UNIVERSITY OF CANTERBURY

MHIT690

MASTERS OF HUMAN INTERFACE TECHNOLOGY

Accurate Wide-Area Tracking for Architectural, Engineering and Surveying Applications

Author:

James HEAD-MEARS

Supervisor:

Dr. Mark BILLINGHURST

Co-Supervisors:

Dr. Adrian CLARK
Dr. Winyu CHINTHAMMIT

October 22, 2013

I, James Head-Mears, declare that this thesis titled, *Accurate Wide-Area Tracking for Architectural, Engineering and Surveying Applications* and the work presented in it are my own. I confirm that:

This work was done wholly or mainly while in candidature for a research degree at this University.

Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

Where I have consulted the published work of others, this is always clearly attributed.

Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

I have acknowledged all main sources of help.

Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date: October 22, 2013

Dedicated to my love Caitlin Rice.

Contents

List of Figures

List of Tables

7

Acknowledgements

I would like to thank my Supervisors Dr Mark Billinghurst, Dr Adrian Clark and Dr Winyu Chinthammit for their continued guidance and support throughout my research. I would like to especially thank Adrian Clark as I am without a doubt that I would not have been able to complete my reseach without his assistance and I will forever be in his debt.

I would also like to thank the CRC for Spatial Information whos contributions enabled me to both complete my research but also to discover new friends and experiances. Not only does the CRC for Spatial Information provides resources for continued research in the Spatial Sector but its members also provide additional guidance and support and I would also like to thank their contribution to my research and the research of others.

Lastly, I would like to thank my friends, family and dearest fiance whos continued support kept me going through it all. It is your support that I hold closest.

Abstract

Augmented Reality (AR) is a powerful tool for the visualisation of, and interaction with, digital information, and has been successfully deployed in a number of consumer applications. Despite this, AR has had limited success in industrial applications as the combined precision, accuracy, scalability and robustness of the systems are not up to industry standards. With these characteristics in mind, we present a concept Industrial AR (IAR) framework for use in outdoor environments.

Within this concept IAR framework, we focus on the improving the precision and accuracy of consumer level devices by focusing on the issue of localisation, utilising LiDAR based point clouds generated as part of normal surveying and engineering workflow.

We evaluate key design points to optimise the localisation solution, including the impact of increased field of view on feature matching performance, the filtering of feature matches between real imagery and an observed point cloud, and how pose can be estimated from 2D to 3D point correspondences.

The overall accuracy of this localisation algorithm with respect to ground-truth observations is determined, with unfiltered results indicating an on par horizontal accuracy and significantly improved vertical accuracy with best-case consumer GNSS solutions. When additional filtering is applied, results of localisation show a higher accuracy than best-case consumer GNSS.

Chapter 1

Introduction

Augmented Reality (AR) is a technology that integrates virtual information with the real world, allowing users to experience and interact with virtual content as they would with real objects (Azuma, 1997). Although the concept of AR was first developed over 40 years ago, it has recently undergone a surge in popularity due to the sudden ubiquity of mobile devices and web technologies capable of running AR applications (Fite-Georgel, 2011).

Intrinsic to the design and use of AR systems is the underlying technologies used to align the virtual content with the real world, and various technologies have been developed to determine the pose of the AR device. The performance of these device localisation technologies often defines the suitability of the system for a particular task. Many existing AR platforms are designed for indoor environments and use computer vision based localisation technologies for determining pose (e.g. Qualcomm Vuforia[1] and Metaio[2]). Recently, a number of AR systems have been developed for use in outdoor environments (e.g. ARKick[3], Layar[4], Wikitude[5], iRiS Augmented Reality[6] and Spyglass[7]), which work by utilising non-visual position and orientation tracking devices present in current generation mobile devices, such as GNSS modules coupled with magnetometers (compass), gyroscopes, and accelerometers, to provide a complete pose solution suitable for low-accuracy applications.

While low-accuracy pose computation is suitable for consumer grade ap-

---

[1]https://www.vuforia.com

[2]http://www.metaio.com

[3]http://refocuslabs.com/app/arkick/

[4]http://www.layar.com

[5]http://www.wikitude.com

[6]http://www.ad-dispatch.com

[7]http://happymagenta.com/spyglass/

plications, industrial applications require higher levels of both accuracy and precision (Fite-Georgel, 2011). While these higher accuracy requirements are difficult to generalise due to the broad scope of tasks undertaken in industry, the accuracy requirements typically range from centimetre to millimetre tolerances. To achieve the accuracy required for outdoor industrial tasks, existing systems often utilise custom hardware such as industrial grade GNSS and inertial sensors (Besbes, Collette, Tamaazousti, Bourgeois, & Gay-Bellile, 2012; Schall et al., 2009; Ong, Yuan, & Nee, 2008), which makes deployment to larger client-bases difficult (Krevelen & Poelman, 2010).

Recent research has attempted to improve the accuracy and precision of pose estimation on consumer-grade devices without the need for additional hardware. The improvement of accuracy and precision has the potential to not only improve user experience, but also allow AR to be utilised in a wider range of industrial tasks, across fields such as engineering, architecture, construction, surveying and GIS.

The process of pose estimation specific to AR, described in this thesis as *AR tracking*, can be separated into two main components, *localisation* and *tracking*. Localisation is the process of determining the current pose within a defined co-ordinate system, while tracking determines the relative change in pose between epochs[8]. The focus of this thesis is primarily on the problem of localisation, and in particular investigate how high-density RGB point-clouds commonly utilised in surveying and engineering can be used to improve the accuracy of localisation for industrial applications.

1.1   Research Motivation

It has been shown that AR can be effectively used for the visualisation of spatial information for industrial tasks (Feiner, MacIntyre, Hollerer, & Webster, 1997; Kato, Tachibana, Tanabe, Nakajima, & Fukuda, 2003; Anagnostou & Vlamos, 2011; A. H. Behzadan & Kamat, 2008), however due to the high accuracy and precision requirements, these systems are only possible with the aid of industrial-grade tracking hardware (Rossler, Rogge, & Hentschel, 2011; Lonsing, 2011; Ababsa, Zendjebil, Didier, Pouderoux, & Vairon, 2012). The use of custom hardware makes it difficult to achieve widespread deployment of these AR systems due to complexity in the assembly and high cost of components. In comparison, mobile devices such as smart phones and tablets are already in wide-spread use for consumer-grade AR tasks, such as adver-

---

[8]In this case the epoch is considered a given sampling time. For example, the epoch can be measured as the refresh time of the users display or as the time of the devices fastest sampling method.

tising and entertainment, however the accuracy offered by the localisation techniques used for these tasks is far lower than that required for industrial tasks.

There has been increased interest in bringing high accuracy localisation to cost-effective consumer-grade devices to overcome issues of scalability for outdoor industrial tasks in AR. Recent research has shown that this is possible to achieve (Gauglitz, Sweeney, Ventura, Turk, & Hollerer, 2012), however localisation using the low accuracy sensors available in these cost-effective devices requires a pre-calibrated environment in order to achieve an accuracy suitable for the tolerances required by industrial tasks. Therefore, it is the aim of this research to develop a localisation technique suitable for AR visualisation of spatial information in outdoor industrial tasks. In particular, the use of colour points clouds, an overlooked source of information that is readily available in many industrial tasks in the development, surveying, construction and engineering fields, is explored as a tool for highly accurate outdoor localisation.

The goal of this research is to provide outdoor localisation techniques which can be used in an AR framework to aid in the visualisation of spatial information for industrial tasks in fields such as urban planning, development, surveying, construction and engineering. The connection between virtual information and physical assets is a key component to many tasks in these areas, and often the user's ability to interpret the relationship between the virtual and physical states is paramount to the success of these tasks.

1.2   Thesis Outline and Research Focus

The focus of this thesis is on the problem of accurate outdoor localisation for AR. Existing research in AR, outdoor localization and industrial processes is discussed in Chapter 2. To give context on how localization fits into a larger AR framework for outdoor industrial applications, a concept Outdoor Industrial AR framework in presented Chapter 3. Following this, the localization algorithm based on 3D point cloud is presented and discussed in Chapter 4. The accuracy of the localization algorithm is evaluated in Chapter 5, and the results are discussed in Chapter 6. Finally, the thesis is concluded and future research directions are proposed in Chapter 7.

This thesis covers the following key research areas:

- Development of a concept Outdoor Industrial AR (IAR) framework.

- Development of a localisation algorithm for Outdoor Industrial AR using 3D point clouds.

- Evaluation of the accuracy of the various processes required for 3D point cloud localisation.

- Evaluation of the localisation algorithm for industrial tasks.

1.3 Summary

This chapter introduced AR as a method for visualising virtual information in the real world, and how that relates to outdoor industrial tasks. It also described localisation as a key enabling technology for AR, and how localisation for outdoor industrial AR is the focus of this research.

Chapter 2

Background

This chapter provides additional information on the areas of research addressed by this thesis and previous relevant related work. Initially the concept of AR is described by providing a brief history on the roots of research in this area. It also defines AR as adopted by this thesis.

Industrial tasks and their requirements are then described with respect to possible applications of AR in outdoor industrial tasks in the engineering, surveying, planning, development and GIS fields. The limitations of previous industrial AR applications are then discussed, specifically focussing on the impacts of hardware and how this affects the adoption rates and continued use of industrial AR systems.

Smartphones and Tablets are then introduced as an alternative for high-accuracy industrial AR tasks and how the primary obstacle for these devices is pose solution accuracy. The task of detecting pose solutions is then described in terms of tracking and localisation that respectfully address relative and absolute pose solutions.

2.1   Augmented Reality

The concept of AR can be traced to the 1960's, with the earliest published work being attributed to Ivan Sutherland (Sutherland, 1968). However it was not until the early 90's that the term Augmented Reality was coined by Caudell and Mizell (1992). Later that decade sufficient research had been completed to warrant AR being recognised as a distinct field of study as well as the first conferences dedicated to AR (Krevelen & Poelman, 2010). In 2001 the International Symposium on Mixed and Augmented Reality (ISMAR) started, which is one of the major AR symposiums for research and industry to come together (Krevelen & Poelman, 2010).
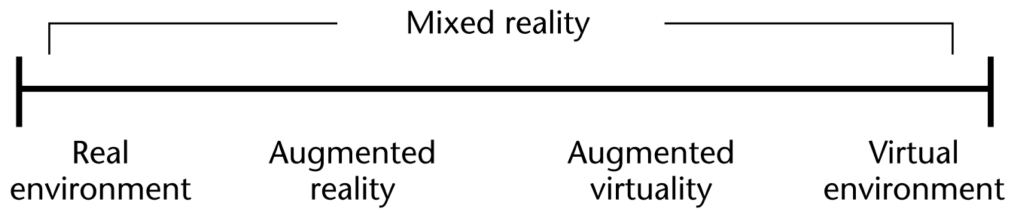
Figure 2.1: Milgram and Kishino's (1994) reality-virtuality continuum, where the position of AR is described in relation to a completely real and virtual environment's.

AR interfaces aim to combine or overlay virtual information onto the real world, thereby enhancing reality (Azuma, 1997; Carmigniani & Furht, 2011). Milgram and Kishino's (Milgram & Kishino, 1994) Reality-Virtuality diagram (See Figure 2.1) demonstrates the relationship that AR has to both Virtual Worlds and Real Worlds. At one extreme, Virtual Environments, commonly known as VR, aim to completely immerse a user in a virtual world and replace reality. Although, the Milgram and Kishino's continuum helps to conceptualise how AR relates to the real world, the definition of AR as presented by Azuma (Azuma et al., 2001) sets the restrictions on which attributes are required for an AR interface. This definition states an AR interface must have the following attributes:

- it combines real and virtual objects in a real environment,

- it is ineractive in real-time; and

- it registers (aligns) real and virtual objects with each other.

This definition does not restrict AR to purely visual systems but also allows the incorporation of additional senses or for senses to be replaced, as with sensory substitution. Visual-AR is currently the most common form of AR, however, there have been several studies of audio-based AR (Bleeker, Lee, & Billinghurst, 2013) as well as haptic AR (Bau & Poupyrev, 2012) in recent years. Visual-AR systems usually, consist of the following:

- display(s) - allows user(s) to view and interact with AR content,

- input device(s) - there are a large number of possibilities interfaces usable as input devices with AR systems,

- and tracking system(s) - positions the AR system relative to its environment and virtual information.

16

These systems are then in turn powered by a computer(s), which manage all of the system overheads (Carmigniani & Furht, 2011). Thus AR systems consist of; user input into the system, tracking calculations by the system, and output from the system back to the user. Each of these functions is equally important for an AR system to work. Nonetheless, it is the tracking stage that is the focus of this research paper.

A tracking process is required by AR systems as it allows the system to display virtual information correctly overlaid on the real world. This is done by providing the systems movement in six degrees of freedom (6DOF), whereby the location is provided as easting, northing, and elevation (x, y, z) variables and the orientation is provided as roll, pitch and yaw values (ω, ϕ, κ).

Research has been working towards a 'Holy Grail' solution for tracking that will provide authentic AR experiences, where the visual cohesion between real and virtual worlds is maintained precisely and realistically in all environments. However, this solution is still not within reach and remains a focus point of continued research. There have been many different approaches suggested that aim to solve tracking as a complete problem (Gauglitz et al., 2012; Wagner, Reitmayr, Mulloni, Drummond, & Schmalstieg, 2010; Pustka et al., 2012). Conversely, we approach tracking as a multifaceted task, consisting of both localisation and tracking.

This allows AR systems to be designed in a modular manner where different components may be interchanged as new techniques are discovered that out perform their counterparts, giving the best AR system for the specified task. This approach also allows easier adaption to different tracking environments as localisation and tracking approaches that may be optimised for specific environments.

The tracking component consists of determining the change in device's pose relative to a previous epoch or its initial start position. Thus, we define *tracking* as the devices ability to determine its position relative to an earlier time. Conversely, *localisation* is the process of positioning the device relative to the environment or coordinate system[9]. There are many different ways that the localisation of a system can be performed.

---

[9]These coordinate systems can be designed on either a local or global scale. For example, WGS84 is a geocentric coordinate system that is utilised by GNSS devices before being converted to local coordinates (Geoscience Australia, 2013) (See Wikipedia Entry: World Geodetic System)

Table 2.1: This table list common tasks, equipment and tolerances of the survey industry. The tolerances listed are only intended as a rough guide for possible AR systems.

Task	Equipment	Tolerance
Bulk earthwork	RTK*	± 20 mm
Mining - Pickup	RTK	± 30 mm
Mining - Model Generation	LiDAR	± 20 - 50 mm
Mining - Stockpile Volumes	LiDAR, RTK	± 1-10%
Level Survey	Automatic Level	± 3-10 mm
Road Survey - As Cons.	Total Station	± 5 mm
Car Components	Photogrammetry, LiDAR	± 10 μ m
Roller Alignment	Laser Alignment	± 50 μ m

* Real Time Kinematic (RTK)

2.2 Tracking for Industrial Tasks

Broadly defining the tracking and localisation requirements of industrial tasks is difficult as each task often differs greatly. Nonetheless, member of an industrial survey[10] company were interviewed to find the accuracy requirements of some of the data capture and set out tasks that they perform. *Fitness for purpose* is a common phrase used within the surveying industry; it is used as an explanation into providing accuracy and precision tolerances without knowing the specifics of the task at hand. Table 2.1 shows the accuracy tolerance of common survey tasks.

The values listed in Table 2.1 are not indicative of all survey tasks but instead can be used as an example of the accuracy requirements that AR systems should aim for. Although, some tracking technologies listed, like Optical an Inertial approaches, in Table 2.3 come within the precision tolerances listed in Table 2.1, precision is not symbolic of a technologies accuracy as localisation plays a large part in this.

2.3 Industrial Augmented Reality

Possible applications that can incorporate AR to some degree are essentially limitless at both consumer and industrial levels with multiple studies on possible applications being conducted in recent time (Carmigniani & Furht,

---

[10]Survey companies work with the capture and generation of spatial information.
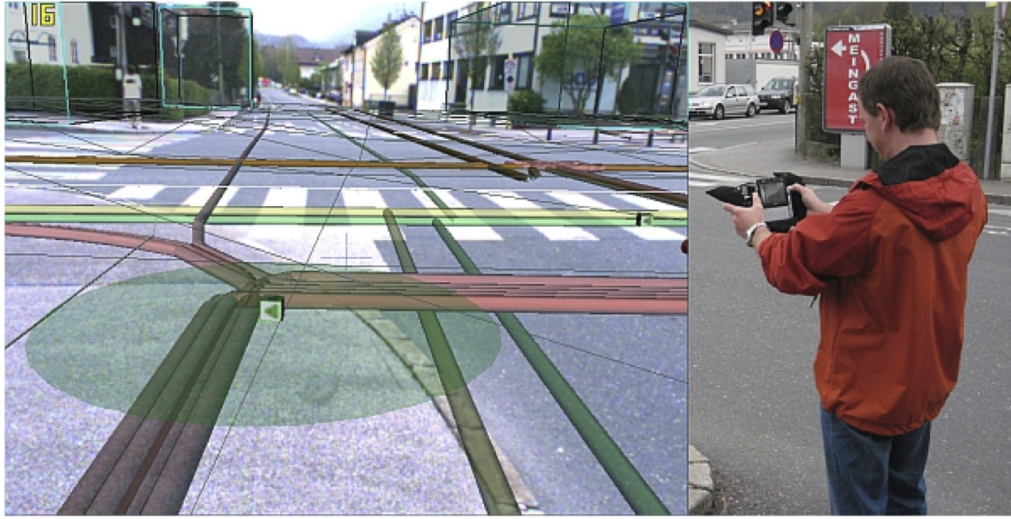
Figure 2.2: Demonstration of possible planning, surveying and engineering type information used in urban planning in an outdoor AR system (Junghanns, Schall, Schmalstieg, 2008).

