SENSOR-ENHANCED IMAGING

BY AIEAT ASSAM

A THESIS SUBMITTED TO THE
UNIVERSITY OF BIRMINGHAM
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

School of Electronic, Electrical
and Computer Engineering
University of Birmingham
December 2012

ABSTRACT

Most approaches to spatial image management involve GPS or image processing. In this thesis, a sensor-focused alternative is explored. It requires user and camera tracking, particularly challenging in indoor environments.

Possible indoor tracking methods are evaluated and pedestrian dead reckoning is selected. A study is conducted to evaluate sensors and choose a combination for pedestrian and camera tracking. Gyroscope and accelerometer offer comparable step detection performance, with gyroscope and tilt compensated compass providing heading data.

Images taken from the same viewpoint are successfully arranged using panorama stitching without any image processing. The results compare favourably to conventional methods. While lacking visual definition of image processing methods, they can complement them if used in tandem.

Sensor compositing and pedestrian tracking are implemented in a unified system. Several methods for fusing compass and gyroscope data are compared, but do not produce statistically significant improvement over using just the compass. The system achieves loop closure accuracy of 91% of path length and performs consistently across multiple participants.

The final system can be used in GPS-denied locations and presents an image content independent way of managing photographs. It contributes to pedestrian tracking and image composting fields and has potential commercial uses (illustrated by an example Android app).

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who made this work possible. First and foremost I would like to thank my supervisor Professor Chris Baber for his support, guidance and help in keeping me on track throughout my research. This work would not have been possible without him.

I would like to thank staff and students of the University of Birmingham for their help and participation in my studies, especially members of the HIT team who had to endure these more often than anybody else.

I am also indebted to current and past staff of Scyron, especially Dr James Cross, for their support and advice.

And lastly I would like to show my heartfelt gratitude to my family, friends and close ones for helping me stay sane (to a degree, some might argue) through the past few years.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF ABBREVIATIONS

ADC – Analogue-to-Digital Converter

ANOVA – Analysis of Variance

ARM – Microprocessor architecture

CCA – Canonical Correlation Analysis

CPU – Central Processing Unit

CSV – Comma-Separated Values file format

DHCP – Dynamic Host Configuration Protocol used for automatic configuration of IP networks

DSLR – Digital Single-Lens Reflex, a camera type

DWT – Discrete Wavelet Transform

EKF – Extended Kalman Filter

ELAN – Multimedia annotation application

EM – Electro-Magnetic

EMX – .NET Micro Framework-based processing module by GHI Electronics

EXIF – Exchangeable Image File format

FEZ – Family of .NET-based development systems by GHI Electronics

FFT – Fast Fourier Transform

FIM – Frequent Itemset Mining

FIR – Finite Impulse Response

FM – Frequency Modulation

GB – Gigabyte

GPS – Global Positioning System

GWAP – Games-with-a-Purpose

HD – High Definition

HDR – High Dynamic Range

HMM – Hidden Markov Model

HOG – Histogram of Oriented Gradients

ICE – Microsoft Image Composite Editor

ID - Identification

IMU – Inertial measurement Unit

ISO – In the context of image capture a speed rating system for measuring the film speed. Adapted to classify sensitivity of digital imaging sensor

JEITA – Japan Electronics and Information Technology Industries Association

JPEG – Joint Photographic Experts Group image file format

LED – Light Emitting Diode

LIDAR – Light Detection and Ranging

LV – Low Voltage

MATLAB – Matrix Laboratory, a high level language and interactive environment for computationally intensive tasks

MEMS – Micro-Electro-Mechanical Systems

MHz – Mega Hertz

NCC – Normalised Cross Correlation

NMEA – National Marine Electronics Association

NTP – Network Time Protocol

OS – Operating System

PC – Personal Computer

PCA – Principal Component Analysis

PDR – Pedestrian Dead Reckoning

PSPP – Statistical analysis software package

RANSAC – Random Sample Consensus

RF – Radio Frequency

RFID – Radio Frequency Identification

RSS – Radio Signal Strength

RSSI – Receive Signal Strength Indication

SAMME – Stage-wise Addictive Modelling with Exponential Loss

SI – System International

SIFT – Scale-Invariant Feature Transform

SIMS – Spatial Image Management System

SLAM – Simultaneous Localisation and Mapping

SR – Super Resolution

SURF – Speeded Up Robust Feature

SVD – Singular Value Decomposition

TFT – Thin-Film Transistor

USA – United States of America

UI – User Interface

UK – United Kingdom

USB – Universal Serial Bus

WLAN – Wireless Local Area Network

WYSIWYG – What you see is what you get

XVGA – Extended Video Graphics Array

1. INTRODUCTION

This research aims to answer the question of how effectively a wearable indoor image spatial arrangement system would work and which sensors would be best for it. There are a number of ways to perform indoor localisation (many of which are reviewed in chapter 2), but not all of them are appropriate for use by people, as opposed to the robot platforms for which many of them are created. Therefore, the choice of sensors will affect the operation of the system and marks an important decision that requires investigating.

Another major issue that this research aims to bring to light is how to handle images taken from the same physical location. When photographing, people might take several photographs, perhaps at different magnification levels. Geo-spatially these photographs are in the same location and would therefore get superimposed over each other when used by conventional photograph-to-map mapping techniques. These photographs, however, could contain complementary information and could form combined information that is not fully available on either of them individually. Investigating ways of fusing the photographs to maximise such information is another aspect of this research.

1.1. RESEARCH QUESTIONS AND CONSTRAINTS

In order to complete the goals of this research, the research questions will need to be answered and prototypes constructed and evaluated in their effectiveness for both localising the photographer and arranging images taken at the same physical location.

The research questions can be summarised in the following form:

A. How can the location at which images are captured indoors be tracked and used to enhance spatial image management?

It is important to distinguish between localisation, positioning and tracking when looking into this topic. The research does not aim to provide absolute coordinates of

the location of capture. Instead, it uses tracking to determine relative positions of capture locations, hence enabling spatially arranging them on the same coordinate plane (non-absolute)

B. How well does sensor-based image compositing compare with visual methods and how can it help spatial arrangement of images?

Image compositing enables arranging images taken from the same viewpoint and has the potential to enhance tracking-based image arrangement from question A. For this research question to be fully explored a way to enhance location-based spatial management has to be devised and its utility investigated.

Both of these topics depend upon a choice of sensors for implementation and are limited by several constraints:

(1) ***A priori* knowledge vs**. **Flexibility** – increasing the amount of *a priori* knowledge about the environment can facilitate the task of tracking, but it also reduces the flexibility of the system. Re-deploying to a new location would require introduction of this data into the system, whereas reduction of the *a priori* knowledge requirement can minimise the number of preparation required before the system can be used. The balance between the two will need to be achieved, ideally favouring flexibility.

(2) **Utility vs. Aesthetics** – while aesthetics of implementation and results produced are an important factor, they have to be weighed against their utility. Greater amount of image data with visible seams in the compositing will have more utility than less complete, but more seamlessly blended alternatives, but may be less visually appealing. This research will focus on the utility aspect, while keeping the aesthetic aspect secondary.

1.2. The Rise of Digital Photography

Digital photography and image capture have been increasing in popularity over the years (GfK Group, 2011). With the arrival of low cost digital camera and cameras on mobile phones (Ito & Okabe, 2003), many households possess image capture capability of some kind, enabling people to generate an enormous amount of image content on the go.

Dedicated digital imaging devices (cameras or phones) are replacing their old analogue counterparts due to their ability to provide instant preview of the capture result and convenience of use, no longer requiring chemical processing of the film and developing of the snapshots. Any person with a personal computer and a printer can nowadays produce simple photographic prints at home. Moreover, some printers allow photographic printing even without the PC, connecting directly to the digital camera or its storage device (memory card) (Camera & Imaging Products Association, 2003).

When higher quality prints are required, these can be conveniently ordered online, without the need to send the camera film for development, utilising instead the quick and efficient online upload facilities.

Digital cameras come in a variety of designs and capabilities. Simple point-and-shoot cameras do not require any special knowledge to use, opting to automate the entire image acquisition process for convenience. They do not require special training or understanding to operate, utilising single button for the entire process and often providing additional magnification capabilities for longer range capture.

More advanced cameras allow optimisation of the imaging process by manually adjusting parameters such as aperture, exposure and sensor sensitivity, thus enabling imaging with a specific depth of field and artistic imaging, such as long exposure night time photographs. High end DSLR (Digital Single Lens Reflex) cameras provide professional photographers

with the capability of producing very high quality photographs, replacing their analogue predecessors in most fields.

It is becoming increasingly difficult nowadays to find a mobile phone that does not include photo capture ability of some kind. While image quality and capture features of such devices can vary a lot between models, they continue to generate ever increasing amounts of image data, often providing convenient ability to upload these directly to social networks via mobile network connection.

1.3. Location-Based Imaging

The rise in popularity of online image sharing sites and social networks has also affected the uptake of digital photography. Whereas conventional photographs need to be scanned before they can be uploaded online, digital ones can be shared immediately (Counts & Fellheimer, 2004). Sites such as Facebook (Facebook, 2011), Flickr (Yahoo! Inc, 2011) and Tumblr (Tumblr, Inc., 2012) focus on sharing images and contain a significant amount of photographs uploaded and shared by their users (Van House, 2007). Photography forms an important part of the tourism experience (Larsen, 2005), ensuring an ever increasing amount of photographic data available on- and offline.

It becomes increasingly important to make it easier for people to locate a specific image among the vast libraries of photographic content available (Bates, 1998). With the increase in popularity of photographic sharing and availability of the positional sensors, such as GPS (Global Positioning System) that come built into many modern smartphones, the practice of embedding additional information inside the photograph metadata has become more common and widely researched (Davis et al., 2005; Aurnhammer et al., 2006). Geo-tagging, for example, involves embedding GPS coordinates into the photographs at point of capture, allowing grouping by location and displaying them overlaid on map, an ability provided by some image sharing sites (Yahoo! Inc, 2011). This allows easily locating specific images by

their proximity to a given location, geo-clustering spatially adjacent images. An example of the simplicity of use of geo-tagging (an image created using a modern smartphone (HTC Corporation, 2011) and then visualised with Picasa (Google Inc, 2013)) is shown on Figure 1. The image contains Chirk Castle in Wales (National Trust, 2011) and has been placed automatically on the map using the embedded GPS coordinates imprinted into the image by the smartphone.

Another example of use of geo-tagging is made available by The Guardian (Guardian News and Media Limited, 2011). It offers people an overview of the London and UK riots of summer 2011 by placing user-submitted photographs on an interactive map showing time-stamped and geo-positioned development of the disturbances via imagery and other sources. The metadata of the photographs allows generation of a geographical timeline of events and thus exposes more information than just the image content alone.

While the sharing of location information alongside the photograph raises privacy concerns (Friedl & Sommer, 2010), it also facilitates collaboration (Chippendale et al., 2009) and social interaction – people can search for photographs taken near a location of interest, for example, before visiting it in person, thus helping them plan their journey better.

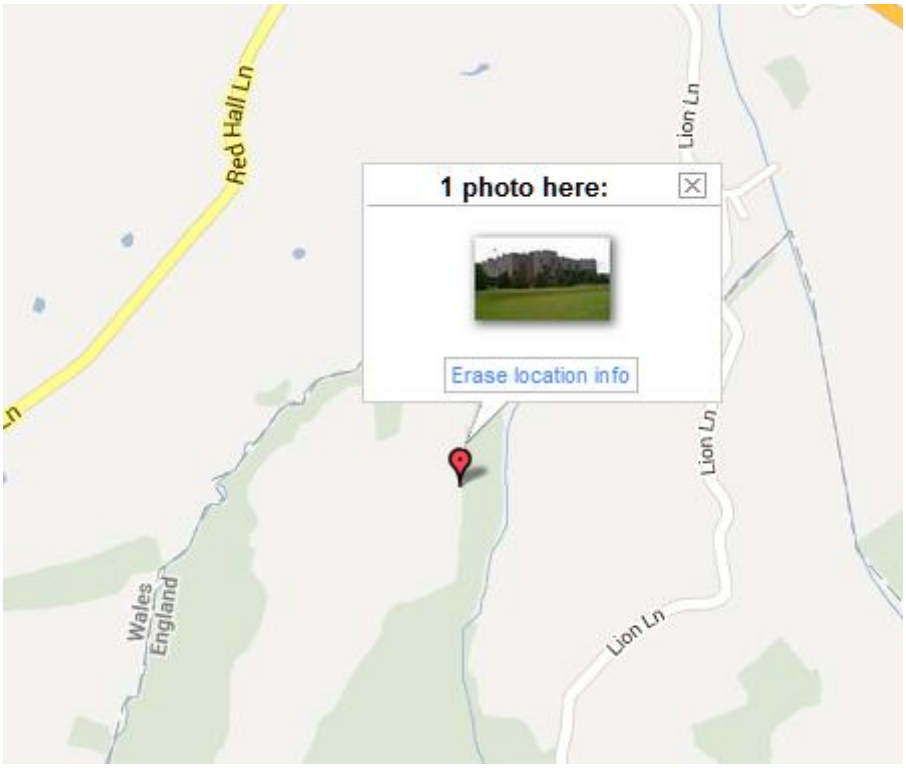


FIGURE 1. EXAMPLE OF A GEO-TAGGED PHOTO PLACEMENT ON THE MAP BY PICASA[1]

1.4. INDOOR LOCALISATION

GPS location provides a good large scale spatial localisation of images, but its typical implementations cannot be used indoors or without clear satellite signal (Nakajima & Tanaka, 2004; Lehtinen et al., 2008; Piras & Cina, 2010). Moreover, spatial image referencing cannot easily handle multiple images taken from the same physical location, since all of them will end up being placed on the same map grid.

This research looks at the issues of indoor (non-GPS) image spatial tagging and the problem of handling multiple images taken from the same physical location. It aims to investigate feasibility of implementing an indoor image spatial referencing system and produce a prototype design to judge its effectiveness. It does not aim to generate comprehensive maps of indoor locations, instead using localisation information as a source for image relative spatial arrangement.

[1] (Google Inc, 2013) Map data: Google

1.5. Outline of Chapters

The topics described above will be investigated in this thesis, which is arranged as follows:

Chapter 2 reviews advances in the field of geo-spatial image tagging and indoor localisation and their relevance and applicability to the topic of indoor spatial image tagging, finally selecting pedestrian dead reckoning as the localisation method for this research.

Chapter 3 looks at the other aspect of image management – handling images taken from the same physical location. It reviews the potential methods of combining images taken from the same viewpoint and the associated research before selecting a method best suited for this research – panoramic image compositing.

Chapter 4 investigates sensors in lieu of choices made in prior chapters both for pedestrian dead reckoning and image compositing, finally selecting optimal sensors for each.

Chapter 5 evaluates a way of compositing pictures without use of image processing. The compositing system is evaluated and compared to the image processing-based stitching approaches in a study and the conclusions about its performance and drawbacks are formed as the result.

Chapter 6 describes technical aspects, software and hardware design of the system combining sensor compositing with pedestrian dead reckoning, explaining design choices and features of the resultant prototype.

Chapter 7 evaluates performance of the system in 6, focusing on the dead reckoning performance, since the compositing subsystem has already undergone comprehensive evaluation in 5.

Chapter 8 revisits the questions asked in this introductory chapter and looks at whether they have been answered or not and to what degree. It summarises the findings of this research and outlines any further work that it could benefit from.

2. PHOTOGRAPHER LOCALISATION

Modern digital cameras embed a lot of additional information into the metadata of the photographs they take. This includes the camera settings, such as focal length, ISO rating and aperture size. Additionally, cameras tend to record items that are not directly related to the camera itself, such as environmental information, e.g. timestamps or geo-spatial coordinates of the camera location (Japan Electronics and Information Technology Industries Association, 2010). Moreover, users can edit special description fields of the photographs to provide plain text information about the image. This can be done directly from within modern operating systems (Microsoft Corporation, 2011). This information provides additional contextual information about the photographs that enables cataloguing and localising individual photographs on a map.

