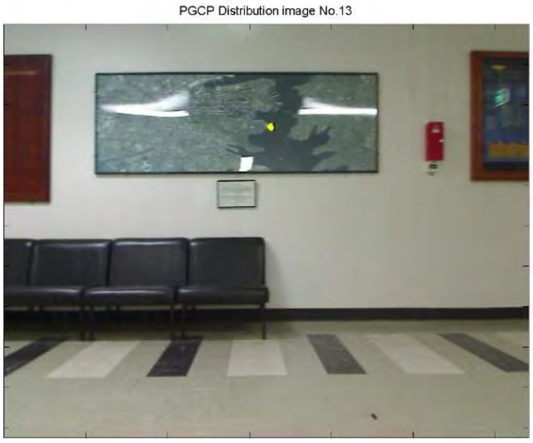


Figure 3. 13 Epoch No. 13 with 1 PGCP

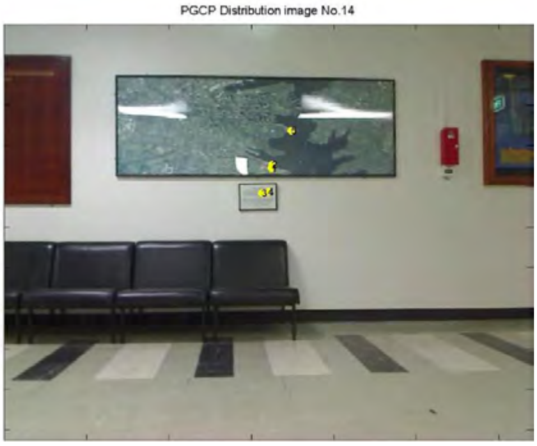


Figure 3. 14 Epoch No. 14 with 5 PGCPs

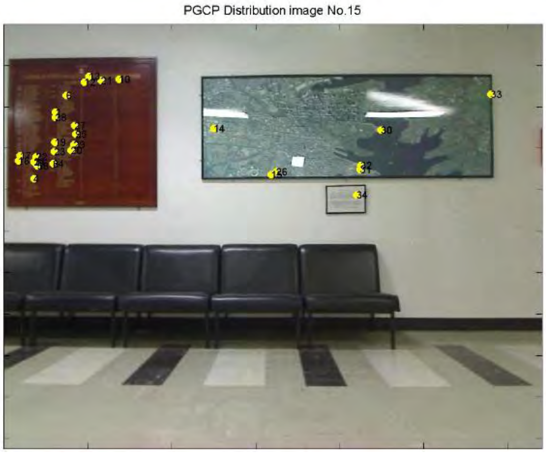


Figure 3. 15 Epoch No. 15 with 34 PGCPs

Comparing the three epochs, we can deduce the main characters of the positioning function. It can be observed that for epoch No. 13, only one PGCP has been generated, which is inadequate to give a solution. In fact, one limitation for the space resection based method is that it requires a minimum number of observations. For epoch No. 14, 5 PGCPs are able to get a positioning result. However, the result largely deviates from the true values. Epoch No.15 has both a better distribution and greater number of PGCPs compared to the other two epochs. It provides a reasonable positioning result. Table 3.9 also shows that epoch No. 15 has much smaller DOP values than those of No.14, which means a better PGCP geometry. Therefore, it is deduced that the accuracy of the positioning result is closely related to the number and distribution of PGCPs. In other words, the geometry of PGCPs is of significant importance to the system. The investigation of the impact of PGCP geometry on positioning performance is introduced in Chapter 6. Since PGCPs are generated from the matched features on the query image and 3D map images, a rich texture of mapping and navigation environment is required for vision-based navigation. The factors that influence positioning performance are further discussed in Chapter 6.

3.4 SUMMARY

In this Chapter, the methodology of the vision-based positioning solution is introduced, including both mathematical model and implementation procedures. By matching the query image with the 3D map, the 3D information is transferred from the map to corresponding SIFT features on the query image. The main contribution is the adoption of geo-referenced SIFT feature points as 3D landmarks for positioning. In this way, the 3D feature points are used as pseudo ground control points and the final positioning result can be resolved based on photogrammetric space resection, which gives highly accurate result in 6 degrees of freedom.

It's worth mentioning that instead of space resection, relative orientation can also be considered to solve exterior orientation of a query image. Relative orientation describes the relative position and attitude of two images with respect to one another. Therefore, the exterior orientation of a query image, which overlapped with certain geo-referenced map image(s), can be calculated based on known parameters of the map image(s). However, to properly control the relative orientation, at least 8-10 well distributed tie points should be measured (Luhmann, 2009). Thus such methodology has normally been used for stereo image analysis. In close-range photogrammetry, in contrast, often involves arbitrary convergent multi-image configurations. Significant rotation and scaling differences challenges the use of relative orientation considerably. For instance, if the overlapping image pair have insufficient spatial ray intersections, uncontrollable model errors may occur. In our application, the query image is supposed to be taken randomly, the position and orientation of which can vary significantly from the map images with overlapping areas. Therefore, we believe such a method is not the optimal choice. The introduction of PGCPs and the use of space resection on the other hand help avoiding the problems. Rather than relying on solely one partner, it takes advantage of multiple overlapping map images. It provides a better geometry and greater redundancy for the least squares solution. Most importantly, the query image can be taken with great diversity from corresponding map images.

It has been noted that for map-based visual positioning like this research, which uses a position-fixing approach, the greater the initial position uncertainty, the longer it will take to perform map-matching. Mismatching can easily sabotage the result if the initial position largely deviates from true values. In this research, the vision-based positioning is performed within certain building using a local coordinate system. If greater areas need to be covered, technologies such as GNSS and WiFi need to be integrated to provide an approximate position, then vision is used to refine the position.

CHAPTER 4
QUALITY CONTROL MEASURES FOR VISION-BASED POSITIONING AND NAVIGATION

4.1 INTRODUCTION

In the research domain of indoor navigation or more generally, navigation in a GPS-denied environment, vision is believed to be the most promising but challenging technologies so far. Compared with other sensors that may hold the promise to supplement satellite-based positioning, vision sensors are cheap, ubiquitous, self-contained and do not suffer from drifting errors. However, available vision-based navigation systems are still inadequate to provide a mature localization function, major problem lies in that stable visual features are difficult to be identified and direct measurement of real world geometry is lost. Vision sensors have a high input rate, which further challenges vision-based navigation systems, especially when a high precision and good reliability is expected to be obtained.

The main purpose of vision-based navigation is to determine the position and orientation of the vision sensor, then mobile vehicle_s(the platform carrying the vision sensor) motion/trajectory can be recovered. Based on the way that self-localization is realized, most available approaches can be grouped into two categories, as has been discussed in Chapter 1: one that relies on the prior knowledge of the navigation environment (e.g. map, or models), and one that does not. Both of these two approaches reply on image matching techniques, and mismatch has become the major error source for vision-based navigation systems. The bottleneck lies in that a vision sensor is inherently fragile against errors while the establishment of such correspondence may easily be sabotaged by input noise, and other error sources. Meanwhile, for the former approach, incorrect object information may also be included at mapping stage (e.g. careless control survey, or image measurement), then the final position result which relies on the map will be affected. Therefore, an outlier detection mechanism especially built for these applications is highly desirable.

Various outlier detection strategies have been developed so far. Some look into all the observations and repeatedly remove the one that fails certain statistical tests (e.g., Data snooping (Baarda, 1968), or making adjustment on their weights (e.g. Huber‘s M-estimators (1964)). Some start with a minimum configuration of observations, and continuously adding samples that meet certain criterion (e.g. RANSAC (Fischler and Bolles, 1981)). Unfortunately, neither of these approaches has provided a good solution for vision-based navigation systems so far. Some recent studies have focused on the potential for a biased pixel measurement being introduced into the navigation solution process (Larson and Craig, 2010). In this research, different outlier detection strategies are compared and evaluated in the context of vision-based navigation. In order to make the best use of their strength and compensate for their weaknesses, a combined use of different outlier detection strategies in a multi-level strategy has been introduced in this Chapter.

4.2 OUTLIER DETECTION STRATEGIES

The classic theory divides error into three different groups: random error, systematic errors and gross errors. Random errors are assumed to be unavoidable, always present in the observations and obey a normal distribution; systematic errors come from the imperfection of functional model that cannot fully describe the reality; gross errors are those errors that occur when a measurement process is subject occasionally to large inaccuracies and have identifiable causes, the corresponding observations are called outliers. Various outlier detection strategies are developed to detect this part of errors and reduce the effect of outliers on final parameter estimation result. In the following context, both their basic principles and the modification this research made to suit the specific requirement for vision-based navigation are discussed. Modification of data-snooping is introduced in Section 4.3.2.1 and modification on RANSAC is given in Chapter 5.

4.2.1 Mean-shift model and variance-inflation model

For geodesy and photogrammetry, outlier detection methods are generally based on least squares adjustment model, which is consisted of two parts: function model and stochastic model. Correspondingly two models are established for ‗gross errors': mean-shift model and variance-inflation model. The former approach regards outliers to be a subset of observations that have the same variance but different expectations compared with ―healthy" observations, thus a shift of the probability distribution of the observations occur. Typical statistical tests based on this assumption include Baarda's data snooping, Pope's τ-distribution test (Pope, 1976) and a generalized outlier detection method (Wang and Chen, 1999). The second approach considers outliers to be observations that have the same expectation but different variance. A good number of outlier detection methods based on variance-inflation model have been developed so far, which includes least absolute values method (Edgeworth, 1987), M-estimators , Generalised M-estimators (Hampel et al., 1986), Danish Method (Krarup, 1980) and so forth. A detailed comparison between these methods on other applications can be found in Knight and Wang (2009). Here data snooping and Huber's M-estimators are explained in details, for they are more generally used and are adopted in our system.

The Gauss-Markov model of observation equations is:

$$\boldsymbol{l}+\boldsymbol{v}=\boldsymbol{A}\hat{\boldsymbol{x}} \quad , \quad \boldsymbol{E}(\boldsymbol{l})=\boldsymbol{A}\boldsymbol{x} \tag{4.1}$$

where $\boldsymbol{v}$ is vector of residuals, $\boldsymbol{A}$ is the design matrix, $\boldsymbol{x}$ is the vector of unknown, and its estimation is $\widehat{\boldsymbol{x}}$, $\boldsymbol{l}$ is the measurement vector. The variance covariance matrix of the measurements $\boldsymbol{\Sigma}$, is given by

$$\boldsymbol{D}(\boldsymbol{l}) = \boldsymbol{\Sigma} = \sigma_0^2 \boldsymbol{Q} = \sigma_0^2 \boldsymbol{P}^{-1} \tag{4.2}$$

where σ_0^2 is a priori variance factor, $\boldsymbol{Q}$ is the cofactor matrix, and $\boldsymbol{P}$ is the diagonal weight matrix. Model (4.1) and (4.2) are the least-squares function model and stochastic model respectively. The solution is a minimization problem:

$$\min: \boldsymbol{v}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{v} \tag{4.3}$$

Then, the least-squares estimation of unknowns is:

$$\begin{cases} \widehat{\boldsymbol{x}} = \left(\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{P} \boldsymbol{A}\right)^{-\mathbf{1}} \left(\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{P} \boldsymbol{l}\right) \\ \boldsymbol{Q}_{\widehat{\boldsymbol{x}}} = \left(\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{P} \boldsymbol{A}\right)^{-\mathbf{1}} \end{cases} \tag{4.4}$$

The procedure for outlier detection is to first determine whether there exists an outlier, then the outlier(s) are identified. The first process is called Global Model Test. The basic idea is that for healthy dataset, the estimated posterior variance s_0^2 is statistically equal to the priori variance σ_0^2 and follows a Chi-square distribution. Provided a significance level, the posterior variance can be tested using a two-tail test (sometimes one-tail test is recommended). If the posterior variance exceeds the critical values, global model test fails and we assume there is outlier(s) in the measurements. More detailed explanation can be found in Teunissen (1990), Wang and Chen (1994).

After outlier has been detected, data snooping is used to identify and remove outlier (s). Assuming there is an outlier $\boldsymbol{\nabla S_i}$ in the $\boldsymbol{i}$th observation, using mean shift model, Eq. (4.1) can be extended to:

$$\boldsymbol{E}(\boldsymbol{l}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e_i} \boldsymbol{\nabla S_i} \tag{4.5}$$

where $\boldsymbol{e_i}$ is a vector of zeros with the ith element equal to 1. Based on Eq. (4.1) and (4.5), a least squares estimation of the outlier $\boldsymbol{\nabla S_i}$ is given as:

$$\boldsymbol{\nabla S_i} = -(\boldsymbol{e_i}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{Q_v}\boldsymbol{P}\boldsymbol{e_i})^{-1}\boldsymbol{e_i}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{v} \tag{4.6}$$

And its variance is:

$$\boldsymbol{D}_{\boldsymbol{\nabla S_i}} = \sigma_0^2(\boldsymbol{e_i}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{Q_v}\boldsymbol{P}\boldsymbol{e_i})^{-\boldsymbol{1}} \tag{4.7}$$

The test statistic is given by

$$\boldsymbol{W_i} = \frac{\boldsymbol{\nabla S_i}}{\sqrt{\boldsymbol{D}_{\boldsymbol{\nabla S_i}}}} = -\frac{\boldsymbol{e_i^T}\boldsymbol{P}\boldsymbol{v}}{\sigma_0\sqrt{\boldsymbol{e_i^T}\boldsymbol{P}\boldsymbol{Q_v}\boldsymbol{P}\boldsymbol{e_i}}} \tag{4.8}$$

If $E(\boldsymbol{\nabla S_i}) = 0$, which means no outlier exists, $W_i|H_0 \sim N(0,1)$. Otherwise the null hypothesis should be rejected. Given confidence level (1-α), if

$$|\boldsymbol{W_i}| > N\left(0,1;1-\frac{\alpha}{2}\right) \tag{4.9}$$

then an outlier is identified by the biggest value of $|\boldsymbol{W_i}|$. For iterative data snooping, the observation associated with the biggest value of $|\boldsymbol{W_i}|$ is detected and removed. The algorithm runs until global model test passes.

Following this theory, an inverse operation has been used in the literature to define the reliability measure. When the alternative hypothesis $E(\boldsymbol{\nabla S_i}) \neq 0$ is accepted, $W_i|H_a \sim N(\delta_i,1)$, where the non-centrality parameter is described by

$$\delta_i = \frac{\boldsymbol{\nabla S_i}}{\sigma_0}\sqrt{\boldsymbol{e_i^T}\boldsymbol{P}\boldsymbol{Q_v}\boldsymbol{P}\boldsymbol{e_i}} \tag{4.10}$$

Given confidence level and the power of the test, one can calculate the lower bound value for the non-certrality parameter δ_0 .Using δ_0 in Eq. 4.10, the lower bound for outlier detection can be calculated:

$$\boldsymbol{\nabla_0 S_i} = \frac{\sigma_0\,\delta_0}{\sqrt{\boldsymbol{e_i^T P Q_v P e_i}}} \tag{4.11}$$

With this test, any outlier that bigger than $\boldsymbol{\nabla_0 S_i}$ can be detected and identified. Therefore it is also referred to as minimal detectable bias (MDB).

For Huber's M-estimators, an iterative robust estimator of the unknowns in the Gauss-Markov model can be obtained based on the M-estimation (Huber, 1981).

$$\text{min: } \boldsymbol{v^T \bar{P} v} \tag{4.12}$$

With the equivalent weights:

$$\boldsymbol{\bar{P}} = \text{diag}(\mathrm{f_i p_{ii}}) \tag{4.13}$$

where fi (i=1,2,…,n) are reduction factor of the weight elements, $\mathrm{p_{ii}}$ are the ith diagonal element of weight matrix. Determination of the reduction factor has been a hot research topic in statistical and geodetic literature. The following reduction function works well in practical situations (Yang et al., 2002):

$$\boldsymbol{f_i} = \begin{cases} \boldsymbol{1} & |\boldsymbol{u_i}| \le \boldsymbol{k_0} \\ \dfrac{\boldsymbol{k_0}}{|\boldsymbol{u_i}|}\left(\dfrac{\boldsymbol{k_1} - |\boldsymbol{u_i}|}{\boldsymbol{k_1} - \boldsymbol{k_0}}\right) & \boldsymbol{k_0} < |\boldsymbol{u_i}| \le \boldsymbol{k_1} \\ \boldsymbol{0} & |\boldsymbol{u_i}| > \boldsymbol{k_1} \end{cases} \tag{4.14}$$

where $\boldsymbol{u_i}$ is standardized residual, $\boldsymbol{k_0}$ and $\boldsymbol{k_1}$ are two constants, usually chosen as 2.0-3.0 and 4.5-8.5 respectively. Some further study can be found in (Förstner, 1983).

When introduced to vision-based navigation society, these methods should be further investigated. In fact, outlier detection strategies which are based on least squares solution share common shortcomings. Least squares adjustment is an optimal parameter estimation method, which work best for data only with normal distributed random noise (without outliers). Otherwise the adjustment process will disperse the observation errors over all observations in the dataset. A smearing effect will take place when gross errors are present. As a result, when several outliers exist in the dataset, the ability for statistic testing to detect and identify each of these outliers will be very limited. In other words, most approaches for outlier detection are based on the assumption that only very few outliers exist. Unfortunately, for vision based navigation application, a high input rate of vision sensors will naturally bring a high percentage of errors into the dataset. At the matching process for self-localization, mismatches will challenge the reliability of vision-based navigation by introducing a large amount of mismatches. Therefore, a direct application of these outlier detection methods into the vision-based domain will lead to a complete failure.

4.2.2 RANSAC

The RANSAC (Random Sample Consensus) algorithm is an algorithm for robust fitting of models. It was introduced by Fischler and Bolles in 1981.It is capable of interpreting and smoothing data containing a significant percentage of gross errors, and it is ideally suited for application in automated image analysis where interpretation is based on the data provided by error-prone feature detectors (Fischler and Bolles, 1981). Unlike the above outlier detection techniques that have their root from surveyors, RANSAC was originated from the computer vision community. Till today, the algorithm has been applied to a variety of applications in computer vision, such as feature matching and image registration.

It is suitable for vision-based navigation application in the sense that it is robust against large proportion of outliers in the input data. More specifically, a vision-based system normally do self-localization based on the matching between real time images (or any visual input) and expectation of the navigation environment in terms of database images or models if a prior knowledge of the environment is available; or a relative position between epochs is obtained via the matching of subsequent images. Either way will need feature recognition function with use of feature-based matching.

And it is always accompanied by a large fraction of mismatches. While obviously previous two approaches can hardly handle this job, RANSAC plays a suitable role. The general procedure for RANSAC to be used in feature based matching is as follows:

- Given tentative matched pairs, randomly choose 4 matches and the homography matrix is computed based on the initial sample.

- Using the computed homography matrix to count the number of inliers.

- Repeat the first two steps for a certain number of times.

- If the number of inliers is a maximum among iterations, the homography matrix and inliers are stored.

- After the certain number of iterations, use the stored inliers to re-estimate homography and the consensus set are treated as correct matches while the ones do not treated as mismatches (outliers).

In the context of SIFT matching (Lowe, 1999), tentative matches are found by searching and locating nearest neighbour between the SIFT descriptors on the image pairs. Figure 4.1 shows the power of RANSAC in removing mismatches, which comes from the system experimenting data. SIFT matching was carried out between map Image No.10 and No.11, a total 185 tentative matches are found by SIFT matching. It could be clearly observed that several mismatches exist (crossed lines are obvious indications of mismatches for the viewing directions of the two images are close to parallel). The RANSAC process retains 13 pairs of matches as correct matches, and all the others have been filtered out as mismatches.

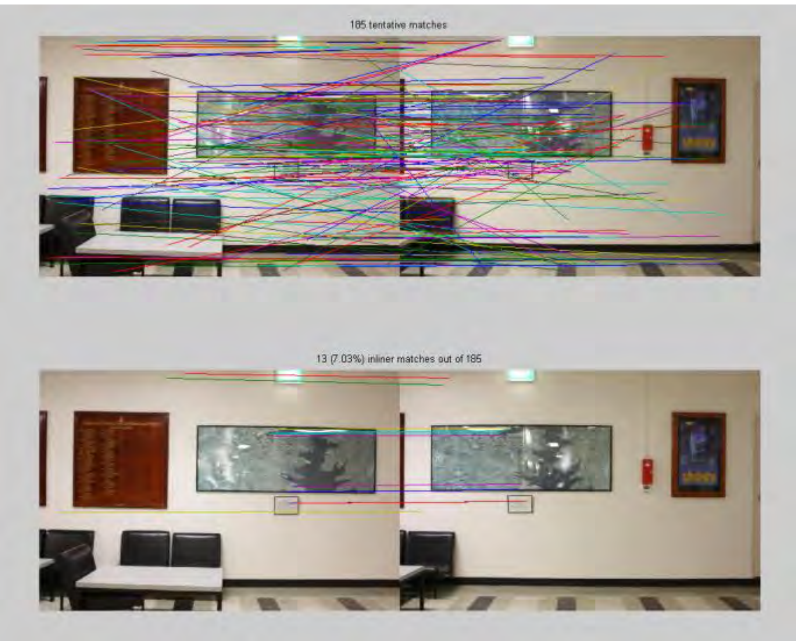


Figure 4. 1 SIFT matching and RANSAC processing, 13 inliers out of 185 tentative matches

Although it can be easily observed that all the remained pairs have close-to-parallel lines, still it had mismatched pairs remained in the dataset (e.g. the lowest yellow line in Figure 4.1). In fact, the RANSAC algorithm is not without its limitations. One problem is that the estimate is only correct with a certain probability, since RANSAC is a randomized estimator. It can be computationally expensive to run many times to ensure the correctness. Otherwise, in difficult matching conditions, it can still include mismatches as inliers. For a vision-based navigation system relies on image matching for object recognition and rough-localization, RANSAC can help filter out most of the mismatches; but for a positioning function that using the measurements of feature points to calculate the exact position and orientation, RANSAC alone can hardly produce satisfactory results. In chapter 5, a method using cross-correlation information to improve RANSAC homography estimation is proposed and discussed.

4. 3 OUTLIER DETECTION MECHANISAM FOR VISION-BASED NAVIGATION SYSTEM

4.3.1 Introduction of the multi-level outlier detection mechanism

Traditionally in the field of mobile robot localization and mapping (using vision sensors), people mainly consider gross errors coming from mismatches between real time visual information and its expectation (model or images). In the field of photogrammetry, on the other hand, gross errors, which may be caused by misidentified image points, or by a careless measurement in the ground control survey, have been of major concern for years. Therefore, vision-based navigation systems, especially those involving photogrammetric methodologies, are complex. Gross errors can come from a variety of sources and can have different characters as well as quantities. By analysing different outlier detection strategies and comparing their pros and cons, we have knowledge of their strength and weaknesses when dealing with problems raised by a vision-based system. In this research, a multi-level step-by-step outlier detection scheme has been proposed for vision-based navigation systems. Along the data flow of the whole system, suitable outlier detection strategies are applied to track and remove possible outliers at each step of the way. This multi-level detection mechanism aims at improving the reliability of the vision-based system by enabling gross errors from difference sources (e.g. image measurement, image matching, ground control survey) to be treated specifically. Besides, for systems using least squares estimation for positioning, chances of convergence failure caused by large scale outliers can be reduced.

Using the principles of photogrammetry as backbone, the vision-based navigation in this research mainly consists of two stages: mapping and positioning. In the first step the 3D map is constructed by geo-referencing images of the target environment. In the second step, images taken by the navigation system in real time are matched with the 3D map (geo-referenced images), thus transferring the geo-information from the map to the real time image for positioning. As much as outlier detection is concerned, two SIFT based image matching and three least squares adjustment processes with use of photogrammetric principles are involved in the system. The chosen strategy is to conduct RANSAC at each time of the SIFT matching process in order to detect and remove most of the mismatches. For each least squares adjustment procedure, an outlier detection strategy (mainly based on Mean-shift Model) is applied to further clean the dataset in order to produce a better estimate of unknown parameters from the least squares solution. The obvious advantage for this scheme lies in that a combined use of different outlier detection strategies can make the best use of their strength and avoid or compensate for their weaknesses. More specifically, the RANSAC algorithm is capable of estimating parameters from a dataset with a significant percentage of gross errors, but it cannot guarantee the correctness of final results because of its random nature. So it is used as a pre-adjustment process to filter out most of the mismatches. At the same time, a cross-correlation method proposed in this research (Chapter 5) is used to improve the performance of RANSAC. Later at the parameter estimation stage using least squares adjustment, more sensitive outlier detection methods (data snooping/M-estimators) is used to guarantee the correctness of the least-squares processing. Its deficiency in face of large fraction of outliers is avoided by the previous processes.

4.3.2 Quality control for 3D mapping

First a mapping procedure is conducted. The quality of the map depends on the accuracy of geo-referencing, therefore, the main objective of quality control at the mapping stage is to detect and remove the outliers for bundle adjustment. Two major input of bundle adjustment are ground control points and tie points. The first dataset come from ground control survey and image measurement of these control points, while the second are common SIFT feature points produced by the matching process.

Rather than throwing all the data directly into the geo-referencing procedure, which is equipped with a gross-error (outlier) detection function, a step-by-step detection scheme is chosen. The aim of outlier detection at the early stage is to alleviate the burden of fault detection in geo-referencing. First, for the ground control points, an iterative data-snooping is used along with soft space resection (introduced in Section 3.2.3) to detect and remove gross errors from the ground control survey and the image measurements of these ground control points. Besides space resection, data snooping has also been modified to suit the requirement of the application, which is further explained in the following section. At the same time, most of the outliers at tie points (SIFT mismatches) are removed by RANSAC at matching stage. It is noted that mismatches left by RANSAC are the most likely error source after previous procedures. In this research, two options are offered. The first one is using the newly proposed method to improve RANSAC homography estimation and outlier detection, which has been applied to the outlier detection for image matching in positioning calculation. More details on such method can be found in Section 4.3.3 and Chapter 5. Since mapping is performed on images with big overlapping areas, the chance mismatches are involved after RANSAC is much smaller compared with the positioning stage. So at final stage of image geo-referencing, as bundle adjustment is also based on least squares, modified iterative data snooping is applied to deal with the small number of mismatches. The overall flowchart is shown in Figure 4.2.

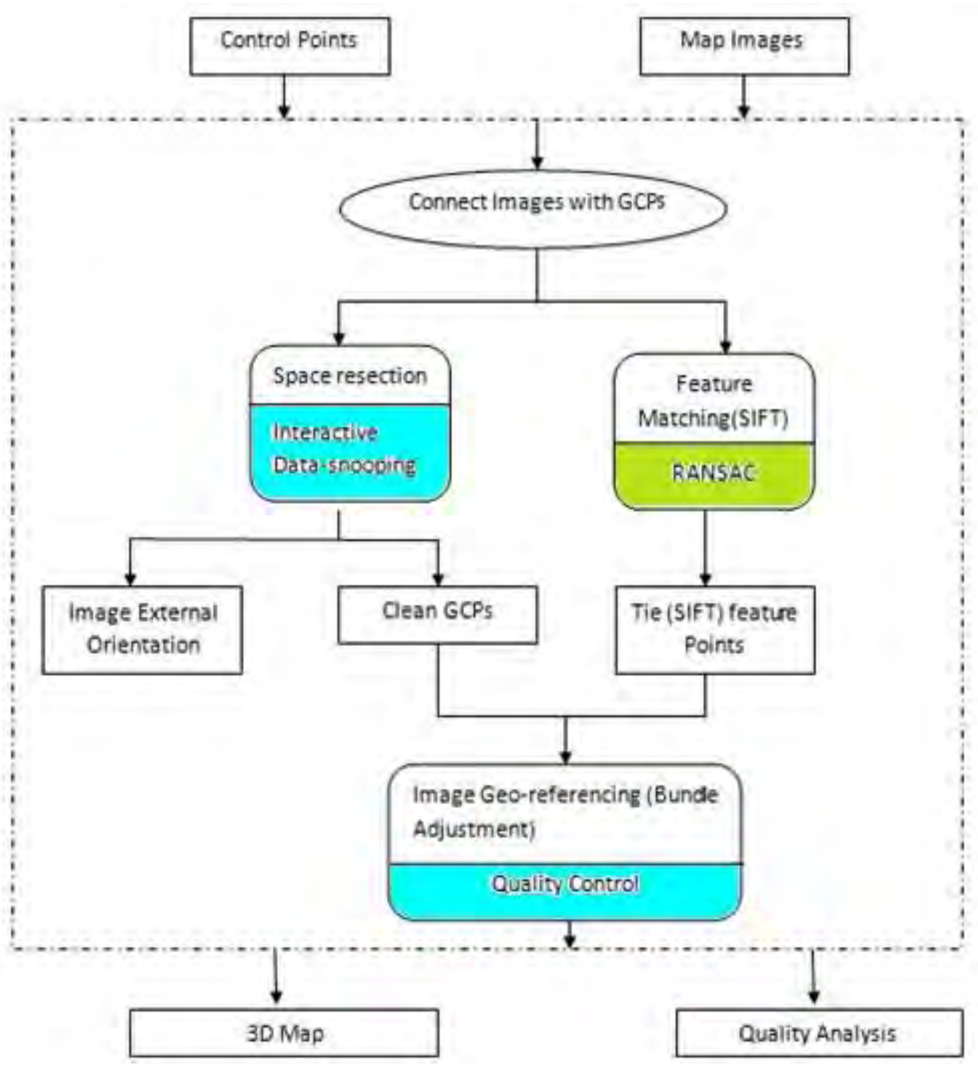


Figure 4. 2 Outlier detection for 3D mapping

4.3.2.1 Extended iterative data snooping

In traditional iterative data snooping, only one outlier is assumed to exist in each adjustment process. In other words, for each iteration, the observation with the biggest w-value is identified as the outlier. In the next iteration, it is removed and the whole adjustment process runs again. It continues until global model test passes and all the w-value is less than the critical value. Such method however, has so far hardly been applied to vision based navigation applications. This is mainly because it works well only when a small number of outliers exist in observations. In vision-based system, on the other hand, mismatches can bring large amount of outliers into the adjustment system. Since RANSAC/ modified RANSAC has been applied at image matching stage, data snooping only needs to deal with small number of outliers. But these observations are sometimes highly correlated. More specifically, here for photogrammetric adjustment which uses the image coordinates or/and their 3D object coordinates as observations in an adjustment procedure, observations of the same point (image 2D coordinates and 3D object coordinates) are correlated. Therefore, the iterative data snooping method is extended in this study by treating the observations from the same point as a unit. When one single observation detected as having gross errors, the whole unit of observations of this point will be removed together.

4.3.3 Outlier detection for positioning

At the navigation stage, another matching based on SIFT is carried out between the real time image and the map images. Mainly when any of the SIFT feature points from the map find its correspondence on the query image, the geo-information it carried can be transferred to its counterpart, which can later serve as pseudo ground control points (PGCPs) for positioning at the final stage. It is critical that mismatches are removed/ avoided during the PGCPs generation. Otherwise the final positioning will be severely affected. On the other hand however, the query image might be taken at significantly different place, angle or lighting condition, or using different devices compared with corresponding map images. Such difference poses great challenge for image matching as well as RANSAC process. Therefore, in this research a new method has been proposed to improve the performance of RANSAC in difficult matching conditions. It has been used for the matching at positioning stage and successfully improved the system performance (Section 5.4.4). More specifically, RANSAC based image matching will be evaluated by a test. Only when the evaluation test is past, the inliers from RANSAC are retained for PGCP generation. The method and its applications are introduced in Chapter 5.

After PGCPs have been generated and at the final positioning calculation, modified space resection (Soft Space Resection) is utilized to calculate vision sensor's external orientation in 6DOF. Meanwhile, extended data snooping are used for outlier detection at the least squares adjustment of space resection. The flowchart is shown in Figure 4.3, with the outlier detection procedures been highlighted.

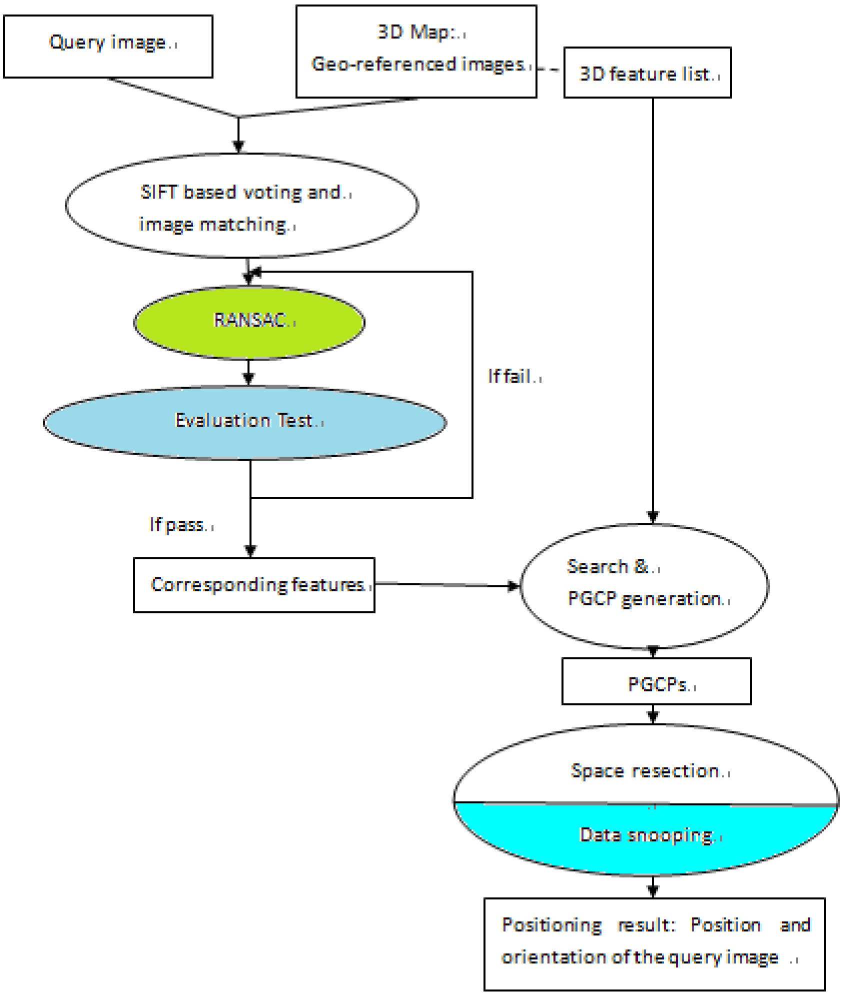


Figure 4. 3 Outlier detection for vision-based positioning

4.4 EXPERIMENTS

Three major experiments are obtained from our system development and explained here. They correspond to the three major parts of outlier detection in the system, namely: outlier detection on space resection and geo-referencing for 3D mapping, outlier detection for vision-based positioning. It is noted that the experiments here mainly discuss the outlier detection algorithm used in the least squares adjustment. For outlier detection during image matching, the RANSAC process and evaluation test proposed in this research are further discussed in Chapter 5.