2011; Krevelen & Poelman, 2010; Ong et al., 2008). Carmigniani and Furht (2011) have identified key industries that could benefit from integration of AR applications, including advertising, education, entertainment, medicine and engineering. Furthermore, industry has been a primary driving force for AR based research (Fite-Georgel, 2011). Industrial AR (IAR) Systems are AR systems designed specifically to aid an industrial processes.

Applications exist for IAR in an outdoors environment in the fields of engineering, surveying, planning, development and GIS, among others. These fields are primarily concerned with visualising spatial information in construction verification and architectural tasks as outlined by Feiner et al. (1997). The first use of outdoor AR for architectural tasks was shown in 1997 (Feiner et al., 1997). This was followed by AR for urban design and planning applications proposed in several papers for both indoor (Kato et al., 2003; Anagnostou & Vlamos, 2011) and outdoor environments (A. H. Behzadan & Kamat, 2008; Rossler et al., 2011; Lonsing, 2011; Ababsa et al., 2012). The results of this research highlight the plausibility of developing outdoor IAR applications for these fields (Rossler et al., 2011).

By using AR on mobile devices it is possible to explore life-sized virtual models of buildings that have not been built yet overlaid on their proposed sites in the real world. The CityViewAR application uses a mobile phone or tablet to allow users to see Christchurch city as it was before beievastated by the 2010 and 2011 earthquakes (Lee, Kim, & Myung, 2012). In this case

Figure 2.3: CityViewAR view of virtual building after its demolition in Christchurch, New Zealand (HIT Lab NZ, 2013).

the phone's GPS[11] and compass sensors can be used to determine a person's position and orientation and display virtual buildings at the correct position in the real world (see Figure 2.3). Applications such as CityViewAR show the potential of mobile AR for supporting urban design and architecture. However, there are a number of technical problems that need to be overcome before AR technology can be useful for the construction process.

In recent years there have been several surveys of the state of IAR (Fite-Georgel, 2011; Ong et al., 2008; Krevelen & Poelman, 2010; Rankohi & Waugh, 2013). Ong et al. (2008) provides a review of AR research and development in manufacturing activities. It also summarises the hardware and software systems designed for IAR applications. Krevelen and Poelman (2010) surveyed the state of technologies, applications and limitations related to augmented reality. They also briefly reviewed frameworks and content authoring tools for AR. Recently, Fite-Georgel (2011) utilised the product life cycle to taxonomically categorise a rubric for the evaluation of existing and proposed IAR solutions. Their research highlights the importance of involving industry in the design process for IAR systems. Fite-Georgel (2011) also believes that, at the time of their research, IAR was still not a reality as there were only two AR systems to be used outside of the lab with only one of these in continual use.

---

[11]In this case GPS refers to a GNSS that utilises satellites from multiple signal providers, for example GPS, GLONASS, Compass and Galileo.

Rankohi and Waugh (2013) recently reviewed AR literature specifically for the construction industry. Their research approached AR from the construction and engineering areas as evident in the following reviewed journals: Journal of Automation in Construction (AIC); Advanced Engineering Informatics (AEI); ASCE Journal of Computing in Civil Engineering (CCE); Journal of Information Technology in Construction (ITCON); Journal of Computer Aided Civil and Infrastructure Engineering (CACIE); ASCE Journal of Construction Engineering and Management (CEM); Emerald Journal of Engineering; Construction and Architectural Management (ECAM); and Emerald Journal of Construction Innovation: Information, Process, Management (CIIPM). They identified key industry values in the following areas:

- contents - current and historic project information

- features - user-friendly interface, integrated into current workflow

- value - system is affordable initially and upkeep with short payback period

Rankohi and Waugh (2013) also found that the majority of AR based papers in the construction and engineering fields focused on field works as the principal user. Furthermore, they found that the primarily visualisation and simulation was the main task of these systems, comparing virtual information with the real-world as a progress indicator. They also discovered that there was a trend of increasing in numbers of mobile AR research in the last decade.

Although not focusing on industry as a whole, the findings by Rankohi and Waugh (2013) can be applied to a wide range of industrial applications. These surveys agreed that IAR is currently not a reality but that the industry is making good progress towards the everyday use of AR in industrial practices. They found that there are still a very limited number of AR applications that are successfully integrated into the industrial process. Fite-Georgel (2011) found that, in general, there are good arguments for the adoption of AR solutions but that the primary reasons for lack of adoption was the scalability of the system and the tracking accuracy and robustness of the system.

Scalability is usually a case of hardware availability, where outdoor AR systems use custom hardware to obtain higher levels of accuracy and precision but make it harder to deploy the AR system on additional platforms. In an effort to address scalability issues, there has recently been a push towards deploying AR systems on a large-scale with the aid of mobile hardware platforms. This approach has been successful for advertising- and entertainment-based AR systems being deployed on the Apple App Store, Google Play Store

and BlackBerry Appworld. Some publications include: ARKick[12] , Layar[13] and Wikitude[14]. However, these applications are targeted towards a consumer market and as such have different requirements to industrial tasks. 3DON[15] and urbasee[16] are two IAR applications that have been designed for the visualisation of architectural work in an outdoor environment but are still limited by consumer-grade positioning sensors.

2.4   Custom Hardware

Some IAR applications use custom hardware because the specific AR task requires an accuracy or precision not achievable with current consumer devices. Research often assumes that perfect registration will eventually be achieved (Livingston & Ai, 2008), either ignoring this component or using custom hardware as a proof-of-concept whilst technology catches up.

Outdoor AR systems commonly use industrial level GNSS systems to aid in pose solution (A. Behzadan & Kamat, 2005; Fong, Ong, & Nee, 2008; Schall et al., 2009; Gleue & Dähne, 2001; Satoh, Anabuki, Yamamoto, & Tamura, 2001; Min, Lei, Wei, & Xiang, 2012; Junghanns, Schall, & Schmalstieg, 2008), including phase-based, differential and RTK solutions. These systems are able to provide a 3D position value with reference to a coordinate system with an accuracy of 20mm horizontally and 40mm vertically in realtime. However, these GNSS devices are often expensive, usually in excess of $5,000 NZD and require either a second GNSS receiver or CORS Network[17] subscription for real-time calculations. Although this price is not outside the scope of some industrial tasks it can be difficult for research projects to acquire these systems on a research budget.

AR systems require additional sensors to provide a full 6D pose solution and commonly use a combination of sensors including accelerometers, gyroscopes and electronic compasses to this affect. Raw observations from these sensors are commonly fused with complex filtering approaches like Kalman Filters and Extended Kalman Filters to determine a pose like in Karlekar,

---

[12]http://refocuslabs.com/app/arkick/

[13]http://www.layar.com

[14]http://www.wikitude.com

[15]http://www.3don.co.uk

[16]http://www.urbasee.com

[17]Continually Operating Reference Station (CORS) Networks provide a correction values to connected GNSS devices allowing RTK based position solutions. They are usually operated commercially but are also government controlled in some parts of the world. Additional information on the New Zealand CORS network can be obtained from the LINZ website: http://apps.linz.govt.nz/positionz/rt/index.php.

Zhou, Lu, et al. (2010) and Oskiper, Samarasekera, and Kumar (2012). Sensor platforms that also provide a 'black box' pose solution can also be used like systems from InvenSense[18] and Xbow[19]. Filtering approaches can also be used to fuse sensor results and computer vision solutions, which have been shown to provide robust relative tracking (Ababsa et al., 2012; Ababsa, 2009a; Williams, Green, & Billinghurst, 2013).

IAR systems have commonly used either a head mounted display (HMD) or *magic lens* approach to deliver virtual content to the system user (Junghanns et al., 2008). HMD's are a promising approach for industrial tasks as they allow the users hands to remain free for conducting tasks whilst still providing AR content. However, in tasks that do not require the users hands to be free the 'magic window' approach can be beneficial as it can overcome some of the limitations presented by HMD's. The *magic lens* approach commonly uses a mobile device to present a video-see-through display drawing digital information onto the display that is usually not visible.

2.5   Mobile Hardware

As the power and portability of personal computers, particularly in the form of smartphones and tablets, has increased there has also been an increase in the suitability of these devices to used for AR applications (Carmigniani & Furht, 2011). Current smart-phones usually contain both front and rear facing cameras, accelerometer, gyroscope, electronic compass and a phase-based GNSS sensor in a portable format. Furthermore, the rear-facing camera on most of these devices allows video-see-through techniques to overlay digital information onto the real environment making suited for video-see-through AR applications. Mohring, Lessig, and Bimber (2004) first showed the capabilities of mobile devices for AR with the accurate tracking of planar targets in 2004.

In addition, the relatively inexpensive cost, mass production and common operating systems of smartphones and tablets allows for the deployment of AR systems on a large scale. This essentially overcomes the issue of scalability as highlighted by Fite-Georgel (2011).

Nonetheless, the sensors and positioning systems in mobile devices are designed to be inexpensive and have a low power consumption, meaning that often the accuracy and precision of these systems is not suitable for a wide range of industrial AR tasks. For example, Figure 2.4 shows the position of a stationary consumer-grade tablet as determined from the on board GNSS

---

[18]http://www.invensense.com/
[19]http://www.xbow.com/
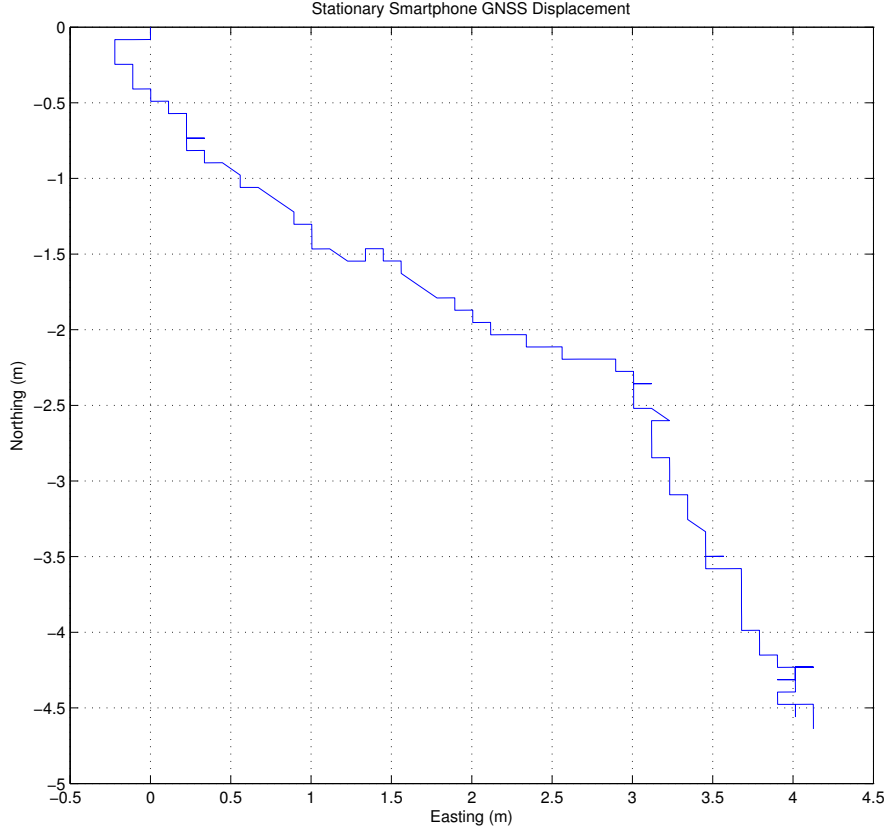
Figure 2.4: Position plot of a stationary tablet utilising on-board consumer-grade GNSS sensors over a period of 30 minutes.

sensor. This plot shows that over a period of 30 minutes the device's relative position has shifted by approximately 6.2 meters. Furthermore, Table 2.2 shows that the absolute error of this device is 9.47 meters when compared with an industrial-grade device that is reporting position to within 0.02 meters. Work by Blum, Greencorn, and Cooperstock (2013) found that smartphone devices, like the Apple IPhone 4, have a meanorientation accuracy of 10 degrees to true north[20].

---

[20]Compasses require a position based correction value to provide a true north bearing. This is due to the miss alignment of the Earth's magnetic poles with respect to its rotation axis, as well as its dynamic nature. In smartphones, correction values are as determined from the World Magnetic Model (http://www.ngdc.noaa.gov/geomag/WMM/DoDWMM.shtml), which is regularly updated (Google Inc, 2014).

Table 2.2: Absolute position comparison of a consumer-grade and industrial-grade GNSS device. Location is determined with respect to the Map Grid of Australia, Zone 55.

Device	Easting - x (m)	Northing - y (m)	Elevation - z (m)
Samsung Galaxy Note 10.1 GSM	526373.323	5252652.619	44.700
Leica Viva GS15	526368.793	5252658.249	38.583
Difference	4.530	-5.630	6.117

2.6 Augmented Reality Pose Solutions

As previously stated, smartphones and tablets contain consumer-grade sensors with low accuracy and precision (Arth, Klopschitz, Reitmayr, & Schmalstieg, 2011; Blum et al., 2013). As a result research often uses custom hardware, capable of higher levels of accuracy and precision, as a proof of concept technique whilst consumer-grade hardware improves. This is especially true with outdoor AR applications as consumer-grade GNSS receivers are not suitable for tracking, requiring these systems to use industrial-grade GNSS receivers for device location (for example: (A. Behzadan & Kamat, 2005; Fong et al., 2008; Schall et al., 2009; Gleue & Dähne, 2001; Satoh et al., 2001; Min et al., 2012; Junghanns et al., 2008)).

Recently research has focused on improving the accuracy and precision of AR systems on consumer-grade mobile devices to provide an immersive and realistic AR experience (Arth, Wagner, Klopschitz, Irschara, & Schmalstieg, 2009). This is because users will have a better AR experience if spatial cohesion is maintained between real and virtual content and poor device pose is the primary reason for reduced cohesion. Rossler et al. (2011) even go as far as identifying pose solution as the most important aspect of an AR system. This is supported by findings from (Mulloni et al., 2012), who found that pose tracking has an impact on user adoption rate of AR applications.

Work by Yi-bo, et al. (Yi-bo, Shao-peng, Zhi-hua, & Qiong, 2008) categorised available technologies into the six generalised fields that can be used for AR pose solutions, including; mechanical, magnetic sensing, GNSS, ultrasonic, inertia, and optical. Each of these technologies can be differentiated using aspects such as accuracy, precision, range, initialisation, time and environmental factors. Carmigniani and Furht (Carmigniani & Furht, 2011) summarised the general quantitative outputs of each of the aforementioned

Table 2.3: Comparison of common tracking technologies as summarised by Carmigniani and Furht (2011). Range: the size of the region that can be tracked. Setup: the amount of time for instrumentation and calibration. Precision: the granularity of a single output position. Time: the duration for which useful tracking data is returned. Environment: where the tracker can be used, indoors or outdoors.

Technology	Range (m)	Setup time (hr)	Precision (mm)	Time (s)	Environment
Optical: marker-based	10	0	10	∞	in/out
Optical: markerless	50	0-1	10	∞	in/out
Optical: outside-in	10	10	10	∞	in
Optical: inside-out	50	0-1	1	∞	in/out
GPS	∞	0	5000	∞	out
Wi-Fi	100	10	1000	∞	in/out
Accelerometer	1000	0	100	1000	in/out
Magnetic	1	1	1	∞	in/out
Ultrasound	10	1	10	∞	in
Inertial	1	0	1	10	in/out
Hybrid	30	10	1	∞	in/out
UWB	10-300	10	500	∞	in
RFID: active	20-100	when needed	500	∞	in/out
RFID: passive	0.05-5	when needed	500	∞	in/out

tracking technologies as shown in Table 2.3. Lopez et al (Lopez, Navarro, & Relano, 2010) reviewed a large number of AR publications with the aim of providing a guide for future research into previous hardware and software approaches. They found that maker-based identification is more common in indoor systems whereas geo-positioning identification (for example GNSS solutions) is more common in outdoor AR. Furthermore, they found that there is no predominant library used for processing of images in AR (Lopez et al., 2010).

In general, tracking systems based on computer-vision are able to provide a more precise pose solutions but often require considerable computational power and initialisation periods (Park & Lee, 2012). Wagner, Reitmayr, Mulloni, Drummond, and Schmalstieg (2008) first demonstrated natural fea-

ture tracking on a mobile device in 2008, since then there have been several commercial (Qualcomm Vuforia[21], Metaio[22], Wikitude[23]) and open-source (ARToolKit[24], Mobile AR Framework[25]) libraries developed that handle AR tracking. Of these systems the computer-vision based libraries are, currently, not suitable for large-scale outdoor environments (Mulloni et al., 2012). In contrast, AR systems that utilise sensor-based tracking solutions are able suitable for large-scale environments. However, in the case of consumer-grade sensors they have poor accuracy, and conversely, industry-grade sensors are highly accurate but with poor scalability.