Typically, spatial information is obtained from global positioning (GPS) data and as such is more suitable for open outdoor areas. When it comes to determining the location of a photograph indoors or in the absence of GPS signal, the potential approaches are varied and each have their own strengths and weaknesses. This section will review a number of the notable ones and their applicability to the task of spatial image management.

Because the field of localisation is very wide and encompasses a number of approaches and sub-fields, categorising it can be done in a number of ways, neither of which will completely segment the research – for example, research can be categorised by sensors used (an approach largely employed in this thesis), but a lot of research brings different distinct sensor types together in one approach, making discrete segmentation impossible.

It is also possible to segment the research based on the algorithms employed. Again, many researchers employ multiple algorithms and some generic algorithms are used in very different approaches, making this segmentation counter-intuitive and artificial.

Looking at the sensor types employed, the following presents a way of looking at the localisation research:

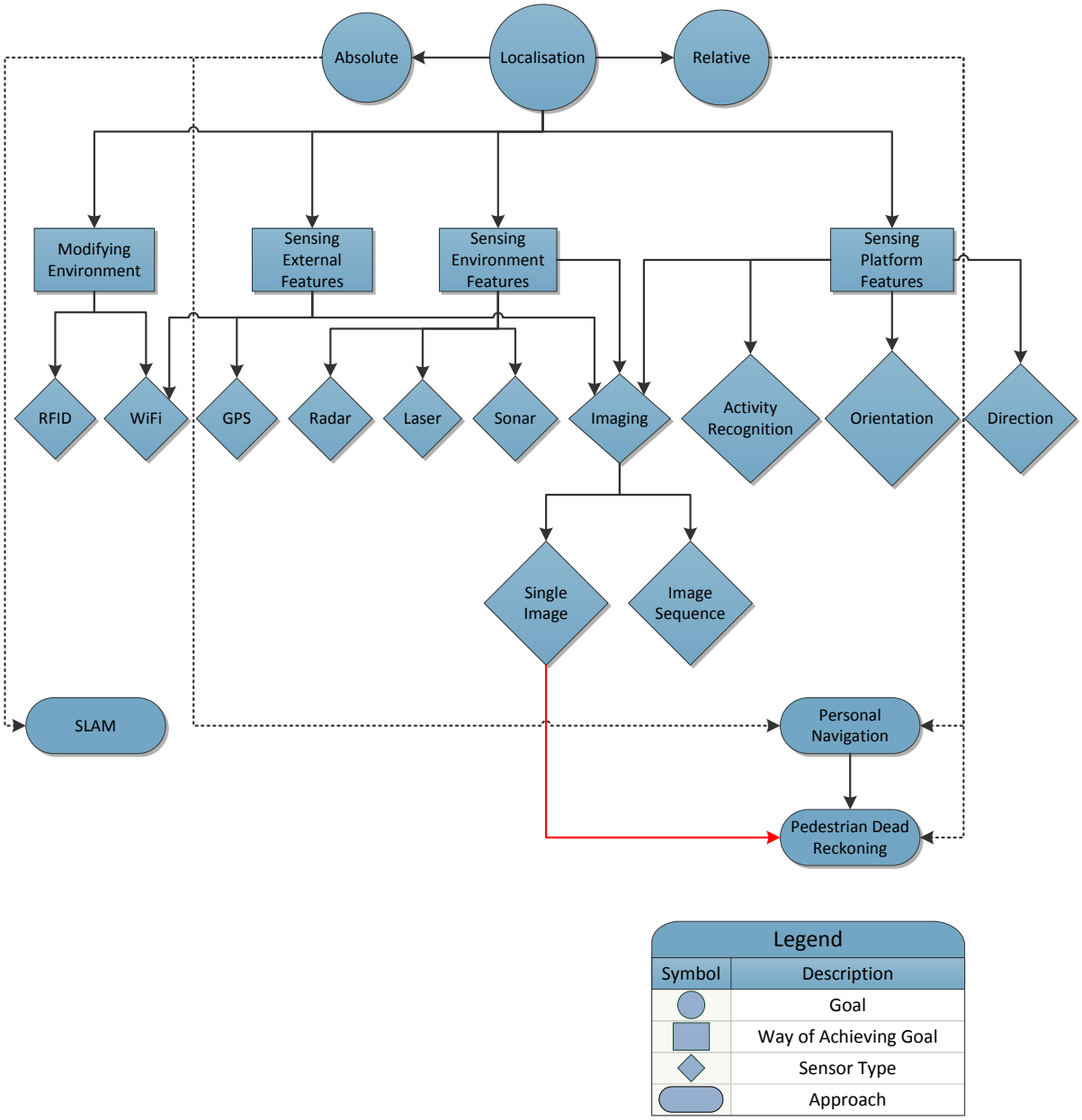


FIGURE 2. LOCALISATION RESEARCH OVERVIEW. RED LINE SHOWS THE LINK THAT THIS RESEARCH AIMS TO EXPAND

Localisation can be done as absolute or relative. Absolute localisation creates more topographical type of map – a map where all the features are presented on a common scale and show direct geo-spatial relations between them. For example, By comparison, relative

mapping creates maps which describe location as sequence of features or relative positions of such features (such as maps generated by Grzonka et al (2010)). A subset of relative mapping would be the dead reckoning-based maps, which map features based on their spatial relation to a certain location (an example of which is described by Foxlin (2005)). Simultaneous localisation and mapping (SLAM) usually focuses predominantly on absolute mapping, whereas a subset of personal navigation, pedestrian dead reckoning focuses on relative mapping.

When performing localisation, some systems rely on modifying the environment where the localisation is performed. These include most RFID-based systems (an extreme example of which is work by Kodaka et al (2009)) and some radio-based ones, reliant on deploying infrastructure into the environment (as described by Papliatseyeu et al (2009) instead of opting to use existing one (such as work by Kuai et al (2010)).

Many approaches attempt to sense features external to the environment. For example, GPS-based systems sense position based on satellite data (capabilities of such sensors are described by Lehtinen et al (2008)), some imaging systems track position of the sun to determine geographical coordinates (an example of this is the work of Cheng et al (2010)) .

Most approaches rely on sensing environment directly. Using range-based sensors such as sonars and lasers allows detecting walls and obstacles in the environment (as used by Ogaz et al (2009)). Radars can be used to perform through-the wall scanning of the environment and build a map of it from the outside, without directly entering the environment mapped (an example of which is research by Chang et al (2009)).

Sensors sensing the platform (be that a human being or a robotic sensor unit) features are very popular due to their flexibility and lack of reliance on outside sources of data. Such systems often sense activities (such as turning or stopping of a human participant, an example of which would be work by Kourogi et al (2010)), orientation (usually provided by

inertial sensors, such as work by Ladetto et al (1999; 2002)) or direction changes (provided by inertial and geomagnetic sensors, such as those used by Feliz et al (2009)).

Imaging sensors are used widely in a variety of scenarios, generating individual images or sequences thereof, such as video recordings. They can be used in detecting external features – such as shadows cast by the sun (as described by Sandnes (2010a)) and day/night cycle (such as research by Jacobs et al (2008)), internal features – such as walls and obstacles (such as those present on 3D maps generated by Furukawa et al (2009)), platform features – such as changes in orientation (an example of this is ceiling pattern tracking used by Fu and Yang (2009)).

When investigating personal localisation, self-contained sensor-based pedestrian dead reckoning (providing relative positioning of a human participant) is a widely used technique. Its relation to non-sequential image sensing is where this research aims to establish an additional link (shown on Figure 2 as a red line).

The rest of this chapter takes a deeper look at notable research in the field of localisation and mapping and is organised by sensor type used.

2.1. GEO-TAGGING SYSTEMS

The concept of geo-spatial image tagging has received much of attention in the recent years. With the global positioning system being widely available in many consumer devices and the prevalence of social networks, researchers focused on ways of organising the media prevalent on the internet into a way that would allow fast search and identification of specific images or videos (Zheng et al., 2010; Van House, 2007; Bates, 1998; Counts & Fellheimer, 2004). Geo-spatial tagging involves putting location data into the photograph metadata. This supports placing photographs on maps and spatially referencing them. Modern smartphones often come with GPS and camera capabilities, allowing embedding location information during the process of photography (HTC Corporation, 2011). Cameras are also emerging on the market that contain built in geo-tagging capabilities (Casio America Inc, 2010).

When used outdoors, location data are usually provided in terms of global positioning system (GPS) coordinates. However, a lot of photographs are taken indoors, and GPS is either unavailable in these situations on most consumer-level products or provides reduced accuracy at higher cost (Piras & Cina, 2010). This section will look at some ways of managing geospatial image information and means of obtaining these data in the absence of satellite signal.

Kim and Chang (2010) enhanced standard GPS-based geo-tagging by providing additional information from an RFID (radio frequency identification) tag reader. By assigning unique RFID tags to camera users and locations, information about photo creator and target object can be added to the photograph metadata. This greatly enhances the information content of the images and allows classifying and searching the photographs for specific people of objects. The use of RFID reader, however, significantly reduces the battery life of the camera. The RFID reader's limited range also complicates object scanning prior to the photographing process. Nonetheless, classification of images is greatly simplified due to the newly embedded information. However, a similar effect could be achieved by simply offering

the user a set of buttons (either physical or on a touch screen) that could be pressed prior to taking the photograph – with the benefit of not requiring modification of objects in the environment or the need to get very close to the object.

The idea that sequences of photographs can yield additional information about their content was investigated by Yuan et al (2010). The authors developed algorithms for the event recognition based on both the image content and the metadata of the images, including geo-spatial location data. The combination of GPS data and recognised visual features provides more comprehensive event recognition than either independently, as demonstrated by the tests performed by the authors using a variety of images from different activities, such as hiking, touring city or ball games.

Cheng et al (2010) combine 3 sources of data – image content, camera direction and its GPS (global positioning system) coordinates – to provide a searchable database of images. The resultant database allows retrieval of similar images by camera facing direction, location and image contents and hence can provide descriptive text from related image tags. The authors show a 38% improvement in precision of correct image retrieval and resultant search-based image annotation when compared with purely image-based methods.

Not all images, however, include GPS coordinates in their metadata and ways of geo-tagging images without such data have been the subject of much research in recent years. Some of this research is directed at the limitations that GPS imposes, such as need to wait for satellite lock acquisition, high power consumption and relatively costly sensors. For example, Sandnes (2009a; 2009b; 2010a; 2010b) used data from the shadow lengths of vertical objects in the scene coupled with the time data that is automatically stored in the image metadata by most commercial cameras (Japan Electronics and Information Technology Industries Association, 2010) to infer the sun elevation. Up to 3 photographs are required to determine the approximate location where the photographs were taken. The approach was

tested and showed to work with several outdoor photographs, identifying the location down to 2° accuracy in latitude and longitude.

Amongst the other approaches investigated by Sandnes is a way of using the timestamps of the photographs taken over several days to infer the location based purely on the assumption that fewer photographs will be taken during the night. This information could define the continent where the photograph was taken. When augmented with further refinements based on the camera exposure settings to allow determination of photographs that correspond to sunset, the location can be inferred. Tests showed that these approaches can identify the continent but have a large degree of error and rely on many assumptions about the photographer's lifestyle and habits. Nonetheless, it does not require any alterations to the camera and can be applied to existing digital photographs, provided they are time-stamped correctly.

A somewhat similar approach was taken by Jacobs et al (2008). Their approach works with static web cameras producing continuous images, thus allowing definition of the day-night cycle that can be referenced against a daylight map in order to provide the approximation for the camera location. Tests on a variety of outdoor cameras showed that the authors are able to localise to within 50 miles and geo-orient the camera to within 5° in azimuth angles. In their earlier paper, the authors investigated a way of using the weather recorded by the webcams to help determine the region where the camera is located (Jacobs et al., 2007). By performing principal component analysis on the images they were able to determine the weather conditions and illumination recorded by the camera. Using accurate camera timestamps and satellite weather maps, Jacobs et al (2007) determined that it is possible to localise the camera to within 44.6 miles. Both approaches, however, rely on web cameras that have static orientation and that produce a stream of images, rather than a single one.

For single image localising (when geographical data is not embedded in the image metadata), Hays and Efros (2008) took a different approach. The authors used a tagged and

labelled image set generated from the image sharing site Flickr (Yahoo! Inc, 2011) in order to identify arbitrary untagged images and their location probability. They tested their approach with a dataset of over 6 million images and 237 test images to be localised and discovered that the median error of their localisation approach is 500km. The authors' approach provides a very coarse localisation of most images, but still shows promise for the future developments as well as a way to use large amount of geo-referenced image data readily available on the Internet.

Another paper using the large image set of Flickr (Yahoo! Inc, 2011) is the work by Singh et al. (2010). In their paper, the authors describe a way of using non-geolocation metadata, such as text description and timestamp tags together with the image content itself to cluster spatially similar images. Singh et al. developed algorithms for converting all of the various mismatched attributes (timestamps, descriptive text tags, image content) into a single space where these could be directly compared and aggregated together. Tests of the approach with various images showed it to enable images to be geo-clustered even when the geo-location tag is missing. The approach is reliant on accurate image metadata, however, which is not always the case, since a human element is involved in setting camera time and defining the textual descriptive tags. Nonetheless, it can help localise images and its precision is likely to only grow as the image set increases, which it does on daily basis as users upload more photographs online.

Another way of using non-geolocation metadata for image localisation is presented by Joshi et al (2010) who used a large database of geo-tagged images linked to the text description tags contained within images. They were able to show that text annotations within tags can be used on their own to infer approximate location of the images. Joshi et al used an earlier generated (Gallagher et al., 2009) set of 1.2 million geo-tagged images from Flickr (Yahoo! Inc, 2011) for the probabilistic location database of annotations. They determined that even

when the description tag does not contain the name of a city or a country, the location the photograph was taken at can be inferred to within 300km radius with over 25% accuracy.

Chen et al (2009), on the other hand, took a completely different approach to geo-tagging images that lack geo-spatial metadata. They chose a game-like approach for the tagging, based on the principles of games-with-a-purpose (GWAP) such as competitiveness and providing incentives for task completion in form of reward schemes. The authors modelled and developed generic design strategies for GWAP-based approaches to geo-tagging and evaluated these using a simulation. The analysis provides a good framework for developing user interactive ways of geo-tagging images. However, the authors did not actually put their simulation results into practice and test the conclusions beyond simulation.

This review has shown a number of approaches that supplement or complement GPS (global positioning system) to define location outdoors. However, for the most part, the approaches rely on GPS or meteorological conditions (Sandnes, 2009a; 2009b; 2010a; 2010b) that preclude their use indoors. While there exist specialised GPS implementations that are able to provide an indoor location fix, their positioning accuracy is reduced when compared with their outdoor counterparts (Piras & Cina, 2010).

This means that accurate indoor localisation, a prerequisite for the indoor geo-tagging, must be achieved using a different approach, as shown by the commercially available indoor geo-tagging cameras that begin to adopt motion sensing for indoor positioning (Casio America Inc, 2010). The next section will look at the means of accomplishing the task of indoor localisation and evaluate them for the use for the purposes of this research.

2.2. INDOOR MAPPING SYSTEMS

Indoor mapping systems have received a significant amount of research and development in recent years. They are often used to help robots and other autonomous units navigate through unfamiliar environments and to perform localisation of pedestrians within an

environment. Simultaneous Localisation and Mapping (SLAM) systems allow an autonomous unit or a person to enter a location and return with a map containing not only the floor plan of the location but often also the geometric shapes of the objects within (Ogaz et al., 2009). While many algorithms analyse the floor level plane based on the assumption that this information is adequate for robotic navigation (Wulf et al., 2004), others (Furukawa et al., 2009) provide full 3D reconstructions of the environment. Additionally, concepts of place recognition applied to many SLAM systems can be linked directly to the non-GPS (Global Positioning System) geo-tagging (Zheng et al., 2010), making it an important consideration to investigate for this research.

This section looks at the work done in the field of indoor localisation and its applicability to wearable mapping devices. The indoor map of the location can be formed either from the inside or outside of the building itself. The more direct mapping of the indoor location involves a robot or a person wearing sensors travelling through the building, while simultaneously mapping it. Alternatively, a through-the-wall scanning system can be applied to estimate the layout of the building floor by scanning it from the outside (Chang et al., 2009; Le et al., 2009).