4.4.1 Outlier detection on soft space resection for mapping

The first experiment was carried out at the mapping stage. Each image collected went through a soft space resection process in order to get their external orientation parameters, and more importantly, detect and remove outliers in the ground control point observations. The extended iterative data snooping was applied. Image 11 (Figure 4.4) was used as an example.

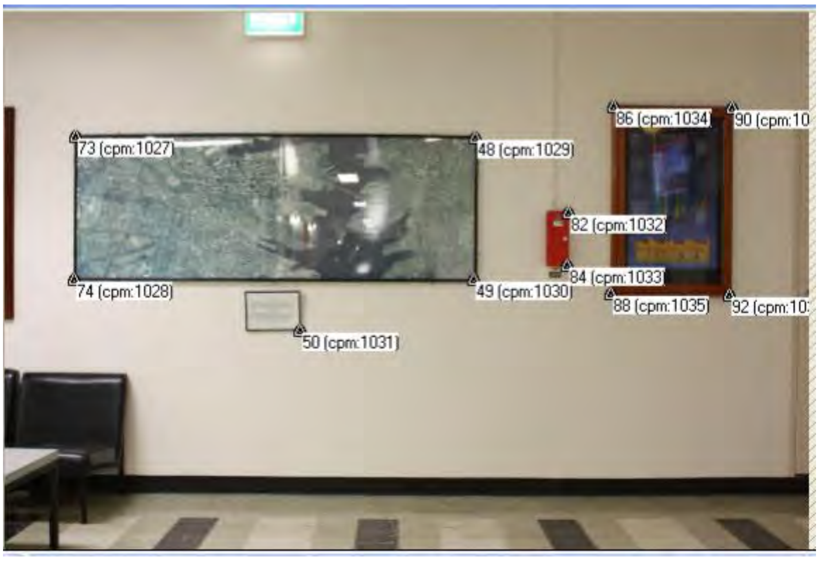


Figure 4. 4 Map image No. 11 with surveyed ground control points

Give confidence level α=0.1% $N(0,1;0.9995) = 3.29$. The extended iterative data snooping was carried out. Part of the result is shown in Table 4.1 with each column representing one iteration.

Table 4. 1 First two and last iteration results of data snooping for space resection

Iterative data snooping result	Iteration ID	1	2	....	7
	F-value	134.8002	114.8105	....	2.512323
Relevant Control Point ID	Observation	w-value			
48	1	-12.303	-9.665	.....	
	2	0.163	-1.097		
49	3	-20.935	-22.194		
	4	3.607	1.245		
..	..	..	..		..
90	19	-19.238	-25.755		
	20	-1.063	-5.193		
92	21	-24.372			
	22	-1.864			
48	23	6.775	3.586		
	24	12.305	9.66		
	25	0.079	-1.077		
49	26	21.103	11.33		
	27	20.954	22.195		
	28	3.466	1.287		
..	..	..	..		..
88	47	-10.524	0.529		-0.21
	48	-10.613	0.671		-0.071
	49	0.429	0.258		0.745
90	50	19.374	25.856		
	51	19.144	25.69		
	52	-1.216	-4.472		
92	53	24.21			
	54	24.364			
	55	-2.046			

In the first iteration, global model test failed. Observation No. 21, 53, 54 had the biggest w values. In fact, they are the measurements of the same point (GCP No.92). Using extended data snooping, GCP No. 92 is removed after the first iteration, which means observation No. 21,22,53,54,55 were all removed, 10 control points were left for the second iteration. The same process carried out iteratively until global model test passed and all the w value is less than the critical value. Finally 7 iterations were carried out until all the outliers have been removed. As a result, 5 control points left out of 11 and treated to be fixed values in next step: bundle adjustment for mapping (Section 2.3.3). For each image, space resection was carried out with the same outlier detection and removal process. The control points left in the end was then be used to calculate exterior orientation of each map image and used in the bundle adjustment for image geo-referencing.

4.4.2 Outlier detection for image geo-referencing

At mapping stage, the extended iterative data snooping is applied to detect and remove possible outliers so as to improve the accuracy of geo-referencing. This experiment was mainly designed to test the power of outlier detection function at this step.

Here the same sample data from Section 4.2.2 is used. SIFT matching was carried out between map Image No.10 and No.11, 13 pairs (No.1001~No.1013) of matched SIFT feature points were left and the rest were removed as mismatches by RANSAC. Then these 13 pairs of corresponding SIFT feature points were used as tie points and put into the bundle adjustment process. The final stage was to calculate the 3D object coordinates of these SIFT feature points (geo-referencing). A report of adjustment result along with accuracy analysis and outlier detection report were produced. Two iterations of data snooping were shown. The geo-referencing results after the first iteration are shown in Table 4.2.

Table 4. 2 Geo-referencing results after the first iteration

Tie point	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013
X(m)	3.230	4.182	3.821	3.094	3.110	3.109	3.231	3.231	3.233	3.233	3.104	3.179	11.28
Y(m)	4.566	2.959	3.235	4.398	4.405	4.411	4.595	4.595	4.646	4.646	4.326	4.614	-0.81
Z(m)	-1.18	-3.02	-2.96	-1.90	-1.86	-1.95	-1.03	-1.03	-1.04	-1.04	-1.89	-1.25	-1.54

Since all matched feature points in this example are on the same wall, it can be noticed easily that tie point No.1013 is wrong. Part of the data snooping result of the first iteration is shown in Table 4.3, in which σ_0 represents the prior standard deviation, σ the poster, f the degree of freedom and F the F ratio. Given the confidence level α=5%, F-test (global model test) fails at 7.2, which indicates certain outlier exists. It can be observed that observation No. 37, 38, 73 and 74 have the same maximum W-value. From the structure of matrices in the least squares solution, we get to know that these 4 observations are the image coordinates (x, y) of same feature point (tie point) No. 1013 on the two matched images No.10 and No. 11 respectively. More importantly, feature point No. 1013 is actually a mismatched feature point. Figure 4.5 shows it with a yellow line connecting its correspondences on two images. The test proves that data snooping method can further detect and remove mismatches in the dataset after the RANSAC process.

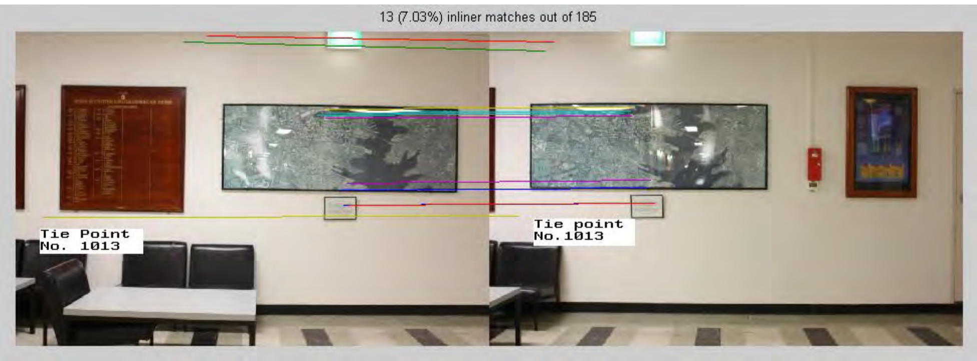


Figure 4. 5 SIFT feature pairs after RANSAC processing, tie point No.1013 has been identified as outlier by data snooping process

Table 4. 3 Data snooping results of iteration 1 for image geo-referencing

Summary of Data snooping result: iteration 1			
σ_0: 0.000025 σ: 0.000067 *f*: 23 F: 7.204570			
Index of internal reliability			
No	L	W	MDB
1	0.000064	5.285000	0.000212
…	…	…	…
37	-0.000002	-8.706000	0.013674
38	-0.000042	-8.706000	0.000529
39	-0.000059	-6.733000	0.000293
…	…	…	…
72	0.000006	0.361000	0.000153
73	0.000002	8.706000	0.010979
74	0.000045	8.706000	0.000496

So tie point No. 1013 was removed (observations No. 37, 38, 73 and 74) and 12 tie points were left the second iteration. The geo-referencing results along with part of the outlier detection report are shown in Table 4.4 and Table 4.5 respectively for the second iteration. And the iteration runs until all outliers have been removed.

Table 4. 4 Geo-referencing results after the second iteration

Tie	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012
X(m)	3.074	6.365	5.186	3.051	3.067	3.088	3.035	3.035	3.040	3.040	3.067	3.008
Y(m)	4.546	2.568	3.052	4.388	4.395	4.401	4.572	4.572	4.620	4.620	4.320	4.590
Z(m)	-1.185	-3.268	-3.100	-1.902	-1.859	-1.946	-1.046	-1.046	-1.053	-1.053	-1.888	-1.26

Table 4. 5 Data snooping results of iteration 2 for image geo-referencing

Summary of Data snooping result: iteration 2
σ_0: 0.000025 σ: 0.000039 *f*: 22 F: 2.446661

The power of the outlier detection function for geo-referencing is also reflected on the precision analysis. In this experiment, σ_0 (the prior) is the standard deviation of observations. Since the bundle adjustment is calculated using equally weighted photogrammetric image coordinates as observations, σ (the poster) reflects the precision of image measurements of control points and feature extraction and matching of tie feature points. So outliers from both sources will influence σ, deteriorate the precision of image coordinates. Comparing σ value between Table 4.3 and Table 4.5, σ has been reduced from 0.000067 to 0.000039. It proves that after the removal of mismatches, the precision of image coordinates has been increased.

It has also been noted that data snooping is suited to data with a small number of outliers like this example. For big number of outliers generated by image matching and left over by RANSAC, a specific method has been proposed and introduced in Chapter 5.

4.4.3 Outlier detection for vision-based positioning

Finally, the outlier detection for the final positioning is discussed. RANSAC with evaluation test will be further discussed in the next chapter. Here the outlier detection has been focused on the final least squares adjustment. Two tests are carried out, one uses global model test with data snooping based on simulated outliers to investigate the nature of the outliers at final positioning, one compare outlier detection strategies used for position calculation.

4.4.3.1 Simulation of outlier detection on positioning solution

At final positioning function, in which a soft space resection model is used, the 3D coordinates of PGCPs are treated as observed unknowns. Therefore, outliers mainly come from the two uncertain inputs of position estimation function: the 2D image coordinates and the 3D object coordinates of the PGCPs. The cause of outliers can be mismatches, or erroneous photogrammetric point determination in the 3D map developments.

This experiment mainly focused on the design and testing of outlier detection module in dealing with erroneous inputs of the positioning function. The procedure was simulated using outliers intentionally inserted into a clean dataset. The image measurement noise level (priory standard deviation) was set to 0.000014 when the PGCP set to 0.00095. Using the one tail global test, F (20, ∞; 0.95) approximately equals 1.57. F-ratio value from the clean dataset was 1.045675, smaller than 1.57 so the global test passed. Figure 4.6 and Figure 4.7 show the Minimal Detectable Bias (MDB) of each observation in the two groups of observations: the image observations and the PGCP coordinate observations. Observation No.1 to No.26 are the image observations while No.27 to No. 65 are 3D coordinates of the PGCPs. It can be seen that these two groups of observations have different levels of MDBs, and the observations within each group have close MDB values.

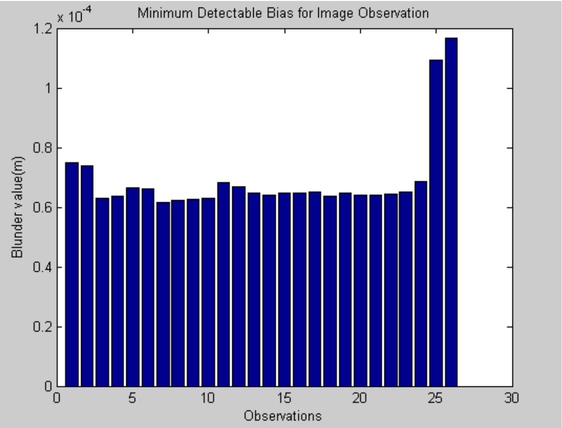


Figure 4.6 MDB for image observations

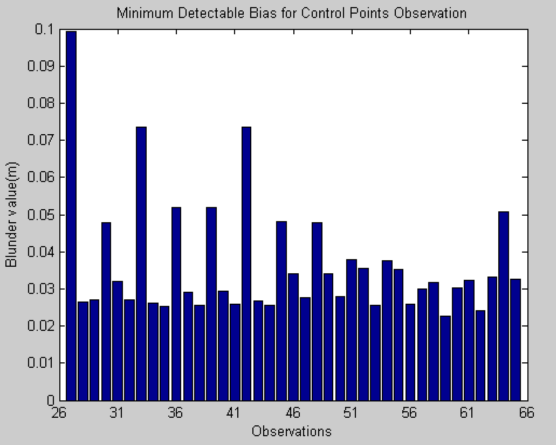
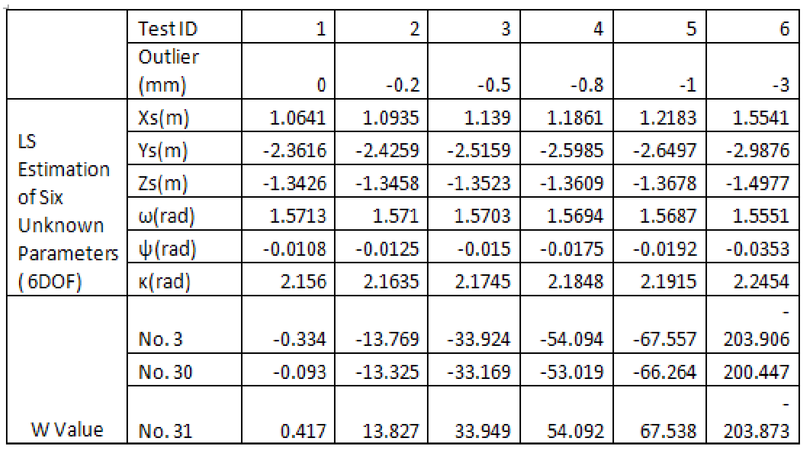


Figure 4. 7 MDB for PGCP observations

In order to simulate the impact of outliers in the first group of observation (image coordinates), outliers with different magnitudes were inserted intentionally into the dataset. The results were shown in Table 4.6 with an increasing magnitude of outliers inserted in the same image observation: observation No. 3, the image coordinate on the x axis of Point 2.

Table 4. 6 W values: outlier Detection in image coordinates

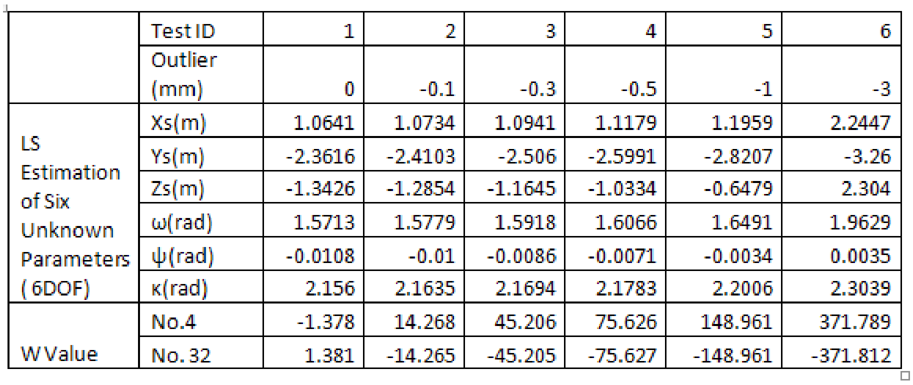


	Test ID	1	2	3	4	5	6
	Outlier (mm)	0	-0.2	-0.5	-0.8	-1	-3
LS Estimation of Six Unknown Parameters (6DOF)	Xs(m)	1.0641	1.0935	1.139	1.1861	1.2183	1.5541
	Ys(m)	-2.3616	-2.4259	-2.5159	-2.5985	-2.6497	-2.9876
	Zs(m)	-1.3426	-1.3458	-1.3523	-1.3609	-1.3678	-1.4977
	ω(rad)	1.5713	1.571	1.5703	1.5694	1.5687	1.5551
	ψ(rad)	-0.0108	-0.0125	-0.015	-0.0175	-0.0192	-0.0353
	κ(rad)	2.156	2.1635	2.1745	2.1848	2.1915	2.2454
W Value	No. 3	-0.334	-13.769	-33.924	-54.094	-67.557	- 203.906
	No. 30	-0.093	-13.325	-33.169	-53.019	-66.264	200.447
	No. 31	0.417	13.827	33.949	54.092	67.538	- 203.873

Firstly, by using the W-statistic to locate an outlier, it was noticed that two other observations (No.30 & 31) together with observation No. 3 all produced big W values. It is noted that observation No.3 is image coordinate of PGCP Point 2-x, when No. 30 and No. 31 correspond to the X, Y value in the object space of the same point (Point 2). The three observations can be highly correlated. By studying the absolute correlation coefficients between the W-statistics for observation pairs of No.3 and No.30, No.3 and No. 31, which is close to 1, it proves the correlation is extremely strong. Secondly, it was observed that when the magnitude of outlier grows, the probability of data-snooping method successfully identifying an outlier increases. According to the result, when the magnitude is greater than 0.8, outlier is always correctly identified by data-snooping.

More tests were carried with outliers in the 3D coordinates of the PGCPs: observation No. 32. The results are shown in Table 4.7. W value indicates either observation No.32 or No. 4 contain an outlier. The high correlation was found between the two observations, which are observations of the same point. Meanwhile, when the magnitude of outlier grows, the probability of data-snooping method successfully identifying an outlier increases.

Table 4. 7 W values: outlier Detection in 3D PGCPs object coordinates



	Test ID	1	2	3	4	5	6
	Outlier (mm)	0	-0.1	-0.3	-0.5	-1	-3
LS Estimation of Six Unknown Parameters (6DOF)	Xs(m)	1.0641	1.0734	1.0941	1.1179	1.1959	2.2447
	Ys(m)	-2.3616	-2.4103	-2.506	-2.5991	-2.8207	-3.26
	Zs(m)	-1.3426	-1.2854	-1.1645	-1.0334	-0.6479	2.304
	ω(rad)	1.5713	1.5779	1.5918	1.6066	1.6491	1.9629
	ψ(rad)	-0.0108	-0.01	-0.0086	-0.0071	-0.0034	0.0035
	κ(rad)	2.156	2.1635	2.1694	2.1783	2.2006	2.3039
W Value	No.4	-1.378	14.268	45.206	75.626	148.961	371.789
	No. 32	1.381	-14.265	-45.205	-75.627	-148.961	-371.812

In summary, observations of the same point, including both image measurements and 3D coordinates, can be highly correlated. Therefore the iterative data snooping has been modified to suit the scenario. It treats the observations from the same point as a unit, when one single observation detected as having gross errors, the whole unit of observations of this point will be removed together.

4.4.3.2 Outlier Detection using data snooping and M-estimators

In the second experiment for positioning, real world dataset is used. The aim is to compare different outlier detection strategies for final positioning. A calibrated video camera (Logitech Webcam Pro2000) was mounted on a moving vehicle with sampling rate at 1 HZ. Its relative position to the vehicle was fixed, which means the experiment was partially controlled: camera height (Z:-0.725m) and two angle of the camera attitude ($\upsilon = 0°, \omega = 90°$) were fixed. The positioning was conducted by extracting image frames from the video and matching them with the 3D map frame by frame. Each frame is an epoch, a position and orientation in 6DOF can be calculated via soft space resection. Baarda‘s data snooping and Huber‘s M-estimators were used respectively at the position calculation using an off-line post processing. Four typical epochs were chosen to illustrate different outcomes.

Firstly the epochs with no outliers been detection are investigated. As shown in Table 4.8, most of the epochs (like No.28) provide a reasonable positioning accuracy, while few epochs (like No.22) deviate from the controlled values as a result of bad PGCP geometry. As has been shown in Tables 4.9, DOP values clearly indicate the different geometric strength. Moreover, it is noticed that when no outliers were detected, positioning result stay consistent regardless of the outlier detection methods used.

Table 4. 8 Positioning results from epoch No.22 and No.28

	Epoch ID	22		28		Controlled
	Outlier	Data	M-estimato	Data	M-estimato	
Positioning result in 6DOF Unit: m &degree	X	-0.386	-0.387	-0.507	-0.507	-
	Y	2.871	2.871	-0.065	-0.065	-
	Z	-2.091	-2.091	-0.724	-0.724	-0.725
	ω	110.409	110.409	89.668	89.668	90
	φ	3.839	3.839	0.745	0.745	0
	κ	-106.513	-106.513	-88.522	-88.522	-
Number of detected PGCPs as outliers		0/8	-	0/20	-	

Table 4. 9 DOP values from epoch No.22 and No.28

	Epoch 22		Epoch 28	
	Data snooping	M-estimator	Data snooping	M-estimator
X DOP	19519	19562	808	801
Y DOP	44642	45031	427	412
Z DOP	49178	49636	372	377
P DOP	69227	69815	987	976
ωDOP	13508	13613	96	98
φDOP	3619	3629	206	203
κDOP	12082	12173	91	90
A DOP	18481	18619	245	242

When outliers were detected, some epochs had consistent positioning results from the two outlier detection methods while the others do not. Epoch No. 27 and No. 31 in Table 4.10 were used as examples. The reason for this is that for epochs in which PGCPs are in small number and geometry is weak, data snooping may remove important PGCP(s) which further deteriorate the geometry, then lead to worse precision and a positioning results further deviate from true values, or even an unconverged adjustment, which is exactly the case for epoch No. 27. The bad DOP values of epoch No.27 is shown in Table 4.11. The M-estimator, on the other hand, retain these points and reduce their weights, the final results therefore outperformed that of data snooping. This is similar to outlier detection for GPS networks. Compared with the robust test, a disadvantage for data snooping and mean-shift model methods in general, is that they remove outlying baselines which in turn deteriorate the shape of the network. If the control points removed by data snooping are located in strong geometric areas, the total PGCPs are in big number, then such influence will be minimized and positioning results produced by the two methods can still be consistent, as epoch No. 31 shown in Table 4.10 and Table 4.11.

Table 4. 10 Positioning results from epoch No.27 and No.31

	Epoch ID	27		31		Controlled
	Outlier	Data	M-estimato	Data	M-estimato	
Positioning result in 6DOF Unit: m &degree	X	-8806.39	-0.505	-0.543	-0.543	-
	Y	1794.963	0.221	-0.990	-0.989	-
	Z	-16080.3	-0.698	-0.882	-0.881	-0.725
	ω	33.288	89.095	91.960	91.960	90
	φ	17.629	0.057	0.286	0.286	0
	κ	-14.527	-88.121	-87.777	-87.777	-
Number of detected PGCPs as outliers/Total PGCPs		1/8	-	1/26	-	

Table 4. 11 DOP values from epoch No.27 and No.31

	Epoch 27		Epoch 31	
	Data snooping	M-estimator	Data snooping	M-estimator
X DOP	10999⍰22	3408	2492	2588
Y DOP	261658830	3471	5854	5974
Z DOP	1934781831	18007	7027	7157
P DOP	2240902045	18653	9480	9676
ωDOP	271354	4837	1800	1808
φDOP	950356	277	393	395
κDOP	201885	780	1582	1592
A DOP	1008745	4907	2428	2441

4. 5 SUMMARY

Vision sensor is inherently fragile against errors. Therefore, any vision-based system requires a robust outlier detection mechanism to ensure a good performance. In this chapter, different outlier detection strategies have been evaluated in the context of vision-based navigation: mainly Baarda‘s data snooping, Huber‘s M-estimator and RANSAC, which are dominating outlier detection methods in the field of photogrammetry and computer vision. The first two methods only work well with very few outliers. RANSAC is able to deal with large percentage of outliers, but has its only limitation. A multi-level operation scheme has been proposed for the system, including both quality control measures for 3D mapping and vision-based positioning. The main contribution is the combined use of various outlier detection methods in a multi-level manner to achieve an improved solution. More specifically, RASANC is used to remove most of the outliers, while data snooping/M-estimator is used at final adjustment process to gurantee the correctness of the input.

Experiments have revealed the nature of the outliers in the system and proved the efficiency of the outlier detection scheme. The simulation test showed that the observations from the same point can be highly correlated, therefore these observations are treated as a single unit in iterative data snooping. Moreover, the experiment proves that data snooping method can further detect and remove mismatches in the dataset after the RANSAC process. Meanwhile, tests also revealed some limitation of current system. For instance, a disadvantage for data snooping and mean-shift model methods in general, is that they remove outliers which in turn deteriorate the geometry. But data snooping is still chosen instead of M-estimator for the reason that divergence may occur when initialization or parameters are not chosen properly for M-estimator. More devoted studies may be required before M-estimator is used for vision-based navigation systems.

CHAPTER 5 ENHANCED RANSAC HOMOGRAPHY ESTIMATION WITH A CROSS CORRELATION TEST AND APPLICATIONS

5.1 INTRODUCTION

Image matching has been a fundamental problem for a variety of applications in photogrammetry, remote sensing, medical imaging, computer vision etc. Typical examples are image stitching and mosaicing, change detection, registration of satellite images, image fusion, vision-based navigation and so forth. Basically, it needs to geometrically align two images with overlapping areas that may be taken from different viewpoints, time, or different imaging devices. The core element is to establish the mapping function between two central perspective images. For many applications, it may be assumed that under most conditions the scene is approximately planar, thus the image transformation can be described by a planar projective homography (Negahdaripour, 2005). The homography transformation, also named projective transformation, has been used as a mapping function as well as a matching constraint for image correspondence. It transfers points from one view to the other so long as they are images of points on the plane. The quality of such homography is critical to the calculation of camera motion and relative orientation, which is widely applied to trajectory recovery based on structure from motion, and 3D reconstruction. When it used as a matching constraint to detect mismatches, it can also improve the quality of image matching. Therefore, despite the variety of intended applications, it is common but crucial task to compute the transformation function between two images that need to be matched and geometrically aligned.

According to the way image matching is performed, homography estimation methods generally fall into two groups: area-based and feature-based. Detailed description of the development in the literature can be found in Gruen (2012). The strength of the area-based methods lie in the fact that they consider global information and take every pixel in the image into account. Bergen et al. (1992) described a hierarchical framework for the estimation of image motion between two images using various models based on the minimization of sum of squares of differences (SSD). The main idea is that the motion estimate from one level of the image pyramid can then be used to initialize a smaller region search at the next level. These methods have been used to register images with pure translation, with more sophisticated motion like a homography, algorithms with better approximation of geometric transformation are adopted. Typical examples include Lucas-Kanade registration (Lucas and Kanade, 1981) and Adaptive Least Squares Matching (Gruen 1985). While it works well on image sequence or stereo pairs with short displacement, one significant problem for the area-based methods, however, is their incapability to deal with images with low overlap, large and complex transformation or significant intensity changes introduced by noise, varying illumination or different sensor types.

Feature-based matching methods, on the other hand, are typically applied when the local structural information is more significant than the information carried by the image intensities. Generally it consists of three steps: firstly distinctive features are detected and extracted from each of the images; secondly features are matched and correspondences are located across images based on feature characters; thirdly, geometric transformation is estimated based on correspondences. Once we have got a set of feature correspondences, motion parameters that best register two images can be estimated using these correspondences. There are two widely used solutions for robust homography estimation with outliers being removed. One is called RANSAC (Random Sample Consensus, Fischler and Bolles, 1981), which has been briefly introduced in Section 4.2.2, and the other is Least Median of Squares (LMS). Feature-based matching combined with RANSAC/LMS has become the dominant approach for image matching and registration in recent years for its invariant nature and capability to handle mismatches. While RANSAC counts the inliers based on the residuals smaller than a pre-defined threshold, LMS replaces the sum with the median of the squared residuals. However, such an approach has yet reached its full potential. The major limitation lies in that both methods start with a random subset of correspondences to estimate a motion model. If the initial selection is erroneous, it will lead to inaccurate or even false estimation of the homography and mismatches will be included as inliers.

Therefore, in this Chapter a method to evaluate and enhance the performance of RANSAC homography estimation is proposed. By integrating cross-correlation information between feature patches, poor estimation can be detected and removed. Moreover, accompanied with RANSAC, this method can largely improve the correctness of image matching and can be applied to a great variety of applications where high quality feature-based matching is used.

5.2 SIFT BASED RANSAC HOMOGRAPHY CALCULATION AND IMAGE MATCHING

Feature-based image matching and registration methods have achieved growing attention because of their ability to tolerate low image overlap and image scale changes (Wu and Fang, 2007). One popular algorithm, SIFT (scale-invariant feature transform) was developed in 1999 (Lowe, 1999) as highly distinctive features that are used to perform reliable matching of the same object or scene between different images. Because of its invariance against image translation, scaling, rotation, and partial invariance to illumination changes and affine or 3D projection, it has been widely used in a variety of applications such as object recognition, robotic mapping and navigation, image stitching, 3D modelling and video tracking.

The SIFT algorithm generates a descriptor for each feature point using local image gradients within the neighbourhood at a selected scale. The good point about it is that the resulting descriptors are highly distinctive since they contain large amount of information and this improves correct matching between features in different images. However, one major problem with it is that it only considers local information and may contain large number of mismatches in the dataset. The reason is that feature-based matching such as SIFT is performed through a Euclidean-distance based nearest neighbour search of feature descriptors (128 dimension vector). And the property of such vector is only extracted from local information. When there are repeated/similar patterns, the features within the region tend either have more than one nearest neighbour or nearest neighbour that from a false area of the matching image because of the similarity of the local patch. This leads to a major challenge for determining the relative transformation: large percentage of mismatches are involved in the point correspondence. Once applied, they will result in inaccurate or even false estimation of the homography, as shown in Figure 5.1.

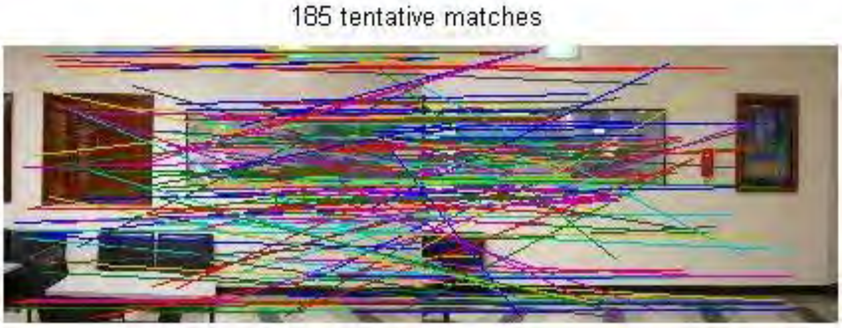


Figure 5. 1 SIFT matching between two images

One common approach is to use RANSAC (the random sample consensus) to filter out mismatched pairs so as to robustly estimate homography and improve correspondences registration robustness. The algorithm runs in several steps: first a number of iterations are performed. Within each iteration, 4 initial pairs are randomly chosen and a homography H is built up on it. Secondly, within each iteration, all other correspondences are then classified as inliers or outliers depending on its concurrence with H. After all the iterations are done, the one with the biggest number of inliers are retained and homography H is rebuilt from the inliers selected by this iteration. Thus mismatches are filtered out as outliers and homography is estimated by the assumed correct matches. An example of RANSAC removing mismatches is shown in Figure 5.2.

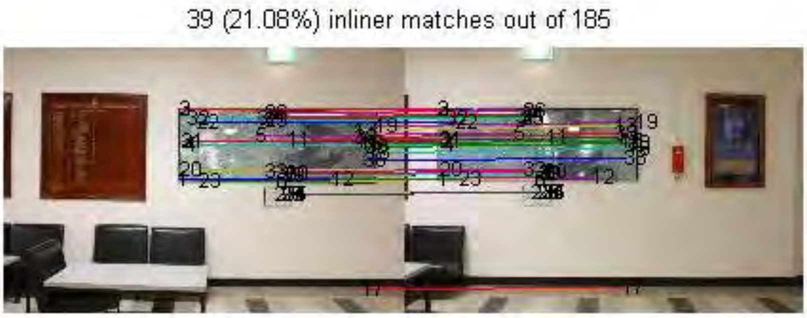


Figure 5. 2 RANSAC process to remove outliers for image matching

However, does this method successfully solve mismatch problem incurred by matching ambiguities? The answer is no. An example has been given in Section 4.2.2, in which a small number of mismatches are involved. Here a worse scenario is illustrated. As shown in Figure 5.3, RANSAC fails to provide a reasonable result. Most of the inliers remained are mismatches (e.g. No.1, No.2 and so forth). Obviously the two matched images have small overlap and repeated patterns can easily be found, such as black frames of the boards, chairs and strip lines on the floor. As a result, a great number of mismatches are produced during SIFT based matching, and RANSAC performs poorly.