There has been an continued focus on hybrid tracking approaches in the last decade that attempt to overcome the shortcomings of each of the individual technologies by combining computer-vision and sensor-based methods, to create robust tracking systems (Satoh et al., 2001; Ribo et al., 2002; Zendjebil, Ababsa, Didier, & Mallem, 2008; Ababsa, 2009a; Gauglitz et al., 2012). Work by Satoh et al. (2001) successful showed that it was possible to correct for sensor drift with the aid of computer-vision techniques as early as 2001. Their system compensated for orientation drift over long periods at a stationary position. Subsequent work by Ribo et al. (2002) was able to demonstrate a real-time hybrid tracking for 6DOF in an outdoors environment. Their system used an Extended Kalman Filter to combine computer-vision and inertial sensor results. However, both these AR systems utilised custom hardware and by today's standard are out-dated.

2.6.1 Localisation

Although there has been a large amount of research into tracking in both indoor and outdoor environments, there has been little work into the localisation of these systems (Arth et al., 2009). Localisation is a sub-component of the device tracking, but is often overlooked whilst being equally important. It is tasked with defining the AR system's position with respect to a coordinate system or datum, in other words to determine the devices absolute position. These coordinate systems can be defined in various ways but usually are defined either with respect to a local (for example target or marker based systems) or global (for example GNSS coordinate systems) datum.

The localisation process can be utilised in different manners, including;

---

[21]https://www.vuforia.com

[22]http://www.metaio.com

[23]http://www.wikitude.com

[24]http://www.hitl.washington.edu/artoolkit/

[25]http://hitlabnz.org/index.php/products/mobile-ar-framework

- initial localisation, where the device's absolute position and attitude is determined at the start of tracking and then relative tracking is conducted until tracking fails, at which time localisation is performed again, or an

- integrated system, where localisation is closely coupled with tracking, like absolute sensors or target based tracking.

Dedicated localisation systems for mobile AR have been developed as early as '92 (Want, Hopper, Falcão, & Gibbons, 1992). This system, utilising infrared (IR) badges, and its successor the Bat system (Addlesee et al., 2001), employing ultrasonic positioning techniques, are among the most unique localisation approaches presented. Since then research has tended to focus on sensor (A. H. Behzadan & Kamat, 2007; Junghanns et al., 2008; Schall et al., 2008; Zendjebil et al., 2008; Ababsa, 2009b; Min et al., 2012), computer-vision (Lorenz Wendt, Bres, Tellez, & Laurini, 2008; Zendjebil et al., 2008; Zhu, Oskiper, Samarasekera, Kumar, & Sawhney, 2008; Arth et al., 2009; Ventura & Hollerer, 2012; Bostanci, Clark, & Kanwal, 2012) or hybrid (Ababsa, Didier, Zendjebil, & Mallem, 2009) localisation approaches, which have had varying degrees of success in different environments.

Sensor-based

In an outdoor environment, sensors that sample absolute readings provide a fast and effective manner of localisation in un-calibrated setting. This allows these systems to operate over wide-areas and often at global scales.

For example A. Behzadan and Kamat (2005) initially published concepts for an outdoor AR system. This was later included into a developed system called UM-AR-GPS-ROVER (A. H. Behzadan & Kamat, 2007), which was designed for the visualisation of construction CAD models. The system employs a sensor-based tracking system whereby the position is determined from a GNSS sensor and the attitude is determined from inertial sensors. Their system used sensors that only provided a relative change in attitude, which is considered uncommon by today's standards.

Junghanns et al. (2008) showcased a mobile outdoor industrial AR solution named 'VIDENTE' that was designed for the visualisation of underground services; such as pipes, electrical & telecommunications cables. The positioning and attitude system was then presented by Schall et al. (2008). The system uses custom hardware in the form of a video-see-through display with industrial grade GNSS, barometer and inertial sensors for positioning. The GNSS results are integrated with the barometer through the use of a Kalman Filter to provide position data, whilst the attitude is determined

through inertial sensor and computer-vision fusion to provide sub-meter accuracy. Nonetheless, the absolute position and attitude of the device are still determined through inaccurate means that would require additional localisation to correct.

Zendjebil et al. (2008) demonstrate a localisation system that comes into effect when visual tracking fails. This system, named the Aid-localisation, initially, estimates absolute pose from GNSS and inertial sensors data before predicting error corrections for future tracking failures. Their approach was able to achieve a mean value of less than 1-degree orientation error and 1-meter positional error in real-world tests. This was later improved by Ababsa (2009b), through the use of structure from motion techniques and the development of an improved extended Kalman filter. However, the GNSS device used in this research is only accurate to sub-meter levels (provided there is an external corrections source) and as such limits the absolute accuracy of the system. Furthermore, their system utilises custom hardware for each of the deployed AR systems making it difficult to scale the system for additional users.

Work by Min et al. (2012) used a custom GNSS and IMU sensors for localisation of the AR system. The localisation solution is then further refined through an interactive approach, whereby the users makes fine adjustments through the use of a gamepad input device reducing localisation error. Due to the low detail of the building models and high accuracy of GNSS sensors used it was not necessary for translational error to be corrected in the evaluation, which may reduce the suitability of this task. Although this system operates on custom hardware a similar approach may be viable for mobile devices, provided the available input is suitable to fine tune 6DOF.

Vision-based

The aforementioned sensor-based localisation systems use industrial and custom hardware and with the current state of technology on mobile devices, vision-based localisation systems are able to achieve a level of precision not achievable by their counterparts. Vision-based systems tended to use targets to localise the system but since then have progressed to using natural feature tracking, like structure from motion (SfM) and feature based matching (Arth et al., 2009). Pure vision-based systems require a pre-calibrated environment to be able to successfully localise, both locally (targets, images, or local datum) or globally (geo-referenced datum).

Lorenz Wendt et al. (2008) investigate the use of SIFT descriptors for the localisation of mobile devices in outdoor environments. Their system employs a pre-calibrated environment of feature descriptors to localise the AR

system. They found that dynamic environments, as well as changing lighting environments decreased the likelihood of successful localisation. Furthermore, they found that increasing the size of the pre-calibrated environment puts additional pressure on the mobile device as storage requirements increase. Although not mentioned in their research, a server-based localisation approach removes this dependency on local data storage.

A novel approach developed by Zhu et al. (2008) utilises a multi-stereo, head mounted stereo front and rear facing cameras, visual system to localise in a pre-calibrated environment. They were able to localise to decimetre level accuracy once the environment has been calibrated. However, this system operates on a custom hardware, making it hard to deploy on a large-scale.

Research by Bostanci et al. (2012) evaluates the potential use of an indoor visual localisation approach in an outdoor environment. They used the OpenCV library to run a SLAM system with BRIEF descriptors to localise the system. Unfortunately, they found that this approach did not work well but that a refined approach produced better results.

Recent work by Ventura and Hollerer (Ventura & Hollerer, 2012), developed a promising system that both localises and tracks the system using only computer-vision within a pre-calibrated environment. Their system extracts panoramas from a video sequence at intervals dependant on motion before an offline reconstruction process generates a map of the environment for use with an online localisation server. Using this technique the tracking system achieved below 25cm positional accuracy and 0.5 degrees for 80% of their test images. Nonetheless the sampling process that generates the localisation map differs greatly from the usual workflow common to engineering, construction and surveying process. Fite-Georgel (2011) identified the increase of industrial workflow as a reason for reduced adoption of AR applications in industry.

Gauglitz et al. (2012) overcame motion limitations in SLAM by calculating both Homography and Essential matrices for each key frame. The correct matrix is selected by GRIK Score evaluation. This technique provides support for both general and rotation-only motion, which, most SLAM techniques do not. Nonetheless, the tracking robustness achieved by this approach is not comparable with PTAM or DTAM but is able to operate in a much larger area (Kleinert & Stilla, 2012; Newcombe, Lovegrove, & Davison, 2011). In their study, Gauglitz et al., identified that further development needs to be put into map feature management systems, specifically in outlier filtering and key frame selection (Gauglitz et al., 2012). Furthermore, they suggest additional research into distinguishing between fore- and background for different motion detection.

Model-Based

Model-based tracking in outdoor environments utilises various characteristics of CAD models to determine the location and attitude of the camera with respect to the CAD model. At the turn of the century the Rockwell Scientific Company demonstrated a model-based visual tracking for outdoor AR (Behringer & Sundareswaran, 2002). Their particular approach employed forward projection of CAD models and the camera feed to determine the optimal location of the camera with respect to the CAD model. This was later followed by work from Reitmayr and Drummond (2006) who used extracted edges from images to track a relative position on a mobile device. Although their research was not focused on localisation, it is still capable of localising with respect to the CAD model and they believe that the integration of a GNSS unit would improve this matching process.

Model-based localisation is potentially the closest aligned to the industrial workflow of engineering, construction and architectural processes, as the use of CAD models is common practice. However, the development of CAD models is a resource expensive process that may not be viable in an environment that is continually being changed, like during a construction process.

Sensor-Fusion

It is possible to circumvent the requirement of a pre-calibrated environment by coupling the vision system with a sensor that provides absolute position to create a hybrid system. Commonly, hybrid systems use GNSS and compass sensors to provide an initial absolute pose estimate before refining this estimate through vision-based approaches.

For example, Zendjebil et al. (2008) utilise a hybrid approach for global localisation of outdoor AR systems. The absolute position is determined with respect to a global datum with the use of GNSS and inertial sensors. Greater focus is placed on predicting the localisation errors for use in quality assessment and recovery when vision-based relative tracking fails (Zendjebil et al., 2008).

Karlekar et al (Karlekar, Zhou, Nakayama, et al., 2010) present a system that uses differences between scene and CAD model contours to correct for initial pose error from sensor solutions before changing to vision based model tracking. However, their system relies on the ability to detect meaningful edges between the scene and CAD model, requiring a substantial amount of the CAD model to be visible at all times.

Ababsa et al (Ababsa et al., 2012) also use a hybrid tracking approach

specifically for geological applications but are dependant on the quality of their CAD models for the accuracy of the vision tracking component. Langlotz et al. (Langlotz et al., 2011) showed that computer vision panorama tracking could be combined with additional sensors to add virtual tags to a scene that appeared fixed in place, although this is restricted to working about the panorama origin.

In summary, localisation approaches can be broadly classified into sensor, vision and hybrid solutions. Both the sensor and vision based approaches have technology specific obstacles that in most cases can be overcome in a hybrid solution. Commonly, sensor based approaches are often less precise than vision based systems but do not require a pre-calibrated environment for localisation. The most promising forms of localisation are those approaches that combine both vision and sensor solutions, essentially being able to overcome limitations in the individual technologies as whole.

In contrast, our localisation approach (to be detailed in chapter 4) uses a sensor-vision localisation approach. This is done by initially determining absolute pose from GNSS and inertial sensors before the absolute pose is refined using computer vision approaches. This primary difference to prior research is the type and source of data used for the computer vision localisation method.

2.7   OpenCV Geometry Reconstruction

Baggio et al. (2012) published a practical guide to using OpenCV that, amongst others, explains the process of recovering 3D geometry using from an image sequence using structure from motion techniques. The method detailed is a standard approach in the computer vision[26] discipline for recovering geometry in non-planar environments. The following list outlines the primary steps in the recovery of scene geometry and camera extrinsic values:

1. Detect and extract keypoint descriptors.

2. Match keypoint descriptors.

3. Solve for the Fundamental Matrix (F).

4. Convert F to the Essential Matrix (E).

---

[26]The field of Photogrammetry utilises other methods of recovering geometry from image pairs, commonly utilise the co-linearity and coplanarity condition in their solutions. For further reading on geometry recovery approaches used in photogrammetry see work by Kraus (2007).

5. Determine extrinsic camera parameters using Singular Value Decomposition (SVD).

6. Reconstruct 3D scene with homogenous coordinates and the camera matrix (P).

7. Determine scale and orientation form existing scene using Perspective-N-Point (PnP) methods.

Epipolar geometry, or the coplanarity condition, is the underlying relationship that allows the difference in relative pose between the images pairs to be determined. It is also worth noting that the approach used by Baggio et al. (2012) uses the first image pair to set an arbitrary orientation and scale before reconstructing the scene. 3D geometry (also known as a space resection) is then used to define all subsequent stereo pairs with respect to the initial arbitrary orientation and scale. The 3D geometry uses the colinearity condition that a ray leaves the camera's origin intersecting a point in the image and a corresponding ground point. Ground points that are common between scenes allow the scale and orientations to be set for further scenes. This is because scenes and geometry reconstructed using image based techniques are unit-less requiring additional constraints to define this.

The reconstruction approaches used in this research differs in key areas that will be further discussed in subsection 4.2.4.

2.8 Error Types

It is a well-accepted fact that all observations in the real world contain both the observation and error. There are several forms of error that, within the school of spatial sciences, are categorised into (Shepard, 1983; Elfick, Fryer, Brinker, & Wolf, 1987):

- Random, and

- Systematic Error

Systematic errors are errors that can be removed through a mathematical relationship or correction factors. For example the position provided by GNSS devices have a large amount of systematic error[27] that can be modelled offline to provide a more accurate position. Random errors are errors that remain once systematic errors are removed. They are normally distributed

---

[27]Table 2.2 demonstrates the magnitude of systematic error that can be found in consumer-grade GNSS devices.

errors that define a sensor's precision or standard deviation for an observation or recording. The only ways to overcome random errors is through the technology advance or by using mathematical procedures to combine additional or complementary observations to provide a more statistically more accurate observation. Gross error can also be classified as random errors but are caused by blunders or mistakes made by users. In this case the localisation process is an automated system which essentially omits this error source but as an example could be viewed as a coding error that produces flaws into the system.

2.9 Summary

This chapter introduced the concept of AR and defined its required characteristics as adopted by this research.

Industrial tasks and their requirements are then described with respect to possible applications of AR in outdoor industrial tasks in the engineering, surveying, planning, development and GIS fields. The limitations of previous industrial AR applications are then discussed, specifically focusing on the impacts of hardware and how this effects the adoption rates and continued of said industrial AR systems.

Smartphones and Tablets are then introduced as an alternative for high-accuracy industrial AR tasks and how the primary obstacle for these devices is pose solution accuracy. The task of determining pose solutions is then described in terms of tracking and localisation that respectfully address relative and absolute pose solutions.

Chapter 3

Outdoor Industrial AR Framework

This chapter outlines a concept AR framework for outdoor industrial tasks. As discussed in section 1.2, localisation is the key focus of this research, and the concept framework is discussed only to provide context for how these localisation techniques fit into a complete outdoor industrial AR framework.

While the concept framework was specifically designed with outdoor industrial applications in mind, it is generic enough to be suitable for any outdoor AR applications in general. The industrial AR framework is designed in a modular manner such that any technique used could be improved independently.

3.1 Framework Concept

The concept Industrial Augmented Reality (IAR) framework was designed to be modular, such that any future advances in technology could easily be integrated to improve the overall performance of the system. The framework is divided into the following basic modules:

- *Data capture* - data is captured from the device sensors at the rate governed by the tracking module which is then passed to the *tracking* module for processing.

- *Tracking* - this module handles device tracking and localisation with correction parameters updated externally to improve localisation.

- *Localisation* - the localisation module calculates correction values based off the devices estimated pose.
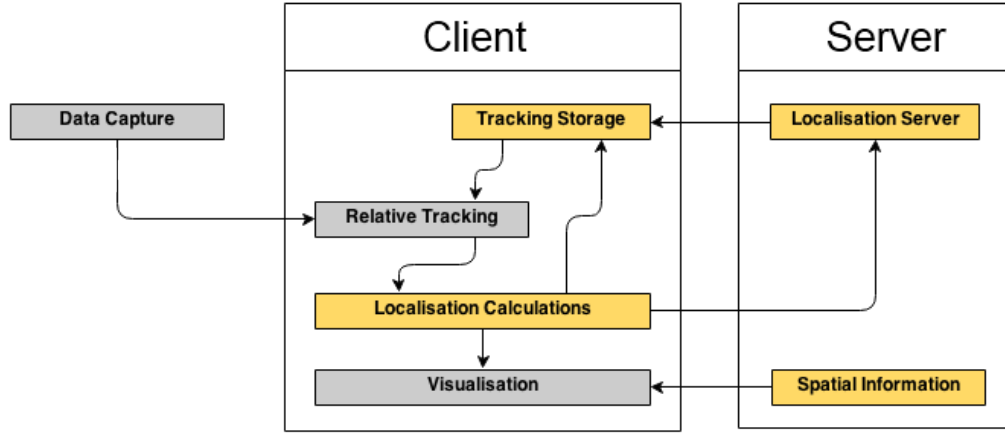
Figure 3.1: Flow diagram dipecting the client/server framework conceptulised in this research.

- *Visualisation* - uses the device pose as determined by the tracking module to render the correct virtual information.

The modular design of the framework allows for it to be build with a client-server architecture, where the user's device (the client) captures data about the environment it is in (e.g. orientation and positional data, images, etc), and then sends this data to a remote server for processing, as shown in Figure 3.1. There are numerous advantages to this, the main ones being the reduction of the computational power and storage space requirements for the client device and the allowance of easy updates to any data used for localisation. There are also significant disadvantages being system latency and startup time to name a few.