2.2.1. MAPPING THE LAYOUT EXTERNALLY

Through-the-wall scanning approaches are an alternative to mapping the building by navigating it. These approaches rely on using the scanners that are able to penetrate the walls of the building to estimate its contents and interior layout. Techniques to accomplish this are largely restricted to the use of radar scanners due to the wall penetration requirements. One example of this is the work by Subotic et al (2008) who presented an iterative method of analysing radar data from external building scans for reconstruction of its internal floor plans. The method was tested on a simulated model of a building and shown to produce results closely matching its internal layout. The approach is reliant on the assumption of consistent building materials throughout and on 90 degree intersection of

walls. The results of applying the same methods on an actual building filled with objects attenuating the radar signals and surrounded by other buildings, rather than isolated for easy fly around by sensors, could vary from those described widely.

A slightly different approach is used by Le et al (2009). They applied ultra-wideband radar to scan the interior of a building by moving the radar in parallel to the two adjacent vertices of it. The authors compared the imaging results produced with those generated via simulation of the same building and produced a map of inside features. Despite the impressive results, the penetration and resolution of the inside building features is relatively low. Additionally, the building scanned is located in isolation from others, a situation that is unlikely to occur in an urban scenario. Moreover, the authors' experiments showed the noise generated from roof reflections and this can be taken to mean additional difficulties when scanning down a multi-storey building.

A way to better identify the interior features is presented by Chang et al (2009) who tested an EM characterisation technique to interpret radar scans of buildings. They have shown that their technique can identify the building interior features such as walls during a fly-by scan and can even detect the presence of furniture. Their testing, however, was performed only on a scaled model of a building that contained a single item simulating furniture. The radar data from a real building full of various objects of furniture and more than one floor will be significantly more complex than the results of their experiment and are likely to require additional analysis for successful characterisation and mapping.

2.2.2. DIRECT MAPPING

The direct mapping approach involves a set of sensors moved around the location and the data collected used to both localise the current position and build/refine an overall map of the environment (known as SLAM – Simultaneous Localisation and Mapping). Most of the research in SLAM is based around the use of autonomous robots. Robots can travel throughout the building in a precise manner with little deviation, covering the same area as

many times as needed for map refinement. They also allow rigid mounting of sensors and accurate pose estimations, making it possible to determine accurately how the sensors are positioned relative to each other and the previous sensor sweep.

Human-centric SLAM approaches, on the other hand, cannot easily provide accurate pose estimation, limiting the sensors that can be employed efficiently and requiring additional processing.

Thrun et al. (1998) present one of the first probabilistic mapping algorithms. Using maximum likelihood estimation, the authors are able to build accurate maps. They tested their approach using a manually controlled robot with the operator generating landmarks. The resultant map, when generated via probabilistic approach is more robust than the map created from raw data. Despite the need for manual tweaking of the resultant map (adding obstacles that weren't correctly mapped), its acquisition is significantly faster than previous manual attempts and these limitations are due to the experimental platform and not the algorithm itself.

Smith et al. (1990) present one of the fundamental papers in SLAM. They modelled uncertainty relationships between different objects and formulated ways of determining in advance if a certain observation is likely to reduce this spatial uncertainty. This permits choosing an action or sequence of actions that can improve the description of spatial relationships. A potential application of these principles is described by Leonard and Durrant-Whyte (1991), who present one of the early ultrasonic SLAM approaches. They designed an autonomous robot localisation algorithm that can utilise features inferred from the data provided by a number of ultrasonic rangefinders to orient itself and construct obstacle map. Despite the theoretical nature of the original paper, its findings proved the basis of a lot of later research into the problem of SLAM (Durrant-Whyte & Bailey, 2006).

Gutmann et al. (1998) compared Markov -based and extended Kalman filter (EKF)-based approaches to localisation of a mobile robot. By introducing varying degrees of noise into the

sensor data they were able to determine that under normal conditions Kalman filter localisation significantly outperforms Markov-based approaches both in terms of accuracy and computational efficiency. However, in presence of stronger sensor noise, Markov-based approaches cope better, suggesting a hybrid approach could be the optimal choice. The authors expanded their research (Gutmann & Fox, 2002) to include particle filter (Monte Carlo Localisation) into the comparison as well as refine the algorithms in lieu of new developments. Their new experiments, employing visual landmark-based localisation found a hybrid Markov-EKF approach and the Monte Carlo Localisation approach to outperform pure EKF when applied to the image sensors.

Guivant and Nebot (2001) proposed a number of optimisations to the standard extended Kalman filter-based localisation. In particular, they designed a way of reducing computational requirements during SLAM by simplifying processing with landmarks local to the sensor platform (in close proximity to it). Having tested their implementation they determined that the proposed enhancements and alterations to SLAM processing result in efficient and accurate map bulding and localisation.

In addition to varied algorithms for SLAM, sensors used also vary in complexity and principle of operation. They range from range sensors, such as sonars and laser rangefinders, optical sensors, such as cameras to radio transceivers and Wi-Fi signal meters. Some approaches rely on a combination of different sensors for a comprehensive result.

2.2.2.1. Radio-Based Sensors

Radio signals are able to penetrate the walls and obstacles inside the building, attenuating differently depending on the materials and structures it encounters. Analysis of the signal parameters allows approximating the location. This and other principles are employed by many researchers for SLAM, including Deissler and Thielecke (2009) who used ultra-wideband radar scans from multiple locations in a room to map its features. Their approach can theoretically detect walls, corners and edges of obstacles. However, due to the way the

feedback path is constructed, the mistakenly detected landmarks will eventually accumulate in the model and cannot be removed once accepted as valid. This makes the need for correct threshold calibration paramount when using this approach and limits its usefulness for large scans involving complex and cluttered locations.

Guo and Filler (2006) developed a way of determining the general shape of an indoor environment (walls, floor and ceiling) by analysing both the direct and reflected signals between radio transmitters on the users, thus being able to estimate accurately the locations of the environment boundaries and the position of the users. While this approach provides a good idea of the overall shape of the location, it requires several radio transmitters (albeit not in a fixed predefined position) and does not provide any information about objects present.

A different way of interpreting radio signals from multiple sources is presented by Fu et al. (2010), who used the approach of compressive sensing (reliant on the fact that certain sparse or compressible signals can often be reconstructed from far fewer samples than required by the Niquist theorem, reducing the storage space). This enabled efficient interpretation of the radio signals from multiple sources and allowed a robot to build a map of the environment and localise itself. The approach is useable with multiple mapping units (robots) but is reliant on predefined radio transmitter sensor positions (unlike the work by Guo and Filler (2006)) and has limited use when faced with a completely new unknown environment.

Another approach reliant on pre-deployed infrastructure is described by Kuai et al. (2010), who use a mobile robot in an environment with multiple RF sensors with known locations to localise the robot relative to them. While the approach is sound for localization purposes, the need for sensor infrastructure and sparseness of the information provided on the mapping results for the corresponding part of the experiment make this approach of limited use for wearable designs.

Taking an altogether different approach to the infrastructure requirements, Kodaka et al. (2009) used a semi-dynamic location containing a large amount of RFID tags and tag readers to allow robots to navigate under changing circumstances. They used a comprehensive lattice of under the floor passive RFID tags with the RFID readers mounted on all of the potential obstacles, such as furniture, moving partitions, robots and even human participants to maintain an up-to-date map of the environment that robots could query in real time, thus enabling them to navigate it efficiently and even follow humans around. The approach's limitation is that each piece of furniture and entity's parameters must be pre-entered into the system and equipped with both the RFID reader and the wireless communications means with the mapping server. Additionally, the RFID tag infrastructure requirements (especially the density of the tags) make the approach highly unrealistic.

The need to deploy radio transmitters can be mitigated by using the existing ones. Wi-Fi access points can be found in large numbers in many locations and can be used as the basis for localising, as shown by Yeh et al. (2009). They developed a method of combining low resolution error prone radio signal maps of Wi-Fi access points signal strengths with the RFID tags located in known points to generate and update an RSS (radio signal strength) map autonomously. The paper does not provide any quantitative comparisons of this approach to the alternative ones and the results are highly dependent upon the "seeding" of the RFID tags on the location. This approach is of limited use in an unknown location and while helpful for the RSS map updating, cannot be easily modified for use in cluttered environments with obstructed paths between the tags.

All of these approaches rely on specialised equipment. By comparison, Zhuang et al. (2010) present a way of using modern smartphones to generate a generic map of the indoor environment. They combine the information from the pre-seeded Wi-Fi access points with the environmental fingerprinting information gathered from the smartphone sensors such as magnetometers, accelerometers and microphone to distinguish various location categories.

Their approach has only been tested in simulation and does not take into consideration incomplete or absent Wi-Fi coverage and signal attenuation and reflection caused by environmental features that is likely to produce a significant amount of conflicting and false readings. These issues can be mitigated by statistical processing of the signals, as shown by Yim et al (2008) who developed an Extended Kalman Filter and K-Nearest Neighbour clustering approach to interpret detected signal parameters of Wi-Fi access points to track a user in a known environment. This approach presents a reasonable degree of tracking accuracy considering the noisy nature of RSSI parameters of the Wi-Fi signal but it relies on prior knowledge of the location (more specifically the intersections) to change Kalman Filter parameters. The approach is useful for areas with multiple access point overlapping but is less so in areas where WiFi signals are fewer in amount.

Another smartphone-based approach is described by Shin and Cha (2010). The authors proposed a system for locating the user based on WiFi fingerprinting. Their approach is implemented on a commercial off-the-shelf smartphone and utilises built in inertial sensors to construct the WiFi topological signal map that is later used to identify user's position based on signal characteristics. The approach proved functional in a building with a large number of WiFi access points, albeit requiring calibration and fine tuning. However, typically the number of access points observed will be less than the 8 detected by the authors in their testing environment. This makes the proposed method useful only in a limited type of location. The automated labelling proposed, reliant on the credit card company interaction, is unlikely to ever occur because of the privacy concerns.

An approach reliant on FM radio is proposed by Papliatseyeu et al. (2009). The authors use a set of FM transmitters (3 for the test environment of the laboratory they chose) positioned at fixed points around the room and an FM receiver carried around by the user. Using received signal strength indication (RSSI) from the transmitters, the receiver is able to determine the approximate distance to them. The test of the proposed design showed

reasonable accuracy (4.5m at 95% confidence level with median 1.3m) and stability of the RSSI fingerprint levels. The system does, however, rely on fingerprinting the location and deployment of the transmitter infrastructure for its use. At the same time, low transmitter cost and general availability of FM receivers in many consumer devices make the approach worth considering for the consumer indoor positioning scenario.

2.2.2.2. Imaging Sensors

Imaging sensors such as 2D and 3D cameras typically rely on finding overlapping features between multiple scans and the knowledge of the sensor positioning change between the scans to construct a representation of the location. For example, Takezawa et al. (2006) used a stereo camera with pre-deployed artificial landmarks at the target location to localise the robot. Their approach does not require the localisation of the landmarks prior to mapping and compared favourably to a laser rangefinder-based algorithm. In their experiment they only used a robot travelling in a straight line in a very simple location and their approach is dependent upon known robot kinematics thus largely unusable on a human being.

Stereo cameras are generally useful for localisation, since they provide the extra dimension that 2D imaging lacks and thus facilitate the spatial recognition of indoor features, as shown by Saez and Escolano (2004) with their mapping approach reliant only on a stereoscopic image generated by a robot-mounted set of cameras. The authors produce accurate floor layouts of the target environments but rely on a rigidly fixed camera with no tilting, only rotation around the gravitational normal. The approach is limited in that it is only usable in uncluttered environments and does not show the environmental features beyond simple walls.

A more interactive mapping approach is presented by Kim et al. (2009a; 2009b) who proposed a purely stereo camera-based method of user-assisted map building. Their method minimises the amount of user interaction by only having the user specify locations of interest and lets the robot complete the map on its own by re-running the path that the user guided it

on. The approach was only tested in an uncluttered corridor with right angled turns and the path taken was very simple and did not involve any branching.

With prior knowledge of the location, mapping can be improved, as shown by Cheong et al. (2007). The authors implemented a vision based localisation system using a stereo camera and prior knowledge of both the topology of the location and the location of objects in it. They developed an approach that uses object information to augment localisation and hence enable the robot to potentially interact with objects. The approach requires stereo camera and on demand adjustment for verification of the localisation hypotheses and as such is restricted to autonomous robots. It also relies heavily on *a priori* knowledge of the environment and static objects therein.

The common problem of imaging approaches, especially multi-camera ones, is the need to retain large amounts of image data. A way to reduce the storage requirements is developed by Konolige and Agrawal (2008) in their work on purely visual map generation. Data collected by a pair of cameras (for stereoscopic vision) is used to construct a map based on a concept of skeleton frames. Instead of maintaining every single frame taken by the camera and thus increase the processing overheads, the authors reduce the frame set, maintaining only enough data to construct highly accurate maps of both indoors and outdoors environments. Despite only testing the approach on the already available image sets rather than generating new ones with their own hardware, the results are similarly impressive in a variety of environments, including very large outdoor paths.

A visually impressive approach is presented by Furukawa et al. (2009) who developed a fully automated system for creating comprehensive 3D models of both exteriors and interiors of the buildings purely from the image data acquired with a camera. The system produces both floor plans and the 3D walkthroughs for the building interior. The system is very computationally intensive, requiring significant post processing for the model generation as well as limiting the model to planes aligned to primary axes. Additionally, it requires the

source images to be taken with multiview stereo in mind, capturing locations from several viewpoints, and is unable to cope with disjointed image sets which are more natural for a casual photographer.

Instead of photographing the location itself, Fu and Yang (2009) chose to photograph the ceiling and thus detect the orientation and movement of the robot. Their approach uses a single ceiling-pointed camera and was tested to work efficiently and be computationally simple. At the same time, the resultant maps are rather simple and the algorithm depends heavily on the static orientation of the camera (so its imaging plane is parallel to ceiling) and presence of right angle intersecting line patterns in camera view. This limits the algorithm to be useable in areas with suspended ceilings and to autonomous robots only.

2.2.2.3. Range Sensors

Sonars and laser rangefinders are both used extensively when mapping unknown locations. With known sensor orientation, it is possible to detect obstacles around the sensor and build a comprehensive map. Wulf et al. (2004) use a custom-made 3D laser range scanner to construct a 3D point cloud for the environment which they then simplify into a virtual 2D model that they use as an input for a line-based SLAM algorithm. The result of this operation is a map of the walls in the target location which is able to cope with cluttered environments. While their approach is good for creating blueprint-like plans of buildings and is able to compensate for the robot motion, the loss of the detail when converting 3D point cloud to its 2D representation means that the information about the objects in the location is lost too.

Another line extraction approach is developed by Prez Lorenzo et al. (2004). They developed a SLAM approach for a multiple sonar-equipped robot, using line extraction with an Extended Kalman Filter for mapping even in partially occluded environments with noise. It is computationally efficient but depends on being able to model the range scan source accurately (identifying the robot pose). On the other hand, Ogaz et al. (2009) used a 2D laser rangefinder sweep results with line extraction to identify geometric shapes in the

environment. Their approach is efficient and identifies obstacles as well as the environment borders by scanning from just a single location, but requires supervision and is only able to find and map a single 2D plane.

Xu et al. (2002) applied clustering and line detection algorithms with adaptive thresholds to the results of the laser range scans to build a map of the location. Their research proves the value of using adaptive thresholding over the constant one but fails to quantitatively show the effectiveness of the approach described for the final result of the map building process. The map presented, while being complete, does not have any reference to the original location to validate if any features were missed or mistakenly detected.

A 3D model of the location can be effectively created from the results of multiple ranging sensor sweeps as shown by Huber et al. (2000). They rely on matching the 2D signatures of free-form 3D surfaces that are consistent between different views of the same 3D surface. Their approach results in a 3D model of the location but requires multiple scan points and a significant amount of processing to identify the 3D features of an indoor environment correctly. Resolution reduction required to allow the matching algorithm to cope with the large amount of scan data makes this algorithm more suitable for the outdoor scenes.