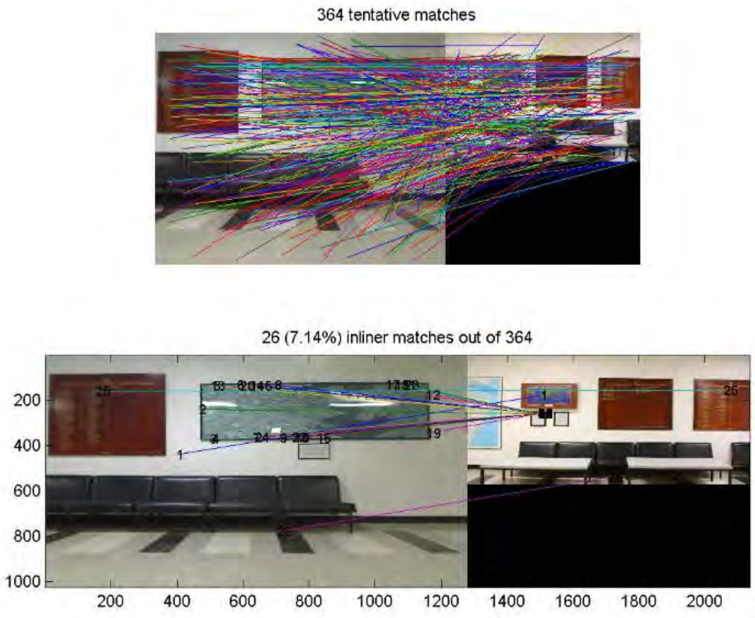


Figure 5. 3 RANSAC process for image matching

RANSAC's failure to remove mismatches mostly due to its random sampling nature. If any of the false matches are selected as the initial putative match in an iteration, and this iteration happens to be the one with biggest number of inliers within limited number of iterations, the inliers selected by RANSAC will have a big chance to be wrong and the final homography estimation will be poor. This is especially true when the two matched images have repeated patterns or low overlap that lead to matching ambiguity, or strong transformation to incur mismatches. Hartley and Zisserman (2003) show that the probability that a sample correspondence is an outlier can be calculated if the proportion of outliers is known. However, in most of the cases we have no idea of the percentage of mismatches involved; or the number of iterations that been able to statistically insure the correctness of RANSAC homography estimation is too high and incurs unaffordable computation load for the application. The major challenge here is that if there are large percentage of outliers in the training samples, any estimation method will have a high risk of failure.

Therefore in this Chapter, an evaluation strategy is introduced to qualify the performance of RANSAC homography estimation. Through such measurement, we are able to distinguish good modelling and poor modelling without the need for priory knowledge of the false rate.

5.3 ENHANCING RANSAC HOMOGRAPHY ESTIMATION WITH A CROSS-CORRELATION TEST

In this study, the author proposes to integrate area-based method into the feature matching process to strengthen the robustness of the matching algorithm against mismatches and noise. More specifically, cross-correlation information is used as an analysis and selection criterion for the matching. Instead of identifying mismatch(es) after it has been generated, it determines how good the homography model (H) is for the two matching images and discard bad H to reduce chances that mismatches are included. The basic idea is to generate patches around each SIFT matched points (named as feature patch) and calculate the normalized correlation coefficient between each pair. Then the significance tests of correlation coefficients are used to qualify the values of the correlation coefficient.

In RANSAC, we use 2-D projective transformation H (planar homography) to approximate the geometric transformation between two images (e.g. I and $\mathrm{I}^{'}$). Any two corresponding SIFT features in images I and $\mathrm{I}^{'}$ that pass RANSAC will comply with the model, which ideally can be expressed as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \\ c1 & c2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{5.1}$$

In (5.1) $\mathrm{p} = [\mathrm{x}, \mathrm{y}, 1]^{\mathrm{T}}$ and $\mathrm{p}' = [\mathrm{x}^{'}, \mathrm{y}^{'}, 1]^{\mathrm{T}}$ denotes the two points expressed using homogeneous coordinates, and $\begin{bmatrix} \mathrm{a1} & \mathrm{a2} & \mathrm{a3} \\ \mathrm{b1} & \mathrm{b2} & \mathrm{b3} \\ \mathrm{c1} & \mathrm{c2} & 1 \end{bmatrix}$ represents homography model H.

$\begin{bmatrix} \mathrm{a1} & \mathrm{a2} \\ \mathrm{b1} & \mathrm{b2} \end{bmatrix}$ parameterized affine changes, $[\mathrm{a3} \quad \mathrm{b3}]^{\mathrm{T}}$ shift parameters and $[\mathrm{c1} \quad \mathrm{c2}]$ projective deformation.

After homography model has been generated by RANSAC processing, a local square patch in image I with a size of $(2\mathrm{w}+1)*(2\mathrm{w}+1)$ centered on p is generated, denoted as $\mathrm{N(p)}$. $\mathrm{N(p)} = (2\mathrm{w}+1)*(2\mathrm{w}+1)$By using the estimated H, $\mathrm{N(p)}$ is transformed into $\mathrm{N(p')}$ and resampled on image I'. Then cross-correlation between the two window patches is calculated using (2), where $\mathrm{G_{uv}}$ and $\mathrm{G'_{uv}}$ represent the intensity values of the two correlation windows, respectively, whereas $\square(\mathrm{G})$ and $\square(\mathrm{G'})$ denote their average intensity.

$$\mathrm{r}(\mathrm{G},\mathrm{G}') = \frac{\sum_{\mathrm{u}=-\mathrm{w}}^{\mathrm{w}} \sum_{\mathrm{v}=-\mathrm{w}}^{\mathrm{w}} \left(\mathrm{G_{uv}} - \square(\mathrm{G})\right)\left(\mathrm{G'_{uv}} - \square(\mathrm{G'})\right)}{\sqrt{\sum_{\mathrm{u}=-\mathrm{w}}^{\mathrm{w}} \sum_{\mathrm{v}=-\mathrm{w}}^{\mathrm{w}} \left(\mathrm{G_{uv}} - \square(\mathrm{G})\right)^2 \cdot \sum_{\mathrm{u}=-\mathrm{w}}^{\mathrm{w}} \sum_{\mathrm{v}=-\mathrm{w}}^{\mathrm{w}} \left(\mathrm{G'}_{\mathrm{uv}} - \square(\mathrm{G'})\right)^2}} \qquad (5.2)$$

In Eq. (5.2), $\mathrm{r}(\mathrm{G},\mathrm{G}')$ varies from -1 to 1, the closer to 1 the higher correlation, the bigger similarity between two patches and greater chance to be correct corresponding points. However, one essential question is: how significant is the relationship that can be treated as correct match? We wish to quantitate the degree of the association so as to differentiate the erroneously matched feature patches.

When the two matched images are identical, the correlation coefficient value will be equal to one everywhere. But generally two matched images are taken with different viewpoint, probably different illumination condition, scale etc. Influenced by these factors, any two correctly matched patches on the images may have different values, not to say mismatches. Such fact poses greater challenge to the problem.

To tackle this problem, the first strategy adopted is to treat each matched patch pair as a sample r, and there is a hypothesized population correlation of correctly matched patches. Then a significance test is performed to test if the sample is from the population of correct matches, or significantly different from the group. Because of the difference between two images, the population correlation coefficient ρ for correct matches should be close to one. Although the cross –correlation may vary theoretically from -1 to 1, most of the sample values lie near ρ with $\rho \neq 0$. The sampling distribution of r is very skewed, as shown in Figure 5.4.

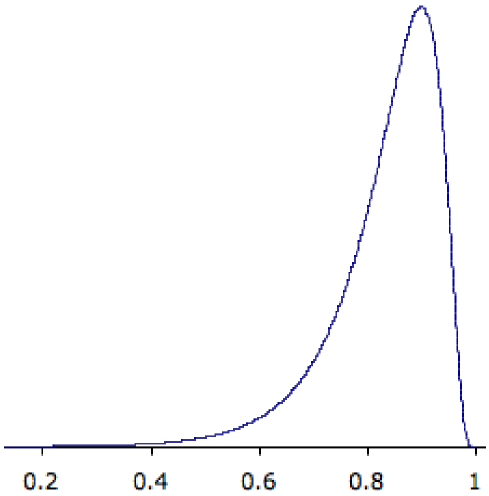


Figure 5. 4 Example of the sampling distribution of r for N = 12 and ρ = 0.90.

Since r is not normally distributed, Fisher transformation is used to form a new statistic with the following formula (Devore, 2012):

$$z_r = \frac{1}{2}\ln\frac{1+r}{1-r} \tag{5.3}$$

The transformed value z_r has an approximately normal distribution. The statistic for testing H_0: $\rho = \rho_0$ is

$$Z = \frac{z_r - \frac{1}{2}\ln[(1+\rho_0)/(1-\rho_0)]}{1/\sqrt{N-3}} \tag{5.4}$$

Now we can use the Z value to determine whether the correlation coefficient r between the two patches is significantly different from the hypothesized population correlation ρ_0 so as to tell if the match is correct. Since the exact ρ is unknown and varies for every pair of corresponding patch, a minimally acceptable value is used.

Rather than deciding if single patch pair is correct or not arbitrarily, we choose to evaluate the image matching based on the overall performance. In other words, the homography model built by the image matching process is evaluated; the model is discarded if the percentage of uncorrelated ―corresponding patches” exceed a certain threshold. Mismatches got involved after the RANSAC process mostly due to the fact that the homography model is incorrect. An obvious benefit for this model-based method over single patch based approach is that it tackles the problem from the root. If the model is poorly estimated by RANSAC, which result in a big number of mismatches left in the dataset, a patch-based approach can hardly guarantee the correct detection of mismatches since the exact ρ is unknown and the minimally acceptable value is based on general understanding of correlation; on the other hand, when a model-based approach is used, the overall performance is considered, the poor model will fail the test with the result being discarded and images re-matched, which ensures the final estimation is based on a correct homography model.

A multi -step strategy has been proposed:

Table 5. 1 The procedure of the cross-correlation test

1) Perform feature based image matching between two images I and I';
2) Use RANSAC process: n pairs of corresponding points(inliers) are found.
   If image matching performed is the 4th time of the same pair, we determine the image pair cannot be correctly matched, break;
3) Generate a square feature patch of N pixels around each feature point ;
4) Calculate the cross-correlation for each feature patch pair: r_i, $i = 1, \cdots (n - 1, n$.
5) From $(r_1, \cdots, r_n)$, obtain the max value r_{max} ;
   If r_{max} indicates its population is at least moderately positively correlated, proceed; Otherwise, current iteration terminates and the procedure goes back to (2).

The statistic test is as follows:

$$z_{\max} = \frac{1}{2}\ln\frac{1+r_{\max}}{1-r_{\max}}$$

We wish to test H_0: $\rho = 0.5$ versus H_0: $\rho > 0.5$ at the significant level of 0.05, then the critical value is 1.645:

$$\text{Z\_maxscore} = \frac{z_{\max} - \frac{1}{2}\ln[(1+0.5)/(1-0.5)]}{1/\sqrt{N-3}}$$

If $\text{Z\_maxscore} \geq 1.645$, proceed
Otherwise goes back to (2);

6) For every feature patch pair i :
   a) Obtain z_i from r_i using Eq.(5.3) ;
   b) If r_i indicates its population is not at least moderately positively correlated, feature patch pair i is treated as a false match. The statistic test is as follows:
   We wish to test H_0: $\rho = 0.5$ versus H_0: $\rho < 0.5$ at the significant level of 0.05.

$$\text{Z\_iscore} = \frac{z_i - \frac{1}{2}\ln[(1+0.5)/(1-0.5)]}{1/\sqrt{N-3}}$$

If $\text{Z\_iscore} \leq -1.645$, the i^{th} match is treated as false match;
Otherwise the i^{th} match passes the test.

7) Calculate the pass rate P_{rate} based on all n pair of matches;
   If $P_{rate} > threshold(0.8\ by\ default)$, we determine the current homography model is correct and the two images are correctly matched;
   Otherwise the current Homography model is false and the procedure goes back to (2).
8) For image pair with correct homography model, delete false matches found in (5).

In this algorithm, two main parameters need to be set in advance. One is the local square patch size N, the other is the threshold for P_{rate} . The choice of target sample size N is critical. If it is too small, the approximation based on Fisher transformation will not be valid. If the patch size is too large, the efficiency of the algorithm will be affected. Meanwhile, N is closely related to the resulting cross-correlation value r_i of the patch pair. When N increases, r_i decreases. For one tail test at level 0.05 and critical value at 1.645, the relationship between N and r_i is shown in Figure 5.5.

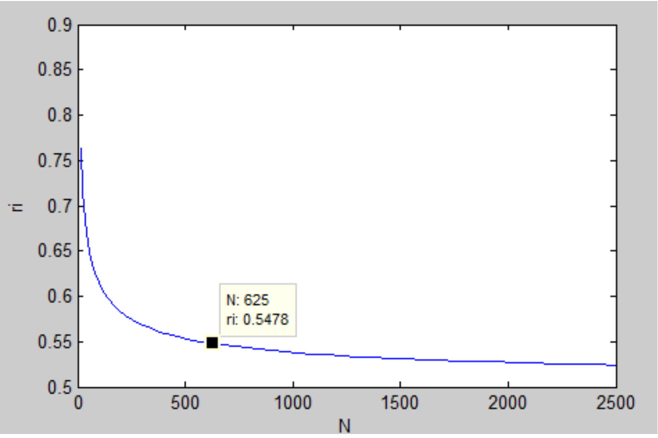


Figure 5. 5 The relationship between sample size N and cross-correlation value r_i

It can be observed from Figure 5.5 that when N is very small, the cross-correlation value r_i changes very sharply. When N increases to certain value, the change of r_i becomes very small and r_i tends to be stable. Since N is a local square patch with a size of $(2\mathrm{w}+1)*(2\mathrm{w}+1)$, in which $(2\mathrm{w}+1)$ is the patch width, therefore it depends on the value of w. When $\mathrm{w} \geq 12$ $(\mathrm{N} \geq 625)$, r_i drops under 0.55 and becomes relatively stable. Considering the efficiency, the recommended w is between 12 and 20. And in this study, the default value for w is 13 and the corresponding N is 729.

In order to determine the suitable $\mathrm{P}_{\mathrm{rate}}$ threshold, experiments are performed. A sequence of images, which consists of 79 images, is matched with another 24 images with/without common areas. The proposed method has been applied with pass rate $\mathrm{P}_{\mathrm{rate}}$ for each matching been recorded. If the test fails at the first condition $(Z_{maxscore} < 1.645)$, the pass rate is at 0. In total 1219 times of matching were

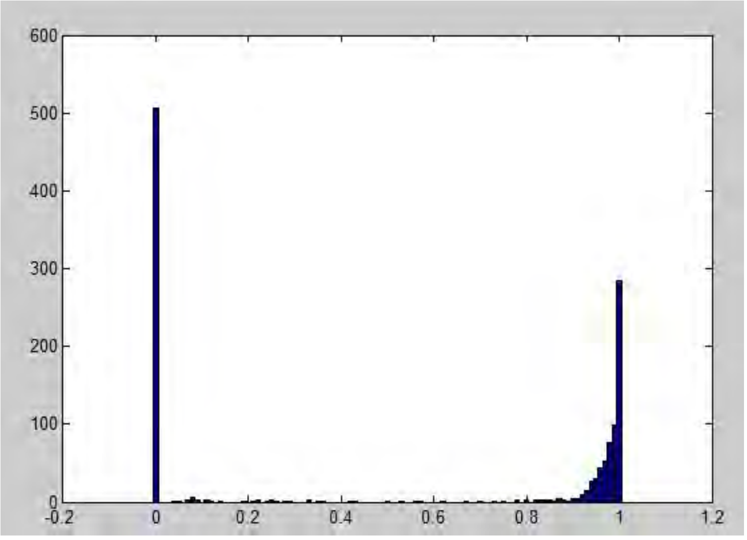


Figure 5. 6 Distribution of pass rate for the 1219 times of image matching

Table 5. 2 Pass rate from 0 to 1 with 0.1 interval

Pass rate interval	0-0. 1	0. 1-0. 2	0. 2-0. 3	0. 3-0. 4	0. 4-0. 5	0. 5-0. 6	0. 6-0. 7	0. 7-0. 8	0. 8-0. 9	0. 9-1
Number of pass rates	519	6	10	5	2	4	2	5	19	647
Distribution of pass rate	42. 6%	0. 49%	0. 82%	0. 41%	0. 16%	0. 33%	0. 16%	0. 41%	1. 56%	53. 1%
Average correct rate	0. 00%	0. 00%	1. 00%	0. 00%	0. 00%	46. 3%	86. 5%	87. 8%	96. 9%	100%

5.4 EXPERIMENTS

In the first two experiments, the methodology proposed in Section 5.3 is tested and evaluated. In the last two experiments, the methodology is applied to certain applications.

5.4.1 Evaluation of the test

In the first experiment, the strength of the test is evaluated. It is focused on how the test performs against the random nature of RANSAC. In other words, the aim is to find out whether the test can tell good RANSAC estimation from bad ones.

Because the inliers RANSAC retained as well as the estimation of homography largely depend on the initial selection, we may get different RANSAC matching results from the same pair of images. Therefore, here the same pair of images is used but run RANSAC with the test proposed several times to get different results. The first test fails to pass because its maximum cross correlation $\mathrm{r}_{\max}$ is 0 (Figure 5.7). It can be observed that the inliers RASNAC retained are indeed mismatches and the homography has been false. The test successfully excluded the bad RANSAC estimation.

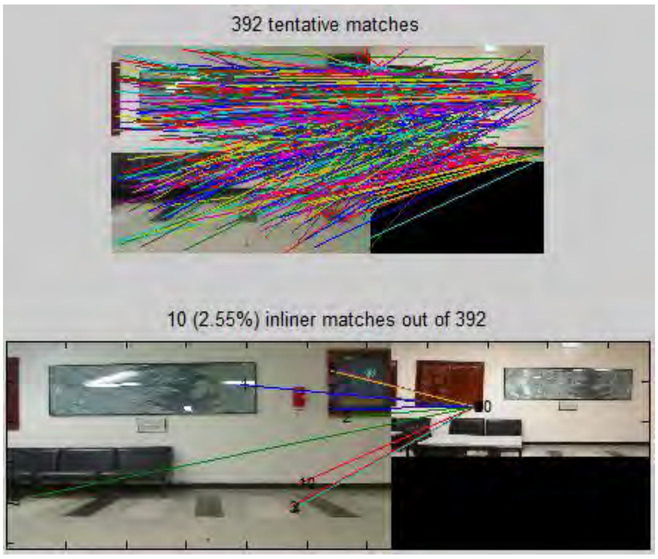


Figure 5. 7 The first evaluation test fail with r_{max} at 0.00

If we run the test again, only very few inliers are retained by RANSAC in the second one, as shown in Figure 5.8. The pass rate is at 40% and mismatches are identified as No. 1 and No. 3-7. The test fails to pass because the threshold 70% has yet been reached. It can be observed that No.1 is obviously mistaken, while the rest have been incorrect rejection. Therefore, the author believes the test has been too ―harsh” on the mismatch identification.

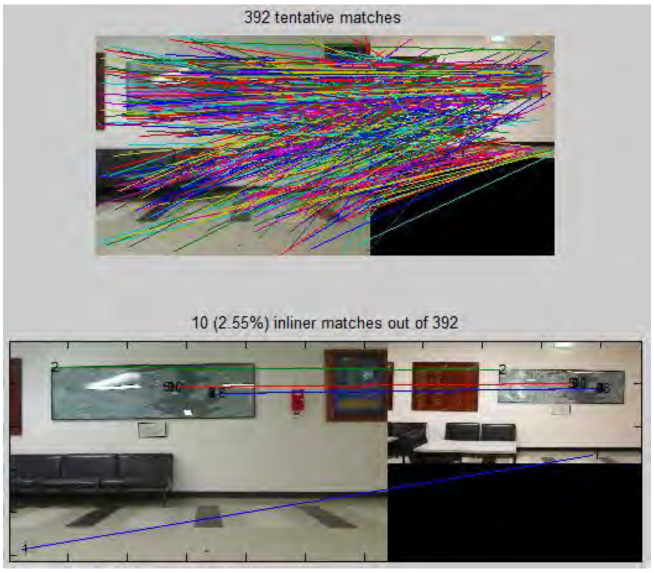


Figure 5. 8 The second evaluation test fail with pass rate at 40%

In the third case, as shown in Figure 5.9, the test passes with 100% pass rate, which means all the inliers are correct matches. It can be observed that the RANSAC indeed provides a satisfactory result.

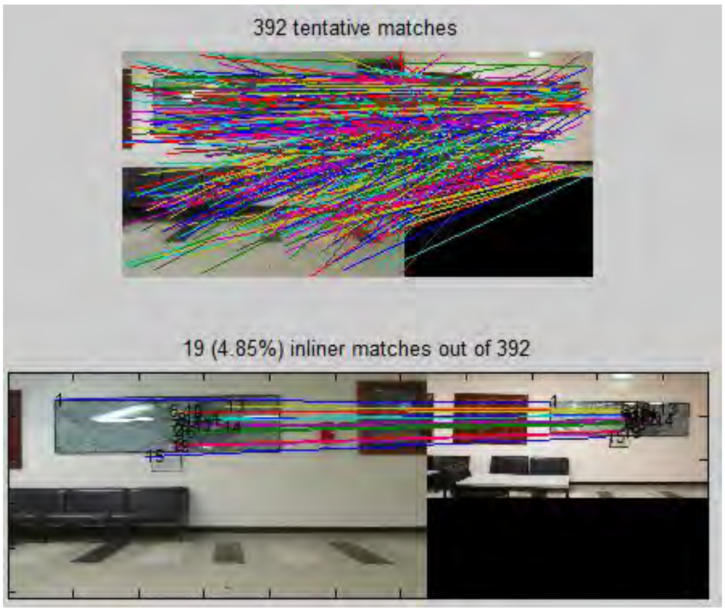


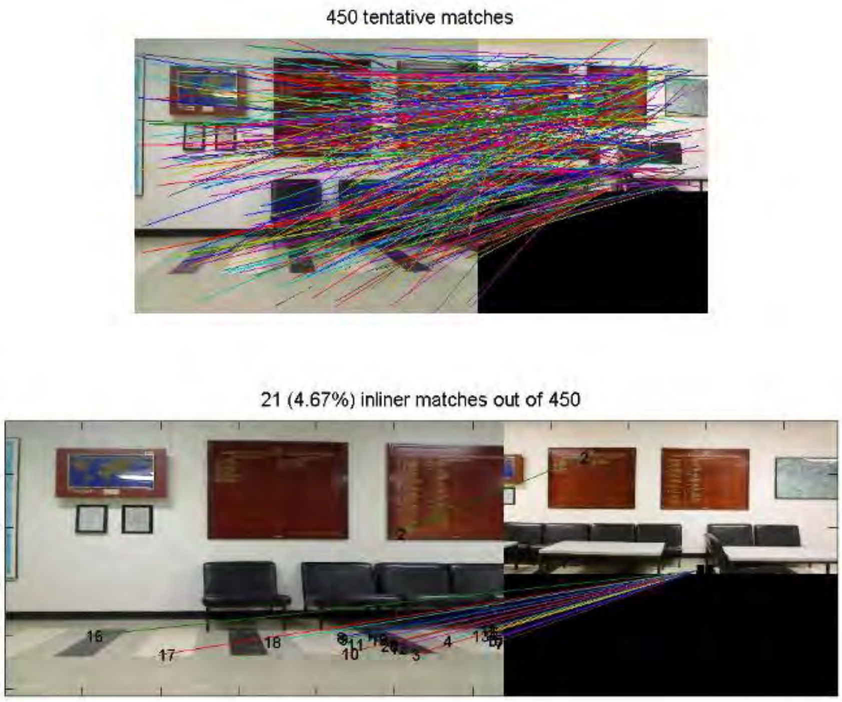
Figure 5. 9 The third evaluation test pass with pass rate at 100%

In summary, the author believes that the test set a harsh bound for the RANSAC estimation. It correctly rejects the false estimation. For estimation that also involves small number of mismatches, it rejects. Only the fine estimation with a high pass rate is retained. In the algorithm we proposed, we repeat the RANSAC process until either a fine estimation is obtained, or after certain times the matching pair are discarded and identified as wrong pair. Disregard the efficiency of the algorithm, the number of tolerable times can be increased.

5.4.2 Capability of the test for difficult image matching scenarios

Using the proposed method, we perform image matching using different image pairs. The focus has been on testing the capability of the proposed methodology against difficult circumstances for matching: matching ambiguity because of low overlap or repetitive patterns, and strong transformation.

Firstly, matching ambiguity result from repetitive patterns is tested. As shown in following two matching pairs in Figure 5.10 and 5.11, both pairs contain rich repetitive patterns: wooden boards and stripe patterns on the floor. It can be easily observed that in the first pair, RANSAC performs poorly since most of the inliers are mismatches. It has failed the test since its maximum cross correlation $r_{\max}$ is 0, which cannot prove its population is at least moderately correlated. The other pair, however, passed the test at a pass rate of 88%. It can be noticed that in the second pair, RANSAC indeed produce a good result.


450 tentative matches

21 (4.67%) inliner matches out of 450

Figure 5. 10 Evaluation test fail with r_{max} at 0.00

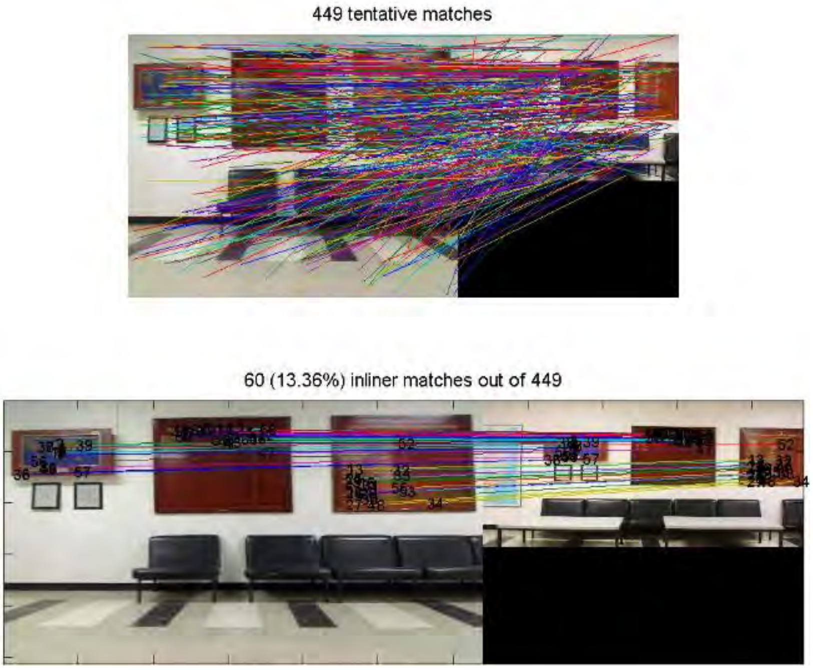


Figure 5. 11 Evaluation test pass: pass rate at 0.88

Secondly, matching ambiguity results from low overlap are tested. As shown in following two matching pairs in Figure 5.12 and Figure 5.13, both pairs have very limited overlapping areas. It can be easily observed that in the first pair, RANSAC performs poorly since most of the inliers are mismatches. It has failed the test since its maximum cross correlation r_{max} is 0, which cannot show its population is at least moderately correlated. The other pair, however, passed the test at a pass rate of 95% with one pair of points (No.1) being identified as mismatch. It can be noticed that in the second pair, RANSAC indeed produce a good result.

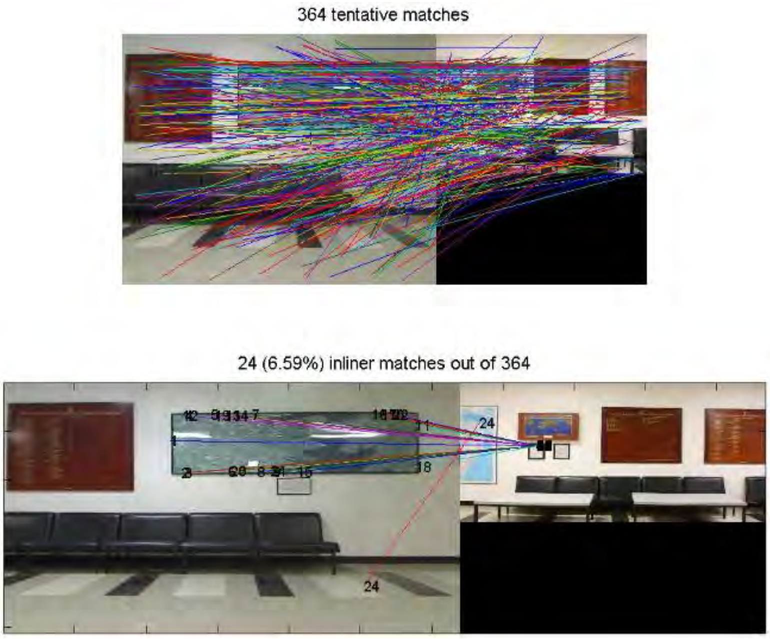


Figure 5. 12 The evaluation test fail with r_{max} at 0.00

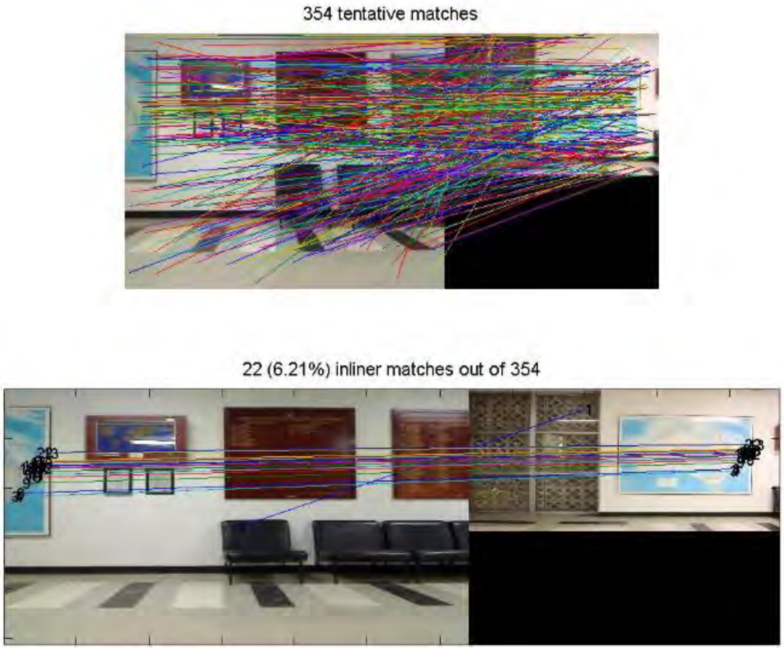
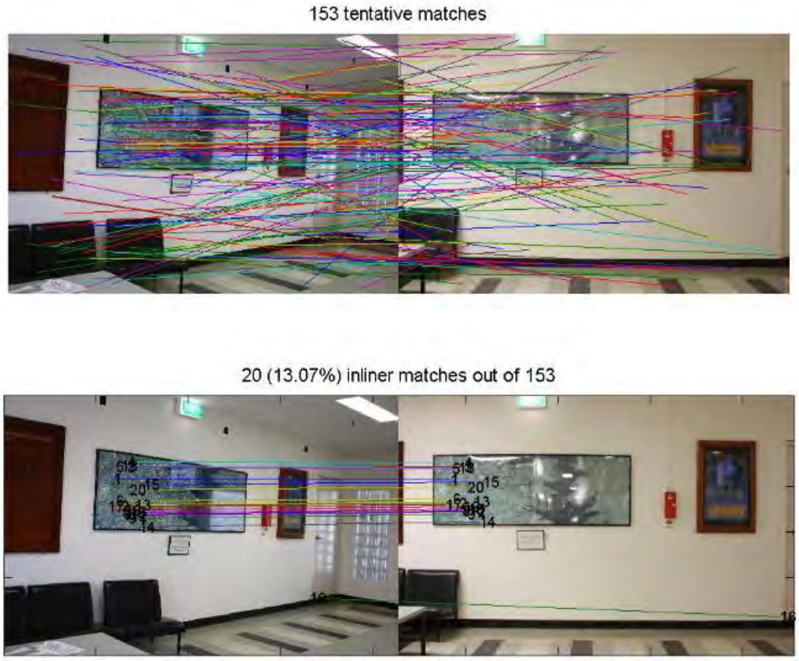


Figure 5. 13 The evaluation test pass at pass rate 95%

In the third test, image matching with RANSAC against large viewpoint changes is tested. As shown in Figure 5.14, the pair has a 50 degree viewpoint shift. It has passed the test at a pass rate of 100%. It can be noticed that RANSAC indeed produce a good result.



153 tentative matches

20 (13.07%) inliner matches out of 153

Figure 5. 14 The evaluation test pass at pass rate 100%

In summary, the test can effectively help prevent poor RANSAC performance and false homography estimation under circumstances like matching ambiguity and dramatic transformation.

5.4.3 Application in image stitching and mosaicing

In the third experiment, the RANSAC based homography estimation method is applied to image stitching and mosaic. A sequence of images is extracted from a video recorded by a moving platform. Adjacent image pairs are matched and stitched to form a mosaic. A total of 77 pairs are matched and 77 mosaics are produced. The performance for each pair is evaluated and the pass rate is shown in Figure 5.15.