The process of localisation in this client-server architecture is as follows:

1. An image is captured from the device's rear-facing camera

2. Key points are extracted from the image

3. Key points sent to the server's localisation module with the most accurate absolute pose solution available

4. The localisation module renders the 3D point cloud to a 2D image using the client's estimated position and orientation

5. The localisation module performs image matching between the generated image and the image captured by the device

6. The pose correction factor is determined from results of image matching and estimated pose

7. The device's pose is updated using pose correction values from the server's localisation module

Figure 4.1 shows a detailed flow diagram describing the operation procedure for the localisation module. Furthermore, chapter 4 explains the intricacies of the localisation process with respect to determining pose between image/point cloud matches.

3.1.1 Concept Design Benefits

The concept IAR framework was designed following the recommendations of Fite-Georgel (2011) and Rankohi and Waugh (2013) to encourage adoption and use outside of academia[28]. In brief, their findings recommend that an AR system needs to be cost effective, scalable, user friendly, and can be integrated into the existing workflows of its target users. The following paragraphs explain how the concept design fits these recommendations, and the benefits that arise from this design.

The aim of a cost effective system is to ensure that there is an overall positive net resource cost when performing a certain task, either by reducing cost or increasing profit. By targeting consumer grade devices, the framework ensures that the initial costs are kept low, meaning that only a small increase in productivity is needed to make the AR framework cost-beneficial. One of the main strengths of AR is the ability to show the relationship between real and virtual information, so AR can increase productivity by allowing for simple and accurate design verification. For example, reworks, the process of fixing or amending previous construction work, is one of the primary causes for construction tasks to go over budget and is usually due to mistakes in the construction design (Love & Li, 2000). The longer that these errors go unnoticed the more rapidly the cost of amendment increases, and any tool which facilitates earlier detection of these errors has the potential to significantly reduce costs. An AR interface can be used to reduce the amount of reworking needed in a construction site.

For an outdoor industrial AR system, scalability relates not just to the number of users but also the robustness of the system to changes in the operating environment. Lighting, scene and temporal changes are examples of environmental factors that a scalable system should be able to account for.

---

[28]See section 2.3 for detailed recommendations given by Fite-Georgel (2011) and Rankohi and Waugh (2013).

In particular, temporal changes can have a large impact on the localisation process as some development projects can be ongoing for periods of months or years. By using a client-server based architecture to perform localisation, costly computation is performed remotely, allowing for lower powered consumer devices to be used. In this case any updates to the localisation data can be performed on the server, resulting in a more robust experience for users despite temporal changes in the data.

To improve the user friendliness of a system, the visible complexity of the system should be reduced as much as possible. In the case of the IAR, this means automating as much of the localisation process as possible to reduce the workload of the user. The server architecture allows for highly complex procedures to be computed automatically without requiring higher powered, more expensive consumer devices.

The framework has been designed to fit the standard workflow for capturing large-scale RGB point clouds, ensuring that additional time spent onsite and costs are kept to a minimum. The data capture process was verified with continual feedback from industrial partners to ensure its conformity. The only way the process differs from the standard workflow is that the point clouds are converted in to Point Cloud Database (pcd) format before being uploaded to the server.

3.1.2 Concept Design Drawbacks

The localisation module was predominantly designed to operate on a server, sending corrections to the mobile device. This approach was justified primarily due to mobile devices being commonly restricted by hardware for the processing and storage of large data sets. Furthermore, server based solutions are able to service multiple clients, reducing storage requirements, hardware cost and data management overheads.

Nonetheless, server-based solutions require some form of connectivity to allow the client to connect to them. For large-scale deployment this is often done through an internet based connection usually in the form of GSM network meaning that there can be issues with using these services in remote environments. Furthermore, there is also the issue of connection latency, whereby additional time must be allowed for in processing requests and data transfer between the client and server. This is also relative in regards to the amount of traffic between the server and client as larger data sets will take longer to transfer between systems.

3.2   Localisation in the Concept Framework

In this concept framework both the absolute position and orientation initially obtained by the system contain systematic and random errors. Furthermore, the image taken by the client device will also contain systematic and random errors, but these are assumed to be negligible[29] when compared with the errors generated by the other sensors in the system. It is the aim of the localisation algorithm to reduce or correct for these errors, thus providing a more accurate and robust IAR framework.

3.3   Summary

This chapter introduces a concept IAR framework for outdoor industrial tasks, based on a client-server architecture. This framework is designed with the focus of providing context for the localisation module described in chapter 4. Benefits and draw backs of the proposed framework are described, and the purpose of localisation within the framework is discussed.

---

[29]The systematic errors that the client image contain, like radial and affine lens distortions, can be modeled with a camera calibration process. This camera calibration process can be automated as it is implausible to require the end user to conduct this procedure due to the technical nature of this task. Furthermore, the intrinsic camera errors found in standard consumer grade devices will have a minimal impact on the localisation system when compared with other errors impacting on the system.

Chapter 4

Pointcloud Localisation

This chapter details the *localisation* module of the Industrial AR Framework explained in chapter 3. In it's most generic sense, this module takes the following inputs: *input pose*, *input image*, *input error* and *control pose*, and produces a *pose solution* and *correction value*, the latter of which being the difference between the *pose solution* and *input pose*.

Figure 4.1 shows the processing order as conducted by this *localisation* module. The points of interest are covered in further detail in this chapter with an evaluation of the *localisation* module conducted in chapter 5.

4.1 System Dependencies

The localisation solution has been developed to operate on Microsoft Windows and Apple OSX operating systems and has been successfully tested in Microsoft Windows 7 & 8.1 and Mac OS X.The CMake[30] build system is used to setup the development environment and dependencies across the different platforms. The system has been written in C++ with shader sections written in GLSL330.

The following libraries are used by the localisation system:

- OpenCV[31] - open source computer vision library used for image processing.

- PointCloudLibrary[32] - open source point cloud library used for processing point cloud data.

---

[30]http://www.cmake.org/
[31]http://opencv.org /
[32]http://pointclouds.org /

Figure 4.1: Flow diagram for *localisation* module showing standard operation procedure.
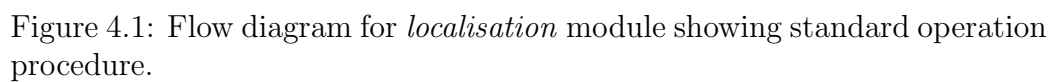
- OpenGL Mathematics[33] - A C++ header wrapper supporting common mathematical functions and GLSL conventions.

- GLFW3[34] - OpenGL library for handling window creation and context handling

4.2   Localisation Approach

The localisation approach used in this research involves matching between an image capture from the client device, and a virtual image created from point cloud data.

Two dimensional image matching is a common method of determining the change of pose of a camera between two different images (Lorenz Wendt et al., 2008; Zhu et al., 2008; Gauglitz et al., 2012). These techniques are limited to only being able to calculate pose from areas where pictures have been captured, limiting their flexibility significantly. In addition to this, the calculated pose is not specific to any existing co-ordinate system, and may not represent the change in pose in real world units.

In this research, point cloud data is used to generate images of virtual viewpoints from any position and orientation. This allows the pose to be estimated given a single image from any position within the limits of the point cloud data. In addition, as the co-ordinate of every point within the point cloud is known in the real world co-ordinate system, the true pose in real world units can be calculated.

The localisation approach works by rendering a virtual image from a viewpoint created using the position and orientation from the device's internal sensors. As these sensors have limited accuracy, the viewpoint will not be exactly accurate, but the virtual image rendered from this location will be able to be matched to the real image captured from the exact location, and the pose estimation updated to the true location.

The following subsections cover methods for capturing point cloud data (subsection 4.2.1), rendering a virtual image from a known inaccurate location (subsection 4.2.2), and localisation using feature matches from the real and virtual image (subsection 4.2.4).

---

[33]http://glm.g-truc.net /0.9.5 /index.html
[34]http://www.glfw.org /

4.2.1 Point Cloud Data Capture

As mentioned in subsection 3.1.1, the methods used to generate the point clouds used by the *localisation* module follows the standard workflow used in the surveying and engineering industry to capture building scale data. This differs from the majority of prior research (Snavely, Seitz, & Szeliski, 2006; Irschara, Zach, & Bischof, 2007; Arth et al., 2009; Takacs, El Choubassi, Wu, & Kozintsev, 2011; Ventura & Hollerer, 2012), who use computer vision based image matching or structure-from-motion approaches generate their localisation point cloud. Point clouds generated using these methods require the environment to be highly textured, and any areas of uniform or repeating textures often lead to regions of either sparse or no coverage in these point clouds. Furthermore, solutions using these techniques require additional information to define the point cloud's scale and orientation (Hartley & Zisserman, 2003).

In comparison to image matching or structure-from-motion approaches, industrial point cloud generation techniques, such as in those used in this research, usually use terrestrial laser scanners (also known as LiDAR). Distance, horizontal and vertical angle observation are observed from the laser scanner, which allows the xyz coordinates to be calculated with 3D geometry and at a known scale. These systems regularly use time of flight or phase-based technologies to determine the distance from the device to the environment with the angular measurements based on a primary horizontal axis and right-angle rotating prism. This approach generates points on a spherical grid from the point of observation. Furthermore, the techniques used by these systems are not affected by texture and lighting conditions, providing a more uniform and dense point cloud. An additional benefit of this process is that LiDAR systems do the majority of their processing online, with only a small amount of work done in an offline environment when two or more point clouds require alignment.

The FARO Focus[35], a lightweight terrestrial laser scanner with an internal RGB camera, was used to generate the RGB point clouds in this research, as shown in Figure 4.2. The laser scanner, along with XYZ and RGB values, also observed point normals and intensity of return values, but these were not utilized in this research. The observed point clouds where aligned using commercial software such that the overall root mean square (RMS) value of the site was less than six millimetres.

---

[35]http://www.faro.com/products/3d-surveying/laser-scanner-faro-focus-3d/overview
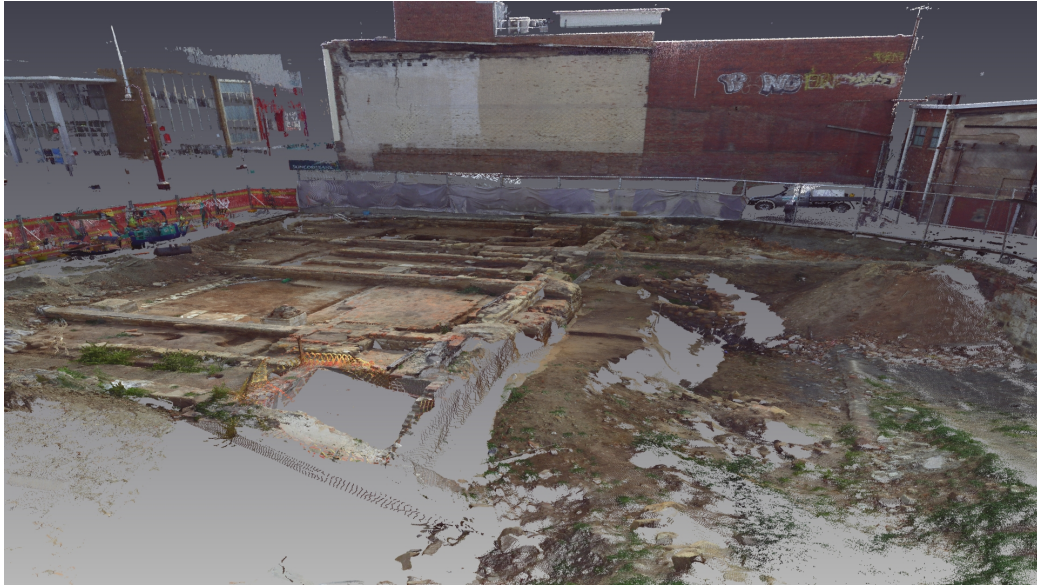
Figure 4.2: An example of the RGB Point Cloud captured by the FARO Focus.

4.2.2 Virtual Image Rendering

Surface Splatting, first presented by Zwicker, Pfister, van Baar, and Gross (2001), is a rendering approach specifically designed for the visualisation of high-density point clouds with no interconnectivity between the points. Surface Splatting approaches were then reviewed by Botsch, Hornung, Zwicker, and Kobbelt (2005) and redesigned to work on modern dedicated GPU's in real-time.

The specific virtual image rendering technique used in this research was inspired by the work of Sibbing, Sattler, Leibe, and Kobbelt (2013), who's focus was to generate photorealistic images that work well using the SIFT feature descriptor. They evaluated a variety of Surface Splatting techniques, both tested and original, and compared the localisation success of each. Their research found that point clouds that only contained XYZ and RGB values were suitable for localisation, but a higher number of correct matches was possible when the point cloud contained additional information like point normal directions. Unlike this research, their system did not evaluate the accuracy of poses calculated, and their approaches required the capture of additional information which is outside of the workflow of industrial processes.

In this research, forward rendering techniques in OpenGL are used to visualise the point cloud. The individual points of the point cloud are repre-

sented using the the OpenGL Point primitive available in OpenGL 3.2 and scaled according to the distance from the viewpoint. The transformation matrices are then updated based on the input pose before the rest of the OpenGL rendering pipeline is executed. At the end of the rendering process, the colour and stencil buffers are recorded for use in the image matching process.

As the rendered images are generated from a geometric scene all points rendered have a corresponding depth value associated with them. This means that the rendered images can now be considered as RGB-Depth images (RGB-D). This allows for additional processing/filtering of rich feature descriptors, which is further described in subsection 4.2.4.

4.2.3   Rendering Parameters

In the localisation pipeline, the standard deviation of error of the input pose calculated using the device's sensors is provided to the *localisation* module. This extra information allows for several optimisations with the aim of improving image to point cloud descriptor matches. If the input pose is normally distributed then there is a high probability that the true pose is located within $\pm$ 2 standard deviations of the input pose for both the orientation and position.

Figure 4.3 shows the overlap between the real (client) image and virtual (server) image when the input rendering parameters are set to the device sensor pose. The error associated with the input orientation is assumed to be 15 degrees, as based on the field measurements of similar devices (Williams, 2013). With the virtual image and client image both featuring a 50° Field of View (FoV), there is potentially a significant amount of non-overlapping image segments.

To reduce the effects that orientation error may have on the image overlap and therefore image matching, different FoV values for virtual image rendering were tested, and an optimal value was determined to maximise image overlap, as shown in Figure 4.4. This evaluation is described in further detail in subsection 5.2.1.

4.2.4   Localisation from 2D/3D Correspondence

The process used by the *localisation* module differs to that of Baggio et al. (2012), outlined in section 2.7, in several key areas. The first difference is that this research uses the Speed-up Robust Feature (SURF) transform (Bay, Tuytelaars, & Gool, 2006) to both detect image key points and to extract corresponding point descriptors from the images. Although other