Another multi-rangefinder approach is presented by Wetherbie and Smith (2001) who developed a computationally simple way of identifying large scale features (such as corridors, intersections and alcoves) using a rule-based approach. Their algorithm relies on data from the sixteen sonar rangefinders positioned to measure ranges in a circle around the robot and the predefined sonar patterns (rules) for the features to be detected. The results show successful detection, albeit highly dependent upon the uncluttered environment and *a priori* knowledge of the features to be expected. Misdetections that occurred during the testing required adjustment of the rules. While the computational simplicity of the approach make it highly desirable for integrating into a handheld device, the limited robustness and

need for an *a priori* location knowledge coupled with the rigid sensor positioning make its use in such a device questionable.

Luo and Yang (2008) focus on complete coverage paths, but also include map building for use in previously unknown environments, mainly for use by cleaning robots capable of obstacle sensing within a limited range. The approach was only tested in a simulation. The resultant map is 2D in nature and has no way of distinguishing between different types of obstacles. The map building is based on complete coverage and as such is mainly useable by the automated units.

Gutmann and Konolige (2000) propose a way of localising a robot in large cyclic environments (where the robot will revisit its prior locations). Their approach enhances existing alternatives by performing the calculation necessary incrementally, thus not increasing in complexity with the increase of robot poses that need correcting for errors. They describe a Local Registration Global Correlation algorithm that is designed to integrate scans into the map and only perform additional processing when closing loop. The authors tested their approach with several laser rangefinder-equipped robots and found it to result in accurate maps of various indoor environments tested.

2.2.2.4. Combination of Sensors

All of the various types of sensors have their own limitations. Therefore, it is often beneficial to use a combination of different sensors to generate a more comprehensive set of data for the SLAM algorithms. Rangefinders and cameras are often combined to produce more robust and reliable mapping approaches, as shown by Ahn and Chung (2007) who used a combination of sonar range data and visual data to build a hybrid map of the environment. Their approach combined the lines and features extracted from the sonar data with those obtained from the imaging and analysing the visual planes to generate a more robust and accurate map than the individual sensors would independently. The map produced as a result does not distinguish between the objects and environment features but the data it

produces is adequate for the purposes of the robot navigation. The sensor mounting considerations make it of limited use in body mounted or wearable designs however.

Another approach is presented by Gallegos and Rives (2010) who use an omnidirectional camera and a 2D laser rangefinder mounted on a robot to generate maps of the environment containing 3D elements, such as vertical lines. They have tested the approach in an indoor environment with satisfactory results. The algorithms used, however, rely heavily upon accurate positioning of both the camera and the laser scanner. Tilting them will interfere with the line detection algorithm, making a human mounted version of this approach questionable. Additionally, the implementation of the line detection algorithm, reliant on detection of radial lines in camera images will result in misdetections when faced with the patterned floors and furniture.

Similarly, Luo and Lai (2010) used robot pose information with the information from the laser range sensors and camera to construct both 2D and 2.5D representations of an indoor environment. Their approach, however, relies heavily on robot pose estimation and while the results produced are very accurate and efficient (partially due to the algorithm selection for optimal alignment choice) they cannot be easily adapted for use on a human being with a typically inconsistent gait.

Rangefinder and camera combination approaches can be effectively used even when rigid mounting is not available, as shown by Naikal et al. (2009). The authors developed two versions of an algorithm for indoor dead reckoning using a laser scanner and a single camera, combining primarily visual odometry for camera images and iterative closest point approach for the laser scans analysis. The approaches were tested both for a 2D movement situation with a rolling platform and a human-mounted backpack. The evaluation used a single loop path in an uncluttered continuous corridor for testing purposes and showed robustness. The analysis results appear affected by the gait characteristic of the operator, but not to a degree that would invalidate their usefulness, at least for the path type provided.

Combining sensors with image data can lead to maps enhanced by texture information, as shown by Li et al. (1999). They used a swivelling camera and sonar unit in order to match the vertical edges detected by the imaging with those obtained from the range data. This allowed building a map of a room that includes both range and texture information. The experimental results show good close range boundary detection that seems to become more erratic at longer ranges. The rigid positioning and slow scanning time (due to the need to perform a full sensor sweep) make this system less usable for human mounted mapping.

Another approach is described by Wu and Liu (2009) who developed a process of using multiple data sources, such as image and LIDAR data to construct a photorealistic textured model of a building exterior. While this technique does not provide any information about the contents of the building, its planar slices can be used as constraints for the indoor SLAM approach. Lamond and Watson (2004) propose a different way of modelling building exteriors, using photographic and laser range data. Their approach combines photogrammetry derived geometric planar data with the fine details captured by a laser range finder to build a complete “hybrid” model of the architectural feature. While the model was only tested on the exterior of a building and is limited to only the objects with planar features, the results appear promising for indoor use in uncluttered environments.

Grzonka et al (2010) use the data from a commercial motion tracking suit (Xsens, 2011) with a combination of dead reckoning and motion recognition to identify paths taken by people inside a building and outdoors, including door open/close events. They used these data for constructing an approximate map of the location, matching the approximate location mapped. The limitation of their approach is in the reliance on an expensive motion capture suit and the closed source post processing algorithms within for detecting the motions. The template based approach to motion identification, while requiring pre-processing, is an efficient way of identifying activities.

A more collaborative approach to mapping is developed by Kleiner et al. (2007) who describe a system for joint mapping of a location by robots and humans. Human participants use dead reckoning based on inertial/magnetic sensors and robots use an approach based on wheel odometry. RFID tags seeded in the area are scanned by both groups and a corrected trajectory obtained, combining the information from robots and humans. The authors tested the approach in an environment with magnetic interference and found it to be producing reasonably accurate results.

2.3. PEDESTRIAN LOCALISATION

Since the research in this thesis explores wearable indoor imaging systems, the use of restrictive sensor positioning (Wetherbie & Smith, 2001) and pre-setup environments (Kodaka et al., 2009) is against the requirements of such a system (outlined in 1.1). Moreover, reliance on imaging sensors for the localisation and mapping approaches limits the environments where such a system can be successfully employed to areas with certain lighting conditions or even specific ceiling patterns (Fu & Yang, 2009).

This research is not aiming to provide comprehensive maps of the indoor environment, but, rather, enable arrangement of photographic images relative to each other (as outlined in chapter 1), and permits a dead reckoning technique to be employed. This approach relies on tracking a person (in context of this research – a camera operator) relative to the starting point throughout their path and thus fulfils the requirements optimally.

This section reviews different means employed for human localisation via pedestrian dead reckoning (PDR) and other personal navigation systems. PDR systems are widely researched for use in satellite-based global positioning (GPS)-denied environments, such as indoors, surrounded by tall buildings or in a canyon. Frequently, emergency situations will require navigation in an environment where conventional navigation means can be restricted (visual processing struggles to function reliably in foggy or smoky environments, satellite or

radio-based navigation can be jammed). In such a situation, a navigation system that is not reliant on external data sources can be highly beneficial.

Pedestrian dead reckoning systems use a known starting location and work by estimating the displacement from this starting location at each point of the journey. They accomplish this by calculating travel direction and distance. Some approaches rely on additional data to augment the path estimation, while others use just the self-contained on-body sensors.

Use of data external to the system itself is occurring frequently in the pedestrian dead reckoning systems. This data can be represented by signals from external transmitters or models of the environment and typically facilitates detection of the dead reckoning parameters.

2.3.1. Pedestrian Dead Reckoning Approaches

Pedestrian dead reckoning approaches estimate distance and direction travelled through wearable sensors and knowledge of human locomotion. Judd (1997) describes one of the early personal dead reckoning modules. His design relies on accelerometers for step counting and 3-axis magnetic compass for heading. Using Kalman filter, the module can fuse dead reckoning and satellite positioning to calibrate user and environment parameters whenever this secondary source of data is available. It has been tested on a long outdoor track and found to help guide the user to the destination waypoint even in the absence of the satellite signal.

Ibarra Bonilla et al. (2011) present a purely inertial pedestrian dead reckoning approach for localising people. Zero crossing method is used for step detection and fusion of accelerometer and gyroscope for yaw change tracking. The approach is validated in a closed loop indoor test but lacks a reliable absolute reference source for comparison, relying instead on user closely following trajectory supplied. Additionally, no information on the number and

nature of participants and size of error is provided. Despite this, the algorithms used for sensor fusion and step detection are helpful for future PDR approaches.

Beauregard (2006) describes a helmet-mounted pedestrian dead reckoning system. Using GPS to assist with neural network-based step parameter tuning, the author showed accurate detection of step occurrence that is able to handle walking speed changes without adversely affecting the accuracy of the localisation (provided network is tuned to multiple walking speeds). The step length estimation algorithm compares favourably to alternatives but the overall system suffers from requiring alignment between helmet sensors and direction of travel that is hard to achieve and maintain outside of laboratory conditions.

Sensors can also be mounted on waist or shoes as shown by Alvarez et al. (2012) and House et al. (2011). Alvarez et al.'s waist mounted pedestrian dead reckoning system shows state of the art accuracy in the location estimation. Of note is the ability to detect the zero velocity stance phase of human walking from the waist, unlike conventional shoe sensor approaches. They use vertical displacement and subject leg length information to obtain accurate step length. The result of this is a PDR system with maximum error of 20% over a long distance walk (178.5m). The approach relies on calibrating gyroscope performance for certain user and track and thus cannot be easily deployed to a new location. House et al. (2011) combine shoe-mounted pedestrian dead reckoning with RFID tag reading to create waypoints to re-set PDR heading errors. By seeding the environment with RFID tags and attaching a passive RFID reader to the shoe, they are able to detect when the user is passing over a known waypoint/tag thus being able to re-calculate dead reckoning parameters and hence improve localisation. Despite the requirement of modelling tag placements, the low cost of passive RFID tags makes this approach a reasonable way of improving on conventional PDR long term accuracy issues.

Another foot-mounted sensor system is described by Feliz et al. (2009), who developed means to supplement existing navigational systems with a pedestrian dead reckoning system

that relies on the foot-mounted accelerometers, gyroscopes, magnetometers (used as a digital magnetic compass) and a barometer for navigation in GPS-denied environments. The authors performed a number of tests on the heading detection methodology, clearly illustrating the advantage of the compass for the outdoors measurements and the advantage of the gyroscope indoors. They have also illustrated ways of using gyroscopes to improve the reliability of the accelerometer-based step detection. The barometer was shown to provide a 3D element to the path reconstruction by adding the ability to detect height. What the work lacks, however, is a test of all of the above combined together into a single system. Such a test would highlight performance of the individual subsystems outside of their own test environment (e.g. barometer outside of a staircase and on a level path).

Stirling et al. (2003) present and evaluate another shoe mounted pedestrian dead reckoning approach. They use gait phase information to help correct in errors caused by movement of the foot during walking. This provides stride length accuracy of about 90%. Resultant dead reckoning provides a path that is an improvement over purely gyroscope-based tracking. However, the accuracy of the dead reckoning results is still inadequate for use without further development.

Ojeda and Borenstein (2007) also use shoe mounted inertial sensors - accelerometers and gyroscopes to perform dead reckoning. They rely on double integrating the accelerometer readouts for position updates and use gyroscopes to estimate sensor orientation. With help of zero velocity updates that occur when foot is stationary during gait phase, the authors are able to reset the accumulated errors and improve tracking. Of interest is the fact that the authors rely on angular rates local minimums to detect zero velocity phase rather than the accelerometer readouts. Having tested their approach on a closed rectangular path, the authors obtained less than 2% error for short paths (104m) which deteriorated to around 5% for a longer (approximately 610m) walk.

Wei Chen et al. (2009) implement a pedestrian localisation system that combines GPS with pedestrian dead reckoning. Using Kalman filter to simultaneously estimate stride characteristics and fuse satellite and inertial position data, the authors achieved localisation that is usable outdoors even when GPS signal is lost. The proposed algorithm produces reasonable results but suffers from compass errors due to tilt and magnetic interference which are only partially solved by filtering. Use of GPS and inertial data from off-the-shelf sensors and filtering the data through common filtering algorithms, such as Kalman or a finite impulse response filter, is also illustrated by Gusenbauer et al. (2010), who use accelerometer and compass from a smartphone device for dead reckoning. Their approach uses GPS for PDR parameter calibration (when available) and includes several efficient algorithms that can be used for the embedded devices. Kalman filter use for combining PDR and GPS results in the algorithm nearly seamlessly switching between modes of operation upon the GPS signal loss when tested outdoors. An *a priori* knowledge of the indoor location was required however to compensate for the errors in the magnetic compass readouts. This makes the approach less useful for previously unknown locations.

Another use of Kalman filter with off-the-shelf devices is presented in research by Jirawimut et al. (2003). The authors developed a method for estimating dead reckoning parameters using an extended Kalman filter. The method requires an absolute measurement source for the initial parameter estimation, provided by the GPS signal. The step parameter detection is reasonable; however the magnetic compass modelling does not cope with the most common source of error – magnetic interference. The design also relies on the availability of the GPS signal for the training phase, making it less useful for fully indoor locations. Gait parameters can be determined without reliance on the GPS, however, as discovered by Kim et al (2004). Their research shows that pattern analysis can be used for step detection, filtering out erroneous steps using standard peak detection. A relationship between acceleration and stride size is experimentally determined and shown to provide accurate results. The heading

determination uses a standard Kalman filter and shows relatively small improvement compared to the unfiltered data. The heading tests are inconclusive, since only a straight line travel with no turns was employed, atypical for the PDR usage scenarios. The step detection algorithm presented relies on a sequence of thresholds that must be exceeded for a step to be detected. This approach requires testing under different walking conditions (slope, speed) for it to be applicable to most walk patterns, rather than normal walking only and it would be beneficial for the authors to reveal the individual variance in the thresholds required for different people.

Dead reckoning path reconstruction can be facilitated when sensor data is augmented with the *a priori* knowledge of the environment. When a model of the environment is available, either in the form of a map or a 3D representation, dead reckoning path can be constrained, reducing the processing overheads as shown by Woodman and Harle (2008) who used a foot mounted off-the-shelf inertial measurement unit (IMU) coupled with a model of the building to track reliably the user through multiple floors without knowing the starting position. The approach used additional data from WiFi signal strength measurements to filter out the incorrect hypotheses about the location. The high degree of accuracy of this approach (localising to within 0.5m 75% of the time and 0.73m 95% of the time) as well as its ability to cope with multiple floors (rather than the conventional 2D mapping typical for PDR) make this approach highly desirable. However, the need for the *a priori* knowledge about the building as well as a 2.5D model of it makes deploying the system in a new location cumbersome.

The need for indoor maps is also raised by Foxlin (2005), who describes a shoe-mounted inertial navigation system for PDR. The system combines inertial, magnetic and GPS sensors (GPS is used outdoors only) mounted in a shoe. It provides reliable 3D tracking both in- and outdoors without requiring prior knowledge of the location. The overall results (0.3% path drift) are very promising for situations where a very accurate 3D position is required outdoors. Despite the success of the indoor test using only the inertial sensors, the author

raises a concern that for precise indoor PDR over longer paths additional beacons or floor plan are required.

Another example of environment knowledge being used to refine dead reckoning is shown by Kourogi et al. (2010) who use a set of waist mounted self-contained sensors, namely accelerometers, gyroscopes, magnetometers and thermometers to perform pedestrian dead reckoning and action recognition in an indoor office environment. The dead reckoning module also contains a barometer and an RFID reader, but their use is not stated in the paper nor shown on the process diagram for the final design. By combining the results of their analysis of the temperature-compensated inertial and geomagnetic data with the prior knowledge of the location they are able to filter out mistakenly detected actions as well as improve accuracy of dead reckoning. Despite ambiguity in the hardware description in the paper (such as the number of angular rate sensors being quoted as both 1 and 3) and the need for *a priori* environment knowledge, the overall results provide a reasonable framework for efficient PDR.