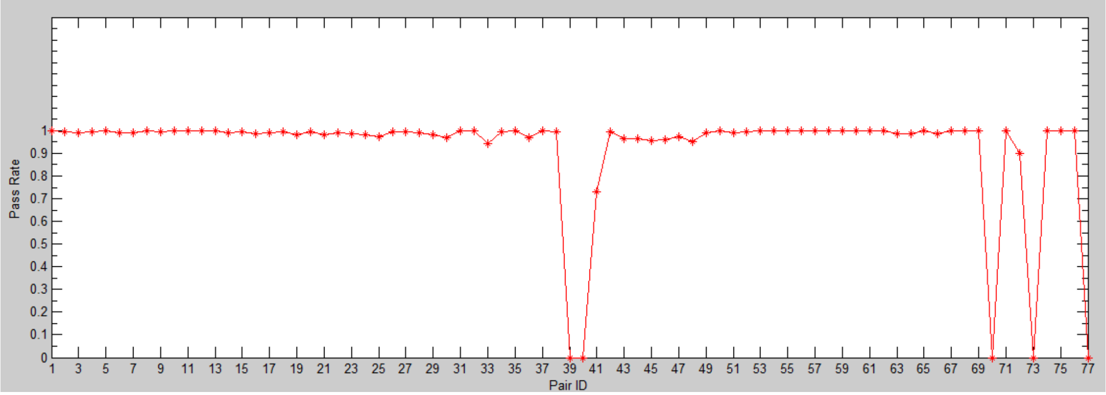


Figure 5. 15 Pass rate of the evaluation from 77 pairs of matching images

It can be observed that most of the pairs past the test with a high pass rate. Five pairs failed. We pick 4 typical pairs, Pair No. 39-42, to illustrate the performance of the test. The test results of the 4 pairs are shown in Table 5.3.

Table 5. 3 Performance of the evaluation test for pair No.39-42

Pair ID	Image Pair ID	Test Result
39	44,45	Fail: r_{max} at 0.33
40	45,46	Fail: r_{max} at 0.00
41	46,47	Fail: pass rate at 73%
42	47,48	Pass: pass rate at 99%

The first two pairs failed the test because their maximum cross correlation r_{max} was too small. To further investigate their image matching performance, the matching and mosaic results are shown in Figure 5.16 and Figure 5.17. It can be easily observed that RANSAC didn‘t provide good estimation thus image stitching failed.

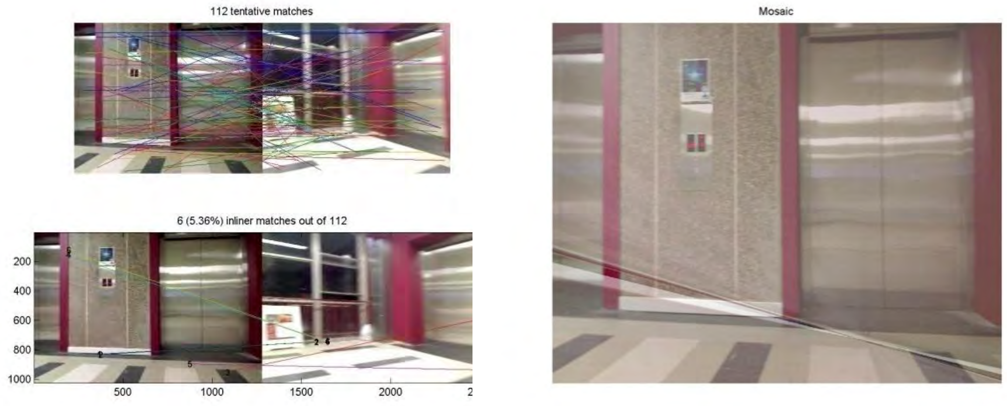


Figure 5. 16 The image matching of Pair No. 39 and its mosaic

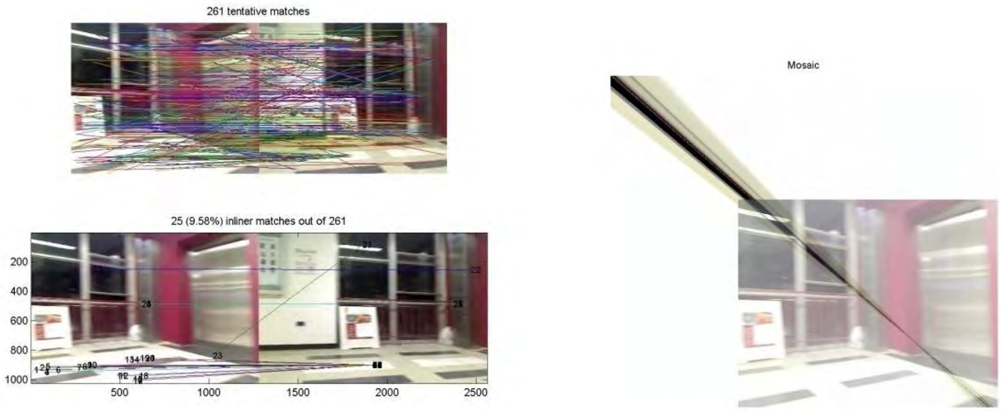


Figure 5. 17 The image matching of Pair No. 40 and its mosaic

Image pair No. 41 and No.42 have different pass rates. The matching and mosaic results are shown in Figure 5.18 and Figure 5.19. It can be observed that both pairs provides reasonable image stitching result (mosaic), with pair No. 42 outperform that of No. 41. Still pair No. 41 failed the test with a pass rate close to 80%. Comparing with two previous pairs, it can be noticed that the test can correctly evaluate the performance of image stitching and mosaic.

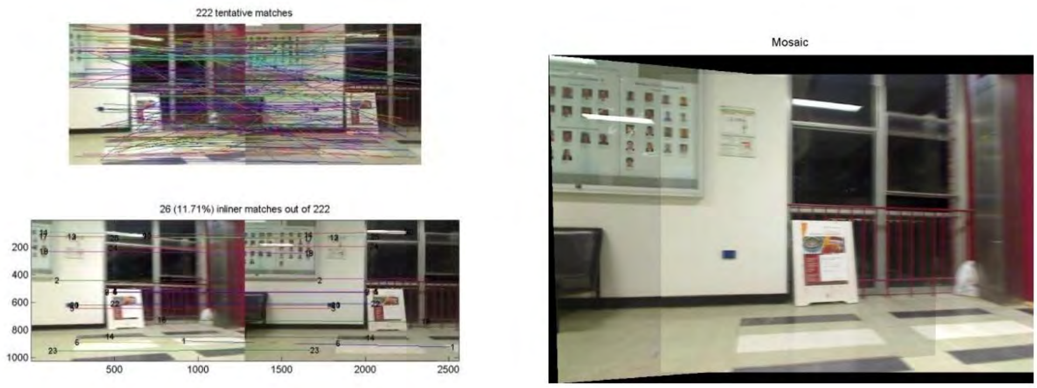


Figure 5. 18 The image matching of Pair No. 41 and its mosaic

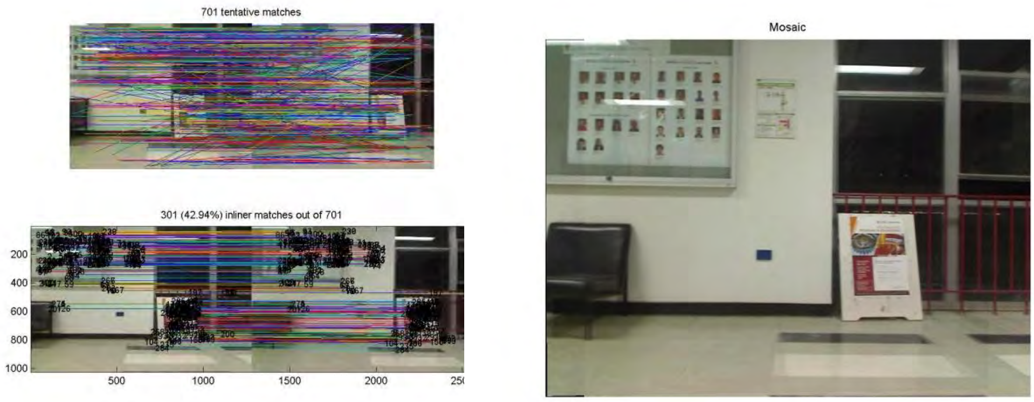


Figure 5. 19 The image matching of Pair No. 42 and its mosaic

5.4.4 Application in vision-based positioning and navigation system

In the last experiment, the method is applied to the vision-based positioning and navigation system. A calibrated video camera (Logitech Webcam Pro2000) was mounted on a moving vehicle with sampling rate at 1 Hz. It moves around a mapped indoor environment, and the system resolves its trajectory by matching the query image extracted from the video recorded by the camera and the pre-stored 3D map. A local orthogonal right-handed coordinate system is used with the Z axis pointing downward. The camera‘s relative position to the vehicle was fixed, which means the experiment was partially controlled: camera height (Z=-0.725m). This experiment was designed to evaluate the performance of the vision-based positioning system before (Test A) and after using the proposed method for image matching (Test B). Therefore the same dataset is used and the positioning calculation ran off-line for twice.

From the video, a total of 83 epochs (frames) were generated and calculated. The calculated 2D trajectories are shown in Figure 5.20 and Figure 5.21 using blue dots. In order to evaluate the positioning results, positioning information generated by a commercial software as reference to investigate the accuracy. More specifically, as GPS cannot be used in indoor areas, we artificially set control points on those image frames and use commercial software PhotoModeler to get their position information as reference, which has been indicated by the red dash lines in Figure 5.20 and 5.21. Within 10m distance, the software can normally achieve centimeter level accuracy. As the parameter Z is controlled (at -0.725m as camera height, axis pointing downwards), the positioning results at Z axis are shown in Figure 5.22 and Figure 5.23 using blue circles and the true values are shown using a red line.

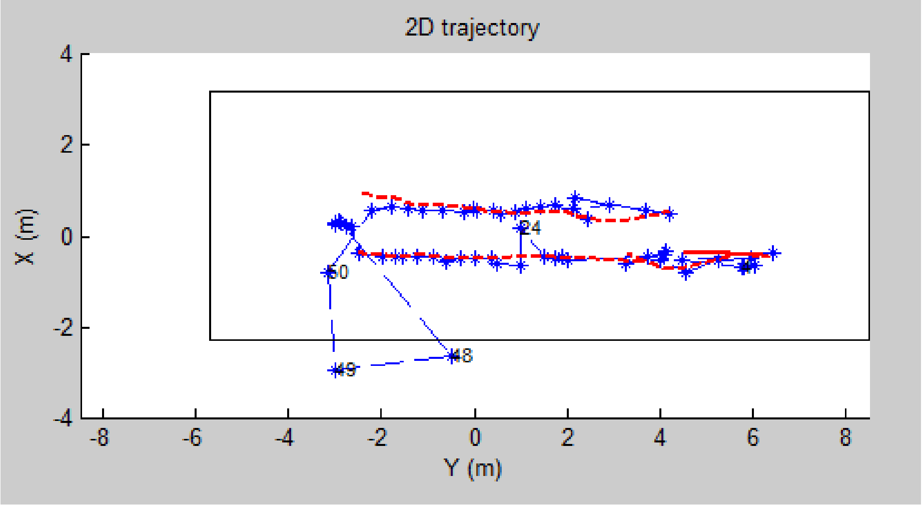


Figure 5. 20 The 2D trajectory of the moving vehicle in Test A (blue dots): without the evaluation test; part of the reference track is shown using red dash line

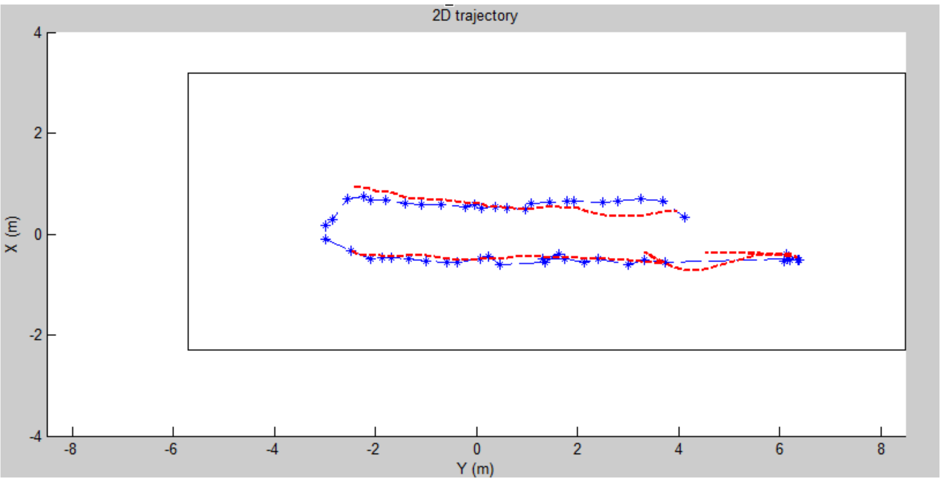


Figure 5. 21 The 2D trajectory of the moving vehicle in Test B (blue dots): with the evaluation test; part of the reference track is shown using red dash line

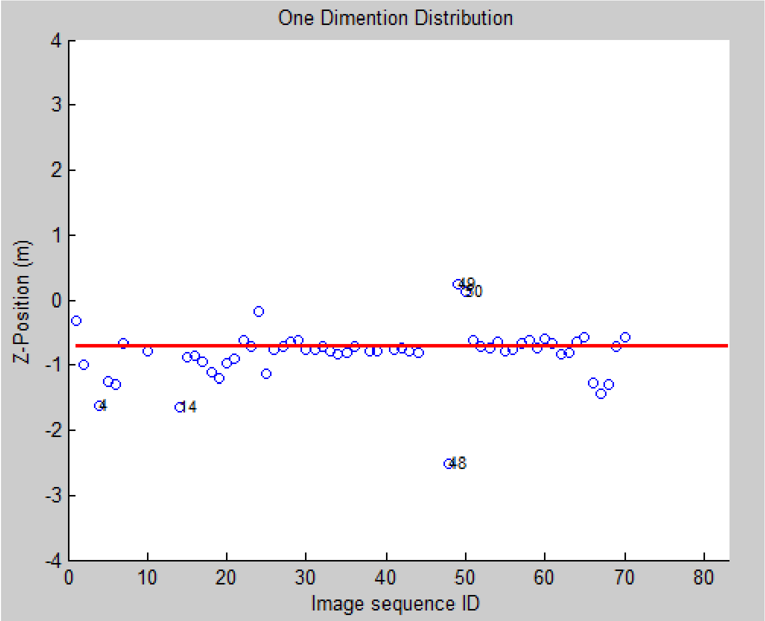


Figure 5. 22 The trajectory at Z axis of the moving vehicle in Test A (blue circles): without the evaluation test; the true value is shown using the red line.

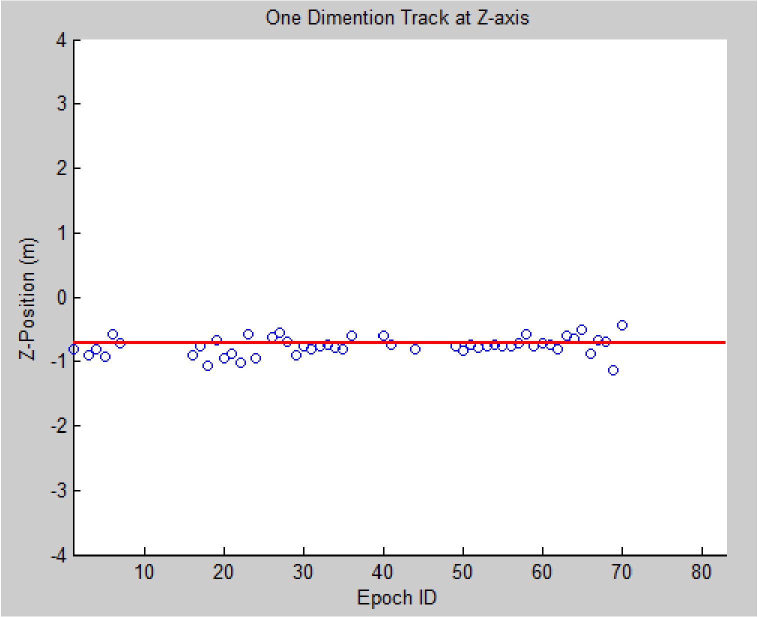


Figure 5. 23 The trajectory at Z axis of the moving vehicle in Test B (blue circles): with the evaluation test; the true value is shown using the red line.

By comparison, it can be observed that the trajectory produced by using the proposed method for image matching has been much improved. As shown in Figure 5.20, without using the method, five epochs (No.4, 24, 48, 49, 50) produce results out of the geometric boundary or deviate from the reference over 0.5m. In Figure 5.22, 5 epochs (No. 4, 14, 48, 49, 50) produce bad results at Z and share 4 bad epochs (No.4, 48, 49, 50) with its X and Y.

Then we investigate the inaccurate results by tracing the pseudo ground control points (PGCPs) back to their correspondences on map images, and discovered that false estimations are the results of mismatches. Taking epoch No. 49 for instance (Figure 5.24 and Figure 5.25), from Test A, the one without using the proposed method for image matching, it generates 5 PGCPs, and their 3D coordinates come from the matched feature points of two map images: map image No. 25 & No.26 (Figure 5.26). However, it is discovered that in Test A PGCP No.1& No.3 were produced by mismatches (red circled), therefore their 3D coordinates are erroneous. By comparison, the matching between epoch No. 49 and its map images in Test B based on the proposed method has also been shown in Figure 5.27. It can be observed that not only no false PGCPs have been generated, but the number of PGCPs have been increased from 5 to 21. Since better PGCPs distribution will benefit the positioning precision, we further investigate the impact of the method on final positioning performance.

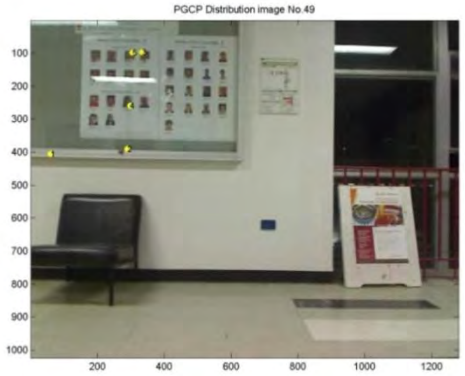


Figure 5. 24 Test A: epoch No. 49 and its PGCPs (yellow dots) without using the proposed method for image matching

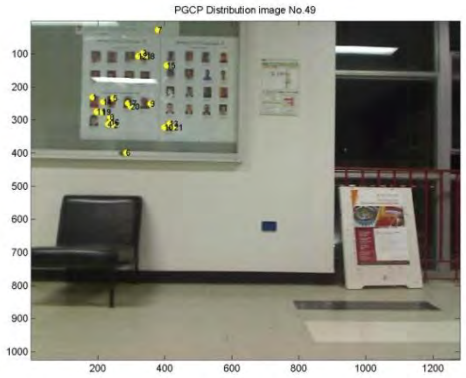


Figure 5. 25 Test B: epoch No. 49 and its PGCPs (yellow dots) with the cross correlation based image matching

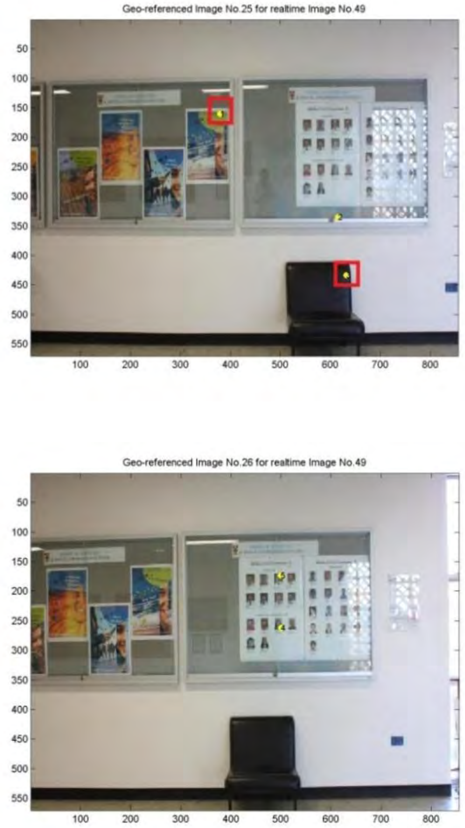

Geo-referenced Image No.25 for realtime Image No.49
Geo-referenced Image No.26 for realtime Image No.49
50
100
150
200
250
300
350
400
450
500
550
100
200
300
400
500
600
700
800

Figure 5. 26 Test A: the corresponding map images No.25 and No. 26 that matched with epoch No. 49 to generate PGCPs. PGCP correspondences are shown using yellow dots.

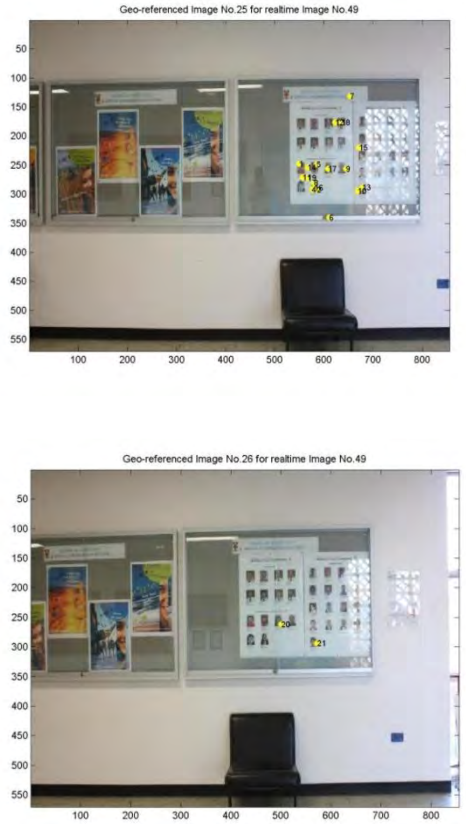

Geo-referenced Image No.25 for realtime Image No.49
Geo-referenced Image No.26 for realtime Image No.49
50
100
150
200
250
300
350
400
450
500
550
100
200
300
400
500
600
700
800
20
21

Figure 5. 27 Test B: the corresponding map images No.25 and No.26 that matched with epoch No. 49 to generate PGCPs. PGCP correspondences are shown using yellow dots.

After excluding the epochs with PGCPs produced by mismatches for the original approach, we compare the performance of the vision-based positioning between Test A and Test B. Using the PhotoModeler results as reference for X and Y, and the controlled value for Z, the accuracy in terms of RMSE from the two tests are shown in Table 5.4. It can be observed that the accuracy has been improved after using the proposed method.

Table 5. 4. Vision-based positioning accuracy with regard to references

	RMSE of positioning results in test A	RMSE of positioning results in test B
X(m)	0.1287	0.1233
Y(m)	0.2289	0.1940
Z(m)	0.2799	0.1406

In summary, the application of the method on image matching for this vision-based navigation system has improved the performance of positioning. Bad positioning results due to false PGCPs from mismatched feature points have been largely avoided. Meanwhile, it can also improve the positioning accuracy. The reason is that the proposed method can help improve the quality of image matching. Poor RANSAC estimations of image matching with a large number of mismatches can be discarded. The retained good RANSAC estimation tends to produce bigger number of inliers with good quality. Therefore, the PGCPs improve in terms of both quality and quantity. And final positioning results are more accurate.

5.5 SUMMARY

Image matching is the fundamental problem for a wide range of applications, and RANSAC has been the most popular strategy for outlier detection and homography estimation in image matching. However, its major limitation lies in that the RANSAC method starts with a random subset of correspondences to estimate homography model and detect outliers accordingly. If the initial selection is erroneous, it will lead to inaccurate or even false estimation of the homography and mismatches will be included as inliers. Therefore, in this research a new method to evaluate and enhance the performance of RANSAC based homography estimation has been proposed. By calculating cross-correlation information between feature patches, bad estimation can be detected and removed. Moreover, accompanied with RANSAC, this method can largely improve the quality of image matching. Experiments have demonstrated that the evaluation test set a harsh bound for the RANSAC estimation. It can correctly identify poor RANSAC estimation and retain only fine ones, and effectively improve the quality of image matching under circumstances like matching ambiguity and dramatic transformation. Moreover, the method has been tested in three applications: image recognition, image stitching and vision-based navigation. Experiments have shown that all three applications benefit from such an approach. For this research, it has improved the performance of vision-based positioning effectively. Its application in the identification of reference images from the database (image recognition) for the vision-based navigation system can be found in Section 3.2.2 and Section 7.3.2. In summary, such an estimation method can be used together with RANSAC for a wide range of applications, and help improve the performance of image matching. Currently the efficiency of the algorithm has not been emphasized. Further research will be focused on this aspect.

CHAPTER 6 EVALUATION OF THE SYSTEM PERFORMANCE

6. 1 INTRODUCTION

The precision and accuracy of such photogrammetric approach of image-based positioning is depending on the precision and accuracy of final space resection process, which is a function of PGCP distribution and measurement accuracy, and any factor that has certain impact on either of these two major components will to certain degree influence final positioning accuracy. Therefore in this chapter, the way that different factors influencing the positioning accuracy are analysed through both mathematical model and experiments, which includes simulations and tests based on real data.

6. 2 MAJOR COMPONENTS DETERMINING POSITIONING ACCURACY

In this section, the two main components that determine the accuracy of the position solution are identified: geometry and measurement accuracy. Both mathematical model and test results are analysed to verify this assumption. Any factor that involved impacts the positioning accuracy is through its influence on these two elements.

6.2.1 Analysis of mathematical models

In Chapter 3 the mathematical model of the vision-based navigation system has been introduced. The least squares based space resection with modification is used for the final positioning resolution as shown in Eq.6.1 and Eq.6.2:

$$\boldsymbol{At} + \boldsymbol{BX} - \boldsymbol{l_1} = \boldsymbol{v_1} \quad , \qquad \boldsymbol{l_1} \sim (\boldsymbol{0}, \sigma_0^2 \boldsymbol{P_1}^{-1}) \tag{6.1}$$

$$\boldsymbol{IX} - \boldsymbol{l_2} = \boldsymbol{v_2} \quad , \qquad \boldsymbol{l_2} \sim \left(\boldsymbol{0}, \sigma_0^2 \boldsymbol{P_2}^{-1}\right) \tag{6.2}$$

in which $\boldsymbol{A}$ contains partial derivatives with respect to the exterior orientation parameters, and $\boldsymbol{t}$ contains the incremental changes to the initial values of external orientation parameters; $\boldsymbol{B}$ contains the partial derivatives with respect to the three coordinates of the (Pseudo) Ground Control Points, and $\boldsymbol{X}$ contains the incremental changes to the initial values of ground coordinates of PGCP.

Combine (6.1) with (6.2):

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{X} \end{bmatrix} - \begin{bmatrix} \boldsymbol{l_1} \\ \boldsymbol{l_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{bmatrix}, \qquad \begin{pmatrix} \mathbf{P_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P_2} \end{pmatrix} \tag{6.3}$$

Corresponding normal equation becomes:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}^T \begin{pmatrix} \mathbf{P_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P_2} \end{pmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}^T \begin{pmatrix} \mathbf{P_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P_2} \end{pmatrix} \begin{bmatrix} \boldsymbol{l_1} \\ \boldsymbol{l_2} \end{bmatrix} \tag{6.4}$$

Substituting some parts of the equation 5 with simple expression:

$$\begin{bmatrix} \boldsymbol{N_{11}} & \boldsymbol{N_{12}} \\ \boldsymbol{N_{21}} & \boldsymbol{N_{22}} \end{bmatrix} \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W_1} \\ \boldsymbol{W_2} \end{bmatrix} \tag{6.5}$$

in which

$$\begin{bmatrix} \boldsymbol{N_{11}} & \boldsymbol{N_{12}} \\ \boldsymbol{N_{21}} & \boldsymbol{N_{22}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A^T P_1 A} & \boldsymbol{A^T P_1 B} \\ \boldsymbol{B^T P_1 A} & \boldsymbol{B^T P_1 B} + \boldsymbol{P_2} \end{bmatrix} \tag{6.6}$$

The covariance matrix of the unknowns is contained in a generalized inverse of the normal equation matrix:

$$\begin{bmatrix} \boldsymbol{Q_{tt}} & \boldsymbol{Q_{tX}} \\ \boldsymbol{Q_{Xt}} & \boldsymbol{Q_{XX}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{N_{11}} & \boldsymbol{N_{12}} \\ \boldsymbol{N_{21}} & \boldsymbol{N_{22}} \end{bmatrix}^{-1} \tag{6.7}$$

Since the quality of space resection based positioning is evaluated by the precision (accuracy) in 6DOF, a relative precision can be measured through the post-adjustment covariance matrix $\mathrm{Q_{tt}}$ of the estimated image orientation in 6DOF (t), regarding σ_0^2 as a variance factor.

$$\boldsymbol{Q_{tt}} = (\boldsymbol{N_{11}} - \boldsymbol{N_{12}}\boldsymbol{N_{22}}^{-1}\boldsymbol{N_{21}})^{-1} \tag{6.8}$$

From Eq. (6.8)

$$\boldsymbol{Q_{tt}} = (\boldsymbol{A^T P_1 A} - \boldsymbol{A^T P_1 B}(\boldsymbol{B^T P_1 B} + \boldsymbol{P_2})^{-1}\boldsymbol{B^T P_1 A})^{-1} \tag{6.9}$$

From Eq. (6.9) we can clearly observe that the final positioning precision &accuracy using a modified space resection model is affected by two major elements: geometry ($\boldsymbol{A}$ & $\boldsymbol{B}$) and the accuracy of measurements: image measurement ($\boldsymbol{P}_1$) and 3D object space coordinates of pseudo ground control points ($\boldsymbol{P}_2$).

In order to further investigate the role of the two components, their influence on the final precision need to be separately evaluated, DOP values are used as the indicator of geometric strength and the estimated standard deviation of observations is used to evaluate measurement accuracy. The calculation of DOP values follows the same way as in the GPS community:

$$\boldsymbol{C_X} = \sigma_0^2(\boldsymbol{A^T A})^{-1} \tag{6.10}$$

in which the part $(\boldsymbol{A^T A})^{-1}$ contains DOP factors in its diagonal elements. More details can be found in Section 2.3.3.1.

The estimated standard deviation of observations is calculated as follows:

$$\hat{\sigma}_0 = \sqrt{\frac{\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v}}{f}} \tag{6.11}$$

in which f represents degree of freedom ($\mathrm{f} = \mathrm{n} - \mathrm{u}$).

6.2.2 Test on the proposed theory

A test has been carried out to test the theory deduction in Chapter 6.2.1. The aim is to verify the deduction that the positioning accuracy is determined by the two components: geometry of PGCPs and measurement accuracy.

In this experiment, real time images are obtained through stable camera stations rather than camera mounted on a moving vehicle. A calibrated CCD camera (Canon EOS4500) is used. In this way, each position of the camera site can not only be calculated by the system but also measured by external tools, total station, with relatively higher accuracy. Therefore the positioning accuracy can be evaluated against reality. The system calculated results, surveyed true value and their difference are shown in Table 6.1.

Table 6. 1 System measured results evaluated against total station results

	Camera Site ID	X	Y	Z
Calculated(m)	3	0.057	1.177	-1.273
	4	0.070	1.844	-1.311
	5	0.067	2.924	-1.261
	6	0.075	5.377	-1.280
Surveyed(m)	3	0.067	1.080	-1.264
	4	0.068	1.885	-1.265
	5	0.062	2.955	-1.264
	6	0.064	5.003	-1.264
Absolute Difference (m)	3	0.010	0.097	0.009
	4	0.003	0.041	0.046
	5	0.005	0.031	0.002
	6	0.011	0.373	0.016

According to Table 6.1, the accuracy of the positioning is between 1-10 centimetre level. Then the impact of geometry and measurement accuracy on positioning precision is analysed. Compare Figure 6.1 and Figure 6.2, it can be observed that the position precision generally follows the trend of DOP, which means geometry has the biggest impact. At the same time, it is noted that they are not exactly the same: e.g. the precision on Z axis drops from epoch 3 to 4, while their DOP values are close. There are only two components that contributing to the final precision: geometry and measurement accuracy. As shown in Figure 6.3, the measurement accuracy from epoch 3 to 4 actually drops as predicted. Although the influence of measurement accuracy is not significant in this case, it still proves the point that the overall positioning precision and accuracy depend on geometry and measurement accuracy. And geometry is the major impact since the final image-based positioning uses the same image-matching algorithm and geo-referenced map, which means the measurement accuracy remains more stable compared with PGCP geometry. It is also noted that the geo-referencing accuracy of the map in different areas may varies slightly, which leads to the variations on measurement accuracy.