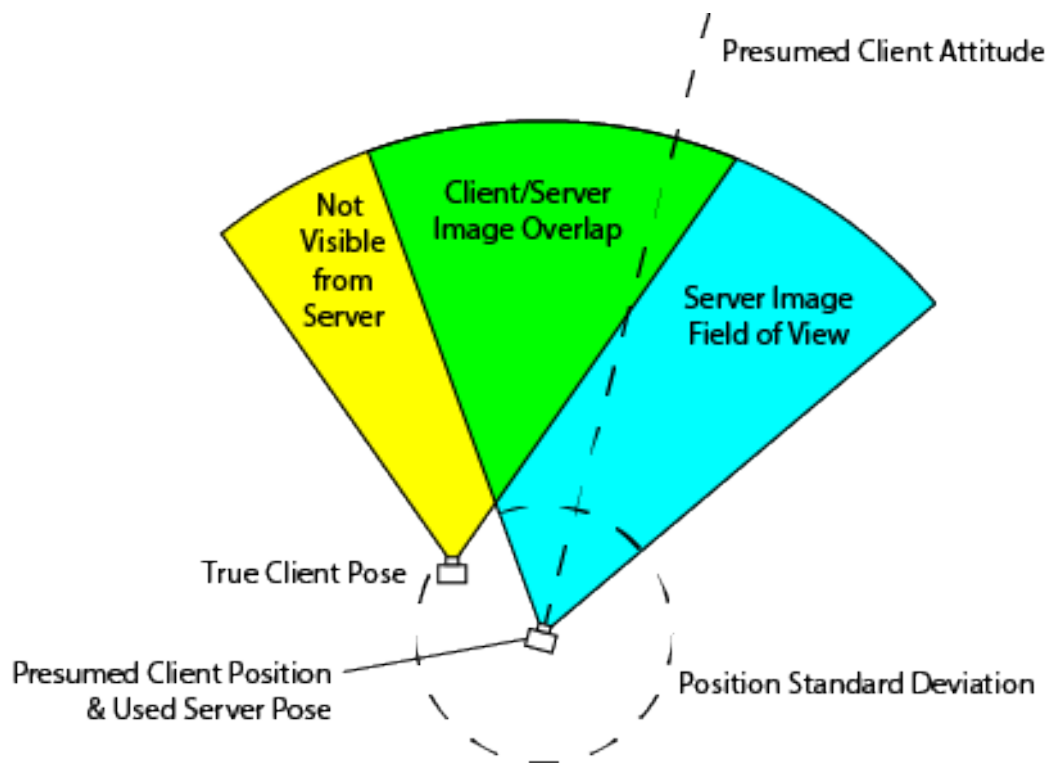Figure 4.3: Example of image overlap between real (client) and server (virtual) images when the virtual image is rendered from the pose estimated by the device's sensors with a 50 degree field of view.
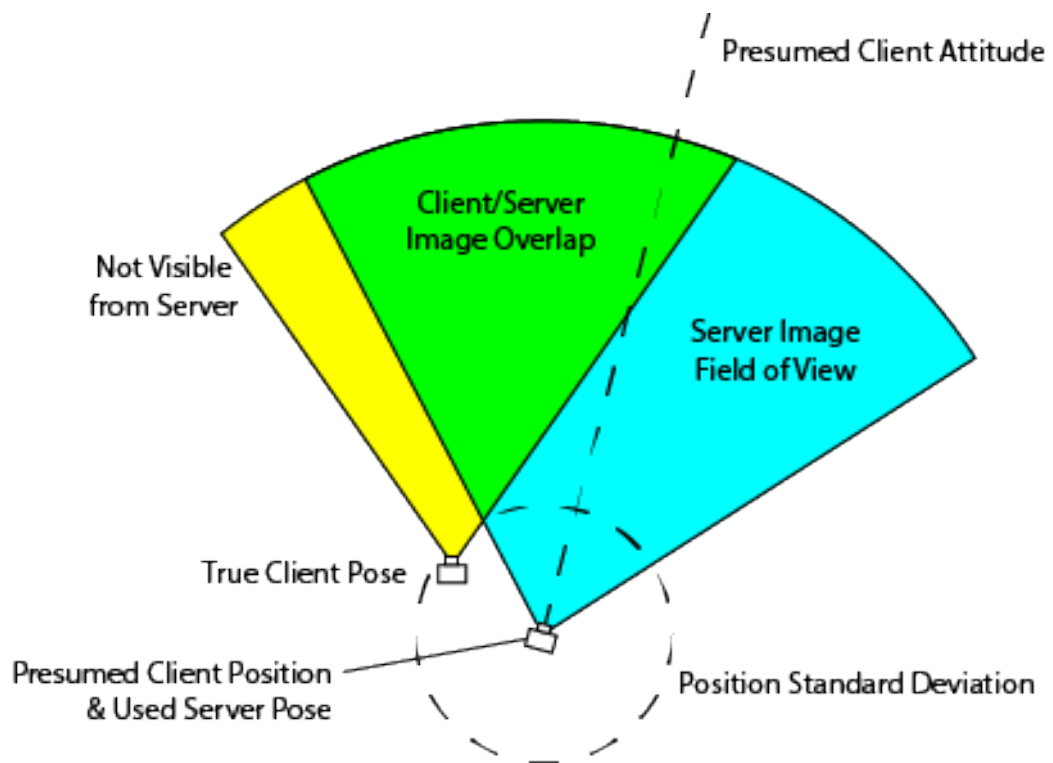
Figure 4.4: Example of image overlap between Client and Server when the server is placed at the presumed coordinates of the client with an increased field of view to account for orientation error.

point descriptors, like FREAK Alahi, Ortiz, and Vandergheynst (2012), have been shown to out perform SURF (Alahi et al., 2012), it was found that these feature transforms did not perform well on the sparse point cloud generated virtual images.

The second difference is in the feature matching stage. The descriptors extracted from the real and virtual images are matched using the Fast Library for Approximate Nearest Neighbours (FLANN) (Muja & Lowe, 2009) k-Nearest Neighbour (k-NN) algorithm. The two best matches for each key point are found, and only matches where the first best match's response is at least 0.5 standard deviations better than the second best match's response are kept. This method effectively filters out feature matches where the is uncertainty if the detected match is truly the optimal.

The remaining differences are the additional filters used to remove any remaining match outliers. The second filter classifies outliers with the aid of a calculated Homography (H) matrix. All keypoints in the real image are transformed using the Homography matrix to determine their ideal positions in the virtual image, and any matching keypoints which fall outside of 1 standard deviation of this position are rejected. It is important to note that the Homography matrix defines the transformation between two planar surfaces, and that while the key points detected do not necessarily lay on planes, there often exist sufficient planar surfaces for in urban environments to still warrant its use.

In section 2.7, an approach for 3D reconstruction of a scene using image pairs was presented, where the reconstructed scene is used to define the scale and rotation of additional feature matches. In this research, the colour point cloud defines the scene and as such the process of scene reconstruction has been completed offline with the use of Terrestrial LiDAR, the advantage of which becomes evident when the co-linearity condition is considered (See Figure 4.5) as the ground coordinates are already known from the virtual RGB-D images. Because of this, steps 3-6 listed in section 2.7 do not need to be completed to determine the extrinsic camera parameters of the client image.

Nonetheless, the Fundamental Matrix (F), a 3x3 matrix that encapsulates the intrinsic geometry between the image pairs, can be used as an additional third filter for the keypoint matches. Image space points in the real image (x_c) and the virtual image (x_s) follow the relationship $x_s^T F x_c = 0$ [36]. Therefore, F can be used to determine the distance between epipolar lines and corresponding point matches to test for and reject remaining outliers,

---

[36]For additional information on the Fundamental matrix see (Hartley & Zisserman, 2003, Chapter 9).
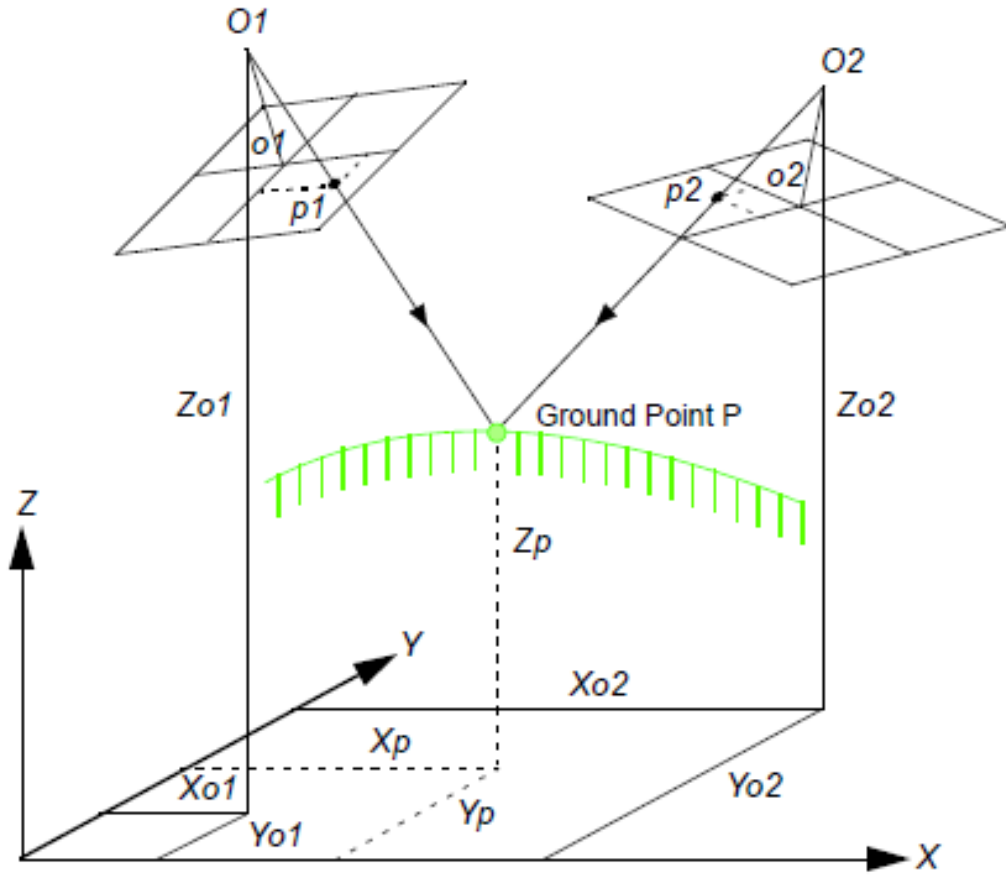
Figure 4.5: Example of the collinearity condition between stereo images. This condition is fundamental in determining the correct pose solution for the correction factor.

further improving the accuracy of the solution.

After the three filters have been employed to remove outliers, the remaining feature matches are then processed using a PnP based solution to determine the extrinsic camera parameters of the client image. OpenCV provides support for the rapid optimisation of PnP solutions given the intrinsic camera parameters. The intrinsic parameters can be determined accuracy using a camera calibration process[37], but can also be approximated at the cost of pose solution accuracy.

---

[37]OpenCV provides tutorials and coding examples to conduct camera calibrations with the use of calibration arrays (http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html).

The focal length and principal point are represented in the intrinsic camera parameters in the following manner:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{array}{l} f_x,\ f_y \text{ - focal length along x, y axis} \\ c_x,\ c_y \text{ - image centre along x, y axis} \end{array} \quad (4.1)$$

While the distortion coefficients are contained within the vector $[k_1, k_2, p_1, p_2, k_3]$, where k_i represents the radial distortion coefficients and p_1 and p_2 are the tangential distortion coefficients along the x & y image axis.

With pose calculation complete, the difference between the pose provided by the device's sensors and that of the calculated pose is then determined, and this difference is used as a pose correction value. This pose correction value is checked against the reported error margins of the device sensors to ensure that an erroneous pose hasn't been computed.

4.3   Summary

In this chapter, a localisation method was outlined based upon image matching between a real image and a virtual image created from a 3D point cloud. The system setup and depenedencies were outlined, and the localisation algorithm was discussed in detail, including point cloud capture, rendering techniques, rendering parameters, and localisation from corresponding feature matches.

Chapter 5

Evaluation of Localisation Algorithm

In this chapter, the localisation algorithm described in chapter 4 is evaluated. First the experimental conditions are described in section 5.1. Following this the effect of the field of view subsection 5.2.1 and Match Filtering subsection 5.2.2 on image to point cloud matching performance is evaluated. The accuracy of localisation in an ideal scenario, using a highly accurate RTK positional estimate, is evaluated in section 5.3, and finally the accuracy of localisation in a real-world scenario, using a less accurate GPS positional estimate, is evaluated in section 5.4.

5.1 Experimental Design

There were two main experimental conditions to be considered for the evaluation the localisation algorithm; The site which would provide the evaluation environment, and the equipment which would be used to simulate our Outdoor Industrial AR device.

5.1.1 Evaluation Site

The aim of the localisation algorithm was to determine pose in outdoor environments which are of interest for industrial applications. For this reason, a real outdoor environment of industrial interest was chosen, the Myer construction site in Hobart, Tasmania, shown as a panorama in Figure 5.1.

The original site was demolished after a fire in 2007, and since this incident the site has undergone archaeological exploration and is now awaiting development of a new shopping centre. This scenario has ideal characteris-

Figure 5.1: A panoramic view of the Myer construction site in Hobart, Tasmania.

tics for an Outdoor Industrial AR application, and the site itself had several attractive qualities as an evaluation site, for example the ease of access and favourable image matching characteristics in the form of large amounts of unique textures with very little repeating patterns. Furthermore, on site vegetation was limited to grass type shrubs covering sections at ground level, increasing the likelihood that ridge tracking points will be visible in both the captured image and 3D point cloud.

The evaluation site was positioned in an *urban canyon*, which ensured long lines of sight and a uniform spread of image key points. These factors lead to an improvement the system's ability to determine the pose, due to the equality of internal angles and distances providing a more robust solution when estimating pose using triangulation of the collinearity and coplanarity condition (Kraus, 2007).

To capture the entire site, colour point clouds were captured with a 4 millimetre tolerance using a Faro Focus 3D terrestrial laser scanner from five different locations, which were chosen to maximise the level of coverage whilst maintaining sufficient overlap to allow alignment of the point clouds. The five individual point clouds where initially aligned using a combination of software packages[38] on an arbitrary coordinate system before further alignment was conducted using historical scans to bring the point clouds onto the MGA datum. Figure 5.2a and Figure 5.2b show the combined point clouds on datum.

---

[38]Faro Scene was used to colourise the scans and export them to iSite for alignment.
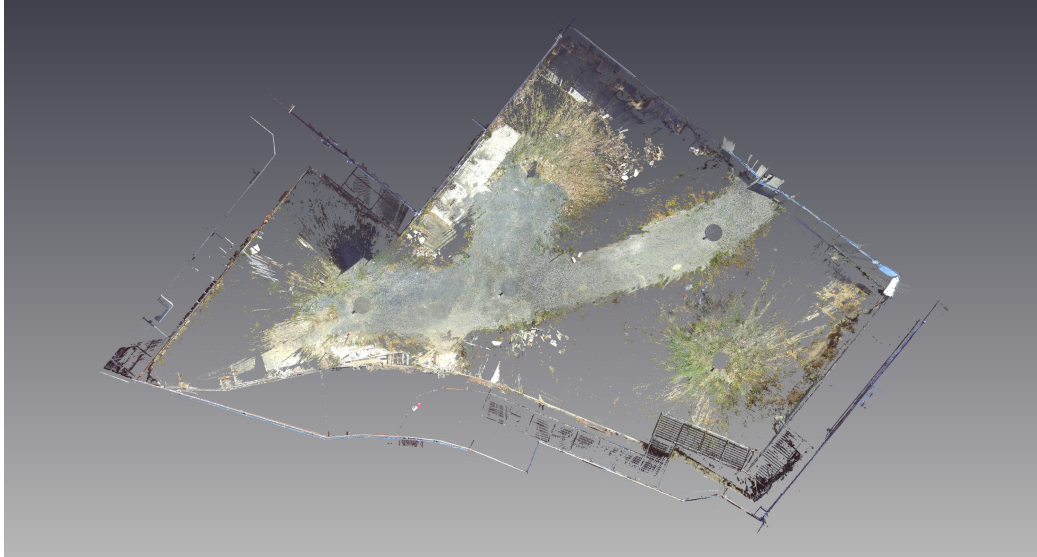
(a) Birds eye view of the evaluation site generated from colour point cloud data.
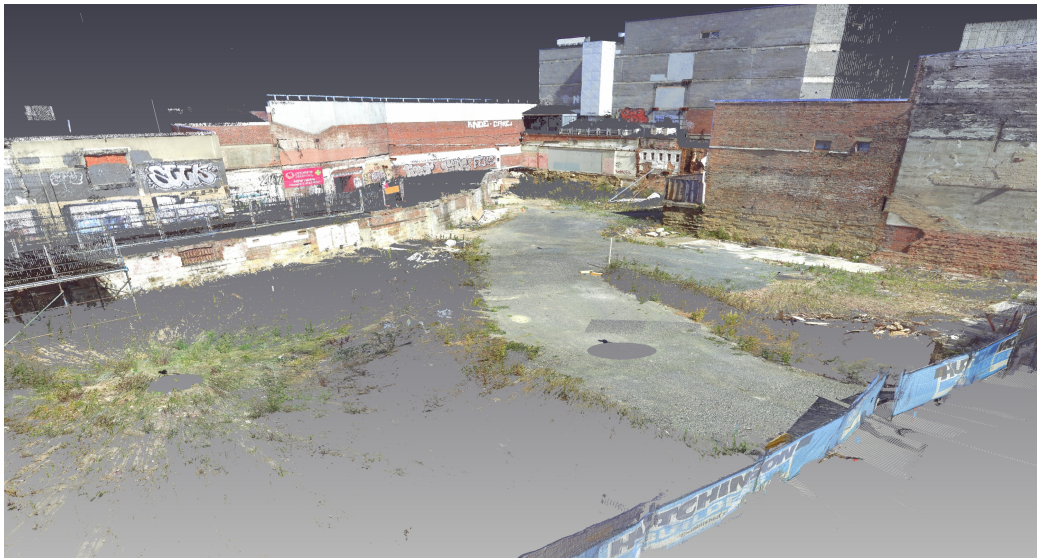
(b) Oblique image of the evaluation site generated from the colour point cloud data.

Figure 5.2: Overview renderings of the Myer test site.

5.1.2 Experimental Equipment

As described in section 1.1, the motivation of this research is to enable localisation on consumer level equipment, such as tablets and smart phones, which is accurate enough for industrial application. For this reason, the device chosen to simulate our Outdoor Industrial AR device was a Samsung Note 10.1, a consumer level tablet equipped with GNSS capabilities, compass, initial sensors and rear-facing camera.

As the localisation algorithm was implemented and tested on a PC, data was captured from the device and stored for later analysis. Images were sampled at a resolution of 640x480 pixels and a rate of approximately 8Hz, and position (from the GNSS device) and orientation (from combination of IMU and Compass) data was time-logged against the image observations with the aid of the Transform flow android data capture tool (Williams, 2013)[39]. Attached to the tablet was a Leica Viva GS15 GNSS device to capture highly accurate RTK values for position as a ground truth comparison [40]. This data was sampled at a rate of 20 Hz, which was then resampled to align with the tablet observations based on overlapping GPS times.

As RTK does not provide a measurement of orientation, it was not possible to record a ground truth orientation for the device, and as such the positional results were used as a primary means of evaluation of the localisation algorithm. In order to provide some degree of orientation evaluation, the orientation determined by the device's sensors was accepted as the ground truth, and a standard deviation of 15° was assumed for all observations (Blum et al., 2013; Williams, 2013). To remove the complexity of multi-axis rotation, rotation around each axis was evaluated independently.