Image data can further benefit a pedestrian dead reckoning system. One example of this is the research by Ishikawa et al. (2009b) in use of an inertial pedestrian dead reckoning (PDR) module coupled with the information from the external sources in a typical indoor infrastructure, such as active radio frequency identification (RFID) tags and security camera footage to dynamically estimate gait parameters and reduce the error inherent in dead reckoning algorithms. The approach proposed has merit, relying on an existing infrastructure. Multiple data sources (inertial sensors, magnetic compass, barometer, RFID reader and trajectory from the cameras – videotrajectory) are fused into a single system and matched to a 3D environment model to track the user. However, the need for existing pre-deployed RFID tags, *a priori* knowledge of both the camera and RFID tag locations as well as static camera orientation limit the usefulness of this design. Security camera footage can also be difficult to access for legal and privacy reasons.

An altogether different approach to image data is taken by Zhu et al. (2007). They developed a system for dynamically building landmark databases to enhance the accuracy of a stereo-camera based pedestrian tracking. The system combines forward and backward facing stereo camera pairs with a low cost inertial measurement unit (IMU). The approach was tested with and without the landmark recognition both indoors and outdoors and was proven to increase visual odometry accuracy with the landmark database enabled. The approach is cumbersome and processing intensive, however, limiting its usefulness with the embedded applications.

Pedestrian dead reckoning systems predominantly rely on inertial measurement sensors for their functionality. These sensors show good results when combined with external sensor data, and can also be used in a self-contained manner. A significant amount of research has been undertaken to evaluate the best ways to use such sensors for a robust and reliable navigation as well as way to improve their accuracy and effectiveness. Ladetto et al. (1999) looked at both accelerometer-based PDR and the use of barometric sensors to augment altitude errors in the GPS positioning. While their analysis is dated, it shows that even early accelerometer based PDR systems could track user location with a degree of reliability (1-2 meter precision). Modern sensors provide greater precision and the results from them are likely to be more accurate as shown by further research. Ladetto et al. later (2002) describe the challenges faced when using the PDR in difficult situations (such as combat) and when the walk is unconstrained by forward walking only. They list ways to overcome these difficulties and design a way of integrating the magnetic and gyroscope azimuth estimates without using integration that is described in further detail in (Ladetto & Merminod, 2002). The work presented contains a lot of tests performed and provides a useful reference for any future development in the field of PDR.

The challenges of developing a self-contained PDR system are further described by Fang et al. (2005) who developed an accelerometer/magnetic compass based PDR that uses

wireless network to offload processing to a dedicated node, rather than keeping it on the PDR module itself. The authors review a number of various dead reckoning sensors to compare their strengths and weaknesses, an overview that provides a number of useful insights into the challenges faced by PDR developers and the ways to overcome them (such as ways to compensate for sensor noise and work around sensor-inherent weaknesses). The final design represents a robust indoor/outdoor pedestrian navigation system that does, however, suffer from inaccuracies in step detection (8% error across multiple subjects) and is prone to errors due to magnetic interference. The reliance on the wireless infrastructure also puts limits on the ease of deployment of the system.

Another source of error compensation algorithms associated with the dead reckoning using self-contained sensors is the work by Judd and Vu (2008), who describe a commercial PDR module as well as provide a high level overview of the algorithms employed to correct for the sources of error and interference. The commercial nature of the module limits the amount of detail provided, but at the same time the summary of the error sources and tests performed can be helpful when designing a custom PDR module. The magnetic correlation approach, in particular, can be useful for closed loop scenarios and is described and tested in a variety of environments.

Chen et al. (2010) performed tests on pedestrian dead reckoning step length modelling approaches, determining that the step length is less important in the PDR than the accurate heading determination. They then developed an extended Kalman filter-based model for accounting for predictable compass errors. The tests performed with the new error model showed its merit for use in outdoor PDR scenarios or to aid the GPS. However, the performance of the model in the indoor environment or an environment with a large number of magnetic interference has not been evaluated.

2.3.2. ALTERNATIVE PERSONAL NAVIGATION APPROACHES

While PDR systems represent the standard way of pedestrian localisation, there are a number of alternative approaches to pedestrian localisation that supplement or replace inertial navigation employed by PDR systems. Using mostly image processing methods, Ishikawa (2009a) created an interactive modelling approach to indoor localising that relies on common features of many such places and user interaction to generate a 3D model. By extracting lines from the user photographs and allowing users to annotate them (determine parallelism inherent in typical buildings), the modeller software can find the vanishing points and hence orientation of the different planes in the model. Through user interaction, each photograph can thus be converted to a local model that are later fused using visual SLAM, PDR and visual feature matching into a global model. Modelling of non-planar surfaces is not covered, thus objects such as spheres will be hard to integrate automatically without importing external 3D shape. Camera images are used for texturing this model and additional viewpoints are presented to the user to facilitate collection of missing photographs required. The author doesn't describe the sensors and algorithms used for pedestrian dead reckoning and the approach is reliant on user guidance, but the resultant modelling is efficient and intuitive, making this approach well worth investigating.

A completely image-processing based system is described Aoki et al. (1999). It relies on training for the environment during which a dictionary of trajectories is auto generated from the chromatic histogram. During navigation walk, hue histograms of camera data are matched to this dictionary to recognise visited locations. The approach recognises short sequences of path walked, referred to in the paper as trajectories. This approach uses hue histograms, thus being able to match locations under different lighting conditions and being resilient to camera angle changes. The training stage required for creation of the dictionary is intensive and relies on careful manual selection of trajectories to train for, but the overall

recognition accuracy of around 76% for paths without gaps in them is very reasonable for a real time system.

Another image processing approach is presented by Davison et al. (2003). The authors use a wearable single camera robot that is able to augment its user's positioning with fine adjustment of camera field of view for improved tracking of visual features. Their approach deals with issues of partial scene occlusion and object repositioning by novel single camera SLAM implementation. The implementation is able to generate sets of visual landmarks and match them using a 3D initialisation over multiple frames. The depth information available about the 2D landmark features allows for a robust mapping. The visual nature of the approach facilitates its potential application in remote expert systems that was presented as part of the system evaluation. The approach, however, is only usable in smaller spaces, such as workspace locations because of the processing constraints of the SLAM used.

Video information can be supplemented by audio data and behavioural analysis to track a person as shown by Clarkson et al. (2000) who used hidden Markov model(HMM)-based statistical analysis of the video and audio data from camera and microphone to recognise coarse locations. Their approach is reliant on behavioural patterns of user to distinguish environments, rather than special environmental features. HMM classification was tested and found to be 85-99% accurate in detecting various activities, such as navigating around a house and outside. Despite a need for a training phase and classification of all the potential activities to be identified, this approach shows great promise for coarse localisation.

Another potential source of additional information to supplement imaging data is range data as described by Wither et al. (2008), who enhance an existing head-mounted vision-based orientation tracking system with range data obtained from a laser rangefinder. This permits them to generate 3D enhanced panoramas when user turns in a spot. These panoramas contain depth information via rangefinder processing and allow better occlusion and annotation. Despite the requirement for accurate and lengthy head sweeping for the

generation of panorama, the results contain depth information that is useful for augmented reality purposes, since they allow modelling the environment from multiple viewpoints and inclusion of partially occluded objects into the scene that is not easy to do without depth data.

Kourogi et al. (2001) propose a positioning method based on registration of input video frames against earlier generated panoramas. By using a set of pre-created location panoramas, video frames can be matched to these in real time to obtain user's position. The key development of this paper is in the image matching algorithm, since typical methods struggle when matching images to cylindrical panorama projections. The authors use inertial tracking of camera rotation angles (Kourogi et al., 2000) to augment purely image matching, thus enabling use in location with limited distinctive visual features.

Several years later Kourogi and Kurata (2003b) use a combination of absolute and relative position estimates fused with a Kalman filter to track people. Image matching for absolute positioning requires generating and registering a library of images for each location prior to use. This information is fused with dead reckoning which does not need the *a priori* modelling. The use of sensor fusion of two very different localisation methods improves on the results of either. The only shortcoming of this approach is the need for the image registration before applying the method. Further details about the fusion algorithm used are provided in another publication (Kourogi & Kurata, 2003a).

Moving away from image processing, (Kourogi et al., 2006), described an enhanced indoor/outdoor pedestrian dead reckoning system that can use active RFID tags and GPS data when available to reduce the error typically associated with dead reckoning. Using environment-seeded RFID readers and making users carry active tags, the authors can detect proximity to known points. Similarly GPS positioning, despite its inherent sources of inaccuracy such as bias and multipath propagation, can work in combination with dead reckoning to improve localisation. Tests show the proposed system to outperform pure GPS and pure dead reckoning localisation.

A different RF-based localisation approach is the one described by Bulusu et al. (2000), who showed a pure RF-based localisation within a network of wireless nodes. Their approach relies on connectivity localisation rather than signal strength and uses ratio of data packets successfully received from a given node as the metric of connectivity. With a dense network of nodes this approach can localise to 87% of ground truth in the experimental setup. While heavily reliant on dependable infrastructure, the low cost nature of the implementation makes it a reasonable alternative to other localisation techniques.

2.4. CONCLUSION

Some of the indoor localisation and mapping techniques reviewed in this chapter are designed for autonomous modules such as robots or vehicles. The principles on which they operate, however, can often be applied to people. The rigid mounting sensors, such as vertically pointed cameras (Fu & Yang, 2009) or rangefinders (Prez Lorenzo et al., 2004) are less suitable to wearable designs. Similarly, approaches requiring complete coverage for mapping (Luo & Yang, 2008), are more suitable for robots that can traverse the area for as long and as many times as is necessary to build the appropriate map. Rigid sensor mounting can also complicate the process of photography that is the reason for the localisation efforts by covering the camera viewing angle or precluding accurate camera operation.

Many of the personal navigation methods rely on *a priori* knowledge of the environment or constrain their operating conditions (such as many of the camera-based ones).

The approaches that seem the most suitable for the wearable design are the ones requiring wireless signal data (Papliatseyeu et al., 2009; Shin & Cha, 2010), image data (Aoki et al., 1999; Davison et al., 2003; Clarkson et al., 2000; Kourogi et al., 2000; Kourogi et al., 2001) and the ones based on pedestrian dead reckoning techniques. The wireless-based approaches rely on existing or deployed infrastructure of wireless transmitters and also require a training phase to generate a signal strength map. This limits their use in a

consumer environment. Image processing methods rely on reasonable lighting conditions or texture information and often require *a priori* data, leaving pedestrian dead reckoning as the best candidate for the indoor geo-tagging localisation component.

There are a number of ways to implement pedestrian dead reckoning and a number of sensors that can be applied. Chapter 4 investigates their applicability and performance for the purposes of this research. Before moving to these issues, chapter 3 returns to the question of managing images taken from the same viewpoint.

3. MANAGING GEO-SPATIALLY OVERLAPPING IMAGES

Analysis of localisation techniques in Chapter 2 showed that there are a number of potential methods for tracking the location of the photographer for geo-spatial image arrangement. However, when several photographs are taken from the same viewpoint (same location), simply placing them on map will cause them to overlap, potentially losing complementary information.

This chapter looks at ways of handling such images and arranging them spatially to reduce potential information loss from purely geo-spatial arrangement.

3.1. INTRODUCTION

The practice of taking multiple photographs and then combining them into a single display can be performed for both aesthetic and practical reasons.

Multiple images of the same scene but with different camera settings, for example, can be combined to create a singular image with the fine details and lighting that none of any of its component parts contain.

It is also possible to combine images taken from a camera that is rotated to generate a composite image (a panorama) that covers a large area whilst retaining the accuracy and resolution of its individual parts.

Another way of arranging photographs is used by David Hockney, a British artist who created photographic composites (which he called “joiners”) by arranging artistically multiple close-up photographs of the target scene (Buse, 2010). This enabled rendering of sensation of motion on an otherwise static canvas (Spalding, 1989). A similar artistic technique was used by Papagianis (2009) but this time with video recordings.

There has been research into automating the process of creating joiners. For example, Zelnik-Manor and Perona (2007) developed an image-processing-based solution to auto-arranging snapshots for the joiners using scale-invariant features (Lowe, 2003) that allow identifying keypoints on images even at different scales.

The techniques for fusing multiple photographs into a single display can be subdivided into two broad categories:

- Fusing photographs to produce a photograph with better quality and photographic parameters than any single component part (achieving super resolution, multi-focus photographs and photographs with high dynamic range)
- Compositing photographs to produce an image that covers a greater area (greater view angle) than any of the component parts. This is typically referred to as creating panoramic photographs

3.2. FUSING PHOTOGRAPHS

There are a lot of different approaches to fusing a number of photographs to create resultant higher quality image (Lukac, 2010). These techniques include:

- Super Resolution (SR) imaging, where fusing multiple low resolution component images produces a high resolution image (or a sequence of high resolution images) (Gunturk, 2010).
- Multi-exposure image fusion, represented by High Dynamic Range (HDR) imaging, where an image is generated that contains a wider range of colour values than any of its single components, reducing or eliminating completely inefficiently exposed areas of the component images (Loscos & Jacobs, 2010).
- Multi-focus imaging, which uses images taken with different focus settings to generate an image with multiple planes of focus (sometimes producing image all areas of which are in

perfect focus). Such an image is typically produced to have all the relevant objects in the scene sharp (Malviya & Bhirud, 2009b).

There are additional photo fusion techniques, such as multi-sensor image fusion (Zou & Liu, 2009), but these commonly require fusing images from various acquisition sources and thus do not apply to the problem that this chapter is looking into.

3.2.1. SUPER RESOLUTION IMAGE FUSION

Super-resolution image fusion allows combining several photographs of the same scene into one that has increased spatial detail, recovering the high frequency information that was lost during the imaging process (Gunturk, 2010). It is used in various fields, such as biomedicine, surveillance and high definition television. A large number of approaches to accomplish super-resolution (SR) exist (Gunturk, 2010), manipulating the images in spatial and frequency domains to produce the fused high resolution results.

Gilman et al. (2010), for example, used a 2D model of an image to produce mapping of source images to the final high resolution fused version. They applied linear least squares optimisation for weighing the source image data against the final fused result, showing their model based approach to outperform a number of SR methods tested against. The authors, however, use a largely synthetic data for their analysis, producing high resolution output from artificially down-sampled source images. It would be of greater benefit to show the improvement using more realistic input data.

A way of using image fusion to create super-resolution and even multi-sensor fused data from space sensors is presented by Said et al. (2009). The authors relied on a learning-based approach for data fusion and attempted to recreate satellite image data of a high resolution band from a set of lower resolution band images. The resultant fused image was compared to the actual high resolution equivalent and shown to be effective as a way of interpolating satellite image data. However, the Bayesian network algorithms that the system

is based on rely on training data for efficiency, which meant that the authors required the actual high resolution example data to pre-train the system before the actual data fusion could commence. This limits the viability of the proposed approach for the needs of this chapter.

Bhushan et al. (2010) proposed a framelet-based method for fusing low resolution, noisy and blurred images into a clear high resolution version. The algorithm takes 3 input images and generates a fused super resolution version with reduced noise and sharper edges when compared to a wavelet-based fusion. The authors tested their approach using small sized (256x256) input images with artificial noise and blur added. Despite promising results with the test set, it would be beneficial to further expand on the nature of modifications made to the test images and validate the algorithm with real-world data.

Super resolution approaches suffer from a degree of noise and error in the output image caused by the inaccuracies when registering (aligning) source images. A way to reduce these distortions is proposed by Hua Yan et al. (2008). By modelling the registration and observation errors, estimating them adaptively and applying the model to the fusion process, the authors were able to show improvement in the image quality and peak signal to noise ratio when compared to alternative fusion methods. The authors used simulated data as well as a realistic image sequence to produce super resolution images, showing the approach they presented to work equally efficiently with both.

Techniques of super resolution can be applied to video data as well as static images, as shown by Fedak et al. (2010). They propose a way of improving motion estimation that is a prerequisite of many super resolution approaches by introducing the quality estimation for the motion vectors that affects the accuracy of the resultant fused image. The authors do not present any evidence to back up their claims of improvement, however, nor do they test their approach beyond simply stating it.