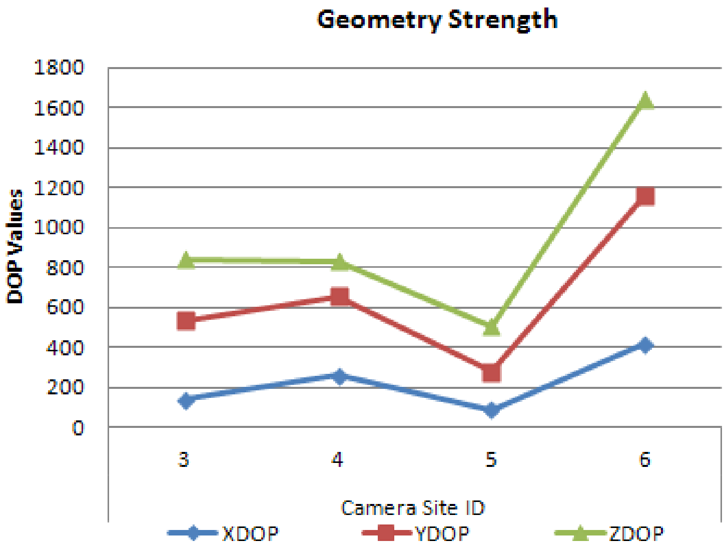


Figure 6. 1 Geometric Strength on the 4 epochs

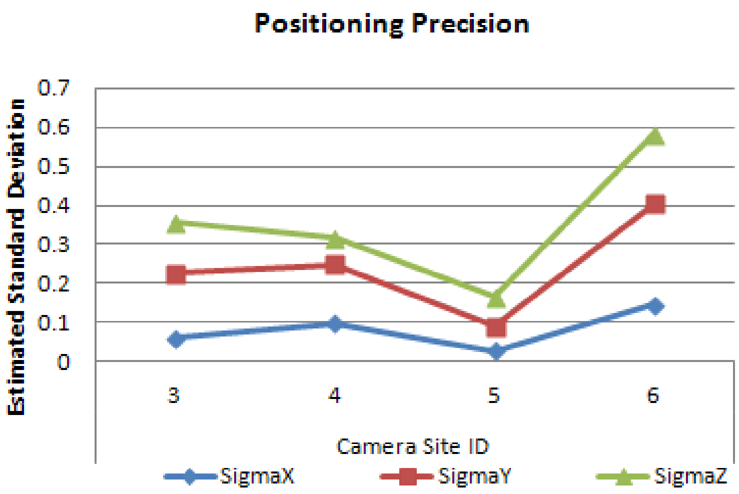


Figure 6. 2 Positioning precision using estimated standard deviation in 3 out of 6 unknown parameters (unit: m).

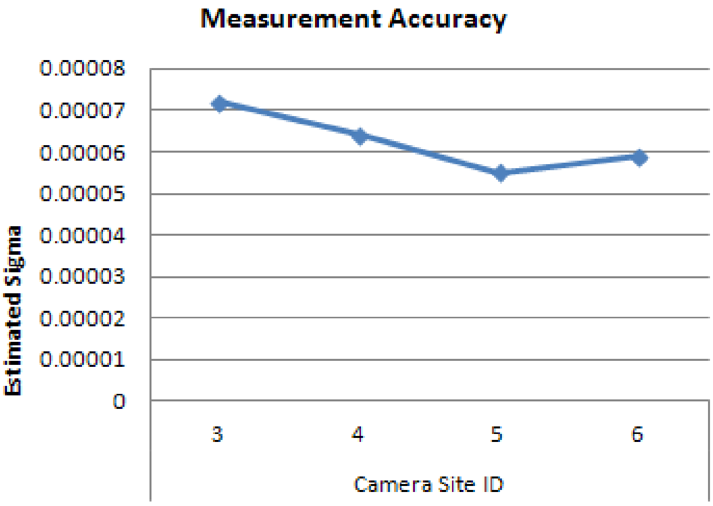


Figure 6. 3 Measurement accuracy on the 4 epochs (unit: m).

6.3 GEOMETRY AND FACTORS INVOLVED

In this section, the major component that determines positioning accuracy, geometric configuration of PGCPs, is analysed.

6.3.1 Geometric impact

This experiment aims at evaluating the geometric impact on final positioning accuracy based on real world test. Vision-based positioning is carried out in the mapped indoor area of the school hallway. A calibrated video camera (Logitech Webcam Pro2000) is mounted on a moving vehicle with sampling rate at 1 HZ. Its relative position to the vehicle is fixed, which means the experiment is partially controlled: camera height (Z:-0.725m) and two angles of the camera attitude ($\omega = 1.57\ \mathrm{rad},\ \upsilon = 0\ \mathrm{rad}$) are fixed. The positioning was performed by extracting image frames from the video and match with the 3D map images frame by frame. Each frame is an epoch; a position in 6DOF is calculated. We took epoch No.20-40 with controlled parameter Z for illustration.

Figure 6.4 shows the calculated Z position between epochs 20-40. In extreme cases, as shown in Figure 6.4 and Figure 6.5, if too few PGCPs are generated, the positioning calculation will fail. Compare Figure 6.4 with the DOP values at Z-axis (Figure 6.6), it can be clearly seen that big DOP values, which means bad geometry, are behind the bad positioning results with low accuracy (e.g. Epoch 22& Epoch 38). The rest of the results are reasonable while their DOP values are below a certain limit. And the absolute accuracy does not follow the exact trend of DOP. Put the 3 figures together, it‘s not hard to observe that a bigger number of PGCPs gives a better chance of good geometry, thus a more accurate positioning result, vice verse. Therefore, it is concluded that the major cause of inaccurate results is bad geometry, and geometric impact plays an important role in the determination of final positioning precision.

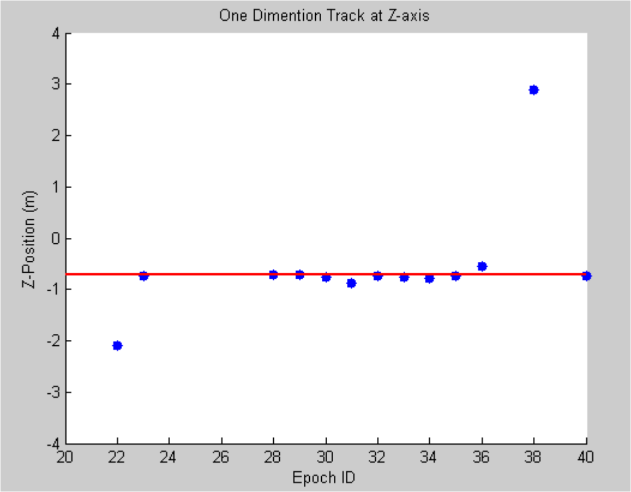


Figure 6. 4 Measured Z position for epoch 20-40

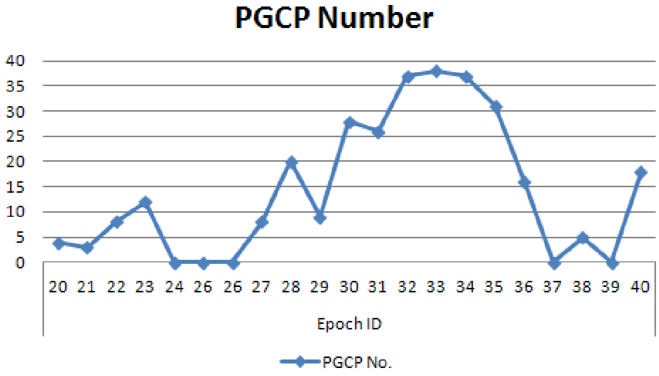


Figure 6. 5 Number of PGCP for epoch 20-40

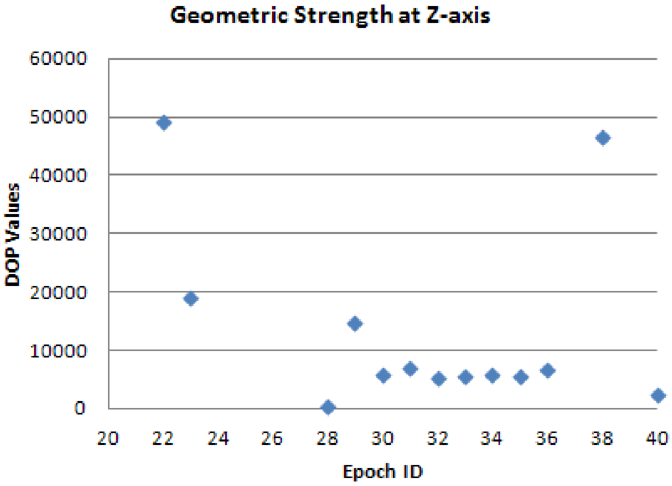


Figure 6. 6 Geometric strength at Z for epoch 20-40

6.3.2 Simulation test on geometric impact

In previous section, the real world test help reveal the geometric impact on the positioning accuracy. Here a simulation test is carried out to further analyse the nature of the relationship between the geometry of PGCPs and system performance. Since the final positioning is based on a least squares solution, the geometry strength also affects internal reliability of the adjustment. This factor has been put into consideration. In this experiment, the PGCPs are simulated using known feature points in the scene (e.g. Figure 6.7). Therefore, their number and distribution can be controlled and set to suit the scenario of the tests.

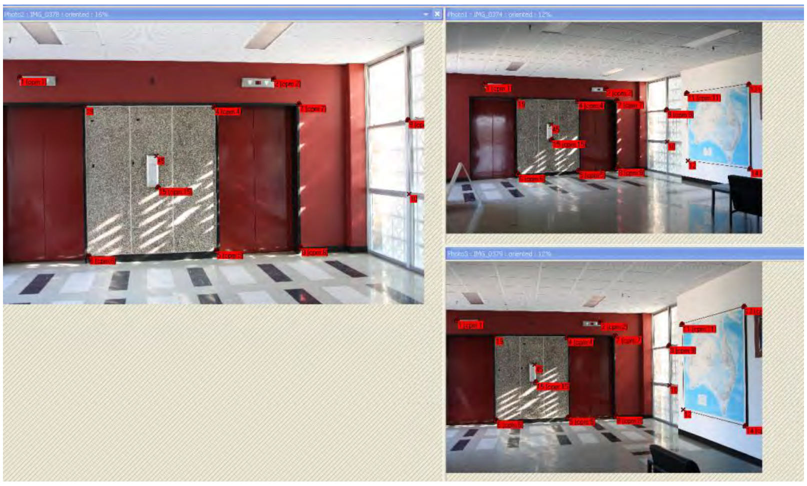


Figure 6. 7 Simulated PGCPs using known feature points in the scene

6.3.2.1 Variation of the number of PGCPs

To reveal the overall relationship between the number of PGCPs and the reliability of the system and precision of positioning, a group of tests were performed on images shown in Figure.6.7 , each tested with 15, 13,11,9,7,5,4 PGCPs respectively. One image with its results was used to show the common phenomena.

Table 6.2 shows the positioning result with the use of this image. It can be seen that the estimation results of the external parameters (6DOF) tend to remain relatively stable with an increased number of PGCPs. The DOP values and the average of controllability values for each set up (e.g. 9 PGCPs) were also calculated. Figure 6.8 shows the variation trend of DOP values with the increase of PGCPs. The figures have further proved that the whole system is unstable with less than 13 PGCPs. The three figures all shows a decreasing trend of the test values (DOP values and average of internal control values), which means with the increase of the number of PGCPs, the precision of positioning is increasing and the internal reliability of the system has been improved. According to the figures, it can also be observed that the increase in PGCP number has more impact on the precision in Z compared with the precision in X and Y.

Table 6. 2 positioning result in 6 DOF

	Number of PGCP	15	13	11	9
LS Estimation Of Six Unknown parameters	Xs(m)	1.4531	1.4446	1.4375	1.393
	Ys(m)	0.0813	0.1066	0.1027	0.273
	Zs(m)	-1.314	-1.3165	-1.2973	-1.3053
	ω(rad)	1.5764	1.576	1.5781	1.5772
	φ(rad)	-0.0136	-0.0129	-0.0133	-0.0129
	κ(rad)	1.9663	1.9632	1.9634	1.944
LS Estimation Of Six Unknown parameters	Number of PGCP	7	6	5	4
	Xs(m)	1.3674	1.359	1.3407	1.3486
	Ys(m)	0.4205	0.5101	1.921	0.7081
	Zs(m)	-1.0514	-1.0869	-2.4409	-0.967
	ω(rad)	1.6048	1.601	1.45	1.615
	φ(rad)	-0.0189	-0.0177	0.006	-0.0161
	κ(rad)	1.9271	1.9171	1.7605	1.8957

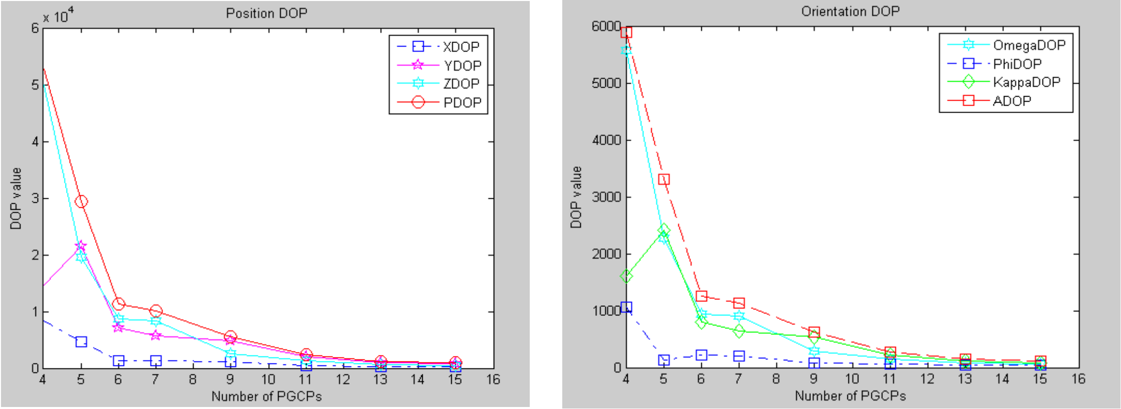


Figure 6. 8 DOP values for position and orientation

Therefore it is concluded that PGCPs should be selected as many as possible to enable an acceptable positioning capability. When the PGCPs obtained for a particular image are not sufficient to provide a stable and relatively precise positioning results, that image for positioning should be rejected.

6.3.2.2 Distribution of PGCPs

In order to investigate how the distribution of pseudo ground control points affect the positioning precision and reliability of the system, two groups of simulation tests were further performed.

For the first group, two sets of PGCPs were choosen, with one set scattered around the image (Figure 6.9) and the other set cantered on a small region located on the image centre (Figure 6.10). Table 6.3 shows the result of one image, 7 PGCPs were used for each set of this case. The estimation results of position and orientation parameters (6DOF) are close to the best results obtained previously with 15 PGCPs, which means the positioning function run successfully and the result is acceptable with both settings. It can be easily observed from DOP values that the precision of positioning is much higher with the scattered PGCPs than with the centred distribution. The internal reliability of the system has not changed much.

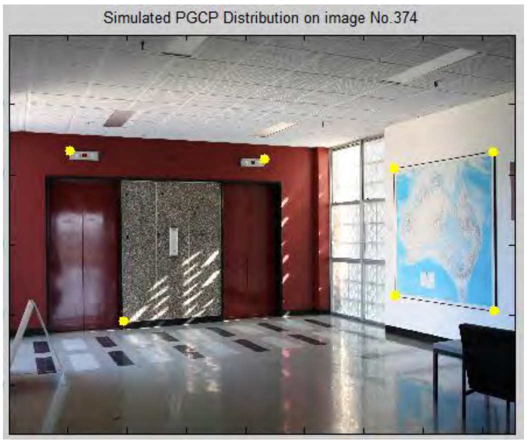


Figure 6. 9 Scattered distribution of simulated PGCPs on image No. 374

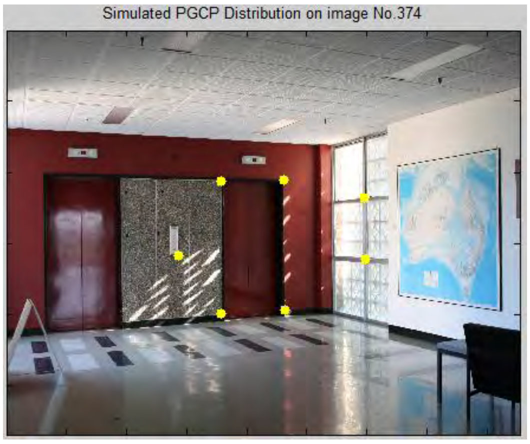


Figure 6. 10 Centered distribution of simulated PGCPs on image No. 374

Table 6. 3 positioning result with scattered and centred distribution

	Distribution	7 PGCP scattered	7 PGCP centred
LS Estimation Of Six Unknown parameters	Xs(m)	1.0952	1.0958
	Ys(m)	-2.446	-2.3501
	Zs(m)	-1.3226	-1.3134
	ω(rad)	1.5733	1.5742
	ϕ(rad)	-0.0115	-0.0115
	κ(rad)	2.1646	2.157
DOP Values	X DOP	748.105225	4441.710041
	YDOP	1889.069529	6149.407959
	ZDOP	1204.459151	4630.356573
	PDOP	2361.983692	8887.306022
	ω DOP	137.002452	451.104166
	φ DOP	70.058549	108.223677
	κ DOP	203.636394	725.128196
	A DOP	255.236465	860.824159
Internal Reliability	Ave ControlV	5.625227	5.514999

The second one aims at investigating how the geometry change of PGCPs, especially from planar to non-planar will affect the positioning precision and system internal reliability. The tests were designed in the way that all three sets had 8 points in common and lay on the same plane. Only one point out of 9 located at different places, with the first test had the point on the same plane (Type 1, Figure 6.11), the second test had the point located on a different plane (Type 2, Figure 6.12),and third test had the point located on the same second plane but with bigger deviation from the optical axis(Type 3, Figure 6.13). The change of DOP values is shown in Figure 6.14.

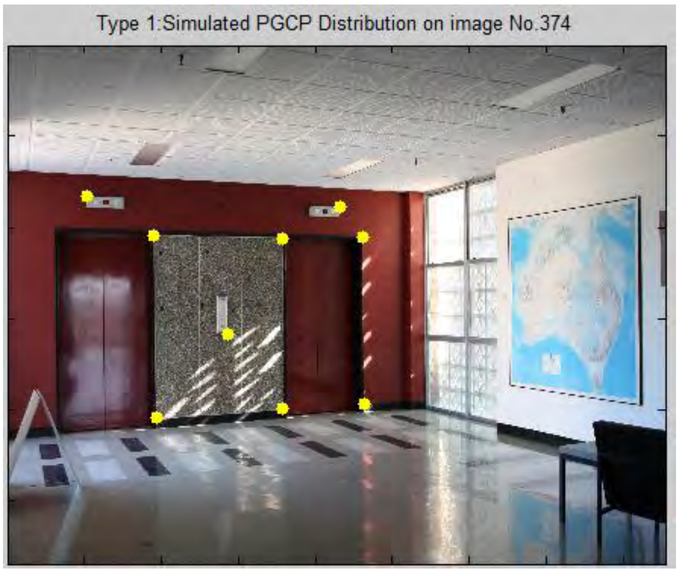


Figure 6. 11 Type 1 distribution: all 9 points lay on the same plane

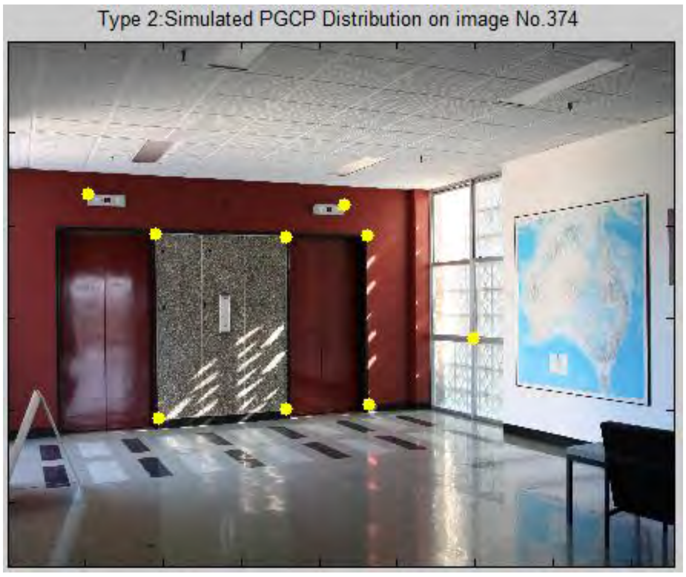


Figure 6. 12 Type 2 distribution: 8 points lay on the same plane, the 9th on second plane

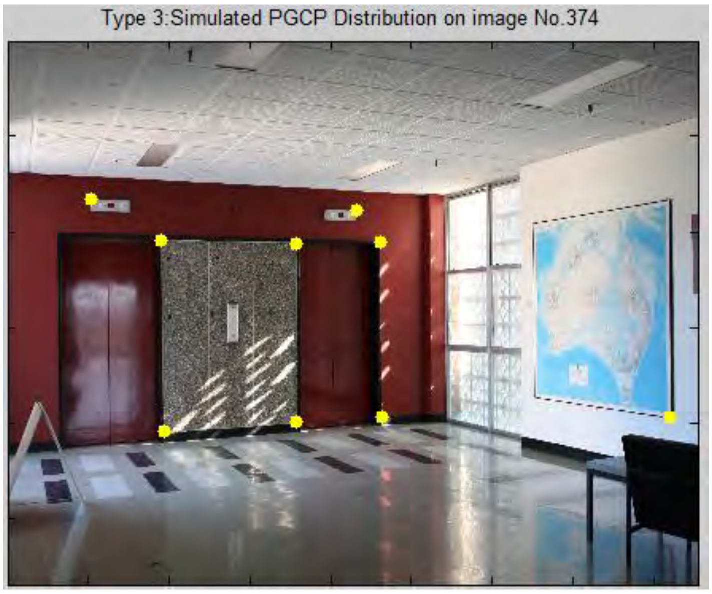


Figure 6. 13 Type 3 distribution: 8 points lay on the same plane, the 9th on second plane

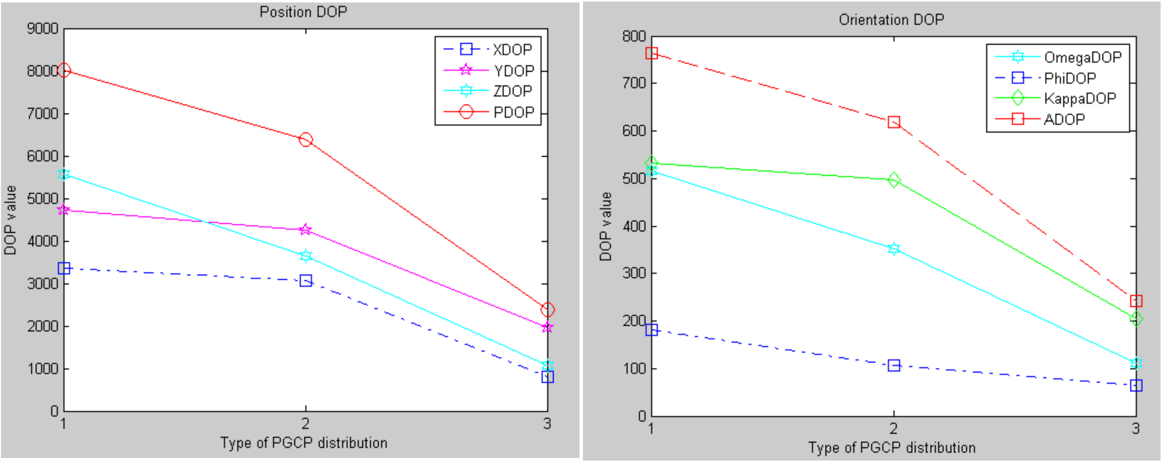


Figure 6. 14 DOP values for position and orientation

It can be observed that a non-planar configuration of PGCPs increases the precision of the positioning result. It shows that the effect becomes more significant with the increasing offset from the optical axis. From Figure 6.13 and Figure 6.14, it can also be observed that the precision in Z is again more affected than that of X and Y, and Omega again being the least affected among the three angle values. It is also noted from the result that internal reliability deteriorates (the average of controllability value grows). This is mainly because the points on the different planes contribute to the geometry largely, thus making it hard to be controlled. It will be difficult to detect any outlier in this observation. In order to improve the precision of positioning and at the same time do not sacrifice system reliability, it is concluded that PGCPs on different planes should be selected evenly.

In summary, the geometry of PGCPs plays an essential role for the system performance. Since PGCPs are produced by matched SIFT feature points, any factor that influence the SIFT matching between query image and reference image(s) will affect the density and geometric configuration of PGCPs, which includes the richness of features, illumination, viewing angle, etc. In the following section, the performance of the system is evaluated with varying image matching conditions. It discusses the impact of these factors on positioning performance through the geometry of PGCPs.

6.3.3 Evaluating the performance of SIFT matching for the vision-based positioning system

A controlled experiment is designed to evaluate the performance of SIFT matching for the vision-based positioning system. Major factors that influence image matching and their impacts on final positioning have been investigated: illumination and viewpoint changes. On top of this, mismatches, which has long been a bottleneck for visual systems have been studied as well.

First, mapping is performed in the target environment. All geo-referenced map images were taken with adequate lighting and viewing direction perpendicular to the wall (mapping area with visual features). Then a calibrated CCD camera (Canon EOS4500) with a fixed focal length at 24.1757mm was used as vision sensor of the positioning system. First Three stable camera sites were deployed facing different mapping areas with X, Y, Z coordinates of the three camera stations surveyed by a total station, and angular changes at each camera site roughly measured. A total 8 pairs of images (16 in total) were taken at the 3 sites, with each pair consisted of one image with adequate lighting and the other covering the same scene but limited lighting. For each image matching process, corresponding pairs before and after RANSAC process (used to reject mismatches) are compared. Furthermore, the number of PGCP generated by each matching process is also studied along with the geometry of these points, which will directly affect the precision of positioning. Finally a manual check for the PGCP locations from the two matched images is performed to verify the correctness of PGCPs generated by image matching. The results are shown in Table 6.4.

Table 6. 4 Performance of SIFT matching in the system

SiteID	Im_ID	Angle	Matches	Inliers	PGCPs	False PGCPs	PDOP	ADOP
1	1	0	578	202	51	1	2399	583
	2	0	515	141	28	0	2575	622
	3	-30	372	103	32	2	1149	223
	4	-30	357	119	34	1	1072	208
	5	50	267	59	2	0		
	6	50	210	37	1	1		
2	7	0	427	145	35	0	2106	530
	8	0	346	96	24	0	2795	701
	9	-20	401	193	59	1	969	235
	10	-20	386	145	60	2	195	50
3	11	0	417	112	11	0	5711	1177
	12	0	295	56	9	0	822	4034
	13	-30	457	146	22	0	372	2133
	14	-30	340	66	6	0	424	2471
	15	20	241	65	5	0	5813	28658
	16	20	206	47	5	0	6624	32951

Firstly, it can be observed with ease, in each pair the image with good lighting condition (e.g. Image No.1) is able to find more common SIFT matched features when matched with reference map image than its counterpart with limited lighting (e.g. N Image No.2). As a result, more PGCPs are generated and a better geometry (smaller DOP values) is provided. This test proves that lighting variation will influence the precision of final positioning by its impact on the geometric strength of the adjustment system. For vision-based positioning and navigation systems alike, which depend on visual information and image matching techniques for localization, one limitation is that illumination changes, which is especially common for outdoor environment, may affect a navigation solution. The reason behind is that most existing local descriptors including the SIFT are based on luminance information rather than color information. Therefore, more robust local descriptors need to be developed.

The third column in Table 6.4 indicates the angular changes of viewpoint at κ (around Z-axis) for each camera site, other angles remain stable. Assuming the viewing direction perpendicular to the mapping area (wall with geo-referenced features) to be 0, a clockwise rotation to be positive changes. It‘s easy to note that the only epochs that fails to give a positioning result is the pair with the most drastic angular change, real time image No.5 (6). Not only does the total number of SIFT matches decreases, the percentage of correct matches filtered using RANSAC has also been reduced. Actually it is the pair with lowest correct rate. As a result, too few PGCPs are generated for positioning.

For better comparison, two epochs with similar coverage of the scene are chosen: No.5 and No.11 shown in Figure 6.15 and Figure 6.16, both of which were taken under amble light. It can be easily observed that Image No.5 include more features, but less SIFT matched points as well as PGCPs are found. Moreover, the correct rate (the 6th column ―Percentage") of No.5 is lower than that of No.11. It indicates that when the two matched images suffer from large viewpoint variation, less SIFT matches will be found, and there‘s a higher chance to generate false matches. As a result, number of PGCPs will decrease, which lead to poor positioning precision. But if we take a look at other images with smaller angular changes, such rule does not apply. The reason is that the performance of SIFT based matching only drops under substantial viewpoint changes.

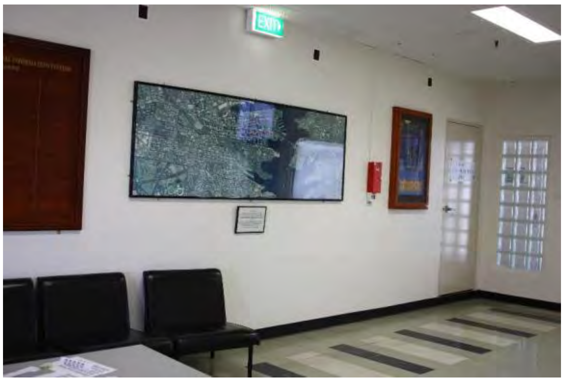


Figure 6. 15 Real time image No.5

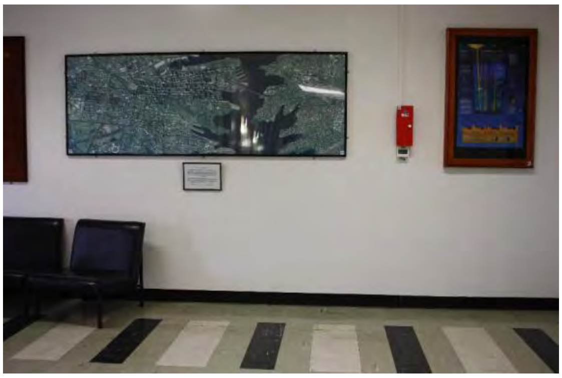


Figure 6. 16 Real time image No.11

Thirdly, it is noticed that after using RANSAC to reject mismatches, there is still a small chance that mismatches been left untreated, which might later generate false PGCP to jeopardise the final positioning process. As shown in Figure 6.17 and Figure 6.18, when the real time query image No.3 is matched with map image No.6, 32 PGCPs are generated from the reliable matches provided by RANSAC. However, 2 false matches were still been spotted during manual check. The outlier detection mechanism introduced in Chapter 4 dealt with these circumstances.

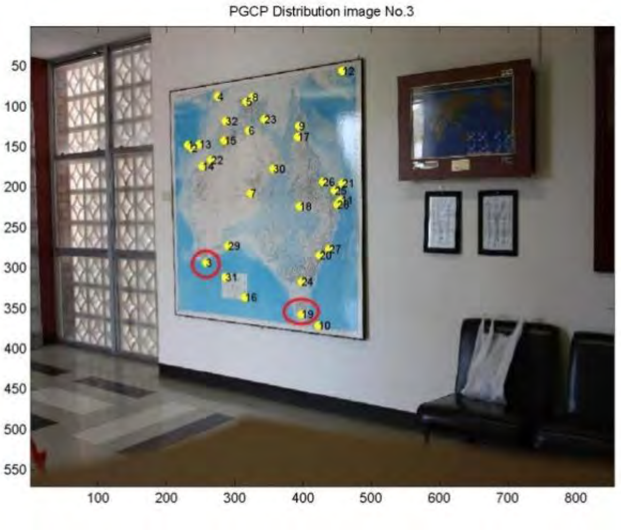


Figure 6. 17 Real time query image No.3 with PGCPs, false PGCPs have been circled.

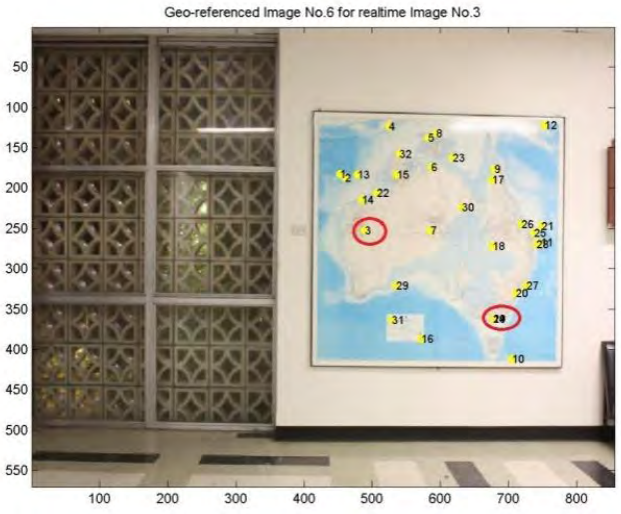


Figure 6. 18 Map image No.6 with correspondences of PGCPs on query image No.3, false correspondences have been circled

In summary three major weaknesses for image matching in the system have been found: invariant feature matching could not deal with drastic illumination changes and large viewpoint shift; Furthermore, mismatches may lead to false PGCPs, which need to be tackled by data snooping during the least square solution of final positioning. These three problems affect the final positioning by changing the PGCPs distribution.

6.3.4 Using ASIFT for viewpoint changes

In order to tackle the problem for unsatisfactory performance of SIFT subject to dramatic viewpoint distinction, some approaches have been recently proposed by some researchers to extend scale and rotation invariance to affine invariance, such as MSER (Matas et al., 2004) and Harris / Hessian Affine (Mikolajczyk and Schmid, 2004). Although these methods have been proved to enable matching with a stronger viewpoint change, all of them are prone to fail at a certain point (Nalpantidis et al., 2009). A better idea is to simulate viewpoint changes in order to reach affine invariance, the most successful algorithm using such method is named ASIFT (affine-SIFT). It is introduced by Morel and Yu in 2009 (Morel and Yu, 2009) to explicitly deal with extreme angle changes (up to 36 and higher). SIFT is only partially invariant to viewpoint changes because it is invariant to four out of the six parameters of an affine transform. Affine-SIFT (ASIFT), on the other hand, simulates all image views obtainable by varying the two camera axis orientation parameters, namely, the latitude and the longitude angles, left over by the SIFT method. Then it covers the other four parameters by using the SIFT method itself (Morel and Yu, 2009).