Data was recorded from a walk around the site, changing the position and the orientation of the device. This data collection stage ran for approximately 4 minutes, and resulted in 1920 samples of images, low accuracy position and orientation measurements, and high accuracy position and orientation ground truth measurements.

5.2 Image to Point Cloud Matching

The first evaluation of the localisation process involved examining the effects of the virtual Field of View (FoV) and Match Filters on the performance of real and virtual image matching.

---

[39]https:///github.com/HITLabNZ /transform-flow-capture-android

[40]Standard accuracy of RTK determined positions are 10mm horizontal adn 20mm vertical.

Virtual images were generated from the point cloud from a view point computed using the position and orientation obtained in the ground truth measurement to ensure that the virtual and real image were as similar as possible, reducing the effect of confounding variables on the results. These artifical images were then matches against the real images captured from the device, using the processes described in chapter 4.

5.2.1   Field of View Computation

subsection 4.2.3 described the effect the virtual camera FoV has on the overlap of real and virtual images, and how this can affect image matching. To calculate this, image matching was conducted between a real image and virtual image rendered at the same position, while the FoV was modified, where a higher number of matches suggests a better FoV. To ensure a minimum amount of overlap between the real and virtual image, a minimum FoV of 70° was chosen for the virtual image rendering.

Due to the high sampling rate of data collection as described in subsection 5.1.2, consecutive observations were located in very close proximity. As such, only every 16th image was used in this test, as this sampled images from a range of different locations around the site.

Table 5.1 shows the number of successful image matches as the field of view increases. Although increasing the field of view increases the real and virtual image overlap, it did not increase the number of successfully matched image pairs. This is likely due to increased feature distortion in the virtual image, as features become more distorted as their distance increases from the image principal point. This problem could potentially be remedied using non-linear projection methods, such as a cylindrical projection approach that maintains conformality, however this was not tested.

Table 5.1 shows the highest number of matches occurred when the FoV of the virtual image (70°) was similar to the FoV of the real image (50°). As discussed in subsection 4.2.3, a lower value FoV for the virtual image would mean a reduction in overlap between the real and virtual images, lowering the localisations robustness to error in orientation computation. For these reasons, a FoV of 70° was used for later experiments.

5.2.2   Match Filtering and Pose Estimation

subsection 4.2.4 introduced four techniques to filter erroneous feature matches from the real to virtual image matches:

- Standard Deviation Filter - Outliers are detected using the correspond-

Table 5.1: Comparison of varying field of view on the ability of the system to determine sufficient descriptor matches between the real and virtual images.

Virtual Image FoV	Number of Real/Virtual Matches
70	76
80	63
90	52
100	20
110	13
120	6

ing distances between the best and second best matches. A match is considered an outlier if the distance of the second best match is within 0.5 standard deviations of the best match.

- Homography Filter - The Homography matrix relating feature points to their matches is calculated, and outliers are determined based on the distance between the homography projected key point and matched key point position.

- Fundamental Filter - The Fundamental matrix relating feature points to their matches is calculated, and outliers are determined based on the distance of the matched key point to the epipolar line projected from its corresponding keypoint.

- Ransac Filter - Ransac filtering uses an iterative approach to detect inliers based on the tolerance parameter and number of iterations completed.

Table 5.2 shows the mean number of keypoint matches remaining after each stage of the filtering process for accepted localisation solutions. Due to reasons described in subsection 4.2.4, the results are shown with and without the Homography filter applied. The filtering approaches are applied sequentially such that the Standard Deviation filter is applied before the Homography, Fundamental and RANdom SAmple Consensus (Ransac) filters. After filtering, any solution with an error greater than 1.5 standard deviations of the ground truth was rejected. To ensure that the inlier results were not affected by pose solutions which were completely wrong.

Table 5.2 shows the number of image match inliers remaining after each filtering stage before pose computation. Table 5.3 shows how many solutions

Table 5.2: Comparison of the number of inliers and outliers remaining after each stage of filtering. Only solutions within 1.5 standard deviations of the ground truth are shown.

Filtering Approach		St. Dev.	Filtering Stages H	F	Ransac	Image Solutions
4 Pass	inliers	110.09	38.70	23.72	20.46	376
	outliers	(89.63)	(18.24)	(3.26)	(0)	(1920)
3 Pass	inliers	110.09	-	26.01	13.20	148
	outliers	(96.89)	-	(12.81)	(0)	(1920)

were found within 1 and 1.5 standard deviations of the ground truth for various pose computation approaches after filtering. In general, the 4 pass filtering approach with Ransac PnP Iterative, Ransac EPnP and EPnP pose computation methods having the highest number of solutions within 1 and 1.5 standard deviations of the ground truth. Of these pose computation methods, Ransac PnP Iterative had the lowest average standard deviation of error, making it the most accurate localisation technique.

5.3 Accuracy of localisation in an ideal scenario

In section 5.2, the Image to Point Cloud matching performance of the localisation algorithm was evaluated. The results shown in Table 5.3 show the number of frames (out of 1920) that were sucessfully localised within 1 and 1.5 standard deviations of the ground truth position. In this section, the accuracy of this localisation is further examined, in terms of the positional and orientation accuracy.

For the evaluation, a 70° FoV was used for virtual camera, and the Ransac PnP Iterative pose estimation was implemented, for the reasons discussed in subsection 5.2.1 and subsection 5.2.2. For each of the 1920 samples, the virtual images were rendered using a viewpoint determined from the position of the RTK ground truth, and the orientation of the device sensor, resulting in a virtual image very similar to the real image. This ideal scenario would result in a localisation solution with no translational component, and a minor orientation component (to correct for error in the device sensor reading).

Table 5.3: Comparison of different pose computation approaches. A successful computation is based on horizontal correction vector being within n GPS standard deviations.

Pose Computation Approach	Filtering Approach	1 std	1.5 std	Ave Std.Dev. (m)
Ransac PnP Iterative	4 Pass	170	365	0.905
Ransac EPnP	4 Pass	165	376	2.269
PnP Iterative	4 Pass	70	171	2.376
EPnP	4 Pass	172	375	1.589
Ransac PnP Iterative	3 Pass	80	148	2.083
Ransac EPnP	3 Pass	78	143	0.953
PnP Iterative	3 Pass	16	27	1.693
EPnP	3 Pass	28	63	1.937

5.3.1 Position Analysis

In this section, the accuracy of the position localisation is analysed with respect to the RTK ground truth values and device sensor's estimates.

Figure 5.3a, Figure 5.3b and Figure 5.3c show the results for the X, Y and Z position estimations respectively. The blue line shows the localisation calculated position, the green dashed line shows the position according to the device sensors, and the red dashed line shows the ground-truth position.

Figure 5.3a and Figure 5.3b show that the device sensor readings (green line) are closer to the ground-truth (red line) than the localisation calculated position (blue line) for the X and Y position estimation. This is reflected in Table 5.4, which shows the mean and standard deviation of error in the localisation calculations, and Table 5.5, which shows the mean and standard deviation of error in the device sensor readings. For both X and Y axes, the standard deviation of error is much higher for the localisation calculated positions than the device sensor. This result suggests that, despite filtering, erroneous images matches are still being included in the localisation computation, and further effort should be made to remove them.

In comparison to the results for the X and Y axis localisation results, the Z axis localisation results show an improvement over the device sensor readings, as shown in Figure 5.3c. This suggests that the Z axis localisation estimation is less susceptible to false positive feature matches than the X and Y axis localisation estimation.
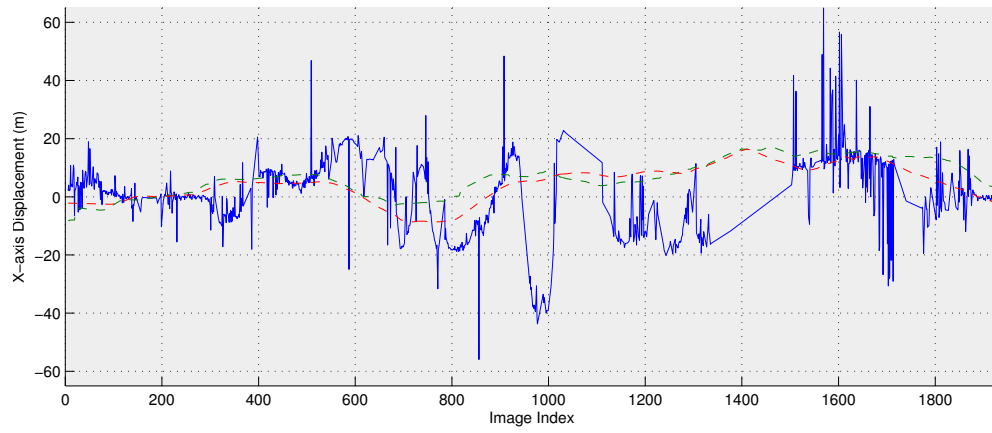
(a) X-Axis Displacement.
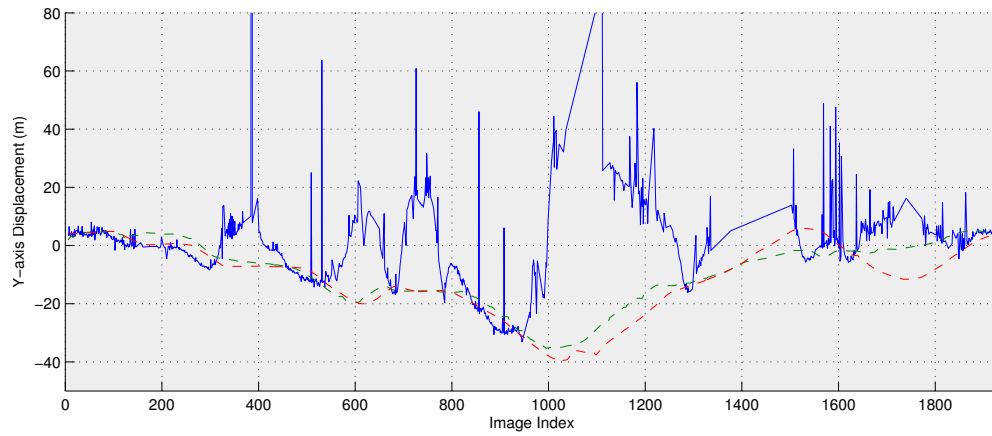
(b) Y-Axis Displacement.
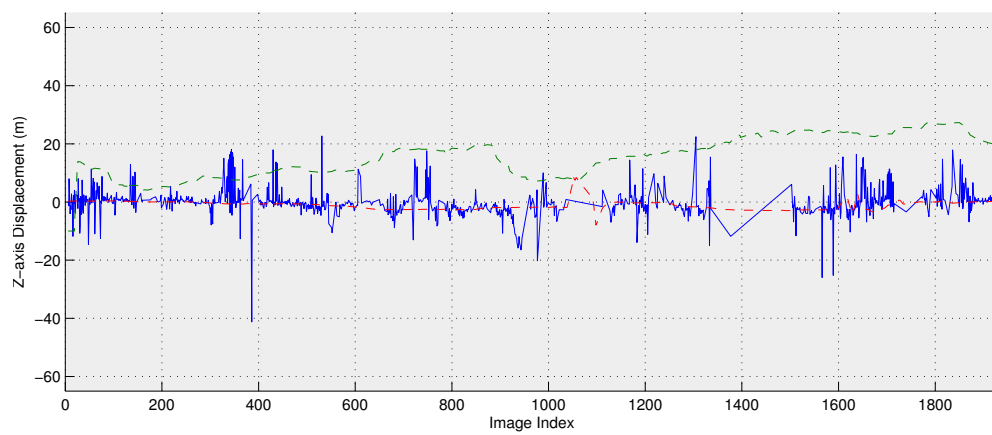
(c) Z-Axis Displacement.

Figure 5.3: Comparison of calculated position (blue), sensor position (green) and ground truth position (red).

Table 5.4: Accuracy and precision analysis of all localisation solutions with mean and standard deviation calculated with respect to the differences to ground truth values.

Component	Mean	Standard Deviation
Δ X-axis	2.369	12.752
Δ Y-axis	-9.246	23.381
Δ Z-axis	-0.743	4.779

Table 5.5: Accuracy and precision analysis of the device sensor positions with mean and standard deviation calculated with respect to the differences to ground truth values.

Component	Mean	Standard Deviation
Δ X-axis	2.551	2.981
Δ Y-axis	- 1.462	3.533
Δ Z-axis	-16.144	6.986

Filtered Position Analysis

The majority of AR capable consumer devices use consumer grade GNSS devices for localisation. In addition to position measurements, these GNSS devices also provide their error margins. From this, the true position can be computed as being $\pm$ 1 standard deviation from the measured position if the errors are normally distributed[41]. This information can be used to filter erroneous solutions by discarding solutions that are not within 1 standard deviation of the device sensor's position estimate.

Table 5.6 and Figure 5.4 show the improved accuracy and precision of filtering erroneous pose solutions by combined axis error. This approach improves the precision about all axes as well as improving the accuracy on the horizontal plane.

Although this additional filtering has improved the precision and accuracy, it has also reduced the number of computed localisations by 84.56%. This means that a higher emphasis is placed on a relative tracking solution

---

[41]This is not always the case with GNSS based observations as there a variety of factors that bias the error distribution like poor geometry, atmospheric errors, clock-bias, signal multi-path and orbital errors to name a few.

Table 5.6: Accuracy and precision analysis before and after filtering by combined axis error.

	All Solutions			Combined Axis Filter		
Comp.	Inliers	Mean	St. Dev.	Inliers	Mean	St. Dev.
Δ Vector	369	18.927	19.193	170	4.621	3.318
Δ X-axis	369	11.368	17.782	170	-1.946	2.515
Δ Y-axis	369	4.553	14.852	170	0.371	3.880
Δ Z-axis	369	1.614	6.102	170	1.180	2.401

Table 5.7: Accuracy and precision analysis of filtering by combined axis error and individual axis error.

	Combined Axis Filter			Individual Axis Filter		
Comp.	Inliers	Mean	St. Dev.	Inliers	Mean	St. Dev.
Δ Vector	170	4.621	3.318	170	4.621	3.318
Δ X-axis	170	-1.946	2.515	490	-0.345	1.860
Δ Y-axis	170	0.371	3.880	524	0.597	1.959
Δ Z-axis	170	1.180	2.401	898	-0.168	1.531

as it will need to be able to maintain tracking accuracy for a longer period of time before a new localisation correction is available.

To increase the number of successful localisation solutions, instead of filtering the solutions based on combined axis error, the solutions were filtered for each axis individually. This way, instead of removing the entire solution when a single axis is outside the error limit, only the value computed for that axis is removed.

Table 5.7 shows the results of the individual axis based filtering, showing a statistical improvement of precision in each axis as well as refining the accuracy in the x- and z-axis solutions. Furthermore, there is an increase in component based inliers by 288%, 308% and 528% respectfully about the x,y and z axis when compared with the combined axis filter.