Generally speaking, most super resolution techniques rely on a set of input images that are sub-pixel shifted relative to each other. Such shift is difficult to accomplish without synthetic data or specialised hardware. And whilst there are ways of accomplishing super resolution with only a single input image (Jing et al., 2010), the technique does not help solve the problem raised in the beginning of this chapter – what to do with multiple images taken from the same viewpoints, since the images are unlikely to be perfectly aligned and sub-pixel shifted.

3.2.2. Multi-Exposure Image Fusion

Multi-exposure image fusion relies on combining information from several images with different exposure settings to generate an image with high dynamic range (HDR). Dynamic range for imaging purposes is “the brightness ratio between black and white pixels visible on the screen at the same time” (Pirinen et al., 2010). Conventional cameras are unable to capture bright and dark areas of the image simultaneously (Loscos & Jacobs, 2010). There exists specialised hardware that is able to overcome this limitation (Canon U.S.A., Inc., 2011), but for a typical consumer camera, capturing HDR images can only be achieved by multi-exposure image fusion.

Piao and Xu (2010), for example, produced algorithms for estimating the optimal exposure times to be used during multi exposure fusion, thus enabling generating a scene-specific set of exposures to produce an optimal HDR image after fusion. Their algorithms generate a set of up to 3 exposure times for a given camera and environment that was shown experimentally to produce an HDR image that optimally highlights most of the areas of the scene. Their approach, however, makes no mention of image re-alignment, leading to the belief that a static camera re-positioning between shots is required. This makes its use for the purposes of this chapter limited.

A different approach is taken by Mertens et al. (2007). Instead of calibrating the fusion process for the camera response and generating an HDR image, the component images are

fused directly, generating high quality low dynamic range image, making the process very computationally efficient and capable of accepting images taken with a photographic flash. Their algorithm estimates a quality measure (based on contrast, saturation and exposure) for each pixel in every image and thus allows blending the final image out of the weighed parts based on this quality. The authors compared (informally) their approach with several alternatives noting its great efficiency and limitations when a highly varied change in brightness over different exposures is detected. The fusion handles flash photographs without need for alterations but does not completely eliminate artefacts due to the flash. The fusion process, however, again relies on perfectly aligned photographs, falling outside of the requirements of this research.

Görmer et al. (2010) developed a real-time processing capable solution for HDR imaging on automotive cameras, for use in vision-based driver assistance systems. They use a standard image fusing transformations, but developed new ways of calculating optimal exposure parameters for a given scene. The approach relies on a special type of camera capable of producing simultaneous multi-exposure, and thus can be used from a moving vehicle. However, the same reliance on a custom hardware imposes limits on the usability of the algorithms designed with non-aligned images. Despite these limitations, the algorithms were shown to produce reasonable results and present a good enhancement to a vision-based driver assistance system at the image acquisition stage, especially in lieu of their real-time processing capability.

Cho and Hong (2004) designed a method of producing an HDR image based on two low dynamic range images, one taken using digital camera default auto exposure setting and the second one being manually set to show under or overexposed regions from the auto-exposure image. The fusion method is shown to produce reasonable quality images, but relies on the photographer manually adjusting the exposure settings for the second picture to reveal over or underexposed regions of the auto-exposed one. This professional knowledge

requirement makes the approach useful for very specific situations and again does not allow for camera movement between shots.

Kong et al. (2007) use genetic algorithms for image fusion. Images are partitioned into blocks determined by genetic algorithms and then blended together to produce an evenly exposed image. The approach is shown to be able to consume a large number of source images at various exposures and produce high quality results. The approach maximises the amount of information in the final composite and iteratively attempts to improve the end result quality. It relies on aligned images but presents a different way of selecting the regions to blend as well as analyse the quality of such a fusion (using entropy, peak signal to noise ratio and gradient information).

3.2.3. Multi-Focus Image Fusion

Multi-focus image fusion tries to enhance a different aspect of the fused image – its sharpness. Limitations of the camera lenses often prevent an entire scene from being in perfect focus, especially for close up scenes. In this situation, fusing multiple photographs that have a different focus plane allows creating a composite that presents a perfectly focused image. An example of this is the work by Malviya and Bhirud (2009b) that analysed several methods of multi-focus image fusion and showed the results from fusing a set of 7 images of the scene. Interestingly, their findings suggest that out of the approaches tested, basic averaging produces the best results according to several quality measurement characteristics. This contradicts their findings in (Malviya & Bhirud, 2009a), however, where the same test using a different set of images resulted in the simple block replace technique producing the best results.

Zafar et al. (2006) preferred to focus on image fusion in the frequency domain only instead. By using the raw results of the discrete cosine transform performed in the camera as a stage in JPEG image format compression, they were able to gain access to lossless image data and fuse it in the frequency domain. The authors tested their approach with both multi-

exposure and multi-focus images and found it to produce high quality fused images in both cases, comparable to those generated by commercial software and spatial domain fusion algorithms.

Yang Sa (2009) proposed a multi wavelet algorithm for fusing multi-focus images. He performs the fusion of a pair of images in a manner that avoids typical inconsistencies associated with different directions of the wavelet coefficients from different images being fused together. The author compared the results of his algorithm with the alternative single and multi-wavelet image fusion approaches, finding it to be better than either of these. What the paper lacks, however, is a comparison with a non-wavelet based fusion, since findings by Malviya and Bhirud (2009a; 2009b) suggest a non-wavelet approach could produce better results.

Saeedi and Faez (2009) used fuzzy logic principles for selecting which of the wavelet coefficients to use for the fusion, classifying the wavelet coefficients obtained via dual-tree discrete wavelet transform (DT-DWT) into in- and out- of focus ones via Fisher classifier. DT-DWT provides finer frequency decomposition than standard discrete wavelet transform. The coefficients from different images are then fused using a fuzzy classifier, able to handle misclassification from the initial Fisher pass. The authors compare their algorithm to several wavelet and non-wavelet based ones, showing it to be outperforming them according to several quality measurements.

Wavelet transform-based methods, however, frequently lead to contrast reduction. A way to overcome this limitation is by use of a Curvelet transform for multi-focus fusion, as implemented by Qiang Fu et al. (2009). Their approach was tested against wavelet transform and Curvelet transform-based methods with different fusion rules and was shown to provide high quality results with better edge resolution. Advantage of the curvelet-based fusion is further highlighted by Xuelong Hu et al. (2010), who validated curvelet-based fusion against

wavelet and their earlier median pyramid method proving it to produce sharper edges and better contrast.

A neural network-based approach is adapted by Siddiqui et al. (2010) to select between and then fuse blocks from various images. They relied on a genetic algorithm-based approach to select optimal block sizes for the input images and pre-trained the neural network using ten pairs of reference multi-focus images (instead of training the system every time, as was commonly used before). The authors evaluated the results of fusion, showing better quality than several alternative algorithms both for blind fusion (when a reference image is not available) and when comparing the fusion result with the reference image.

3.2.4. CONCLUSION

All of the reviewed techniques for image fusion rely on the source images having large areas of overlap (often near 100%) and sometimes very specific camera settings. Whilst they could be of great use when dealing with the photographs made by specialised hardware (Canon U.S.A., Inc., 2011), a typical consumer camera-produced photographs will require static cameras (use of a tripod) and availability of manual or semi-manual per-photograph settings as well as specialised photography knowledge to be usable.

All of this makes photograph fusion methods suboptimal for the needs of this work, where the photographs are likely to be misaligned and could even be taken at different magnification factors, since most consumer digital cameras nowadays come with at least the most basic “zoom” functionality. Thus it is important to look at the alternative way of dealing with single viewpoint photographs – compositing.

3.3. COMPOSITING PHOTOGRAPHS

Image compositing (image stitching) arranges and alters several photographs in a way that allows creating an image that contains combined visual information from its component parts, effectively increasing the angle of view to encompass the areas covered by its parts. There

are many ways of compositing photographs into panoramas, absolute majority of which rely on the image processing techniques.

A prerequisite of image compositing is that the images partially overlap. An image that does not have overlapping regions with other ones will be impossible to add to a composite using conventional image processing techniques. The photographs for the composite are usually taken with the same camera magnification settings, however, there exist techniques (Brown & Lowe, 2007) for compositing even the images with different scale factors.

There are a number of challenges faced by image stitching approaches, including the aforementioned scale difference between images. Photographs could be taken with slightly different camera settings, affecting the contrast and lighting of the images. Stitching these will produce visible seams where different photographs meet. Photograph compositing also requires deforming them in order to keep the relevant features aligned between the photographs, since individual photographs rarely cover the exact same planar surface (Koo et al., 2009). Therefore a surface such as sphere, cylinder, cube or plane is often selected and photographs are projected onto it for compositing. These and other challenges has been the subject of research in recent years, aiming to discover ways of automating the compositing process and improve its performance and visual quality.

Jia and Tang (2008), for example, investigated means of preventing structure and intensity misalignments in composited images by performing deformation of the images in a 1D plane. Whereas conventionally structure matching was performed in 2D, the authors derived algorithms that allow computing optimal partitions between pairs of overlapped images in 1D, significantly simplifying the ambiguity of structure matching. This allows them to compute a deformation vector that provides smooth transition of both the intensity and structure between two images. Tests of the approach showed it to produce very seamless results. Authors were also able to illustrate the capability of the new algorithms to handle dissimilar

images with varying textures. Dissimilar textured images are still aligned well, even though the results do not look natural, showing the sheer robustness of the approach.

Another image transformation method is investigated by Koo et al. (2009). They researched ways of automatically selecting the correct surface type to use for the stitching, affecting the warping functions required. The algorithm proposed by Koo et al. is able to automatically select a suitable surface and warp the images onto it. It is experimentally shown to produce visually pleasing composites without need for user interaction or target reference plane being contained in one of the images. This algorithm presents a reasonable enhancement to any image stitching software that currently relies on user input for target surface selection.

For the actual stitching of warped images, Koo et al use a combination of various methods, including scale invariant features (SIFT) – based stitching. These are one of the key algorithms in image compositing. Scale invariant features are local image features that tend to remain invariant during rotation, scale and partially illumination change and were described by Lowe (1999). The SIFT approach performs staged filtering on image to detect maxima and minima of difference-of-Gaussian points corresponding to key points in image. These are then converted into feature vectors and used to identify object models. These object models can then be used to align and transform source images during composition.

Scale invariant features are used by a number of different image compositing approaches due to their robustness, allowing identifying matching features on overlapped images. One example of SIFT-based compositing is described by Brown and Lowe (2007). By using SIFT matching across multiple images, the method described is able to stitch photographs that are not aligned along a single row and can automatically handle noisy pictures. The authors showed that SIFT-based stitching can match images despite rotation, illumination and magnification (zoom) differences between them and produce seamless panoramic composites. Moreover, introduction of noise images and images from separate panoramas into the source image pool does not break the compositing process. Separate panoramas

are created and noise images are rejected. The process is fully automated and presents a great advancement in non-interactive image compositing.

A different automatic stitching algorithm is presented by Li et al. (2008), who again use the scale invariant features (SIFT) for the basis of their compositing. This method is designed for work with image sequences and is shown to produce reasonable results even when the images are out of sequence and contain noise images. The blending and warping of images appears to distort the images to a greater degree than the approach outlined by Brown and Lowe (2007), but the overall results look reasonable and the compositing process is computationally simple, using weighted average blending.

Xing and Miao (2007) have developed their own approach to SIFT-based compositing. Instead of using interpolation for the image coordinates, they use coordinates of the images after the transformation for stitching, which allows them to reduce the number and complexity of necessary calculations. The approach provides stitching capability for images with different illumination due to the use of SIFT features. It does not, however, contain any colour equalisation or blending, therefore the transform result contains seams in the stitching area. The authors compare their processing to compositing using software application without stating which application they compare to, making it difficult to validate their findings. Additionally, the transform calculation uses the first of two images as reference plane and warps the second image onto it. This can have suboptimal results, as shown by Koo et al. (2009).

The false edges introduced by the stitching process and evident in many approaches can be minimised, as shown by Zomet et al. (2006). The authors performed a number of tests and analyses on various stitching approaches available at the time and compared their results to determine which one minimises the number of false edges. They developed a framework for image stitching in the gradient domain and showed it to outperform optimal seam, pyramid blending, feathering and Poisson editing approaches. Tests on the images produced very

smooth results. However, the gradient-based approach presented is computationally complex and is significantly slower than the other approaches evaluated, leading Zomet et al. to suggest using it only if the faster approaches fail.

A new stitching algorithm based on image edge feature pixel identification and extraction is proposed by Zhuang et al. (2009). Instead of searching the entire image for corresponding features, only the image edges are analysed, reducing the computational requirements of the process. The algorithm was tested with an image and shown to enable stitching. However, it has trouble distinguishing true edges from edges caused by image noise and thus requires further refinements before it is applied to complex scenes. The authors quote computational efficiency of the algorithm but fail to provide any numbers to justify these claims.

Kang (2010) relies on motion estimation to perform image stitching. He uses block matching based on phase correlation between images, using phase information of the images to match image blocks. This makes the matching process insensitive to brightness changes between images. Once the image blocks are matched, he determines the motion vector between a pair of images and then filters out outlier motion vectors to arrive to the main motion vector required to stitch images. Weighted average method is then applied to the overlap region and the images are stitched and seamlessly blended together. This method is highly reliant on large overlapping regions between images but Kang's tests on a large image sequence show it to have merit. The block (instead of feature) matching nature of the approach, however, requires the images to have the same magnification factor (zoom) and is likely to fail when the images are rotated significantly relative to each other.

Tang and Jiang (2009) proposed using energy maps as the basis for image compositing. They use pre-aligned images (that can be aligned using any other stitching or image registration techniques) and intelligently calculate the optimal seam that will cause the least distortions and artefacts in the resultant composite. Using pixel energy as a reflection of how noticeable the region is to naked eye, the authors calculate an optimal stitching line, that

skirts the objects that end to be more noticeable to people. Tang and Jiang tested their approach with a number of different difficult scenes (including moving crowd and book pages) showing it to require relatively little processing power (8.9 seconds to stitch 6 230x220 pictures on an Intel Pentium 1.7 GHz CPU) and produce results with visibly reduced ghosting, blurring and visible seams.

Another approach for calculating the optimal boundary along which to perform compositing is published by Cao et al. (2010). The authors developed approaches that analyse gradient and curvature of a pair of partially overlapped images and determine a seam boundary that would minimise the errors in both. Their work takes into account gradient and curvature and can therefore overcome photometric inconsistencies between input images that pose problem for standard minimum error boundary approaches to stitching. The authors tested their approach on a number of grey-scale and colour images, pre-aligned using SIFT-based registration, and discovered that their method provides better performance that standard minimum error boundary approaches and handles photometric inconsistencies more robustly. The algorithms presented by Cao et al. are sound and produce good selection of seam line, but lack any image blending for a truly seamless stitching as suggested by the title of their paper.

A number of image compositing approaches focus on making them computationally simple and efficient to enable their use on mobile phone devices. Modern mobile phones often come with built in cameras and have reasonable processing capabilities (HTC Corporation, 2011), making it possible to implement the compositing directly on the device, as shown by Xiong and Pulli (2010b; 2010a). The authors developed fast compositing algorithm for sequential image stitching. This algorithm relies on the images being ordered sequentially and thus can reduce the memory footprint by only storing the current image pair in memory, rather than the entire sequence, which is important for mobile devices with limited resources. The approach generates high resolution and high quality composites despite the limitations of the hardware. The compositing does not compromise on quality and is able to maintain low

resolution version of the image for display purposes, while storing high resolution composite in the phone storage, thus able to composite very large panoramas without running out of memory. The memory usage of the new approach was compared to global panorama stitching and shown to be a significant improvement. Stitching was tested on both indoor and outdoor scenes successfully, showing merit in use of dynamic programming and colour correction for fast stitching.

An image stitching algorithm with even lower processing requirements is published by Chang et al. (2011). Their approach is designed for use on embedded devices and again relies on SIFT features for image matching. The authors chose to reduce the complexity of SIFT feature matching and location by using a down-sampled image for processing and only processing specific image blocks. Simple blending and pixel repairing techniques help reduce the appearance of seam in the resultant composite without significant impact on performance. When tested on an embedded system, Chang et al discovered that an XVGA-sized image (1280×1024) could be produced in 0.6 seconds and have visual quality perceptually on par with that generated by Adobe Photoshop software.