In this research, ASIFT is applied to replace SIFT in order to achieve a more robust positioning result against viewpoint variation. At both mapping and positioning stage, ASIFT based image matching is used in the same way SIFT is utilized. In order to evaluate its performance and compare it with that of SIFT, datasets from the same controlled experiment is used.

Figure 6.19 and Figure 6.20 show the matching between real time query image No.5 with map image No.10 using SIFT and ASIFT respectively. Under dramatic view changes, ASIFT produce more reliable matches (inliers), while although SIFT get as many tentative matches, most of which are mismatches and filtered out by RANSAC. When dealing with images without much angular difference, however, ASIFT had a rather unstable performance, as shown in <Figure 6.21 and Figure 6.22>, and <Figure 6.23 and Figure 6.24>, with the former group favours SIFT and later one favours ASIFT. The reason for it lies in that the author of ASIFT has already included an outlier detection mechanism (ORSA) in ASIFT algorithm. When the angular change is high, SIFT tends to produce more mismatches thus outperformed by ASIFT in terms of both number of inliers and PGCPs. On the other hand, if the angular change is low, because of the pre-filtered mechanism of ASIFT, the inliers of ASIFT largely depends on the combined effect of ORSA and RANSAC. In order to further compare the performance of the two algorithms for the system, ASIFT is used without RANSAC at positioning stage. The result is shown in Table 6.5.

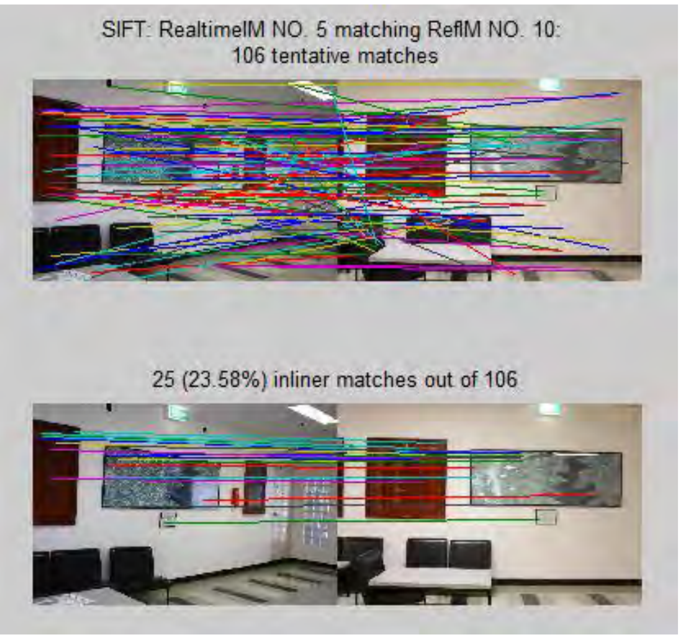

SIFT: RealtimeIM NO. 5 matching RefIM NO. 10:
106 tentative matches
25 (23.58%) inliner matches out of 106

Figure 6. 19 Matching between real time query image No.5 with map image No.10 using SIFT+RANSAC, dramatic angular change, 25 reliable matches (inliers).

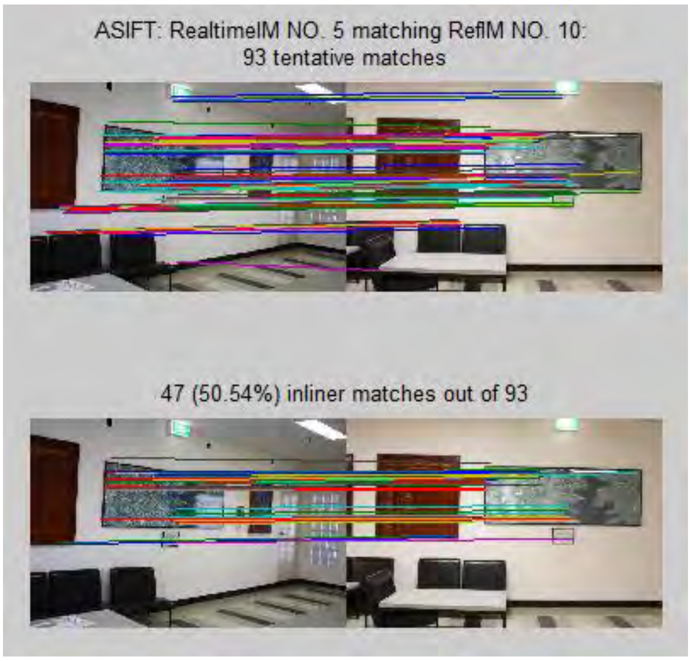

ASIFT: RealtimeIM NO. 5 matching RefIM NO. 10:
93 tentative matches
47 (50.54%) inliner matches out of 93

Figure 6. 20 Matching between real time query image No.5 with map image No.10 using ASIFT+RANSAC dramatic angular change, 47 reliable matches (inliers).

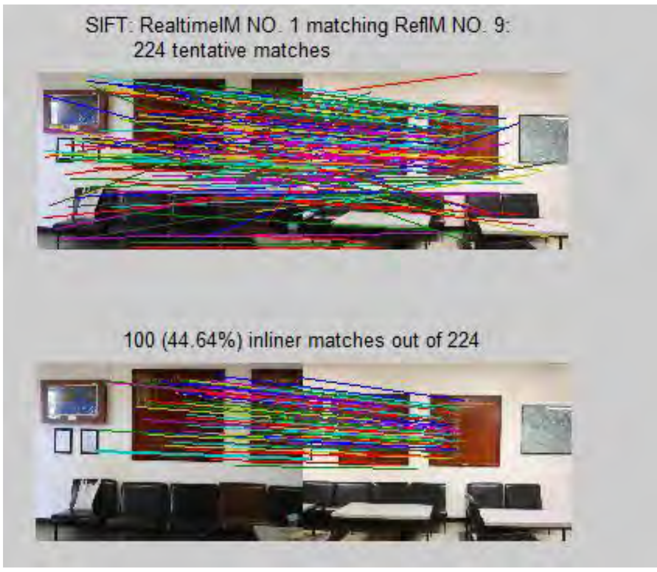

SIFT: RealtimeIM NO. 1 matching RefIM NO. 9:
224 tentative matches
100 (44.64%) inliner matches out of 224

Figure 6. 21 Matching between real time query image No.1 with map image No. 9 using SIFT+RANSAC, small angular change, 100 reliable matches (inliers).

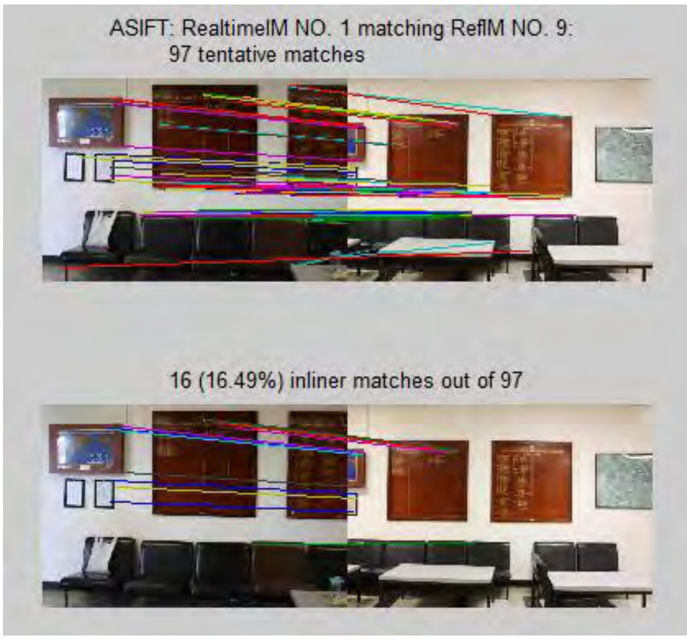

ASIFT: RealtimeIM NO. 1 matching RefIM NO. 9:
97 tentative matches
16 (16.49%) inliner matches out of 97

Figure 6. 22 Matching between real time query image No.1 with map image No. 9 using SIFT+RANSAC, small angular change, 16 reliable matches (inliers).

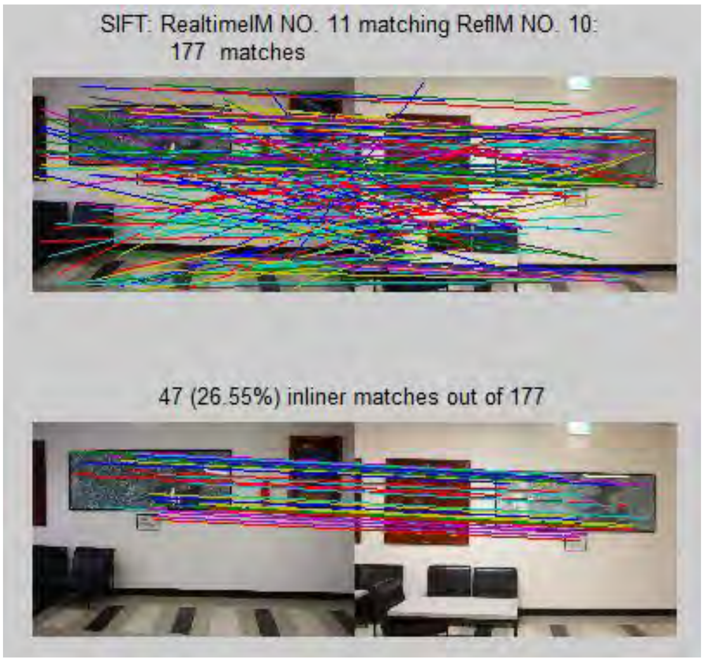

SIFT: RealtimeIM NO. 11 matching RefIM NO. 10:
177 matches
47 (26.55%) inliner matches out of 177

Figure 6. 23 Matching between real time query image No.11 with map image No. 10 using SIFT+RANSAC, small angular change, 47 reliable matches (inliers).

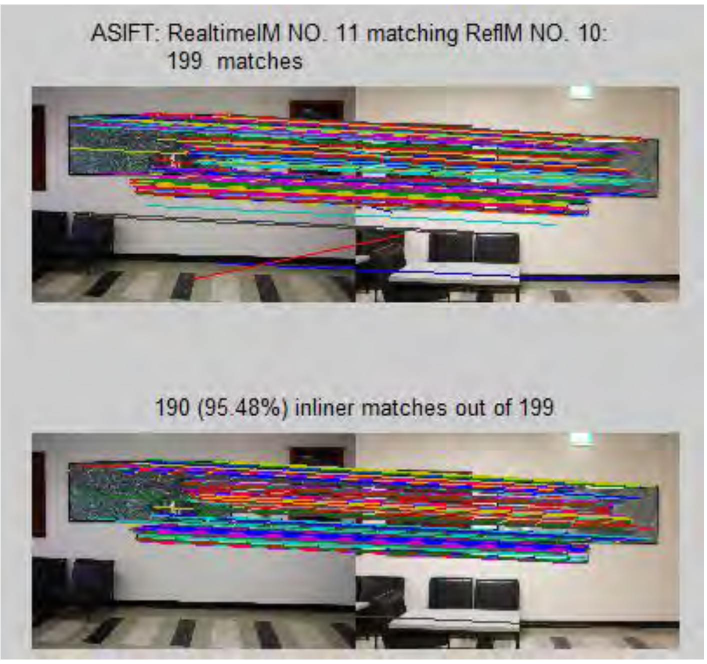

ASIFT: RealtimeIM NO. 11 matching RefIM NO. 10:
199 matches
190 (95.48%) inliner matches out of 199

Figure 6. 24 Matching between real time query image No.11 with map image No. 10 using ASIFT+RANSAC, small angular change, 190 reliable matches (inliers).

Table 6. 5 Performance of ASIFT matching in the system

SiteID	Epoch_ID	Angle	Matches	PGCPs	False PGCPs	PDOP	ADOP
1	1	0	296	13	3	2112	516
	2	0	199	7	4	62	4651
	3	-30	218	22	2	1353	277
	4	-30	150	11	2	2584	512
	5	50	381	14	0	1761	369
	6	50	163	3	0		
2	7	0	257	17	1	670	165
	8	0	134	11	1	856	217
	9	-20	178	12	1	1746	405
	10	-20	151	13	4	2144	534
3	11	0	716	15	0	9370	1913
	12	0	325	9	0	8838	1907
	13	-30	471	10	2	12801	2401
	14	-30	185	3	0		
	15	20	407	21	0	9490	1914
	16	20	260	12	0	21657	4358

Comparing Table 6.4 and Table 6.5, an obvious improvement happens on epoch No.5, one with big angular change. Using ASIFT it is able to produce a positioning result with reasonable number of PGCP. But epoch No.6, image with the same view as No.5 but limited lighting, is still unable to get a result. Compare every pair of adjacent images, it is easy to deduce that ASIFT shares the same shortcoming with SIFT: being sensitive to illumination changes. For epochs with low angular change, ASIFT has yet outperform SIFT in terms of the geometry (represented by PDOP and ADOP) provided by PGCPs. It has also been observed that the correct rate of PGCPs has yet reached 100% for in the two tables, which means both RANSAC for SIFT and ORSA for ASIFT have left some mismatches untreated. The reason behind is that all

RANSAC-like (e.g. RANSAC and ORSA) methods have the same bottleneck: when mismatches are near their epipolar line, no matter how far away they are from their true correspondences, it is hard for these mismatches be detected. This test even revealed the lower correct rate of PGCPs from ASIFT. Considering the fact that ASIFT is more computationally expensive than SIFT, for real time applications like vision-based navigation, it is important to reduce the complexity of the algorithms been used. Therefore we conclude that ASIFT is less efficient than SIFT when dealing with low angular change. We only choose to use ASIFT as a backup plan when SIFT fails to get a result because of dramatic viewpoint changes.

6. 4 MEASUREMENT ACCURACY AND FACTORS INVOLVED

In this section, the second component that determines final positioning accuracy, measurement accuracy, is discussed. It is the overall accuracy indicator of observations in the system, which mainly comes from two groups: image measurements (the image coordinates of PGCPs) and 3D coordinates of PGCPs. The first group of observations are produced by SIFT feature extraction that have consistent accuracy, and the ground coordinates of PGCPs are provided by indirect geo-referencing.

In order to investigate how the accuracy of these two groups of observations influence measurement accuracy which further affect final accuracy of a position solution, Monte Carlo simulation is used. Monte Carlo simulation is a well proven and efficient way to investigate the numerical properties of a complex mathematical model with respect to artificial noise in the input data (Robert and Casella, 2002). Noise is added with due regard to statistical distributions and typical noise levels so that the resulting output data varies realistically (Luhmann, 2009). Here, Monte Carlo simulation is used to add noise to these two groups of observations respectively, and final positioning results along with statistical analysis report are calculated. The formula follows:

$$\mathrm{P_m} = \mathrm{P_0} + (\mathrm{nRNG} * \mathrm{s_p}) \tag{6.12}$$

With $\mathrm{P_m} =$ randomly modified paramete; $\mathrm{P_0} =$ input value of parameter P ; $\mathrm{nRNG} =$random value that follows standard normal distribution; $\mathrm{s_p} =$ standard deviation/ noise level. Therefore, parameter P is modified by adding noise $(\mathrm{nRNG} * \mathrm{s_p})$ at the level of $\mathrm{s_p}$.

6.4.1 Influence of image measurement accuracy

First image measurement noise was simulated and added to the input of image coordinates for final space resection. The original input image coordinates $(\mathrm{P_0})$ obtained by SIFT matching algorithm have a measurement accuracy at 0.00004 $\mathrm{m}(\sigma_0)$, the noise level added to the input has been set varies from 0 m to 0.00008m at an interval of 0.00001m. At each noise level, the simulation runs 5000 times. The measurement accuracy (Sigma estimated) is calculated using the mean value of results from the same noise level. Figure 6.25 proves measurement accuracy decreases with the increasing noise level at image coordinates, which means the image measurement accuracy affect the measurement accuracy positively. Figure 6.26 illustrates the variation trend of DOP values with the increasing noise level, and it remains at a stable certain value, which proves the geometry factor is not relevant. In order to investigate the impact of image measurement noise on final positioning accuracy, the inner precision (standard deviation) of 6DOF measurement within the simulation results is calculated at each noise level. It proves that 6DOF precision decreases with decreasing image measurement accuracy (Figure 6.27 and Figure 6.28).

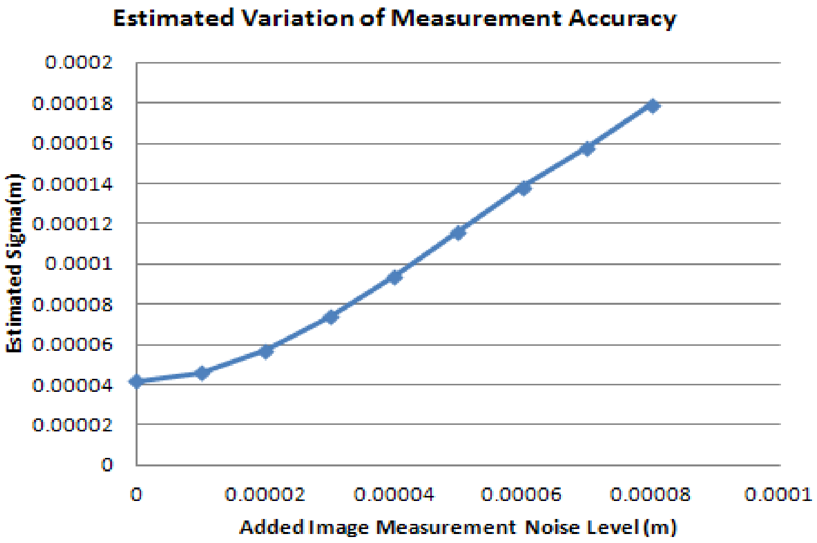


Figure 6. 25 Variation of measurement accuracy at different image measurement noise level

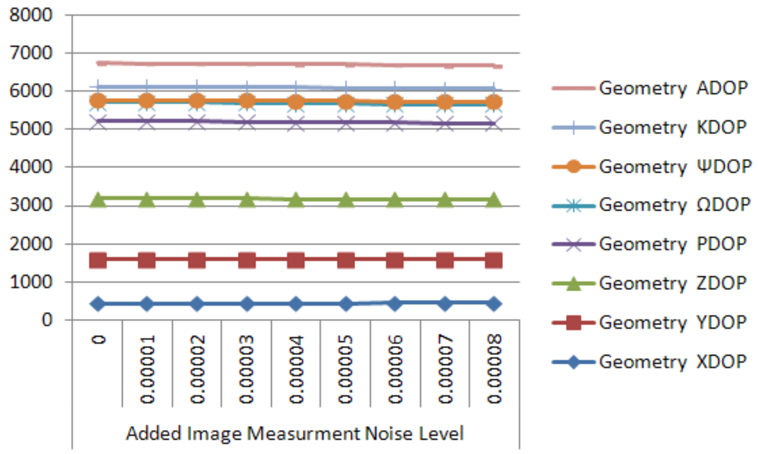


Figure 6. 26 Geometric strength (DOP values in 6DOF) at different noise level

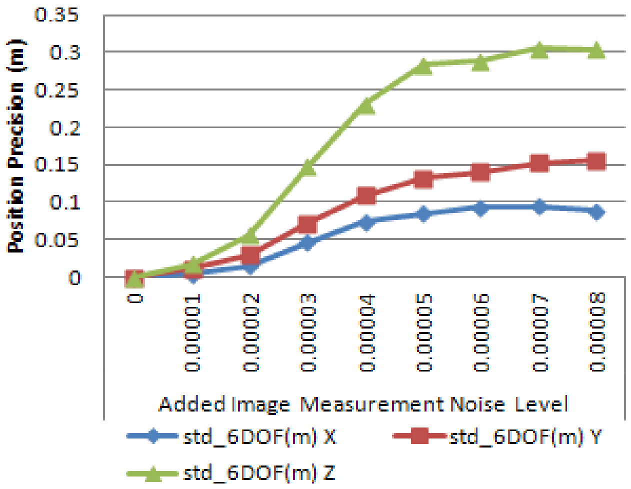


Figure 6. 27 Variation of position precision at different image measurement noise level

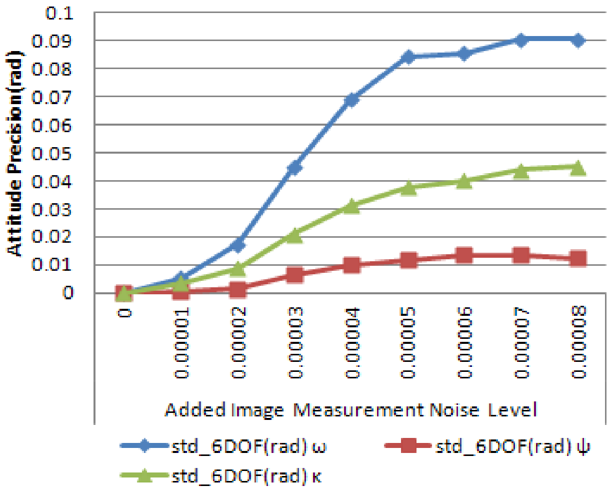


Figure 6. 28 Variation of attitude precision at different image measurement noise level

6.4.2 Influence of geo-referencing accuracy

Another factor that influences the final positioning accuracy is the precision and accuracy of 3D coordinates of PGCPs, which is determined by geo-referencing accuracy of the map images. In this experiment, noise is added to the 3D coordinates of PGCPs. The original input of PGCPs 3D coordinates have an accuracy at 0.03m, the noise level added to the input has been set varies from 0 m to 0.03m at an interval of 0.005m. At each noise level, the simulation runs 5000 times. The measurement accuracy (sigma estimated) is calculated using the mean value of results from the same noise level. Figure 6.29 proves measurement accuracy decreases with the increasing noise level at object coordinates of PGCPs, which means geo-referencing accuracy of the map images affect the measurement accuracy positively. It is also observed that DOP values remain relatively stable with the increasing noise level, which means the geometry is not affected.

In order to investigate the impact of 3D object coordinate noise on final positioning accuracy, the inner precision (standard deviation) of 6DOF measurement within the simulation results is calculated at each noise level. As shown in Figure 6.30 &6.31 6DOF precision decreases with decreasing 3D coordinates‘ accuracy. It can be observed that geo-referencing accuracy of the map images actually exerts certain amount of influence on final positioning accuracy via measurement accuracy. Therefore, in order to achieve a higher accuracy with such approach, the accuracy of bundle adjustment (indirect geo-referencing) need to be improved.

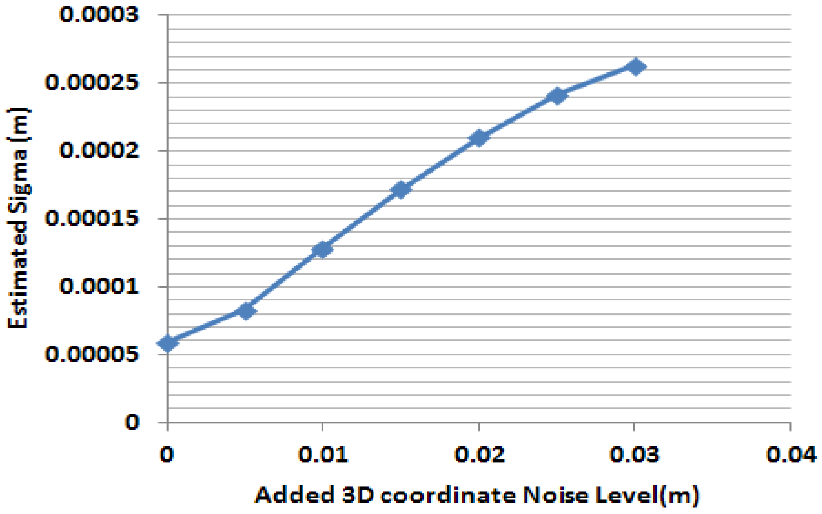


Figure 6. 29 Variation of measurement accuracy at different noise level of 3D

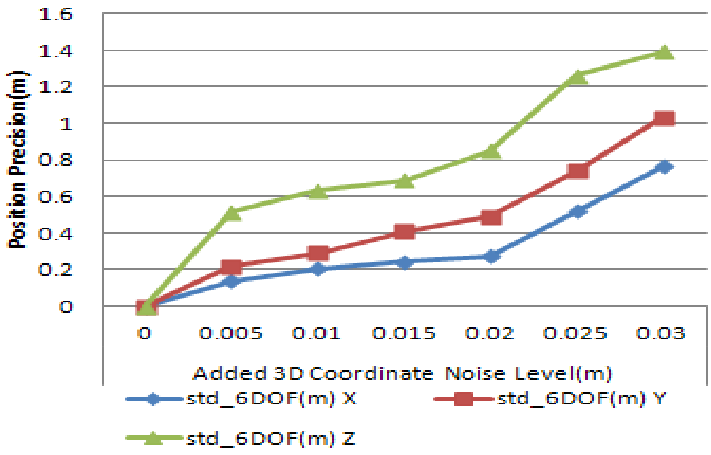


Figure 6. 30 Variation of position precision at different noise level of 3D coordinates

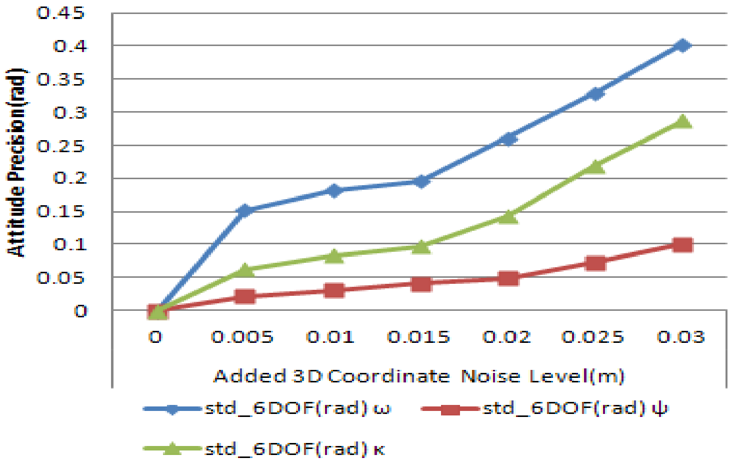


Figure 6. 31 Variation of position precision at different noise level of 3D coordinates

6. 5 SUMMARY

In this Chapter, the vision-based navigation system has gone through a numerical evaluation with a focus on its accuracy. The accuracy of such positioning method is currently between 1-10 centimetre levels, which the author believes has yet reach its fullest potential. Thus factors that affect its final accuracy are analysed. Two major components, geometry of PGCP and measurement accuracy are identified through the analysis of function model and practical performance.

The nature of the impact of PGCP geometry on positioning has been analysed through real world data as well as simulation. It is concluded that the major cause of inaccurate results is bad geometry, and geometric impact plays an important role in the determination of final positioning precision. Therefore, the generation of PGCP with an even distribution on the query image is of significant importance for the system performance. Factors that contribute to the PGCP generation need to be investigated. Since the generation of PGCP is closely related to the image matching procedure, the system has been evaluated against various image matching conditions, like viewpoint changes and illumination changes. Possible improvements have been discussed.

In the later part of the Chapter, factors that influence measurement accuracy for final positioning are also identified and analysed: image measurement and geo-referencing accuracy at mapping stage. Their influence is simulated by inserting artificial noise and the consequent impacts are evaluated by using Monte Carlo simulation. It is observed that the accuracy of map image geo-referencing can exert a substantial effect on final positioning accuracy.

In summary, this Chapter carries out an in-depth and thorough evaluation on the nature of the vision-based positioning and navigation system. Factors that influence the positioning performance (mainly accuracy) have been identified and discussed. Some limitations have been revealed. The system largely depends on the generation of PGCPs as 3D landmarks for positioning. Therefore, it cannot function well in environment lack of stable visual features. Meanwhile, the improvement of image measurement accuracy as well as geo-referencing accuracy can benefit the system performance. Future improvement of the system can be based on these aspects.

CHAPTER 7
OUTDOOR EXPERIMENTS AND DISCUSSION

7.1 INTRODUCTION

Ubiquitous positioning is considered to be the main goal for today‘s Location-Based Services (LBS). While satellite-based navigation has achieved great advances in the past few decades and has been applied to both military and civilian applications in a mature manner, positioning and navigation in GPS challenged areas has remained a largely unsolved problem and thus is currently receiving growing attention. The availability of rich imagery of large parts of the earth‘s surface under many different viewing conditions presents great potential, both in computer vision research and for practical applications (Snavely et al., 2008). The author believes it can also bring enormous opportunities to the field of navigation and location-based service. Images associated with vision sensors have been researched for positioning and navigation purposes since early last century. They are superior to many other techniques because they can operate both indoors and outdoors. In recent years the low cost built-in sensors on mobile devices (e.g. smartphone), especially high resolution cameras have placed greater demand for a breakthrough in their applications for location-based services.

In previous Chapters, vision-based navigation has been mainly focused on indoor areas. In this Chapter the aim is to extend the function to outdoor environment so as to fill in the gap for satellite based navigation systems. However, the significant difference between outdoor and indoor environments has divided the early stage of the research into two different groups. For outdoor vision-based navigation system, the traditional approach is to match the real time query image with the reference images in the database. Whenever a match is found, the position information of this reference image is transferred to the query image and used as user position. This is essentially an object-recognition and image retrieval problem. A great variety of work has been done to address the location recognition aspect by using different image matching techniques (e.g. Schaffalitzky and Zizzeman,2002;Goedeme and Tuytelaars, 2004). A further improvement is to calculate the relative position between the query view and the identified reference view to obtain more accurate position estimation. In 2006 Zhang and Kosecka first used a wide-baseline matching technique based on SIFT features to select the closest views in the database, then the location of the query view was obtained by triangulation. In Robertson and Cipolla (2004) the orientation of the sensor was also estimated since the pose of the query view is obtained from plane-to-plane transformation. Building façade was used as dominant plane. On the other hand, indoor visual navigation has been considered a quite different field. Related research has mainly focused on robotic visual SLAM (e.g. Davison, 2003), and significantly different methodologies like structure from motion and stereo viewing (Elinas et al., 2006) are adopted, which are not suitable/attainable to be extended to LBS for common users. The major reason for it lies in that indoor and outdoor environments create different scenarios, requirements, and sometimes contradictory conditions to the visual system. In terms of size indoor environments are limited to certain buildings while outdoor positioning requires regional or even global coverage. In terms of accuracy indoor positioning obviously poses greater challenge. In terms of sensors used to assist vision sensor, satellite based navigation system can only cover outdoors and WiFi is more likely to be used for indoors. In terms of vision-based algorithms and methods, greater diversity can be found. For instance, the use of stereo vision to extract depth information is more suitable for indoors since the range for depth detection is limited by the baseline. And the visual features are normally different: in outdoor environment artificial landmarks (e.g. buildings, road signs) feature primarily the edges and corner points, while for indoor environment features are richer in the shapes and textures. However, despite of all these factors, vision is by its nature capable of working in a complementary manner to satellite-based technology. Therefore, it is high time that a consistent framework of image-based visual navigation technology to be developed, which is capable of filling the gap in satellite-based system deficiencies, providing coverage from outdoors, urban canyons to indoor environments.

Seamless vision-based positioning covering both indoor and dense urban environments has therefore attracted growing attention nowadays, especially for the research of navigation system that used on mobile platforms. A common approach is to use visual device as the main or complementary sensor that collaborate with other sensors on board. In Przemyslaw and Pawel (2012) the authors present an algorithm for estimating a pedestrian location in urban environment, and include data from GPS, inertial sensors, probability maps and a stereo camera. In (Chen et al., 2012) the authors utilize the GNSS and map-matching for outdoor positioning; multiple sensors as well as signals of opportunity is adopted for indoor environment. From the perspective of vision-based navigation, the major difference between these two approaches lies in that the former uses stereo camera to directly extract depth information, while the second is based on single camera and query image matching. Following these two main streams, many vision-related navigation solutions designed for GPS-degraded environment can be found (e.g. Ruiz-Ruiz et al., 2012; Pei et al., 2011; Karimi, 2011; Jaspers et al., 2012; Chowdhary et al., 2013).

In this Chapter a hybrid vision-based positioning system is presented, which extends indoor visual navigation to outdoor environment. Since GPS can achieve good positioning resolution in open areas, the vision-based system is designed to function in places where artificial landscape is available, such as urban canyons. Such nature can mitigate the deficiency of satellite-based systems. It mainly uses visual input to match with geo-referenced images for image-based positioning resolution, and also takes advantage of multiple sensors onboard, including GPS receiver and a digital compass to assist visual methods in various aspects.

7.2 IMAGE GEO-REFERENCING AND MAPPING FOR OUTDOOR ENVIRONMENTS

During the mapping stage, image feature geo-referencing is the core process. Different types of image features are used to cater different scenarios. In outdoor urban environments, artificial landmarks (e.g. buildings, road signs) feature primarily the edges and corner points. Therefore they are first surveyed in the field with Map Grid of Australia (MGA) coordinates derived. Then images are collected and Harris corner detector is used to exact the corner features from these images. Among the big number of features points extracted, the ones that have been surveyed are identified from the feature list and associated with images with both 2D image pixel coordinates and 3D surveyed coordinates been recorded (e.g. Figure 7.1 and Table 7.1). For landmarks that have a certain volume in the space, such as buildings, images of façades are geo-referenced. Each building is described by the contour information (corner coordinates and edges they intersected) and the geo-referenced images, including both the outdoor façades and the indoor environments.