Utilising an individual axis filtering and correction approach leverages the strengths of 3D geometry computation. Geometry impacts both the accuracy and precision of triangulation based solutions, especially when there are acute angles at the point of interest. This is often the case with image
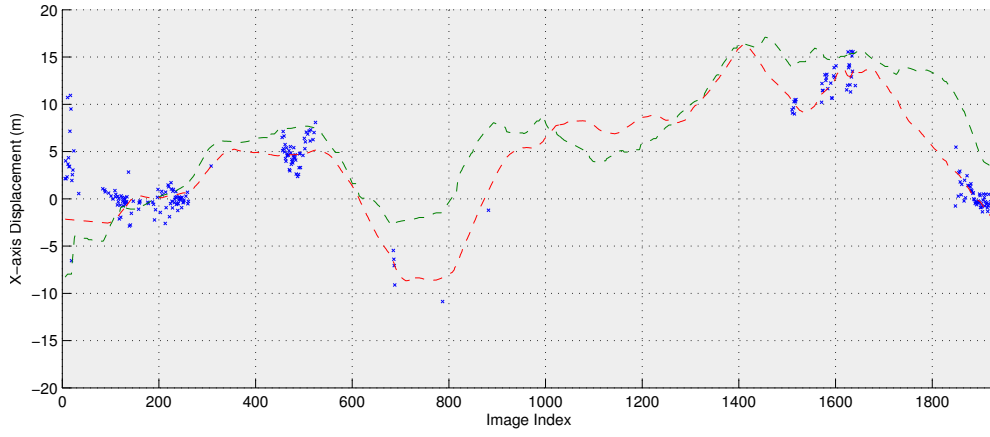
(a) X-axis displacement
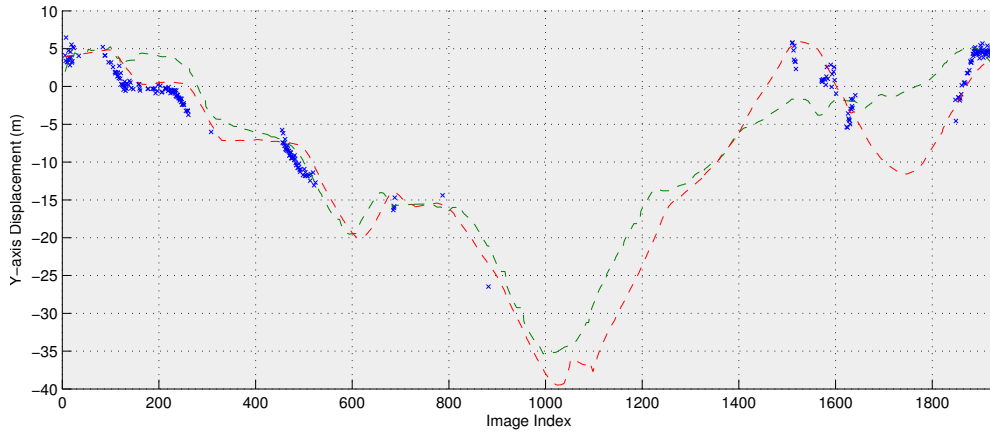
(b) Y-axis displacement
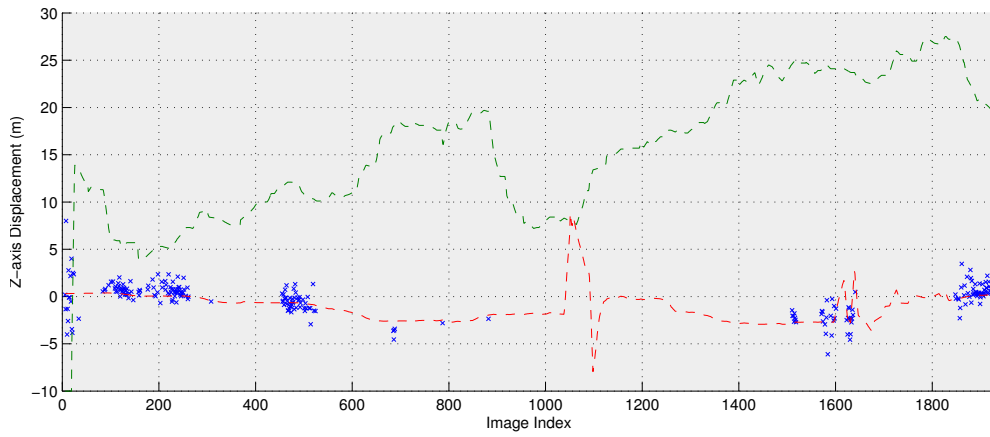
(c) Z-axis displacement

Figure 5.4: Localisation values after Combined Axis Error filtering (blue) compared to the device sensor estimate (green) and ground-truth (red)
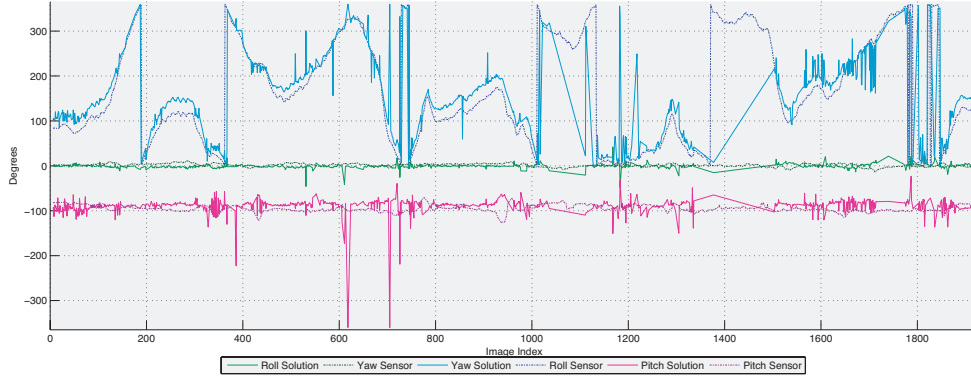
Figure 5.5: Comparison of sensor and solution based orientation components.

based solutions, as image matches are restricted to the image plane and consequently the internal angles from the projected rays cannot have an angle larger than the camera FoV. This results in reduced accuracy and precision along the camera's forward facing axis, conversely improving accuracy and precision in the camera's *Up* and *Right* axis. This is evident in Figure 5.3c, where the position solution is relatively stable when compared with the x- and y-axis solutions, due to the z-axis remaining mainly parallel to the image *Up* axis. Further verification of this can be found in Table 5.6 where the z-axis has been consistently more accurate than the initial position estimate regardless of outliers.

5.3.2 Orientation Analysis

As discussed in subsection 5.1.2, RTK was used as a ground truth to evaluate the localisation computed position. Unfortunately, RTK does not provide a orientation estimation, and thus there is no ground truth to quantitatively compare the localisation computed orientation too. To give an indication of the accuracy, Figure 5.5 show the results for the roll, pitch and yaw orientation calculations for the localisation compared to that measured by the device's sensors. The blue, magenta, and green lines represent the yaw, pitch and yaw for the localisation computation respectively, and the corresponding dashed lines are the corresponding device sensor orientations.

To provide a qualitative evaluation of the localisation orientation results, Figure 5.6 shows a set of six real images (1st column), the corresponding virtual images rendered using the orientation given by the device sensor (2nd column), and the virtual images an orientation correction factor calculated

through localisation has been applied (3rd column).

5.4 Accuracy of Localisation in a real-world scenario

In section 5.3, the localisation algorithm was tested using virtual images rendered using the RTK ground truth position and the device sensor orientation. This produced virtual images which were very similar to the real images, with an ideal localation solution with no translational component and a minor orientation component.

Consumer grade GNSS solutions are commonly used as the primary method of localising mobile AR applications, as opposed to the highly accurate RTK sensor used in section 5.3. To create a more realistic scenario, a further evaluation was run, this time using the device sensor position and device sensor orientation to compute the virtual image rendering viewpoint. By doing this, error is introduced in both the position and orientation of the rendering viewpoint, mimicking the noisy data the system would have to deal with in a real world scenario. The same parameters and ground truth were used in this example as in section 5.3.

5.4.1 Accuracy Assessment

Table 5.8 displays the accuracy of the localisation computation when the virtual image was rendered using the device sensors position and orientation after combined axis filtering based on device sensor error. Similar to results seen in the ideal scenario (Table 5.6), the filtering of solution for combined axes based on the error range of the device sensors improves the accuracy and standard deviation of the solution at the cost of reducing the number of localisation solutions. Table 5.9 shows the results with the individual axis filter, and as in Table 5.7, these results provide higher accuracy and precision than the device sensors position estimate (Table 5.5). These results show that the localisation algorithm can provide a position based correction value that will improve the device sensor position when correctly applied.

(a) Input Image 1   (b) Virtual Image 1   (c) Corrected Image 1

(d) Input Image 2   (e) Virtual Image 2   (f) Corrected Image 2

(g) Input Image 3   (h) Virtual Image 3   (i) Corrected Image 3

(j) Input Image 4   (k) Virtual Image 4   (l) Corrected Image 4

(m) Input Image 5   (n) Virtual Image 5   (o) Corrected Image 5

(p) Input Image 6   (q) Virtual Image 6   (r) Corrected Image 6
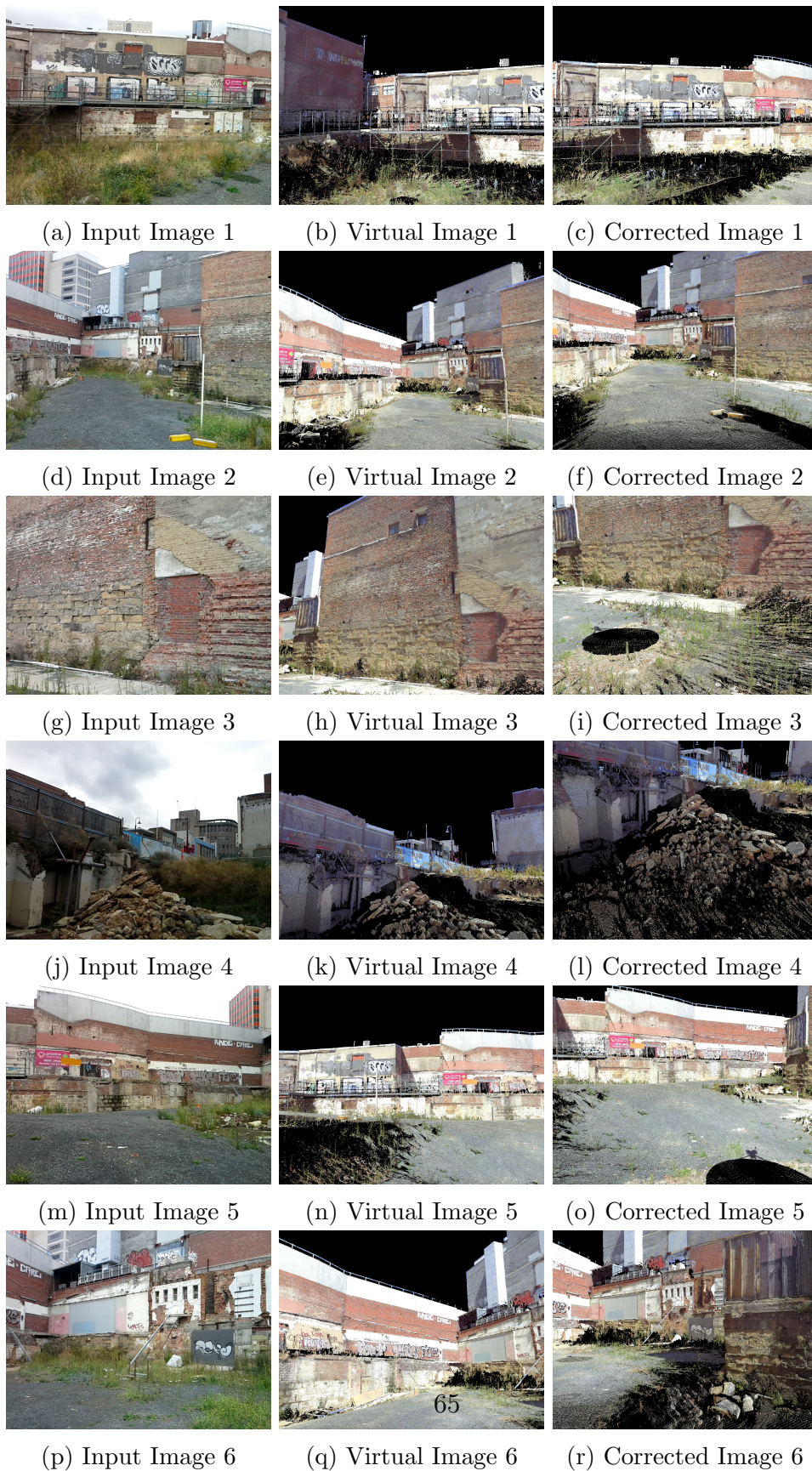
Figure 5.6: Example comparison of input, virtual and corrected images.

Table 5.8: Accuracy and precision analysis of Unfiltered and Combined Axis Filtered localisation solutions with rendering viewpoint set by consumer grade GNSS device and collection of IMUs.

	All Solutions			Combined Axis Filter		
Comp.	Inliers	Mean	St. Dev.	Inliers	Mean	St. Dev.
Δ Vector	369	18.927	19.193	63	2.866	0.866
Δ X-axis	369	11.368	17.782	63	0.696	1.454
Δ Y-axis	369	4.553	14.852	63	1.706	1.441
Δ Z-axis	369	1.614	6.102	63	0.449	1.085

Table 5.9: Accuracy and precision analysis of Combined Axis Filtered and Individual Axis Filtered localisation solutions with rendering viewpoint set by consumer grade GNSS device and collection of IMUs.

	Combined Axis Filter			Individual Axis Filter		
Comp.	Inliers	Mean	St. Dev.	Inliers	Mean	St. Dev.
Δ Vector	63	2.866	0.866	63	2.866	0.866
Δ X-axis	63	0.696	1.454	152	1.090	2.103
Δ Y-axis	63	1.706	1.441	152	0.863	1.943
Δ Z-axis	63	0.449	1.085	273	0.337	1.533

Chapter 6

Discussion

This chapter provides further interpretation of the results found in chapter 5.

6.1   Image to Point Cloud Matching

In section 5.2, two components of the Image to Point Cloud matching component of the localisation process were evaluated, the effect of FoV on image matches, and the performance of various match filters and Pose Estimators.

Increasing the FoV of the virtual image will increase the overlap of the real and virtual images (See subsection 4.2.3) resulting in increased robustness to orientation error, however it also increases feature distortion, reducing the number of feature matches. Experimentally, it was found that a virtual image FoV of 70° yielded the highest number of matches.

Four match filtering approaches were outlined in subsection 4.2.4, and evaluated in subsection 5.2.2. All of these filters reduced the number of outlier matches, at the cost of some inlier matches. Of most interest though is the effect of the Homography filter. Traditionally, a homography is used to describe the geometric relationship between planar surfaces. However, it was found in this research to have a significant effect on the number of correct image matches found in the dataset. It is speculated that this is because of the number of planar surfaces present in the environment (e.g. building faces). It is possible that, in the cases where image matches occurred across non-planar surfaces, this might reduce the robustness of the localisation system due to degrading the solution geometry, however no evidence was found to support this. In fact, it was found that the the 4 Pass filtering approach, which included the Homography based filter, generated the more inlier solutions then the when the Homography matrix was excluded.

The methods used to determine the pose also yielded difference accuracy

results. The OpenCV library provides functions that allow for the pose estimation as well as optimisation of said functions. A short analysis was run on the different pose estimations solutions available (See section 5.2) on a series of 120 images. From this analysis it was found that the Ransac based PnP Iterative solution generated the most number of solutions within 1 standard deviation of the ground truth solution, and had a lower average standard deviation of error.

6.2   Accuracy of Localisation

Although the Iterative Ransac PnP solution provided the highest level of accuracy, erroneous solutions were still determined by this method. As such the filtering system was employed to filter bad solutions by comparing the combined positional error to the error bounds of the devices low accuracy position and orientation sensors. It was found that this significantly reduced the number of erroneous solutions. Accuracy was further improved through by comparing individual axes to these error bounds, and only discarding individual erroneous axis measurements. In practice, this approach would only work if the input values were normally distributed.

The accuracy of the solution was unaffected when the error was normally distributed. Overall it was found that the accuracy of the localisation system increased so that it was on par with horizontal consumer grade GNSS solutions when erroneous solutions were removed by combined axis filtering. In addition, the individual axis filtered localisation results (See Table 5.7) were better than the device sensor's accuracy and precision by a magnitude of 8 and 4 respectively (See Table 5.9).

As previously mentioned in subsection 5.1.2, the device's positional sensor was a consumer grade GNSS device. As such the non-Gaussian error of the input value could be caused by a majority of factors, including a non-localised geoid model, which was used to convert from an ellipsoidal height to mean sea level or altitude. Although the vertical position is reported by the device manufacturer as altitude, it is still unknown as to the model and parameters used in the conversion from ellipsoidal height to mean sea level or altitude.

The device sensor data was captured in favourable conditions leading to a best case resolution in terms of accuracy. In less favourable environments, such as obstructed sky-view, the localisation server may become even more beneficial.

A much higher accuracy was expected from the localisation approach based on the associated data used in the point cloud (rms: 0.008 m) and the pixel ground resolution for the average projected rays (0.044 m). It is

68

unclear as to the cause of the large errors in the the calculated solution but an overall accuracy was estimated to be 1:50. There is potential that the low resolution input imagery coupled with poor pose computation input, caused by the majority of matched points falling within an acute angle, result in the poor localisation accuracy.

Despite this, the accuracy provided by the localisation process was still able to offer a improvement in accuracy over device sensor values (See Table 5.7). With improved filtering, it is believed the localisation system will be able to provide a more accurate solution than consumer grade GNSS devices.

Chapter 7

Conclusion

This research has presented a concept IAR framework for use in outdoor environments. This framework has been designed to follow recommendations as outlined by Fite-Georgel (2011) and Rankohi and Waugh (2013) to encourage adoption and use outside of academia.

As part of this IAR framework, a localisation module was successfully implemented which provides a novel localisation solution utilising RGB point-clouds, an overlooked source of data that is becoming readily more available for a majority of engineering, planning and design tasks. This localisation system used image-to-pointcloud correspondence to image match and compute the camera pose.

The localisation module's accuracy was assessed in an ideal scenario and found to provide an accuracy improvement over consumer-grade GNSS height without additional filtering of the pose solutions. With additional filtering of the pose solutions, improvements were seen across all axes. The localisation module was further evaluated in conditions designed to mimic the real-world operation of the module. This involved the application and update of the localisation correction value to an initial position estimate obtained from a low accuracy consumer GNSS device. It was found, that at this stage, the rudimentary pose solution filtering system was not robust in the detection of outlier solutions.

7.1   Future Research

Point clouds generated from computer vision approaches differ significantly from point clouds generated by LiDAR systems in areas specifically relating to density and descriptor tags. As such the majority of prior research on 2D-3D image-point cloud matching cannot be applied to LiDAR based point

clouds. There has been very little research into generating imagery from Li-DAR based point clouds for the purpose of image matching to date, with only one other work by Sibbing et al. (2013) focusing on this area. Therefore, this area of research remains largely unexplored, with many possible applications in improving the realism of the virtual imagery.

Initial attempts at using keypoint detection and description approaches that have been reported as previously highly robust (e.g. FREAK) were found to have a sub-par performance when used in with the virtual imagery generated by the localisation module. Further research in how these keypoint algorithms could be tuned to apply to spare data sets could be conducted.

Furthermore, due to the novel process of generating virtual imagery from 3D point clouds, additional research into different perspective projection approaches may improve the robustness of the system. Cylindrical projection approaches warrant particular interest as panoramic imagery could be generated with minimal distortion about the horizontal equator. This projection approach could be extended to utilising conformal projection approaches to maintain shape at the cost of scale.