3.4. CONCLUSION

As discovered in the previous subsections, there are a number of different approaches to handling multiple images taken from the same viewpoint. These images can be combined together to either improve the visual quality and details of the combined image, such as via multi-focus or super resolution image fusion. Or, alternatively, the photographs can be stitched to form panoramic views that provide a greater field of view than individual photographs.

Image fusion techniques require the source images to be registered and perfectly aligned, typically relying on nearly complete overlap between the source images, which is difficult to achieve with a handheld camera or without specialised hardware. They also require

consistent magnification between the images, denying the users free use of the zoom technology. This makes the image fusion approaches a suboptimal choice for the purposes of this research.

On the other hand, the image stitching (compositing) approaches require only partial overlaps between the source images and can handle images which are at different magnifications, lighting or even rotated relative to each other, when using feature-based stitching, such as SIFT-based (Lowe, 1999) approaches.

The common limiting factor of all the image stitching approaches remains, however, the need for the source images to be overlapping. During normal photographing, camera operator will not necessarily concentrate on providing overlaps between different photos. If the camera user is trained to maintain overlapped photographs, it is unlikely that he will be able to do this efficiently without a fixed camera position and some sort of real time feedback on the areas covered by photographs already. Such instant feedback is not normally available in consumer digital cameras, making the process of generating panoramas less straightforward. Whilst sequential overlaps would be fairly easy to track with an image processing software (where overlapping photographs are taken one after another, gradually covering greater area), if the photographs are taken out of sequence, the processing requirements will grow considerably, since all of the images will have to be retained in memory and analysed. Xiong and Pulli (2010b; 2010a) illustrated the high memory cost of handling multiple images in their research. There is a need for a more robust system for providing real-time feedback to the image compositing, which will allow photographers to realise complex panoramas using images with different camera settings and orientation.

Image processing techniques rely on identifying matching features in separate photographs for compositing. Even when using scale invariant features (SIFT) image processing will only be able to arrange photographs after an overlap is available. While SIFT helps solve the problem of different magnification levels and rotation in the images, the problem of need for

an overlap remains unsolved. This is where additional sensors can provide information to augment or even potentially replace image processing. Cheng (2007) patented a way of automating panorama generation using gyroscopes or other rotary sensors. As the camera turns around its vertical axis, it automatically takes a snapshot once the rotation exceeds the threshold. Another similar invention, patented by Hubel and Genetten (2005), presented a similar device for automated panoramic image triggering, this time based on an optical mouse sensor optionally replaced by a gyroscope. Mlgaard (2004) patented a way of using partial orientation information to help detect vibrations in camera. Ishikawa (2009a) used sensor information about camera orientation to help guide users to fill in missing fragments of the 3D model his system generated.

These patents and designs show the potential of using additional sensors to automate the process of image capture for composite panoramas. Considering the orientation tracking requirements of Pedestrian Dead Reckoning (which was chosen as the photographer tracking mechanism after research analysis in chapter 2), camera-mounted orientation sensors might be able to provide the necessary instant feedback to the camera user or even perform the entire compositing process without the need for any image processing. After all, camera orientation is directly related to the area captured by it, and by relating all of the images in the same coordinate plane provided by the orientation sensors, it becomes feasible to expect the individual photographs to align correctly to each other, facilitating or even rendering unnecessary the image processing-based compositing.

For this to be possible, information about the camera orientation must be available. By knowing the relative orientation of the camera between shots, its field of view can be estimated and the resultant images aligned to correspond to their arrangement in the original scene being photographed.

Both camera orientation tracking and pedestrian localisation require a set of sensors and processing algorithms to be determined that would provide adequate accuracy and efficiency

for the purposes of this research. Chapter 4 looks at possible sensors and algorithms and determines their suitability for this research.

4. SENSOR SELECTION STUDY

Research described in chapters 2 and 3 highlights that there is a large number of sensors that can be used to achieve both photographer localisation and sensor-based image compositing.

For the pedestrian dead reckoning approach, chosen in 2.4 for the purposes of this research, two types of data should be obtained from sensors – distance and direction of any displacement occurring when the photographer moves around the environment. Such distance is normally determined via step counting and direction obtained by tracking yaw (heading) angle.

For the compositing of photographs via sensor data, the camera orientation should be determined. This consists of determining 3 Tait-Bryan angles – pitch, roll and yaw (Diebel, 2006). A way of determining pitch and roll angles has been standardised as estimation of angle between the sensor and gravity normal determined via standard accelerometer sensors that provide very accurate results when the sensor is stationary (which camera can be assumed to be when taking photographs under normal conditions) (Luczak et al., 2006; Eric Tseng et al., 2007; Miles, 1986).

Yaw (heading) estimation both for compositing and pedestrian dead reckoning is more difficult and has several different potential data sources described in this chapter.

Localisation for pedestrians has its own nuances and limitations, largely based on the nature of human gait and the degree of randomness and variability in gait parameters for different people. This chapter looks at ways that the pedestrian localisation can be accomplished and investigates what sensors would be required.

4.1. Potential Data Sources

The work outlined in chapters 3 and 4 highlights a number of potential data sources that can be utilised for the pedestrian dead reckoning and compositing within the constraints of this research (outlined in 1.1). The pedestrian dead reckoning requires both heading and distance travelled to be accurately tracked. Distance travelled can be approximated by using pedometry (step counting), knowing that the stride length for normal walking is related to the walking speed (step frequency) (Feliz et al., 2009) and as such tends to remain reasonably constant for the same person.

4.1.1. Sensors Used for Turn Detection

The heading detection is an essential part of the pedestrian dead reckoning and is also required for camera orientation detection for sensor compositing. Its accuracy is even more important to the successful dead reckoning algorithm than step detection (Chen et al., 2010). This makes the selection of sensors for heading determination an important factor affecting the success or failure of the overall system.

In a typical building the corridors tend to form right angled turns (Kim et al., 2009a; Kim et al., 2009b), facilitating heading estimation by constraining it to detect the occurrence of a turn rather than its magnitude. Without this constraint, the system needs to continuously estimate the heading magnitude. However, during walking human beings tend to choose the shortest path to a destination, not necessarily following the right-angled turns in corridors precisely. This makes unconstrained turn detection method more robust and usable in wider selection of locations.

4.1.1.1. Gyroscopes

A gyroscope (angular rate sensor) is designed for detection of the rotation of the sensor around its axis of sensitivity. This makes it a natural candidate for heading change tracking. Gyroscopes do not provide the absolute heading; rather they give the rate of heading change

at the sampling instance (first derivative of the turn angle). This means that the heading itself can be obtained by integrating the gyroscope readings.

Gyroscopes suffer from a number of sources of errors (Fang et al., 2005), affecting the accuracy of their readouts and making those accurate only for short term measurements (Feliz et al., 2009). Gyroscopes are very sensitive to temperature and typically require temperature compensation (that comes built in with higher end models). They suffer from bias drift and measurement noise requiring extensive signal processing to compensate for, sometimes more than an embedded microcontroller can provide.

Integrating a noisy gyroscope signal leads to errors accumulating and quickly reducing the accuracy of the measurement (Feliz et al., 2009), making them less used for continuous measurement systems and typically employed to augment another way of heading measurement (Ladetto et al., 2002).

4.1.1.2. Magnetometer (Digital Compass Sensor)

Magnetic compass sensors are designed to provide the absolute heading measurement relative to the geomagnetic North. They accomplish this by measuring the magnetic field using magnetometers and hence inferring the direction the sensor is facing.

Magnetic compass sensors come in standard and tilt compensated variety. Standard magnetic compass sensors require the sensor to be positioned parallel to the surface of the Earth. This is caused by the fact that the lines of geomagnetic flux are not parallel to the surface of the Earth and the magnetic inclination (angle between the horizontal plane and the lines of the magnetic flux) causes the tilt in the compass to be incorrectly detected as the alteration in the heading.

Tilt compensated magnetic compass sensors include a tilt detection mechanism and are thus able to accurately detect the heading even when the compass is not level with the horizontal plane.

Magnetic compasses, however, are vulnerable to magnetic interference from large metal structures that are present in the indoor environment, such as support beams and household appliances. This makes them less useful for indoor use (Feliz et al., 2009) . Additionally, the uneven nature of human gait tends to lead to mistakes in heading detection for both standard and tilt compensated compasses. The standard ones cannot handle tilting and the tilt compensated ones rely on accelerometers for tilt angle calculation. During walking, accelerometers are affected by the acceleration due to stepping and forward momentum, thus reporting incorrect tilt angles leading to inaccuracies in the overall heading estimate (Chen et al., 2010).

4.1.1.3. Global Positioning System

GPS-based heading sensing relies on a clear signal from a number of satellites (minimum of 4) in geo-synchronous orbit of Earth. Using time difference analysis on the signals from multiple satellites (whose positions are geographically static), the position of the receiver can be determined. Change in position can then be used to determine the direction of travel.

GPS could replace the use of the pedestrian dead reckoning, since GPS already provides full location information. GPS accuracy, however, is highly dependent on environmental conditions and the type of the GPS receiver in use. Some technologies, such as assisted, differential GPS or wide area augmentation system help improve the inherent error in the GPS positioning lock. The number of satellites locked also improves the results.

Even the best GPS receivers, however, can have an error in their location estimation of several metres and are typically designed for outdoor use. Commercial grade GPS receivers were shown to provide an accuracy of 3m in open outdoor environment (Khoomboon et al., 2010; Lehtinen et al., 2008), deteriorating to 20m in urban and semi-urban environment and dropping further to over 150m when only the minimum amount of 4 satellites were locked (Lehtinen et al., 2008). Whilst there is a number of ways to improve GPS accuracy (Khoomboon et al., 2010; Nakajima & Tanaka, 2004; Venkatraman et al., 2010) and research

exists into the high sensitivity GPS implementations usable inside buildings (van Diggelen, 2002; Schon & Bielenberg, 2008; Jianjun et al., 2009), their use indoors remains limited (Piras & Cina, 2010), thus making GPS a suboptimal choice for the purposes of this research.

4.1.2. Sensors Used for Step Detection

Step detection forms the second half of the pedestrian dead reckoning. By detecting and counting steps and knowing the heading, the path taken by the users can be recovered from a known starting position.

4.1.2.1. Accelerometer

Accelerometers are widely used for the step detection (Jirawimut et al., 2003; Kourogi et al., 2010; Fang et al., 2005; Ladetto et al., 2002; Foxlin, 2005; Judd & Vu, 2008). During a step, accelerometers measure a vertical impact of the foot colliding with the ground. A simple peak detection algorithm (Gusenbauer et al., 2010) or a more advanced implementation such as zero-crossing algorithm (Ibarra Bonilla et al., 2011) can be used to detect the step occurrences with a reasonable degree of accuracy.

Using pattern analysis, the step detection can be made even more robust (Kim et al., 2004) and even enable detection of lateral movement, such as sidestepping (Ladetto et al., 2002).

The accelerometer can be mounted on the torso of its user or even held in hand (Gusenbauer et al., 2010), provided the multi-axis accelerometer is used.

Accelerometer use allows for the zero velocity updates to improve step detection and gait parameter determination (Chen et al., 2010; Ladetto et al., 2002; Naikal et al., 2009; Jimenez et al., 2009). A foot mounted accelerometer can be analysed in the periods of rest corresponding to the time in the gait when the shoe lies flat on the ground and as such is not experiencing acceleration aside from the gravity normal.

4.1.2.2. Gyroscope

A leg or foot mounted gyroscope can offer a reliable way of detecting steps and distinguishing between the type of travel that occurs (walking running and moving up/down the stairs) (Lim et al., 2008). Gyroscope is able to detect the movement of the leg corresponding to the walk rather than the impact of the foot on the ground.

The relative merit of gyroscope and accelerometer based step detection varies depending on the techniques and processing involved (Lim et al., 2008; Jimenez et al., 2009) indicating the need to re-evaluate the two methods in context of this research.

A peak detection algorithm can be efficiently utilised to deduce the number of steps taken and the times of their occurrence.

4.2. STUDY AIM

Because there are many different sensors that can be used for step detection and heading tracking, a study needs conducting to determine their relative effectiveness and performance.

The aim of this study is to gather data from a pedestrian equipped with sensors to determine the best way of tracking actions such as walking and turning. Study relies on inertial and geomagnetic sensors for the detection of walk actions and uses absolute reference provided by a video recording to evaluate sensor data post-processing effectiveness.

4.3. STUDY DESIGN

Each participant walks a rectangular path around a table (180 x 120 cm) at their natural walking speed. They walk 10 circles counter-clockwise around the table. A video camera is used to capture their walk and provide absolute reference:

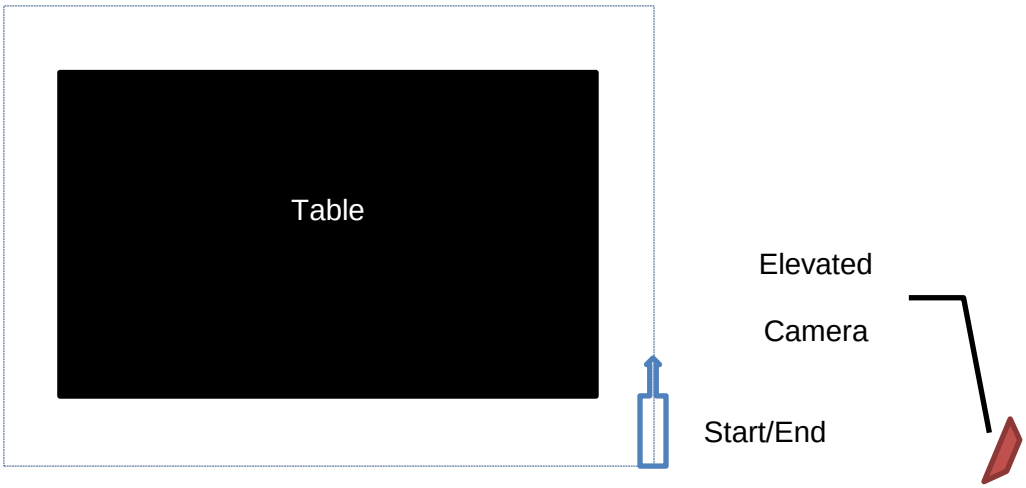


FIGURE 3. STUDY PATH LAYOUT. A RECTANGULAR TABLE IS USED AS THE BASIS FOR THE PATH. AN ELEVATED CAMERA IS USED TO PROVIDE ABSOLUTE REFERENCE DATA

Whilst walking, each participant wears a recording module that records the readings of sensors mounted on the participant.

Turns taken by the participants are determined from the torso mounted compass and gyroscope and compared to each other with the video camera serving as the absolute reference. Similarly, steps taken by each participant are counted on the video recording and the results compared to the outputs of step detection algorithms run on leg gyroscopes (positioned on lower leg, as this provides better detection results (Tong & Granat, 1999)) and the torso accelerometer to determine best sensor for detecting either activity. Torso was chosen as the mounting point for accelerometer to minimise the tilt of the sensors during walking (Ceccato et al., 2009; Stirling et al., 2003).

Each participant thus produces 10 sets of data for analysis and 8 participants that have been used for this study therefore make 80 sets of data.

4.4. STUDY PROCEDURE

The study took place at the University of Birmingham and involved 8 participants, who undertook the study separately without seeing or knowing about the study procedure beforehand.

Each participant was led into the room where the study took place one at a time. They were then told that they needed to walk around the table at their normal pace for 10 cycles counter-clockwise starting from the position indicated as starting. The participant was then shown a typical walk performed by the study conductor who made one circle around the table starting and ending at the position closest to the camcorder.

Participant had the recording module strapped to them, making sure the straps were tightened comfortably and did not restrict their movement. They were then given the camera module with the category selection button pad, led to the starting position and oriented in the direction of travel. Camera module, more specifically, its number pad, acted as a start button, which was explained to the participant.