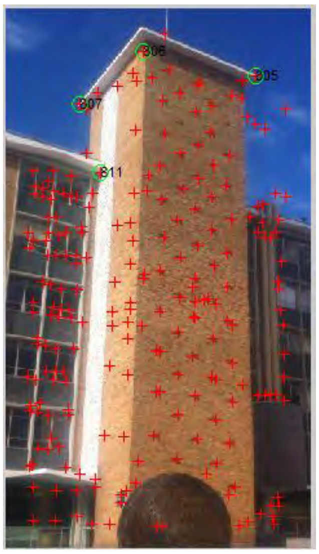


Figure 7. 1 Reference image No.15: Corner points have been extracted by Harris corner detector and shown with red crosses; 4 geo-located corner points have been identified and shown in green circle.

Table 7. 1 Geo-located corner points for reference image No.15,

X, Y in pixel and Easting, Northing and Height in meters

RefIM	PointID	X(pixel)	Y(pixel)	Easting(m)	Northing(m)	Height(m)
15	807	83	108	336512.9	6245551	63.9
15	811	106	184	336511.4	6245548	56.9
15	806	154	49	336511.9	6245543	63.9
15	805	280	76	336519.8	6245542	63.9

7.3 OUTDOOR POSITIONING

In urban canyons or indoor areas, GPS positioning accuracy can be degraded because the signal may suffer from blockage, multi-path effects, etc. For single point positioning (SPP) used on people‘s mobile devices, the accuracy can be 10s meters or worse. Therefore, image-based methods are used to mitigate the deficiency. However, if solely replace GPS with vision-based methods, retrieving images from a large image database that covers the whole navigation route will be time consuming and the computation load is not affordable for mobile devices. Therefore, a multi-step solution is proposed: firstly GPS data is used to narrow down the search space; then a voting strategy is used to find reference images corresponding to the query view among the localized image space; finally, a hybrid technique is proposed that uses the measurements from GPS, digital compass and visual sensor onboard to calculate the final positioning result in 6DOF for outdoor environments. The overall outdoor positioning procedure is shown as follows:

- ***Step1***:Take query image with GPS and compass measurements;
- ***Step2*** : Use GPS data to localize image space;
- ***Step3:*** Retrieve from the candidate image space the reference image(s) that contain the scene corresponding to the query image. If no correspondence is found, go to step 1 with enlarged search space. If yes, continue;
- ***Step4:*** Outdoor positioning resolution.

7.3.1 Using GPS to localize image space

The 3D maps are storedin the form of reference images and their geo-information. We treat each building as a record and describe its contours with line segments which contain coordinates of both ends (corners). When a user is navigating (walking or driving) through the space, images are taken when position information is required.

Whenever a query image is taken with its GPS position tagged, the initial position is given by the GPS tag and the initial orientation is given by the digital compass onboard (e.g. P for query image No. 3 in Figure 7.2). A circle will be generated with the center at current GPS tagged position, and the radius (r) determined by a threshold (n), which is a certain magnitude of the horizontal precision of the GPS reading (σ). By default n=1.

$$r = n * \sigma \quad (7.1)$$

Then the system will search for landmarks (corner points or line segments) appearing within the circle. The mobile device will load the images related to the landmarks that have been found. For buildings, the algorithm calculates the shortest distance between center point and the building line segments. If this distance is shorter than the radius, the segment line must cross with the circle. The corresponding building(s) images will be chosen and form an image space for further processing. Figure 7.3 illustrates the calculation process for user position P, and the resulting image space is shown in Figure 7.4. By using GPS information to narrow down the search space, the query image will need to match with a small image space rather than a whole image database at later stage.

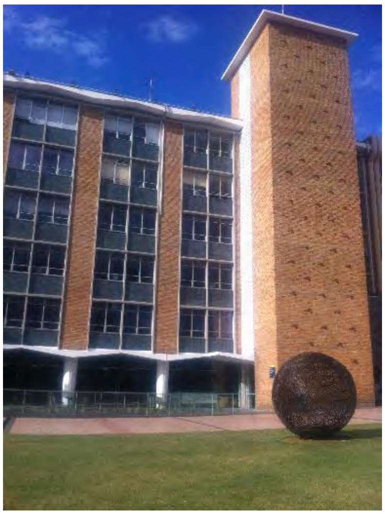


Figure 7. 2 Query image No. 3 with GPS tag information: Zone 56, Latitude: -33 ° 55' 5.40120",Longitude:151°13'52.79880 ", Altitude: 33.92m; digital compass measurement: 34° NE

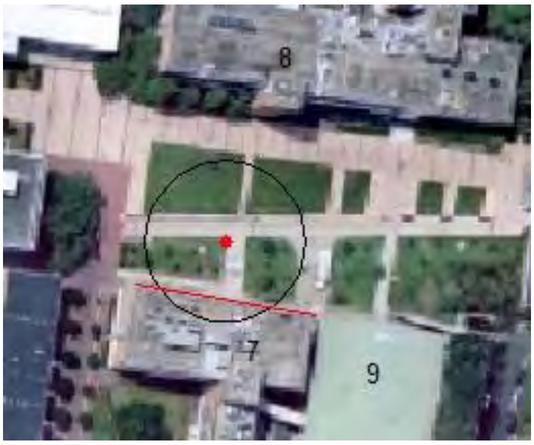


Figure 7. 3 Given a GPS position data from mobile device at P shown with red dot, a circle is drawn with 20m radius representing the search space. Line segment from building No. 7 crossed with the circle, so reference images of building No. 7 are chosen for image space

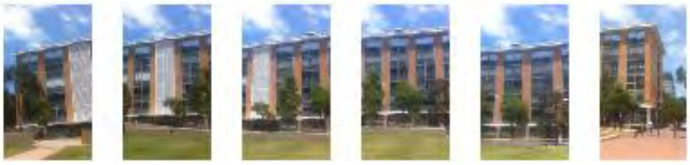


Figure 7. 4 Image space created for P including façade images of Building No.7.

7.3.2 Image Retrieval using SIFT-based Voting Strategy

The goal of this process is to retrieve, in the candidate image space, the reference images with the scene corresponding to the query image. The procedure is similar to the voting strategy introduced in Section 3.2.2. In the system this process identifies the target building/local environment and prepares the corresponding reference images for outdoor & indoor navigation. The SIFT features are extracted from the query image (e.g. Figure 7.5) and matched with the reference feature database generated from the candidate image space (e.g. Figure 7.6).

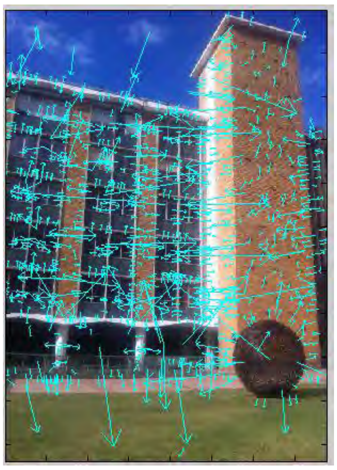
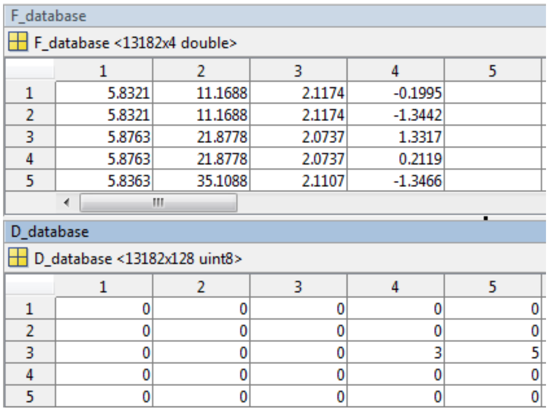


Figure 7. 5 SIFT features extracted from the query image No.3



F_database

F_database <13182x4 double>

	1	2	3	4	5
1	5.8321	11.1688	2.1174	-0.1995	
2	5.8321	11.1688	2.1174	-1.3442	
3	5.8763	21.8778	2.0737	1.3317	
4	5.8763	21.8778	2.0737	0.2119	
5	5.8363	35.1088	2.1107	-1.3466	

D_database

D_database <13182x128 uint8>

	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	3	5
4	0	0	0	0	0
5	0	0	0	0	0

Figure 7. 6 Reference feature database generated for the candidate image space: 13182 features from 20 reference images in the search image space.

The difference of image retrieval for outdoor positioning lies in that if no corresponding image is found within current candidate image space, the procedure goes back to previous step and enlarges the search space (the threshold n) to recalculate the crossed landmarks. And the images in previous image space are removed and new image space is produced. In the example of query image No.3 at point P, image space in Figure 7.4 is removed because no corresponding image has been found, and new image space in Figure 7.7 is generated after the search space has been enlarged.

The reference images with greater numbers of votes obviously have higher chance of containing common scene with query image. Therefore, by ranking in descending order of the number of votes, the top m (5 in this case) reference images are chosen as ones corresponding to the query scene and retrieved from the candidate image space. Specific building(s) that covered by the query view can also be identified. One problem however, is that we still cannot be certain if the top ranked images containing common view with the query image. The evaluation test introduced in Chapter 5 is adopted to remove images with no common areas.

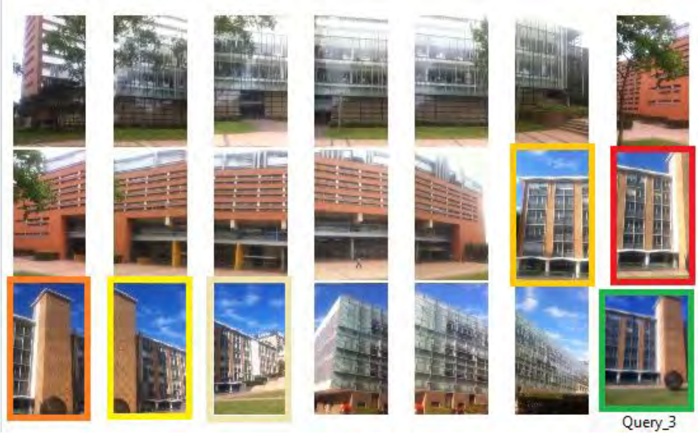


Figure 7. 7 Image retrieval: 5 top ranked images have been identified from the image space with 20 reference images. Query image has a green border when the 5 top voted reference images have borders from dark red to light yellow, the darker the colour the higher rank it has (which indicates greater relevance). All top voted reference images indicates the same target building, BD No.8

For the top 4 reference images, the tests all past with high pass rates. The results are shown in Figure 7.8 –Figure 7.11 .

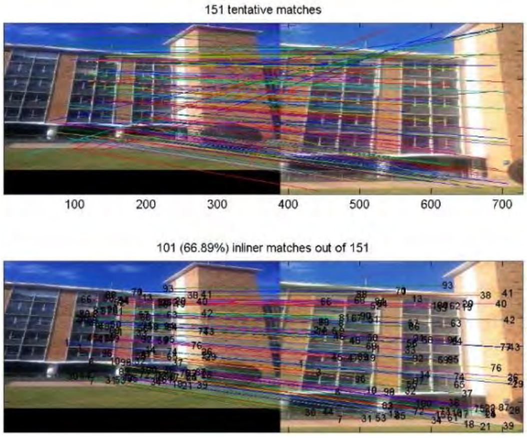


Figure 7. 8 Query image3 match with reference image 14 pass rate at 98.8%

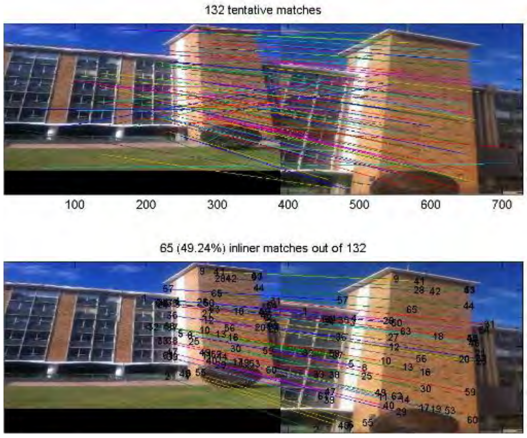


Figure 7. 9 Query image3 match with reference image 15 pass rate at 100%

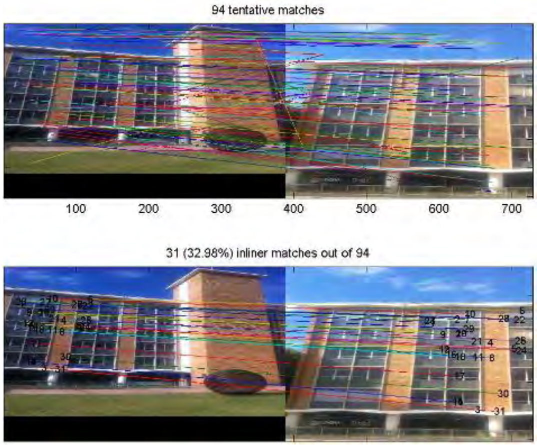


Figure 7. 10 Query image3 match with reference image 13, pass rate at 89.7%

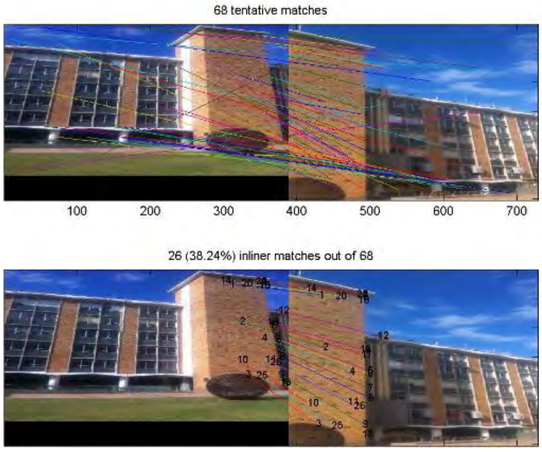


Figure 7. 11 Query image3 match with reference image 16, pass rate at 95.8%

The 5th ranked image, reference image No.17, however failed to pass the test. Its maximum cross correlation r_{max} is 0. As shown in Figure 7.12 , we can easily observe that although reference image No.17 and the query image include the same building, the two cover different parts and have no common areas. The image matching with RANSAC performed poorly, retaining 9 mismatches as inliers. Using the cross correlation information to evaluate the RANSAC process, the algorithm successfully identified the failure of the image matching and removed the reference image No.17 from the corresponding image list.

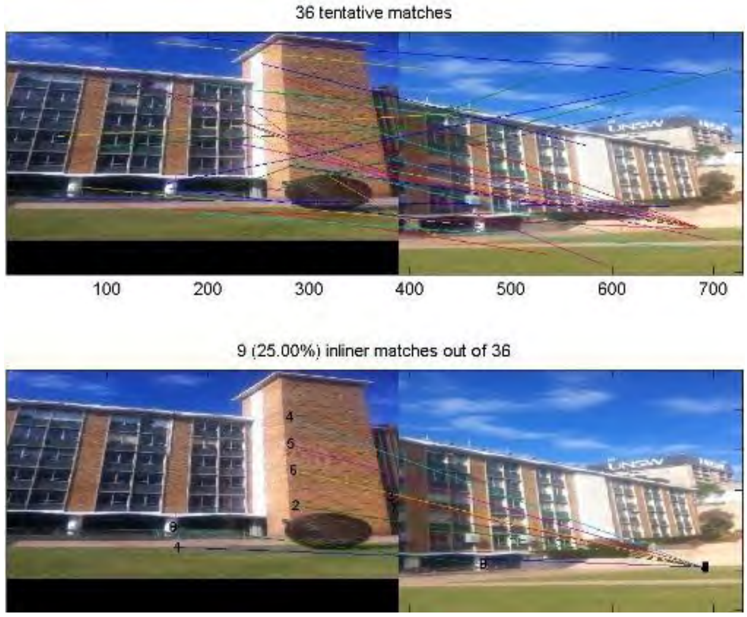


Figure 7. 12 Evaluation test fail for reference image No.17: r_{max} at 0.00

7.3.3 Outdoor positioning resolution

This section introduces a hybrid technique that uses the GPS as well as the digital compass measurements, and image-based positioning technology for outdoor positioning. In fact it deals specifically with urban environments with artificial landmarks.

After reference images have been identified in the previous step, positioning is carried out based on the matching between the query image and identified geo-referenced images. Since the outdoor reference images are geo-referenced through the corner features, to ensure corners to be matched in the query image as well as to strengthen the robustness of matching, the author apply a combined use of the Harris corner detector and the SIFT descriptor (referred by Harris/SIFT method). Firstly, the Harris corner detector is used to extract corner features from the query image and SIFT descriptors are computed at the positions detected by Harris detector on the query image. In the meantime, SIFT descriptors are also generated for geo-located corner features on the reference image that to be matched. Then feature matching is carried out between the two images based on SIFT descriptor matching. RANSAC is used to remove mismatches. As shown in Figure 7.13, 15 pairs of correct matches are found, among which 4 are geo-located corner points (No. 4, No.8, No.13, No.15) that are identified in Figure 7.1 and Table 7.1. Therefore, the 3D geo-locations of these points are transferred from reference image No. 15 to the query image, which can be then used as PGCPs for positioning resolution. More PGCPs can be generated by matching the query image with all the corresponding reference images selected by previous step. The given query image obtained 6 PGCPs as illustrated in Figure 7.14.

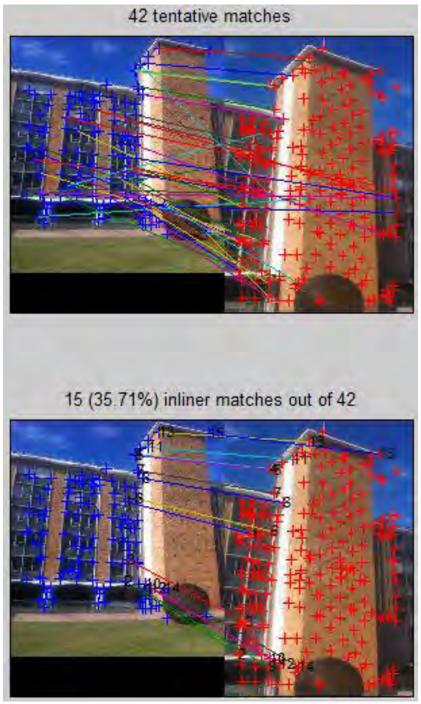


Figure 7. 13 Query image matching with reference image No. 15 using Harris/SIFT method; Harris corner features are tagged by blue and red crosses respectively, and matched corner features using SIFT descriptor matching are shown by lines

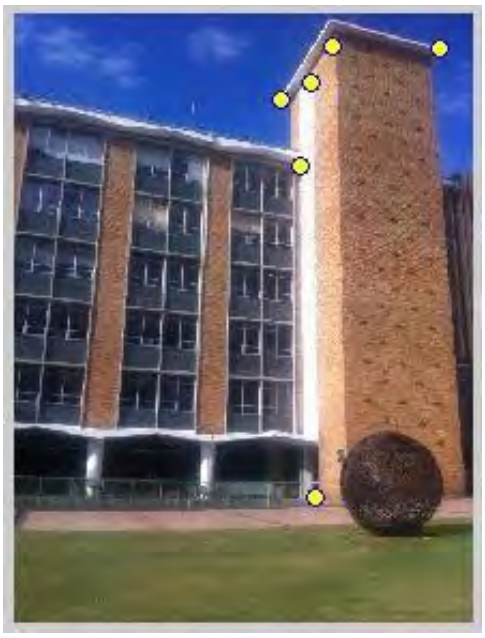


Figure 7. 14 PGCPs generated for the query image No.3, which are shown with yellow dots.

After enough PGCPs have been generated, the methodology introduced in Chapter 3 is used to solve the position and orientation of the query image. Although space resection based on a least squares solution can provide relatively accurate result, it requires a good initial value for the least squares adjustment to converge. This is where the raw GPS and digital compass measurements come into play. Normally standalone GPS or AGPS is used on mobile devices, which can only provide either low accurate positioning or unstable performance in dense urban area. Therefore, the GPS provides the initial positions and compass chip on mobile devices registers magnetism in three dimensions, which gives initial orientation values. By using vision-based positioning, user position and camera orientation in 6DOF can be achieved. For the query image, the computed position is shown in Figure 7.15.

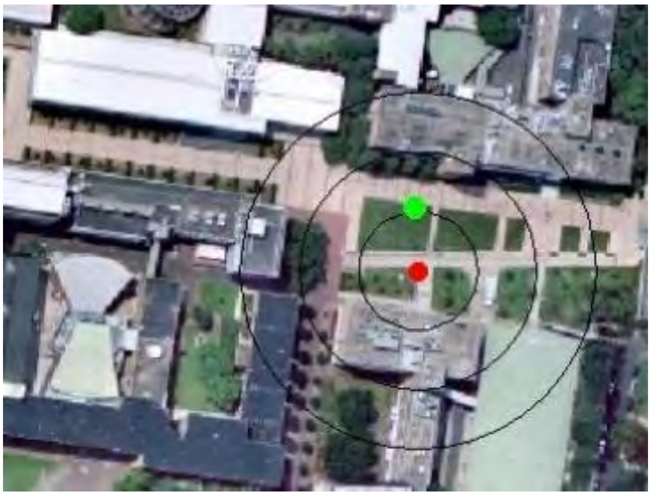


Figure 7. 15 Positioning result for the query image shown with green dot. The red dot indicates the location determined by the GPS, with the black circles show the process to enlarge the search space (1-3 times of its horizontal precision).

7.4 EXPERIMENTS

In the experiment, a positioning test in outdoor environment was carried out. A path on the university campus is chosen. It has 7 pre-surveyed GPS ground control points (GCPs) on the way, as well as buildings on both sides, which can simulate the situation of urban area. Images of the building facades as well as indoor environment (test scene) were recorded in the database and geo-referenced 3D maps were generated for positioning. Then a user walked along the path. The position of each epoch when images were taken and the trajectory are resolved based on the image-based system developed. The data is post-processed using Matlab 2011a and an orthophoto of the UNSW campus.

During the outdoor test, a user holding a mobile handset walked along the path and took (query) images for self-localisation. The device used was an iPhone 4 smart mobile phone, which integrates a backside-illuminated 5 megapixel rear-facing camera with a 3.85 mm f/2.8 lens, and employ an assisted GPS system (A-GPS) and a specialized integrated circuit chip as the iPhone's digital compass for navigation. The 'user' passed by each of the 7 GPS GCPs and took images of the environments on the 7 sites, and randomly took another 11 images along the path. Totally 18 epochs were resolved.

Firstly, the performance of outdoor image-based positioning with its accuracy was investigated by calculating the user positions and orientations at the 7 GPS GCPs through the proposed method and compared them with surveyed true values. The accuracy of vision-based method and standalone GPS is also compared. The 6DOF results are shown in Table 7.2 and the accuracy revealed by RMSE in Table 7.3. It can be seen that the GPS measurements in the urban environment are poor, around 20m in the experiments. By using the proposed method, the accuracy has been improved to around 10m in the test.

Table 7. 2 System calculated positioning results in 6DOF for the 7 GCPs

Epoch ID	Easting(m)	Northing(m)	Height(m)	Omega (degree)	Phi (degree)	Kappa (degree)
1	336269.08	6245563.38	26.67	-89.30	1.44	71.35
2	336291.39	6245554.18	23.78	-119.22	-1.25	96.66
8	336435.40	6245546.05	30.97	-123.64	-5.74	117.53
11	336478.50	6245533.39	36.99	-103.93	-159.27	-64.17
12	336522.79	6245543.21	49.73	-139.19	173.80	-136.67
13	336562.16	6245522.26	44.43	-92.81	-12.54	-80.99
18	336511.91	6245409.58	48.33	58.16	153.63	12.12

Table 7. 3 RMSE of GPS measurements and system calculated positions using surveyed values as true values.

RMSE	Easting(m)	Northing(m)	Height(m)
GPS measurements	20.37	19.59	21.00
Calculated	8.43	10.31	7.20

Secondly, the overall trajectory, including 18 epochs, is calculated and shown in Figure 7.16 (horizontal) and Figure 7.17 (vertical). It can be easily observed that horizontally the raw GPS measurements present a few jumps (e.g. epoch No.2) and intersected track, which are not true; while the vision-based method provides a trajectory closer to the true trajectory. Meanwhile, the height information provided from GPS deviate largely from true values, while the system results are much improved.

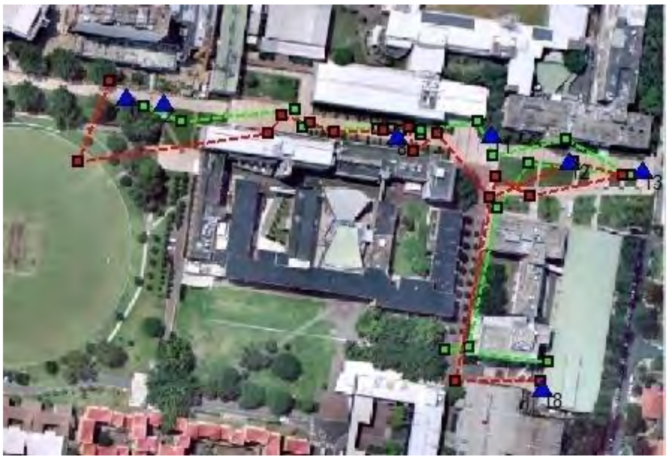


Figure 7. 16 Red dash line shows the trajectory obtained from the build-in GPS receiver, while green dash line shows the calculated results; blue triangles represent the true GCPs that user passed by.

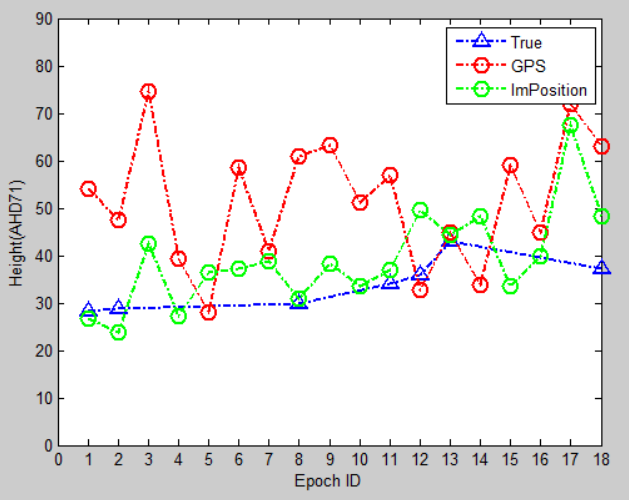


Figure 7. 17 Height: blue icons represent true values; red ones are altitude measured by the device; green ones indicate the calculated results.

Thirdly, the study investigates the theoretical precision for the vision-based position solution in 6DOF by using their estimated standard deviation, as shown in Figure 7.18 and Figure 7.19. It can be observed that most of the epochs have a 0-5m standard deviation on each direction, while the orientation standard deviation mostly between 0-10 degrees. The theoretical precision is consistent with the accuracy evaluated by the 7 check points. Moreover, it is noticed that certain epochs have very low precision (large standard deviation) compared with other epochs (e.g. Epoch No. 1, No.2, No.10, No.13). The reason behind is poor geometric distribution of PGCPs on query images. Such nature has also been found and further investigated in the indoor experiments. The main contributing factors that determine the PGCPs geometry are the geo-referenced 3D feature density of the reference images, the quality of image matching and most importantly the covered scene of the query image. Therefore one possible way to improve the outdoor positioning performance is to include greater number of corner features with better distribution when taking the query image. In other words, such vision-based method performs the best in areas where artificial landmarks are sufficient (like deep urban canyons), which is a complementary character for satellite-based navigation system.

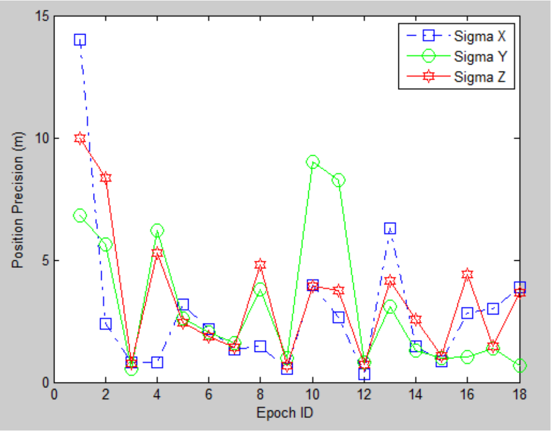


Figure 7. 18 Position precision for the 18 epochs

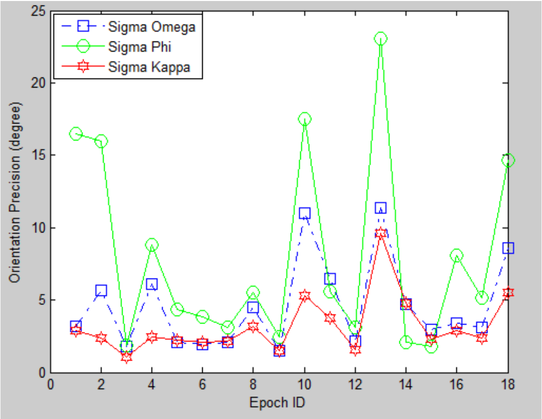


Figure 7. 19 Orientation precision for the 18 epochs.

In summary, for SPP used on people‘s mobile devices in urban canyons, the accuracy can be 10s meters or worse, and varies significantly depending on the signal. Vision-based methods, on the contrary, can provide stable results with relatively better accuracy as long as enough visual features are covered by the query image.

7.5 SUMMARY

This Chapter has presented a comprehensive system that adopted hybrid vision-based method with combined use of onboard sensors (GPS, camera and digital compass) to achieve a seamless positioning from indoor to outdoor environments. Using the same strategy, geo-referenced images are used as 3D maps for vision-based positioning. Therefore the outdoor positioning share the same nature with the outdoor system: the geometry of PGCPs essentially determines the positioning accuracy.

Experiments have demonstrated that such a system can largely improve the position accuracy for areas where GPS signal is degraded (such as in urban canyons). It also reveals the major challenge for such system, that is, it largely depends on the texture of the view. For outdoor environment, shortage of texture because of poor lighting condition may poses tremendous challenge to the system. Future research will be focused on these aspects.

The author believes that the system has potential to overcome the deficiency of satellite based solution, since it targets at GPS challenged environment and works especially well at places with buildings/ artificial landmarks and indoors. The required hardware, single camera integrated with GPS receiver and digital compass, can be easily found on people‘s mobile devices (smart phones etc). With the boom in LBS and growing attention to geo-spatial techniques for everyday life (e.g. GPS-tagged image), we hope such technologies can bring the vision based techniques for position and navigation to a new level and finally achieve ubiquitous positioning.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1 CONCLUSIONS

Navigation technology is booming along with the growing demands from consumers. The major challenge today is to provide positioning capability in GPS-degraded environments, such as indoors and urban canyons. Vision sensor is regarded to be highly promising because of its ubiquitous and self-contained nature. Although vision-based positioning and navigation system has been investigated and developed for over two decades, the early research has been limited by the hardware and image processing capability. Today both cameras and underlying platforms have been advanced dramatically. Build-in high resolution cameras as well as many other sensors (e.g. compass) have been adopted on people‘s mobile devices for daily life. All these technologies can be potentially used for navigation. Moreover, significant progress has been made in image processing algorithms and related research. Therefore, given the rich literature of vision-based navigation and the recent progress in terms of both software and hardware, it is high time for a breakthrough in the research domain.

Most existing vision-based navigation systems depend on the exploitation of one or more cameras. Like other navigation technology, vision-based approaches can also be divided into two categories: position-fix and dead reckoning. The former method is based on the matching between query image and pre-stored information of the environment to realize self-localization. This study gives such approaches the name map-based approach. The later one determines the camera motion through the analysis of sequence of images. It is therefore a mapless approach. This method has been limited to short distances since it suffers from drifting errors and requires other sensors to calibrate. In this research, we mainly focused on the map-based visual navigation.

Currently, they are still far from mature to supplement GNSS. Major challenges have been identified in four aspects: mapping, poisoning accuracy, reliability, and coverage. To respond to these challenges, this research introduced a vision-based navigation system that aiming to address these problems and provide a comprehensive navigation solution.

8.1.1 Reality-based 3D map

The map used for vision-based navigation has gone through a series of developments. It is one of the major components for the navigation system. But it has yet been fully investigated. The literature of this topic has been reviewed in Chapter 1. It started with 3D models, however the full 3D reconstruction lacking details limits the development of this approach. Then 3D models are replaced by appearance based approaches where images are used to ―memorize” the navigation environment. Navigation map in the form of images can retain the details of the environment and save the effort for 3D reconstruction. The major shortcoming of pure image based map, however, lies in that it can only provide 2D information of the surface and rough 3D location information. It can hardly support positioning which require high accuracy. Recently, 3D image feature based visual positioning has increased the interest of researchers as it not only takes advantage from the appearance based approach, but also been able to provide 3D geometric information of the environment for pose estimation.

In this research, the author has further investigated this idea and proposed a new concept of 3D map. The new 3D map is defined as a sum of geo-referenced points with three dimensional (3D) local or global coordinates that are overlapped on images of the environment. Users of the 3D map will have the benefits of geo-referencing with 3D coordinates as well as realistic visualization. One basic function of the 3D map is for positioning and navigation. The main difference between this approach to the available image-based navigation methods lies in the fact that the images are geo-referenced, which means they themselves can give absolute position information(local or global) in 3D, functioning like a sensor (eg.GPS). Such a 3D map is the foundation of this research. It enables positioning to be resolved in high accuracy with 6 degrees of freedom. Whenever a new query image is taken, it can be matched with the geo-referenced map images. Then common feature points can be used to transfer the 3D information from the map to the query image and used for position solution.

In Chapter 2, the 3D map has been introduced with its development procedure. Range-based and image-based 3D mapping have been the two main stream methods to obtain 3D geometric information for map construction. In this research, image-based method from the field of photogrammetry has been adopted. The essential part is image geo-referencing, and an indirect geo-referencing method is used. Multi-image matching has been introduced into the mapping process, and bundle adjustment is used to calculate the 3D object coordinates of the feature points. The 3D map has been implemented covering an indoor testing field. Its accuracy and capability to support vision-based positioning have been evaluated through experiments. Experiments evaluated both the theoretical precision and absolute accuracy of geo-referencing, as been at centimetre level, and the positioning result from the 3D map with better geo-referencing accuracy is more accurate.