The pose estimation functions provided by the OpenCV library allowed for some optimisation parameters in the form of allowable projection error, number of iterations and minimum inliers as well as providing an inlier list along with the calculated solution. By optimising these inputs, it is possible that a better pose estimation could be computed in the face of outlier points.

References

Ababsa, F. (2009a, October). Advanced 3D localization by fusing measurements from GPS, inertial and vision sensors. In *2009 ieee international conference on systems, man and cybernetics* (pp. 871–875). IEEE.

Ababsa, F. (2009b). Toward a Real-Time 3D Reconstruction System for Urban Scenes Using Georeferenced and Oriented Images. In *2009 second international conference on computer and electrical engineering* (Vol. 1, pp. 75–79). Dubai: IEEE.

Ababsa, F., Didier, J.-Y., Zendjebil, I., & Mallem, M. (2009). Marker Less Vision-Based Tracking of Partially Known 3D Scenes for Outdoor Augmented Reality Applications. In L. Bolc, J. Kulikowski, & K. Wojciechowski (Eds.), *Computer vision and graphics se - 22* (Vol. 5337, pp. 218–227). Springer Berlin Heidelberg.

Ababsa, F., Zendjebil, I., Didier, J.-Y., Pouderoux, J., & Vairon, J. (2012, July). Outdoor augmented reality system for geological applications. In *2012 ieee/asme international conference on advanced intelligent mechatronics (aim)* (pp. 416–421). IEEE.

Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., & Hopper, A. (2001). Implementing a sentient computing system. *Computer*, *34*.

Alahi, A., Ortiz, R., & Vandergheynst, P. (2012, June). FREAK: Fast Retina Keypoint. In *2012 ieee conference on computer vision and pattern recognition* (pp. 510–517). IEEE.

Anagnostou, K., & Vlamos, P. (2011, May). Square AR: Using Augmented Reality for Urban Planning. In *2011 third international conference on games and virtual worlds for serious applications* (pp. 128–131). IEEE.

Arth, C., Klopschitz, M., Reitmayr, G., & Schmalstieg, D. (2011, October). Real-time self-localization from panoramic images on mobile devices. In *2011 10th ieee international symposium on mixed and augmented reality* (pp. 37–46). Basel: IEEE.

Arth, C., Wagner, D., Klopschitz, M., Irschara, A., & Schmalstieg, D. (2009, October). Wide area localization on mobile phones. In *2009 8th ieee*

international symposium on mixed and augmented reality (pp. 73–82). IEEE.

Azuma, R. (1997). A Survey of Augmented Reality. *Media, 6*(4), 355–385.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Baggio, D. L., Emami, S., Escriva, D. M., Ievgen, K., Mahmood, N., Saragih, J., & Shilkrot, R. (2012, March). Exploring Structure from Motion with OpenCV. In *Mastering opencv with practical computer vision projects* (Vol. 11, pp. 129–160). Birmingham, UK: Packt Publishing.

Bau, O., & Poupyrev, I. (2012, July). REVEL. *ACM Transactions on Graphics, 31*(4), 1–11.

Bay, H., Tuytelaars, T., & Gool, L. (2006). SURF: Speeded Up Robust Features. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *Computer vision eccv 2006 se - 32* (Vol. 3951, pp. 404–417). Springer Berlin Heidelberg.

Behringer, R., & Sundareswaran, V. (2002). Model-based visual tracking for outdoor augmented reality applications. In *Proceedings. international symposium on mixed and augmented reality* (pp. 277–322). IEEE Comput. Soc.

Behzadan, A., & Kamat, V. (2005). Visualization of Construction Graphics in Outdoor Augmented Reality. In *Proceedings of the winter simulation conference, 2005.* (pp. 1914–1920). IEEE.

Behzadan, A. H., & Kamat, V. R. (2007, July). Georeferenced Registration of Construction Graphics in Mobile Outdoor Augmented Reality. *Journal of Computing in Civil Engineering, 21*(4), 247–258.

Behzadan, A. H., & Kamat, V. R. (2008, December). Simulation and visualization of traffic operations in Augmented Reality for improved planning and design of road construction projects. In *2008 winter simulation conference* (pp. 2447–2454). IEEE.

Besbes, B., Collette, S. N., Tamaazousti, M., Bourgeois, S., & Gay-Bellile, V. (2012). An interactive Augmented Reality system: A prototype for industrial maintenance training applications. In *Mixed and augmented reality (ismar) 2012, ieee international symposium on* (pp. 269–270).

Bleeker, T., Lee, G., & Billinghurst, M. (2013). *Ego- and Exocentric interaction for mobile AR conferencing.*

Blum, J., Greencorn, D., & Cooperstock, J. (2013). Smartphone Sensor Reliability for Augmented Reality Applications. In K. Zheng, M. Li, & H. Jiang (Eds.), *Mobile and ubiquitous systems: Computing, networking, and services se - 11* (Vol. 120, pp. 127–138). Springer Berlin Heidelberg.

Bostanci, E., Clark, A. F., & Kanwal, N. (2012, July). Vision-based user tracking for outdoor augmented reality. In *2012 ieee symposium on computers and communications (iscc)* (pp. 000566–000568). IEEE.

Botsch, M., Hornung, A., Zwicker, M., & Kobbelt, L. (2005). High-quality surface splatting on today's GPUs. In *Proceedings eurographics/ieee vgtc symposium point-based graphics, 2005.* (pp. 17–141). IEEE.

Carmigniani, J., & Furht, B. (2011). Augmented Reality: An Overview. In B. Furht (Ed.), *Handbook of augmented reality* (pp. 3–47). New York, USA: Springer New York.

Caudell, T., & Mizell, D. (1992). Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the twenty-fifth hawaii international conference on system sciences* (Vol. ii, pp. 659–669 vol.2). IEEE.

Elfick, M. H., Fryer, J. G., Brinker, R. C., & Wolf, P. R. (1987). Theory of Measurment and Errors. In *Elementary surveying (si adaptation)* (7th ed., pp. 13–38). Artarmon: Harper & Row.

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In *First international symposium on wearable computers digest of papers* (pp. 74–81). IEEE Comput. Soc.

Fite-Georgel, P. (2011, October). Is there a reality in Industrial Augmented Reality? In *2011 10th ieee international symposium on mixed and augmented reality* (pp. 201–210). IEEE.

Fong, W. T., Ong, S. K., & Nee, A. Y. C. (2008, September). A differential GPS carrier phase technique for precision outdoor AR tracking. In *2008 7th ieee/acm international symposium on mixed and augmented reality* (pp. 25–28). IEEE.

Gauglitz, S., Sweeney, C., Ventura, J., Turk, M., & Hollerer, T. (2012). Live tracking and mapping from both general and rotation-only camera motion. In *Mixed and augmented reality (ismar) 2012, ieee international symposium on* (pp. 13–22). Atlanta, Georgoa, USA: IEEE.

Geoscience Australia. (2013). *World Geodetic System 84.*

Gleue, T., & Dähne, P. (2001). Design and implementation of a mobile device for outdoor augmented reality in the archeoguide project. In *Proceedings of the 2001 conference on virtual reality, archeology, and cultural heritage - vast '01* (pp. 161–168). New York, New York, USA: ACM Press.

Google Inc. (2014). *Android Geomagnetic Field API.*

Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision* (2nd ed.). New York, USA: Cambridge University Press.

Irschara, A., Zach, C., & Bischof, H. (2007). Towards Wiki-based Dense

City Modeling. In *2007 ieee 11th international conference on computer vision* (pp. 1–8). IEEE.

Junghanns, S., Schall, G., & Schmalstieg, D. (2008). VIDENTE: What lies beneath? A new approach of locating and identifying buried utility assets on site. In *Proceedings of the 5th international symposium on lbs & telecartography (lbs08)* (p. 3). Salzburg: LBS.

Karlekar, J., Zhou, S. Z., Lu, W., Loh, Z. C., Nakayama, Y., & Hii, D. (2010, October). Positioning, tracking and mapping for outdoor augmentation. In *2010 ieee international symposium on mixed and augmented reality* (pp. 175–184). IEEE.

Karlekar, J., Zhou, S. Z., Nakayama, Y., Lu, W., Loh, Z. C., & Hii, D. (2010, July). Model-based localization and drift-free user tracking for outdoor augmented reality. In *2010 ieee international conference on multimedia and expo* (pp. 1178–1183). IEEE.

Kato, H., Tachibana, K., Tanabe, M., Nakajima, T., & Fukuda, Y. (2003). A city-planning system based on augmented reality with a tangible interface. In *The second ieee and acm international symposium on mixed and augmented reality, 2003. proceedings.* (pp. 340–341). IEEE Comput. Soc.

Kleinert, M., & Stilla, U. (2012, November). On sensor pose parameterization for inertial aided visual SLAM. In *2012 international conference on indoor positioning and indoor navigation (ipin)* (pp. 1–9). IEEE.

Kraus, K. (2007). *Photogrammetry - Geometry from Images and Laser Scans* (Second ed.). New York, USA: Walter de Gruyter.

Krevelen, D. V., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *The International Journal of Virtual Reality*, *9*(2), 1–20.

Langlotz, T., Degendorfer, C., Mulloni, A., Schall, G., Reitmayr, G., & Schmalstieg, D. (2011, August). Robust detection and tracking of annotations for outdoor augmented reality browsing. *Computers & Graphics*, *35*(4), 831–840.

Lee, D., Kim, H., & Myung, H. (2012, November). GPU-based real-time RGB-D 3D SLAM. In *2012 9th international conference on ubiquitous robots and ambient intelligence (urai)* (pp. 46–48). IEEE.

Livingston, M. A., & Ai, Z. (2008). The effect of registration error on tracking distant augmented objects. In *Mixed and augmented reality, 2008. ismar 2008. 7th ieee/acm international symposium on* (pp. 77–86). Cambridge.

Lonsing, W. (2011, October). Architectural models in urban landscapes: Synthesis of markers and virtual structures. In *Mixed and augmented reality (ismar) 2011, ieee international symposium on* (pp. 109–110).

IEEE.

Lopez, H., Navarro, A., & Relano, J. (2010, September). An Analysis of Augmented Reality Systems. In *2010 fifth international multi-conference on computing in the global information technology* (pp. 245–250). Valencia, Spain: IEEE.

Lorenz Wendt, F., Bres, S., Tellez, B., & Laurini, R. (2008). Markerless Outdoor Localisation Based on SIFT Descriptors for Mobile Applications. In A. Elmoataz, O. Lezoray, F. Nouboud, & D. Mammass (Eds.), *Image and signal processing se - 50* (Vol. 5099, pp. 439–446). Heidelberg: Springer Berlin Heidelberg.

Love, P. E. D., & Li, H. (2000, June). Quantifying the causes and costs of rework in construction. *Construction Management and Economics*, *18*(4), 479–490.

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, *E77-D*(12), 1321–1329.

Min, S., Lei, L., Wei, H., & Xiang, R. (2012, December). Interactive registration for Augmented Reality GIS. In *2012 international conference on computer vision in remote sensing* (pp. 246–251). IEEE.

Mohring, M., Lessig, C., & Bimber, O. (2004). Video See-Through AR on Consumer Cell-Phones. In *Third ieee and acm international symposium on mixed and augmented reality* (pp. 252–253). IEEE.

Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *In visapp international conference on computer vision theory and applications* (pp. 331–340).

Mulloni, A., Grubert, J., Seichter, H., Langlotz, T., Grasset, R., Reitmayr, G., & Schmalstieg, D. (2012). Experiences with the Impact of Tracking Technology in Mobile Augmented Reality Evaluations. In *Mobilehci 2012* (p. 4). San Francisco: ACM.

Newcombe, R. A., Lovegrove, S. J., & Davison, A. J. (2011, November). DTAM: Dense tracking and mapping in real-time. In *2011 international conference on computer vision* (pp. 2320–2327). Barcelona: IEEE.

Ong, S. K., Yuan, M. L., & Nee, A. Y. C. (2008, May). Augmented reality applications in manufacturing: a survey. *International Journal of Production Research*, *46*(10), 2707–2742.

Oskiper, T., Samarasekera, S., & Kumar, R. (2012, November). Multi-sensor navigation algorithm using monocular camera, IMU and GPS for large scale augmented reality. In *Mixed and augmented reality (ismar) 2012, ieee international symposium on* (pp. 71–80). IEEE.

Park, J. J., & Lee, D. (2012, November). Digital map based pose improvement for outdoor Augmented Reality. In *Mixed and augmented reality*

(ismar) 2012, ieee international symposium on (pp. 309–310). IEEE.

Pustka, D., Hulss, J., Willneff, J., Pankratz, F., Huber, M., & Klinker, G. (2012). Optical outside-in tracking using unmodified mobile phones. In *Mixed and augmented reality (ismar) 2012, ieee international symposium on* (pp. 81–89).

Rankohi, S., & Waugh, L. (2013, August). Review and analysis of augmented reality literature for construction industry. *Visualization in Engineering*, *1*(1), 9.

Reitmayr, G., & Drummond, T. (2006, October). Going out: robust model-based tracking for outdoor augmented reality. In *Mixed and augmented reality (ismar) 2006, ieee/acm international symposium on* (pp. 109–118). IEEE.

Ribo, M., Lang, P., Ganster, H., Brandner, M., Stock, C., & Pinz, A. (2002, November). Hybrid tracking for outdoor augmented reality applications. *IEEE Computer Graphics and Applications*, *22*(6), 54–63.

Rossler, T., Rogge, S., & Hentschel, C. (2011, September). A case study: Mobile augmented reality system for visualization of large buildings. In *2011 ieee international conference on consumer electronics -berlin (icce-berlin)* (pp. 311–314). IEEE.

Satoh, K., Anabuki, M., Yamamoto, H., & Tamura, H. (2001). A hybrid registration method for outdoor augmented reality. In *Augmented reality (isar) 2001, proceedings ieee and acm international symposium on* (pp. 67–76).

Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., & Schmalstieg, D. (2008, June). Handheld Augmented Reality for underground infrastructure visualization. *Personal and Ubiquitous Computing*, *13*(4), 281–291.

Schall, G., Wagner, D., Reitmayr, G., Taichmann, E., Wieser, M., Schmalstieg, D., & Hofmann-Wellenhof, B. (2009, October). Global pose estimation using multi-sensor fusion for outdoor Augmented Reality. In *2009 8th ieee international symposium on mixed and augmented reality* (pp. 153–162). IEEE.

Shepard, F. A. (1983). Engineering Surveying problems and solutions. In *Engineering surveying* (2nd ed., p. 21). Edward Arnold.

Sibbing, D., Sattler, T., Leibe, B., & Kobbelt, L. (2013, June). SIFT-Realistic Rendering. In *2013 international conference on 3d vision* (pp. 56–63). IEEE.

Snavely, N., Seitz, S. M., & Szeliski, R. (2006, July). Photo tourism. *ACM Transactions on Graphics*, *25*(3), 835.

Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Fall joint computer conference* (pp. 757–764).

Takacs, G., El Choubassi, M., Wu, Y., & Kozintsev, I. (2011, July). 3D mobile augmented reality in urban scenes. In *2011 ieee international conference on multimedia and expo* (pp. 1–4). IEEE.

Ventura, J., & Hollerer, T. (2012, November). Wide-area scene mapping for mobile visual tracking. In *Mixed and augmented reality (ismar) 2012, ieee international symposium on* (pp. 3–12). Atlanta, Georgoa, USA: IEEE.

Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., & Schmalstieg, D. (2008, September). Pose tracking from natural features on mobile phones. In *Mixed and augmented reality (ismar) 2008, ieee/acm international symposium on* (pp. 125–134). IEEE.

Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., & Schmalstieg, D. (2010). Real-time detection and tracking for augmented reality on mobile phones. *IEEE transactions on visualization and computer graphics*, *16*(3), 355–68.

Want, R., Hopper, A., Falcão, V., & Gibbons, J. (1992). *The active badge location system* (Vol. 10).

Williams, S. (2013). *Real-Time Hybrid Tracking for Outdoor Augmented Reality* (Masters Thesis). University of Canterbury.

Williams, S., Green, R., & Billinghurst, M. (2013). Hybrid Tracking using Gravity Aligned Edges. In *Computer-human interaction new zealand.* Christchurch.

Yi-bo, L., Shao-peng, K., Zhi-hua, Q., & Qiong, Z. (2008, October). Development Actuality and Application of Registration Technology in Augmented Reality. In *2008 international symposium on computational intelligence and design* (Vol. 2, pp. 69–74). Wuhan: IEEE.

Zendjebil, I., Ababsa, F., Didier, J.-Y., & Mallem, M. (2008, November). Hybrid Localization System for Mobile Outdoor Augmented Reality Applications. In *2008 first workshops on image processing theory, tools and applications* (pp. 1–6). IEEE.

Zhu, Z., Oskiper, T., Samarasekera, S., Kumar, R., & Sawhney, H. S. (2008, June). Real-time global localization with a pre-built visual landmark database. In *2008 ieee conference on computer vision and pattern recognition* (pp. 1–8). Anchorage, AK: IEEE.

Zwicker, M., Pfister, H., van Baar, J., & Gross, M. (2001). Surface splatting. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques - siggraph '01* (pp. 371–378). New York, USA: ACM Press.