8.1.2 Vision-based positioning and navigation

After the 3D map has been built, the vision-based positioning and navigation system proposed in this research was introduced in Chapter 3. Its extension to outdoor environments is given in Chapter 7. The main methodology contributions of the system are listed as follows:

1) This research proposed the use of geo-referenced image feature points as 3D landmarks for positioning. The main idea is that by matching the query image with the 3D map, the 3D information is transferred from the geo-referenced map images to corresponding image features on query image for self-localization.

2) The positioning result is calculated based on photogrammetric space resection. Such an approach has rarely been found in previous research. Here 3D feature points are served as pseudo ground control points. The methodology of space resection has been revised according to the needs of the system. More specifically, the 3D information transferred from the map is treated as pseudo observations so as to receive adjustment during the least squares process. In this way, the system only requires a single camera but is able to give highly accurate position information in 6 degrees of freedom, which is superior to many similar systems. Experiments have shown that in indoor environment the positioning accuracy is around 10-20cm even using low resolution cameras.

3) Seamless vision-based positioning covering both indoor and dense urban environments has attracted growing attention nowadays, especially for navigation systems that can be built on mobile platforms. Therefore in this research, the system has been extended from indoor to outdoor environments. A comprehensive system that adopted a hybrid vision-based method with combined use of built-in sensors (GPS, camera and digital compass) has been presented. The consistent framework consists of the use of geo-referenced images as 3D map and space resection for position solution. For outdoor it adopts multiple sensors to assist the position solution. More specifically, a multi-step solution has been proposed: firstly GPS data is used to narrow down the search space; then a voting strategy is used to find reference images corresponding to the query view among the localized

image space; finally, a hybrid technique is proposed that uses the measurements from GPS, digital compass and visual sensor onboard to calculate the final positioning result in 6DOF. Experiments have demonstrated that such a system can largely improve the position accuracy in areas where GPS signal is degraded (such as in urban canyons).

8.1.3 Accuracy analysis of the vision-based navigation system

The accuracy and reliability of the system are the major concerns of this research. The later aspect will be further concluded in Section 8.1.4. The accuracy of such photogrammetric approach of vision-based positioning is depending on the precision and accuracy of final space resection process. Based on both mathematical model and experiments, it has been identified that the final positioning accuracy is a function of PGCP distribution and measurement accuracy. Any factor that has certain impact on either of these two major components will to certain degree influence final positioning accuracy. Therefore in this research, the ways that different factors influencing the positioning accuracy have been analysed.

PGCP geometric distribution has been analysed first. DOP values are adopted to evaluate its strength. Experiments based on real world data reveals that the major cause of inaccurate results is bad PGCP geometry, and geometric impact plays an essential role in the determination of final positioning precision. The simulation tests have been used to further analyse the nature of the relationship between geometry of PGCPs and system performance. It is observed that PGCPs should be selected as many and evenly distributed as possible. Otherwise, insufficient number and too centred distribution of PGCP may lead to the failure of the least squares adjustment, since it cannot converge.

Since PGCPs are produced by matched image feature points, any factor that influence the feature-based image matching between the query image and reference image(s) will affect the density and geometric configuration of PGCPs, which includes the richness of features, illumination, viewing angle, etc. Therefore in this research, the performance of the system has been evaluated with varying image matching conditions. A controlled experiment has been carried out to evaluate the performance of SIFT matching for the system. Major weaknesses such as viewpoint changes and mismatches incurred by image matching in the system have been identified. Experiments revealed that these factors could lead to insufficient or false PGCPs. ASIFT has been introduced into the system to deal with viewpoint changes, and outlier detection mechanism has been used to detect and remove false PGCPs.

The second component that determines final positioning accuracy is measurement accuracy. It is the overall accuracy indicator of observations in the system, which mainly comes from two groups: image measurements (the image coordinates of PGCPs) and 3D coordinates of PGCPs. The first group of observations are produced by feature extraction, and the ground coordinates of PGCPs are provided by indirect geo-referencing. Monte Carlo simulation has been used to investigate the impact of the accuracy of these two groups of observations influence measurement accuracy, which further affect final accuracy of a position solution. It has been concluded that feature extraction has a consistent pre-determined accuracy, while geo-referencing accuracy varies for the 3D map. Therefore, the improvement on geo-referencing accuracy will benefit the positioning accuracy.

8.1.4 Quality control of the vision-based navigation system

Vision sensors are cheap, ubiquitous, self-contained. However, it is also inherently fragile against errors. It has a high input rate but can only capture the two dimensional information of the 3D world- direct measurement of real world geometry is lost. For vision-based navigation systems, one essential element is to find feature correspondences between images, which can be reference map image and query image, consecutive image frames and so forth. The major challenge lies in that stable visual features are difficult to be identified, and the establishment of feature correspondence may easily be sabotaged by input noise, mismatches and other error sources. Although a variety of outlier detection strategies have been developed in the literature, none of them has been able to provide a good solution for vision-based navigation system. Data snooping and M-estimator developed by the geodetic field are based on the assumption that only very few outliers exist, which is inadequate to address the visual related problems. RANSAC has been the most popular method for outlier detection in the field of computer vision. It performs well in face of high percentage of outliers. However, the major limitation lies in that it starts with a random subset of correspondences. If the initial selection is erroneous, it will lead to inaccurate or even false estimation of the homography and mismatches will be included as inliers.

Therefore, to address these problems and strengthen the reliability of the system, two major approaches have been taken in this research. First, a multi-level operation scheme of outlier detection has been proposed and implemented for the system. It includes both quality control measures for 3D mapping and vision-based positioning. The main contribution is the combined use of various outlier detection methods from different fields in a multi-level manner to achieve an improved solution. More specifically, RANSAC is used to remove most of the outliers, while data snooping is used at final adjustment process to guarantee the correctness of the input. Experiments have revealed the nature of the outliers in the system and proved the efficiency of the outlier detection scheme.

Secondly, a method to evaluate and improve the performance of RANSAC based outlier detection and homography estimation has been derived. It integrates intensity-based method into the feature matching process to strengthen the robustness of the matching algorithm against mismatches and noise. More specifically, cross-correlation information is used as an analysis and selection criterion for the matching. Instead of identifying mismatch(es) after it has been generated, the method determines how good the homography model (H) is for the two matching images and discard bad H to reduce chances that mismatches are included. The basic idea is to generate patches around each SIFT matched points (named as feature patch) and calculate the normalized correlation coefficient between each patch pair. Then the significance tests of correlation coefficients are used to qualify the values of the correlation coefficient. According to the correlation coefficient generated by each pair of the feature patch correspondence, the homography model built by the image matching process is evaluated: bad model is discarded, retaining only the good ones. Accompanied with RANSAC, experiments prove that this method can largely improve the correctness of image matching and can be applied to a great variety of applications where high quality feature-based matching is used, like object recognition and image stitching. The method has been applied to two main parts of the vision-based navigation system: identifying the reference images from the database, and final positioning. Experiments have demonstrated that such a method can effectively detect and discard mismatched reference images, and largely improve the positioning accuracy.

8.2 RECOMMENDATIONS FOR FUTURE RESEARCH

The following recommendations can be made for future studies:

1. Reality-based 3D map is a newly proposed concept in the literature. In the context of this research, it uses geo-referenced images to support vision-based navigation. It is still an immature technique, which can be improved and extended. In terms of

geo-referencing accuracy, currently it limits to centimetre level. The sparse feature points are geo-referenced instead of every pixel of the image. Dense matching can be introduced into the 3D mapping process to increase the accuracy of the map. Meanwhile, the visualization of the map is currently limited to single reference images, which can be stitched to provide a panoramic view of the navigation environment. Moreover, such mosaic can be used as street view to support geometric measurement. A greater area will be covered. The ultimate goal is to develop a geo-referenced 3D map, which is realistic, image-based, enabling geometry measurements and can support various geo-location services.

2. Based on the test results of this research and similar approaches in the literature, a vision-based positioning and navigation system can provide positioning information in areas where GPS signal is degraded, including both indoors and outdoor urban canyons. Still there are some limitations and gaps of this research that can be further investigated and addressed:

    1) For outdoor environments, the shortage of texture because of poor lighting conditions may pose tremendous challenges to the system. A possible way is to develop more robust feature descriptor for image matching.

    2) In the case of changed landscape, which is more likely for indoor environments, such as change of posters or movement of furniture, such an approach will suffer from incorrect results due to mismatches. Therefore, the 3D map need to be updated when changes have been made, or complementary sensors need to be integrated.

    3) For indoor positioning, another limitation is that a GPS signal is not available. The current approach is to use the previous GPS data to identify the building and load all the map images of the interior of the building for indoor navigation. For further investigation, this research can incorporate WiFi signal and used it

in the same way GPS has been used for outdoors: provide rough location to help narrow down the search space of map images and initial value for space resection. Besides, the use of WiFi can also reduce the chance of misidentified locations. As a matter of fact, vision-based positioning is best to be used as a refinement on rough positions achieved by other sensors, since the query image need to match with reference images from a large database. The greater uncertainty the rough estimation, the less efficiency and accurate the matching will get. Further research will focused on the integration of other sensors with vision to improve the current approach.

4) Current research has mostly been performed off-line using a matlab platform for post-processing. Further research will also take efficiency of the algorithms into account, and move to mobile platform, such as smartphones. Algorithms developed or adopted in this research for positioning will be modified accordingly. For instance more efficient image matching algorithms, such as SURF might be used.

REFERENCES

Aboelmagd, N., Tashfeen B. Karamat, and Georgy J.,2013. Fundamentals of Inertial Navigation, Satellite-Based Positioning and Their Integration. Springer.

Aguilera, D.G, Lahoz J. G., 2006. Laser scanning or Image-based Modeling? A Comparative through the Modelization of San Nicolas Church. International Archives of Photogrammetry and Remote Sensing, Volume XXXVI, B5, Dresden

Alhwarin, F., Wang, C., Ristic-Durrant, D., & Gräser, A., 2008. Improved SIFT-features matching for object recognition. In Visions of Computer Science-BCS International Academic Conference . pp. 179-190.

Antonis A.A., Dimitri P.T., and Cedric G., 2004. Biomimetic centering behavior for mobile robots with panoramic sensors. IEEE Robotics and Automation Magazine, vol. 11, no. 4, pp. 21–68.

Baarda, W. ,1968. A Testing Procedure for Use in Geodetic Networks, Publications on Geodesy. vol. 2, no. 5, pp. 1-97, Netherlands Geodetic Commission.

Barron, J.L., Fleet, D.J., Beauchemin, S.S.,1994. Performance of optical flow techniques. International Journal of Computer Vision 12(1), 43–77.

Bay H., Tuytelaars T., and Gool, L. V., 2006. SURF: speed-up robust features”, 9th European Conference on Computer Vision, pp. 404-417.

Bergen, J. R., Anandan, P., Hanna, K. J., & Hingorani, R., 1992.. Hierarchical model-based motion estimation. In Computer Vision—ECCV'92(pp. 237-252). Springer Berlin Heidelberg.

Berretti, S., Bimbo, A.D., Pala, P., Amor, B.B., Daoudi M., 2010.A Set of Selected SIFT Features for 3D Facial Expression Recognition., 2010 20th International Conference on Pattern Recognition (ICPR), pp.4125-4128, 23-26 Aug. 2010,doi: 10.1109/ICPR.2010.1002

Bonin-Font, F., Ortiz, A., & Oliver, G., 2008. Visual navigation for mobile robots: A survey. Journal of intelligent and robotic systems, 53(3), 263-296.

Borenstein J., Everett H.R., and Feng L.,1996. Naviagtion Mobile Robots: System and Techniques.Wellesley, Mass.: AK Peters.

Boris R., Effrosyni K., and Marcin D.,2008. Mobile museum guide based on fast SIFT recognition. The 6th International Workshop on Adaptive Multimedia Retrieval, pp. 26-27.

Brown,D.C. ,1976. The bundle adjustment – progress and prospects. International Archives of Photogrammetry, 21(3), ISP Congress, Helsinki, pp.1-33.

Brown, M. A., 2005. Multi-Image Matching using Invariant Features, Doctoral dissertation, The University of British Columbia, pp. 16-17.

Bujakiewicz A.,Podlasiak P.,Zawieska D.,2011. Georeferencing of Close Range Photogrammetric Data. Archives of Photogrammetry, Cartography and Remote Sensing, Vol. 22, pp. 91-104.

Chatila R. and Laumond J.-P., 1985. Position Referencing and Consistent World Modeling for Mobile Robots. Proc. IEEE Int'l Conf. Robitics and Automation, pp. 138-145, Mar. 1985.

Chen, R.; Wang, Y.;Pei, L.; Chen, Y, 2012. Virrantaus K.3D Personal Navigation in Smartphone Using Geocoded Images. GPS World. October 2012 issue, pp36-42.

Chowdhary G., Johnson E., Magree D., Wu D., Shein A., 2013. GPS-Denied Indoor and Outdoor Monocular Vision Aided Navigation and Control of Unmanned Aircraft. Journal of Field Robotics, accepted, Jan 2013.

Christensen H.I., Kirkeby N.O., Kristensen S., and Knudsen L., 1994. Model-driven Vision for Indoor Navigation. Robotics and Autonomous Systems, vol. 12 pp. 199-207.

Cobzas, D., Zhang, H., and Jagersand, M., 2003. Image-based localization with depth-enhanced image map. In International Conference on Robotics and Automation.

CoorK.H. K.H. Sharkawi, M.U. Ujang and A. Abdul-Rahman, 2008. 3D navigation system for virtual reality based on 3D game engine. The International Archives of Photogrammetry, Remote Sens Spat Inf Sci, 2008, Vol.XXXVII, Part B4, 513-518

Coors, V., C. Kray, K. Laakso, and C. Elting, 2004. Presenting route instructions on mobile devices. Book Chapter in ―Geo-Visualization".

Davison, A. J., 1998. Mobile Robot Navigation Using Active Vision. Thesis: University of Oxford.

Davison, A.J. and Murray D.W., 2002. Simultaneous localization and map-building using active vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7):865 – 880.

Davison, A.J., 2003. Real-time simultaneous localisation and mapping with a single camera. Proceedings of the 9th International Conference on Computer Vision, Nice, 2003.

Devore, J. 2012. Probability & Statistics for Engineering and the Sciences. CengageBrain. com.

Edgeworth F.Y., 1987. On observations relating to several quantities. Phil. Mag. 24 (1887), pp. 222–223 5th Series.

Elinas, P., Sim, R., Little, J.J., 2006. SLAM: Stereo Vision SLAM Using the Rao-Blackwellised Particle Filter and a Novel Mixture Proposal Distribution. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2006), Orlando, Florida, USA, May 15-19, 2006.

Faig, W., 1985. Lecture Notes on Aerial Triangulation and Digital Mapping, Monograph 10, School of Surveying, the University of New South Wales, Australia.

Fischler M.A. and Bolles R.C.,1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6):381C395, 1981.

Förstner, W., 1983. Reliability and discernability of extended Gauss-Markov models. Deutsche Geodätische Kommission (DGK) Report A, No. 98, 79–103.

Forstner, W., 1986. A Feature Based Correspondence Algorithm for Image Matching. International Archives of Photogrammetry, Vol. 26-III, Rovaniemi, Finland, 1986.

Fukatsu, S., Kitamura, Y., Masaki, T., Kishino F., 1998. Intuitive control of –bird's eye" overview images for navigation in an enormous virtual environment. In Proceedings of the ACM symposium on Virtual reality software and technology (pp. 67-76). ACM.

Gil, A., Reinoso, O., Ballesta, M., Pedrero, J.M.,2007. Emerging Technologies, Robotics and Control Systems - Volume 2, pp.108-13.

Goedeme, T. and Tuytelaars T.,2004. Fast wide baseline matching for visual navigation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC. 27th June - 2nd July, 2004; pp. 24 – 29.

Griffin, D., 2011. How does the global positioning system work. In GPS World Public Forum (Vol. 26).

Groves, P.D.,2007. Principles of GNSS, inertial, and multi-sensor integrated navigation systems. Artech House, 2007.

Groves, P. ,2013. Future Trends in Integrated Navigation. Inside GNSS, April 2013

Gruen, A.W., 1985. Adaptive Least Squares Correlation: A powerful Image Matching Technique. South Africa Journal of Photogrammetry Remote Sensing Cartography, pp. 175-187.

Gruen, A. 2012. Development and status of image matching in photogrammetry. The Photogrammetric Record, 27(137), 36-57.

Guilherme N. D. and Avinash C.K.,2002. Vision for Mobile Robot Navigation: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 24, No.2, Feb 2002.

Hampel FR, Ronchetti EZ, Rousseeuw PJ, Stahel WA, 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

Hartley R. and Zisserman A.,2003. Multiple View Geomerty in Computer Vision. Cambridge University Press, second edition.

Harris C.G. and Pike J.M.,1987. 3D positional integration from image sequences. Image and Vision Computing, 6(2):87–90, 1987.

Harris, C.; Stephens, M., 1988. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151.

Hofman-Wellenhof B., Legat K., and Wieser M.,2003. Navigation: Principles of Positioning and Guidance. Austria: Springer-Verlag Wien, New York, USA, 2003.

Hrabar, S., Sukhatme, G. S., Corke, P., Usher, K., & Roberts, J. 2005. Combined optic-flow and stereo-based navigation of urban canyons for a UAV. In 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 3309-3316.

Huang, H., Maire, F., & Keeratipranon, N., 2005. A direct localization method using only the bearings extracted from two panoramic views along a linear trajectory. In Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005) (pp. 201-206). Springer Berlin Heidelberg.

Huang, T. and Netravali, A.,1994. Motion and structure from feature correspondences: A review. Proceedings of IEEE, 82(2): 252–268.

Huber, P. J.,1964. Robust Estimation of a Location Parameter.” Annals of Mathematical Statistics 35:73—101.

Indyk, P. and Motwani, R., 1998. Approximate nearest neighbors:towards removing the curse of dimensionality. In: STOC '98:Proceedings of the thirtieth annual ACM symposium on Theory of computing, ACM, New York, NY, USA, pp. 604–613.

Jaspers H., Schauerte, B., Fink. G.A., 2012. SIFT-based Camera Localization using reference objects for application in multi-camera environments and robotics. Proceedings of the International Conference on Pattern Recognition Applications and Methods,Portugal, 2012.

Jazayeri, I., 2010. Image-based modelling for object reconstruction. PhD thesis, Engineering - Geomatics, The University of Melbourne.

Jung, I. K., Lacroix, S., 2003. High resolution terrain mapping using low attitude aerial stereo imagery. Proceedings of the Ninth IEEE International Conference on Computer Vision , pp. 946-951.

Karimi H., 2011. Universal Navigation on Smartphones, Spring New York, 2011.

Kidono, K., Miura, J., Shirai, Y.,2002. Autonomous visual navigation of a mobile robot using a human-guided experience. Robot. Auton. Syst. 40(2-3), 121-130.

Kitanov, A.; Bisevac, S.; Petrovic, I., 2007. Mobile robot self-localization in complex indoor environments using monocular vision and 3D model, Advanced intelligent mechatronics, 2007 IEEE/ASME international conference on , vol., no., pp.1-6, 4-7.

Knight, N. and Wang J., 2009. A Comparison of Outlier Detection Procedures and Robust Estimation Methods in GPS Positioning". The Journal of Navigation 62(04): 699-709..

Kosaka A. and Kak A.C.,1992. Fast Vision-Guided Mobile Robot Navigation Using Model-Based Reasoning and Prediction of Uncertainties. Computer Vision, Graphics, and Image Understanding, 56(3):271–329.

Krarup T., Juhl J. and Kubik K., 1980. Götterdämmerung over least squares adjustment. 14th Congress of the Int. Soc. Of Photogrammetry, Hamburg, Vol B3: 369-378.

Kuhl, A.,2004. Comparison of Stereo Matching Algorithms for Mobile Robots. M.Sc. Thesis, Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, Ilmenau, Germany, 2004.

Lang F. and Forstner,W.,1995. Matching techniques. In Second Course in Digital Photogrammetry. Landesvermessungsamt NRW.

Larson, Craig D., 2010. An Integrity Framework for Image-Based Navigation Systems. Doctoral Dissertation. Air Force Inst of Tech Wright-Patterson AFB Oh School of Engineering and Management. Jun. 2010.

Lisowska, P.Use of Digital Photogrammetry for Architectural Inventory. Eng. Diploma thesis, Warsaw University of Technology

Li X., Wang J., Li R., Ding W., 2011. Image-based positioning with the use of geo-referenced SIFT features. Proceedings of the Incorporating the International Symposium on GPS/GNSS (IGNSS 2011), Sydney, Australia, November 2011.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. 1999. The proceedings of the seventh IEEE international conference on computer vision. Vol. 2, pp. 1150-1157.

Lowe, D., 2004. Distinctive Image Features from Scale Invariant Key points. International Journal of Computer Vision. pp. 91-110.

Lucas B. and Kanade T.,1981. An iterative image registration technique with an application to stereo vision. In Proc. Seventh International Joint Conference on Artificial Intelligence, pages 674-679, Vancouver,Canada, Aug. 1981.

Luhmann, T., Robson,S., Kyle, S., Harle, I.,2006. Close Range Photogrammetry: Principles, Methods and Application; Whittles Publishing, Scotland, UK, 2006; pp. 206

Luhmann, T., 2009. Precision potential of photogrammetric 6DOF pose estimation with a single camera. ISPRS Journal of Photogrammetry and Remote Sensing, 2009, Volume 64, Issue 3, Pages 275-284.

Manessis, A., Hilton, A.,Palmer, P., McLauchlan, P., Shen, X. , 2000.Reconstruction of scene models from sparse 3D structure. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2666-2673.

Matas, J., Chum, O., Urban, M., Pajdla, T.,2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22 (2004) 761-767

Meijers, M., S. Zlatanova and N. Pfeifer, 2005. 3D Geoinformation Indoors: Structuring for evacuation.Proceedings of Next Generation 3D City Models, Bonn,Germany, pp. 11-16.

Mikolajczyk, K., Schmid, C.,2004. Scale & a_ne invariant interest point detectors. International Journal of Computer Vision 60 (2004) 63-86

Mikolajczyk K. and Schmid C., 2005. A Performance Evaluation of Local Descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 10, October 2005.

Moravec H. P.,1977. Towards Automatic Visual Obstacle Avoidance, Proc. 5th International Joint Conference on Artificial Intelligence, pp. 584.

Morel J. M. and Yu G., 2009. ASIFT: A new framework for fully affine invariant image comparison," SIAM J. Imag. Sci., vol. 2, no. 2, pp. 438–469,Apr. 2009.

Nalpantidis L., Chrysostomou D. and Gasteratos A., 2009. Obtaining Reliable Depth Maps for Robotic Applications with a Quad-camera System. ICRA09 Workshop on Safe navigation in open and dynamic environments Application to autonomous vehicles.

Negahdaripour, S., Prados, R., & Garcia, R., 2005. Planar homography: accuracy analysis and applications. In IEEE International Conference on Image Processing, 2005. ICIP 2005. (Vol. 1, pp. I-1089).

Nalpantidis L., Chrysostomou D. and Gasteratos A., 2009. Obtaining Reliable Depth Maps for Robotic Applications with a Quad-camera System. ICRA09 Workshop on Safe navigation in open and dynamic environments Application to autonomous vehicles, 2009

Niste´r D., Naroditsky O., and Bergen J , 2004. Visual Odometry, Proc. IEEE Conf. Computer Vision and Pattern Recognition.

Ohno, T., Ohya, A., & Yuta, S., 1996. Autonomous navigation for mobile robots referring pre-recorded image sequence. In Intelligent Robots and Systems' 96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on (Vol. 2, pp. 672-679). IEEE.

Pei L., Chen R., Liu J., Liu Z., 2011. Kuusniemi H.; Chen Y.; Zhu L. Sensor Assisted 3D Personal navigation on a Smart Phone in GPS Degraded Environments. Proceedings of the 19th International Conference on Geoinformatics, Shanghai, China, June 24-26, 2011.

Pope, A.J., 1976. The statistics of residuals and the detecion of outliers. Technical" Rep. TR-NOS-65-NGS-1, National Ocean Survey,Rockville, Md. 1976.

Przemyslaw, B.; Pawel, S.,2012. Enhancing positioning accuracy in urban terrain by fusing data from a gps receiver, inertial sensors, stereo-camera and digital maps for pedestrian navigation. Sensors, 2012, 12, 6764–6801.

Rakkolainen, I., Timmerheid, J., and Vainio,T., 2000. A 3D city info for mobile users",Proceedings of the 3rd International Workshop in Intelligent Interactive Assistanceand Mobile Multimedia Computing (IMC'2000), Rockstock, Germany, pp. 115-212.

Remondino, F. and Ressl, C., 2006. Overview and experiencesin automated markerless image orientation. IAPRS, Vol. 36, Part 3, pp. 248-254.

Remondino, F.,2006. Image-based modeling for object and human reconstruction (Doctoral dissertation, Politecnico di Milano).

Rivlin, E., Shimshoni, I., Smolyar, E., 2003. Image-based robot navigation in unknown indoor environments. Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems ,Vol. 3, pp. 2736-2742.

Robert, G. Casella, 2002. Monte Carlo Statistical Methods, Springer, New York.

Robertson, D. and Cipolla, R, 2004. An image-based system for urban navigation. Proceedings of the British Machine Vision Conference (BMVC), Kingston University, London,7-9th Sept 2004.

Royer, E.,Bom,J.,Dhome, M., Thuillot, B.,Lhuillier, M., Marmoiton, F.,2005. Outdoor autonomous navigation using monocular vision. Proceedings of IEEE International Conference on Intelligent Robots and Systems, pp.3395-3400.

Royer, E., Lhuillier, M., Dhome, M., & Lavest, J. M.,2007. Monocular vision for mobile robot localization and autonomous navigation. International Journal of Computer Vision, 74(3), 237-260.

Ruiz-Ruiz A.J., Lopez-de-Teruel P.E. and Canovas O.,2012. A multisensory LBS using SIFT-based 3D models. Proceedings of the International Conference on Indoor Positioning and Indoor Navigation, Sydney, Australia, 13-15th November, 2012.

Sabe K., Fukuchi M., Gutmann J.S., Ohashi T., Kawamoto K., and Yoshigahara T..2004. Obstacle Avoidance and Path Planning for Humanoid Robots using Stereo Vision. Proceedings of the International Conference on Robotics and Automation, New Orleans, April, 2004

Santos-Victor J., Sandimi G., Curotto F., and Garibaldi S. 1993. Divergent Stereo for Robot Navigation: learning from Bees. Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp 434-439.

Schmid C. and Mohr R.,1997. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):530–535, May 1997.

Schaffalitzky, F. and Zisserman, A., 2002. Multi-view matching for unordered image sets. Proceedings of the Seventh European Conference on Computer Vision (ECCV‘02). Copenhagen, Denmark; 27 May -- 2 June 2002; pp. 414–431.

Se S., Lowe D., and Little J.,2001. Vision-based mobile robot localization and mapping using scale-invariant features. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA),pages 2051–2058, Seoul, Korea, May 2001

Se, S., Lowe, D., and Little, J. 2002. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. International Journal of Robotic Research, 21(8):735–760.

Sharkawi, K., Ujang, M.U.,Abdul-Rahman, A, 2008. 3D Navigation System for Virtual Reality based on 3D Game Engine. ISPRS Congress Beijing 2008, 513-518.

Skog, I., Handel, P., 2009. In-car-positioning and navigation technologies: a survey. IEEE Trans Intell Transp Syst 10(1): 4-21.

Snavely, N., Seitz, S. M. and Szeliski, R.,2008. Modeling the world from internet photo collections. International Journal of Computer Vision, 2008, 80(2): 189–210.

Teunissen, P.J.G., 1990. Quality control in integrated navigation systems. Aerospace and Electronic Systems Magazine, IEEE , vol.5, no.7, pp.35-41.

Tewinkel,G.C., 1958. Panel-Future of Analytical Aerial Triangulation. Photogrammetric Engineering, March 1958

Trucco, E.; Verri, A., 1998, Introductory Techniques for 3-DComputer Vision, Prentice Hall.

Tsubouchi T. and Yuta S. , 1987. Map-assisted Vision System of Mobile Robots for Reckoning in a Building Environment. Proc. IEEE Int‘l Conf. Robotics and Automation, pp. 1978-1984.

Turk M. and Pentland A.,1991. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1): 71-86.

Van Driel, J. N. ,1989. Three dimensional display of geologic data. Three dimensional applications in Geological Information System. Taylor & Francis: London, 1-9.

Wang C., Su Y., Hong C., 2008. A 3D virtual navigation system integrating user positioning and pre-download mechanism. Proceedings of the World Academy of Science Engineering and Technoloy. v. 30: 172-176.

Wang J, and Chen Y. 1994. On the reliability measure of observations. Acta Geodaet et Cartograph Sin. pp 42–51 (English edition).

Wang, J. and Chen, Y. 1999. Outlier detection and reliability measures for singular adjustment models. Geomatics Research Australasia, 71,57-72.

Wu, C.,2012. VisualSFM: A Visual Structure from Motion System. Available online: http://www.cs.washington.edu/homes/ccwu/vsfm/ (Accessed on 15 December 2012).

Wu, F., & Fang, X. ,2007. An improved RANSAC homography algorithm for feature based image mosaic. In Proceedings of the 7th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision. World Scientific and Engineering Academy and Society (WSEAS) (pp. 202-207).

Yang, Y., Song, L., and Xu, T., 2002. Robust estimator for correlated observations based on bifactor equivalent weights. J. Geodesy, Berlin, 76(6–7), 353–358. 2002.

Yuan, L., and H. Zizhang.,2008. 3D indoor navigation: A framework of combining BIM with 3D GIS." 44th ISOCARP Congress.

Zabih R. and Woodfill J.,1994. Non-parametric local transforms for computing visual correspondence. Third European Conference on Computer Vision, Stockholm, Sweden, May 1994.

Zhang G., Dong Z., Jia, J., Wong T. T., and Bao H., 2010. Efficient non-consecutive feature tracking for structure-from-motion. Proceedings of the European Conference on Computer Vision (ECCV '10).

Zhang W. and Kosecka J., 2006. Image based localization in urban environments. Third International Symposium on 3D Data Processing, Visualization and Transmission, University of North Carolina, Chapel Hill, USA, June 14-16, 2006.

Zhang Y., Cao J., Lou L., and Su B., 2012. Appearance-based mobile robot navigation using omnidirectional camera. Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery, pp.2357-2361, May 2012

PUBLICATIONS DURING PHD STUDIES

Referred journal papers:

1) **Li X.**, Wang, J. 2013. Image Matching Techniques for Vision-based Indoor Navigation Systems: A 3D Map Based Approach. Accepted for publication by *Journal of Location Based Services* (on 8th of August, 2013).

2) **Li, X**., Wang, J., & Li, T. 2013. Seamless Positioning and Navigation by Using Geo-Referenced Images and Multi-Sensor Data. *Sensors*, 13(7), 9047-9069.

3) **Li X**., Wang J., Knight N., & W. Ding 2011. Vision-based Positioning with a Single Camera and 3D Maps: Accuracy and Reliability Analysis. *Journal of Global Positioning Systems*. 10(1): 19-29.

4) Yuan Z., **Li X**., Wang J., Yuan Q., Xu D., Diao J., 2011. Methods of 3D map storage based on geo-referenced image database. *Transactions of Nonferrous Metals Society of China*, 21,s654-s658

Referred conference papers:

1) **Li X.**, Wang J., 2012. Image Matching Techniques for Vision-based Indoor Navigation Systems: Performance Analysis for 3D Map Based Approach. Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN 2012), Sydney, Australia, 13-15th November.

2) **Li X**., Wang J., 2012. Multi-image Matching for 3D Mapping in Vision-based Navigation Applications. Proceedings of the Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), Finland, Oct. 2012

3) **Li X**., Wang J., 2012. Evaluating photogrammetric approach of image-based positioning. Proceedings of the XXII Congress of the International Society for Photogrammetry & Remote Sensing (ISPRS2012), Melbourne, August 2012

4) **Li X**., Wang J., Li R., Ding W., 2011. Image-based positioning with the use of geo-referenced SIFT features. Proceedings of the International Symposium on GPS/GNSS (IGNSS 2011), Sydney, Australia, November 2011.

Abstract referred conference papers:

1) **Li X**., Wang J., Liu W. & R Li., 2013. Geo-referenced 3D Maps: Concept and Experiments. The 8th International Symposium on Mobile Mapping Technology, Tainan, Taiwan, 1-3 May.

2) **Li X**., Wang J., Yang L., 2011. Outlier Detection for Indoor Vision-Based Navigation Applications. Proceedings of the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2011), Portland, OR, September 2011.

3) **Li X**., Wang J., Knight N., Olesk A., Ding W., 2010. Indoor positioning within a single camera and 3D maps. IEEE Proceedings of Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010,Oct.

Conference Presentations:

1) Image Matching Techniques for Vision-based Indoor Navigation Systems: Performance Analysis for 3D Map Based Approach. The 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN 2012), Sydney, Australia, 13-15th November.

2) Multi-image Matching for 3D Mapping in Vision-based Navigation Applications. The Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), Finland, Oct. 2012

3) Evaluating photogrammetric approach of image-based positioning. The XXII Congress of the International Society for Photogrammetry & Remote Sensing (ISPRS2012), Melbourne, Australia, August 2012

4) Image-based positioning with the use of geo-referenced SIFT features. The International Symposium on GPS/GNSS (IGNSS 2011), Sydney, Australia, November 2011.