The camcorder was then turned on and finely adjusted to make sure the entire table was visible, especially the area where participant's legs would be on the far side of the table from the camcorder.

Participant was told to press the top button on the category selection pad when they were ready and to come to a stop once they've completed the last circle (which would be counted by the study conductor and verbally propagated to the participant).

At this stage, participant pressed the start button, initiating the recording and producing an audible tone recorded by camera for synchronisation purposes.

Participant began walking around the table while the conductor remained near the camera to adjust it if required. When participant started their last cycle, they were reminded to stop and hold once they were back in starting position.

Once a participant completed the last circle, camcorder was turned off and then the recording module was powered off and unstrapped from the participant.

Sensor recording log was the transferred to PC storage and marked with participant identification and the study continued with the next participant until all 8 have completed it.

4.5. SENSOR RECORDING MODULE DESIGN

The sensor recording module reads data from the sensors and stores it on a removable storage to facilitate transfer of data to the desktop for post processing and analysis. .NET Micro Framework was chosen as the embedded platform for the implementation of this module. It allows development and debugging using high level object oriented programming language (C#) and data structures used on embedded device are in many cases directly portable to the full desktop version of C# applications. The design of the recording module is summarised on Figure 4:

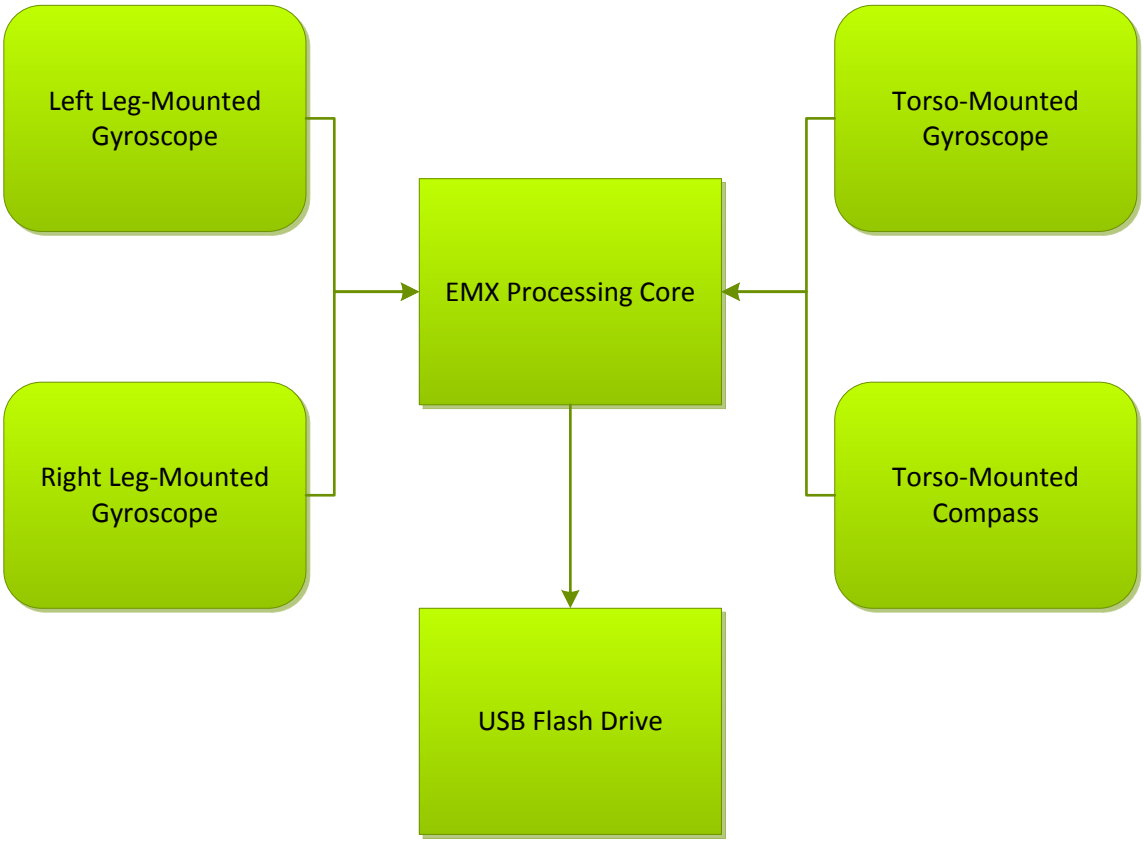


FIGURE 4. DESIGN OF THE SENSOR RECORDING MODULE

The GHI Electronics EMX module was used as the main processing unit for the recording module, in the form of FEZ Cobra development system. This module is based on the LPC2478 microcontroller by NXP Semiconductors and offers enhanced framework features and support including ability to access removable storage and operate with file systems recognisable by both desktop computers and the embedded platform. This significantly simplifies development and iteration of the software design.

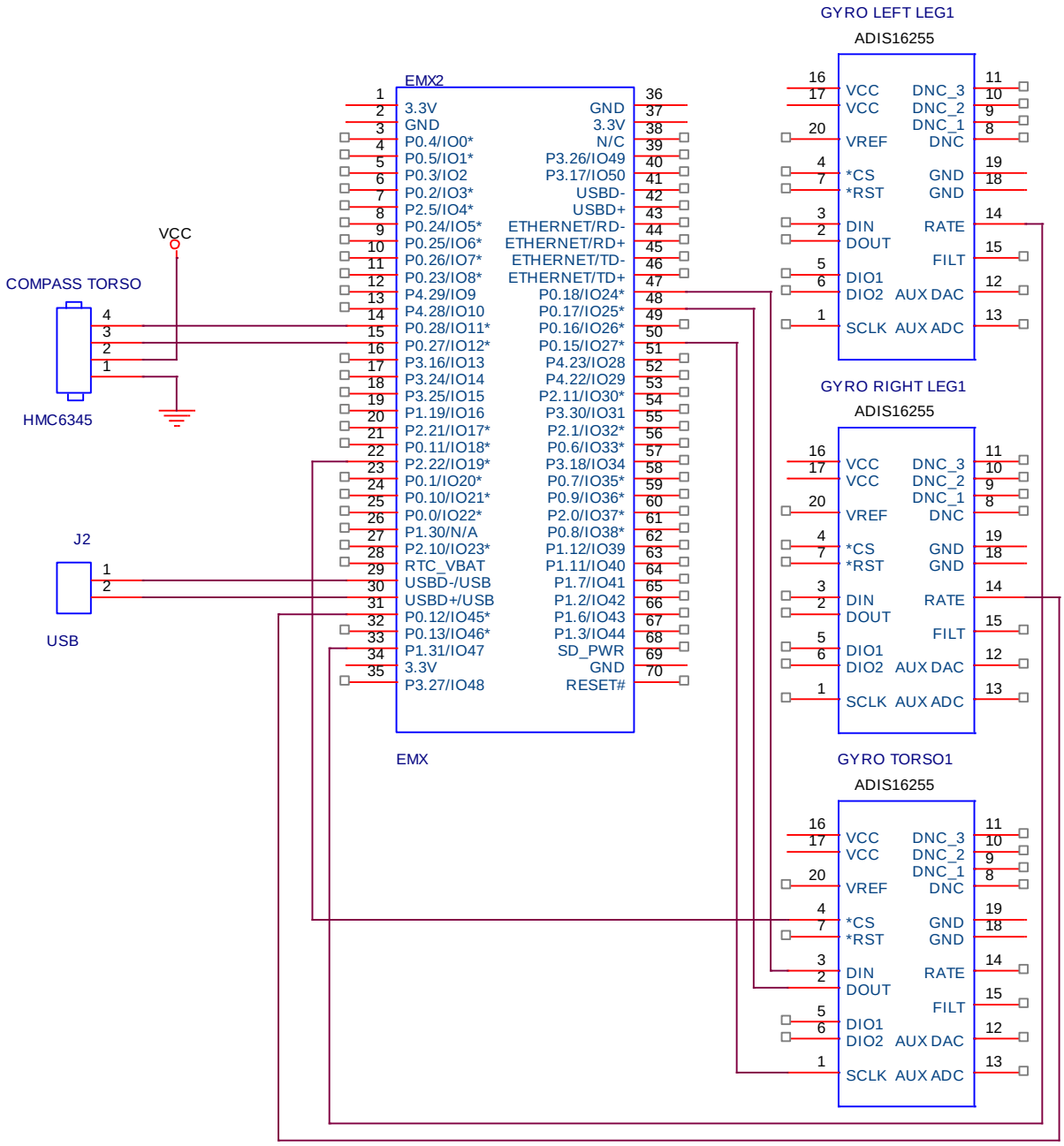


FIGURE 5. SENSOR RECORDER OUTLINE SCHEMATIC. ONLY THE MAIN CONNECTIONS ARE SHOWN.

The module is powered by a 12V Lithium Polymer battery via a switched mode step down power supply based around a DimensionEngineering DE-SWADJ regulator providing greater power efficiency than the use of conventional linear regulators would. The power supply noise generated by the switched mode regulator (inherent to these kinds of power supplies even after filtering) was determined experimentally to be low enough as to warrant the use of

this type of component, which is also confirmed by reviewing datasheets of the relevant components.

EMX module is connected to the sensors represented by gyroscopes, accelerometers and magnetic compass. Angular rate data is provided by 3 ADIS16255 digital gyroscopes – two mounted on participant's legs and one on the torso heading tracking module (attached to the back of the neck). The advantage of this digital gyroscope is in the high accuracy and stability of its readings and its ability to output internally integrated angle reading in addition to angular rate data. This is of limited use on leg mounted gyroscopes for step counting, but is essential in heading tracking of the torso unit.

Compass and accelerometer readings are provided by HMC6343 tilt compensated compass on the torso module. This compass can output tilt compensated heading as well as readouts of 3-axis accelerometer it relies on for tilt compensation as well as raw readouts from the 3-axis magnetometers used for geomagnetic field detection.

HMC6343 is connected to the EMX processor through I^2C bus and outputs all of its readouts on request. Heading tracking ADIS6343 gyroscope is connected to SPI bus of the EMX controller and can send both raw angular rate and the integrated angle reading. During testing, however, it was discovered that multiple ADIS16255 gyroscopes cannot be connected to the same SPI bus because of incompatibility in the current ratings of devices and controller with the lengths of wires required for mounting sensors on legs and neck. Several alternatives were considered, including use of external SPI bus controllers and use of secondary SPI bus on the EMX module. However, because only the angular rate reading of the leg gyroscopes is needed, these were instead connected to the analogue-to-digital converter (ADC) modules of the EMX microcontroller. ADIS16255 are able to output angular rate reading in different forms, with analogue output being one of them and this proved to be the most robust way of dealing with leg gyroscopes.

Data from sensors is stored on a USB (Universal Serial Bus) storage device. Initial design used a human-readable comma-separated values (CSV) file format for storing sensor data. . Initial testing showed that the writing speed of the USB flash drive is the bottleneck for the sensor recorder. Since the maximum normal human step rate is 3Hz (Fang et al., 2005) and a single peak in the leg-mounted gyroscope reading corresponds to a step, the minimum required sampling rate is 6Hz to avoid missing out on potential steps.

The turning rate is slower than the walking rate, at less than 0.7Hz in average person (Hase & Stein, 1999), falling to under 0.4Hz in elderly people (Skrba et al., 2009), suggesting that minimum sampling rate of 6Hz should be adequate to detect turn occurrence.

Tests with the sensor platform found the sustained parse rate to be a minimum of 5.18HZ and maximum of 8.62Hz, averaging 8.13Hz. The .NET micro framework performs garbage clearing regularly that slows the recording to 5.18Hz. On average, however, the recording rate is sustained around 8Hz mark.

This suggests that the target 6Hz rate is not achievable reliably because of the specifics of the platform and required reconsidering the storage format.

Close analysis of the performance of the different subsystems on the microcontroller revealed that string processing required for CSV data file generation to be the reason for low sustainable recording rate.

Altering the data storage format to a simple binary file allowed the data recording rate to increase to average of 30Hz, which surpasses the minimum requirement for the detection of both steps and turns.

4.6. SENSOR PROCESSING.

Different sensors require different processing to extract information from them. Since the purpose of this study is to determine the best way of detecting step and turn occurrences, choice of the processing method is part of its goal.

4.6.1. STEP DETECTION

The simplest and most straightforward method of step detection for both accelerometer and gyroscope is threshold-based peak detection. Because of the cyclic nature of human gait, peaks in acceleration or angular rate will correspond to the step occurrence.

Peak detection algorithm moves through the sensor data until it exceeds a specified calibrated threshold. At this stage algorithm considers a peak to have occurred and switches to watching for sensor data to fall below this or different threshold (depending on implementation), at which stage algorithm resets and searches for the next peak (Kourogi & Kurata, 2003b).

An alternative to peak detection is a zero crossing algorithm. This algorithm estimates average magnitude of sensor signal and subtracts this value from the sensor data. The task of step detection then becomes task of detecting when sensor data crosses magnitude of zero in positive or negative direction (depending on sensor mounting and implementation) (Ibarra Bonilla et al., 2011).

In case of gyroscopes, both of these approaches need to be implemented on sensors from both legs and results combined. In case of accelerometer, a further nuance is possible. In addition to investigating just the Z (vertical, closest aligned to gravity) component of acceleration, peak detection and zero crossing can be applied to the magnitude of total acceleration that is defined as:

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2} \qquad \text{(Eq. 4.1)} \quad [1]$$

Where a_x, a_y and a_z are accelerometer readings and $|a|$ is the magnitude of total acceleration (normalised acceleration).

[1] (Ibarra Bonilla et al., 2011)

During each walk along the path, each participant takes a number of steps that can be calculated from the video recording for absolute reference. Each algorithm will then attempt to estimate the number of steps from sensor data and their results compared to reference figure to quantify performance

4.6.2. HEADING TRACKING

There are two main methods for heading tracking that are evaluated in this study – use of magnetic compass and use of a gyroscope. Both have very different strengths and weaknesses that are summarised in Table 1. In addition to tilt compensated compass, a non-tilt compensated 2D compass will be used. Since torso mounting of the heading tracking sensor platform reduces the theoretical tilt experienced during standard locomotion to a maximum of 5 degrees (Ceccato et al., 2009), the fact that 2D compass is not affected by acceleration due to walking might give this compass an advantage over tilt compensated 3D one (Chen et al., 2010). Because HMC6343 exposes magnetometer readouts directly, the values of two axes perpendicular to the gravity normal will be used to estimate heading as follows:

$$\alpha = \tan^{-1}\frac{m_y}{m_x} + \beta \qquad \text{(Eq. 4.2)}$$

Where m_x and m_y are horizontal components of the magnetic field and β is the magnetic declination (angle between magnetic and true north). The value for magnetic declination is known for specific geographical coordinates and is approximately 2 degrees in Birmingham, UK (Finlay et al., 2010; National Oceanic and Atmospheric Administration, accessed 2012).

TABLE 1. HEADING TRACKING METHODS COMPARISON

Data Source/Method	Strength	Weakness	Potential Use
Gyroscope	Not affected by acceleration or	Short term accurate but drift makes	Can be used to validate readings from alternative

	magnetic interference	readout increasingly erratic. Requires integration to obtain heading information	data sources and in integrated form as source of heading data. Needs absolute reference to reset the integration and drift errors
Tilt compensated magnetic compass	Very accurate when affected by only gravity acceleration	Readings suffer in presence of magnetic interference and when walking due to tilt compensation algorithm being unable to distinguish between acceleration due to gravity and due to platform movement	Using acceleration magnitude detector, acceleration-caused errors can be reduced. Reading are very reliable when user is stationary, such as when taking a photograph
Purely magnetometer-based compass heading estimation	Is not affected by acceleration errors	Can only produce approximate direction due to the platform tilt due to walking errors. Affected by magnetic interference	Can be used in place of tilt compensated compass when the sensor platform is affected by accelerations due to walking

Heading tracking will not be constrained to 90 degree increments and therefore cannot be processed in a discrete manner like the step detection. Instead, two attributes of the heading tracking have to be looked at – turn detection accuracy and heading stability during walking.

There are 4 turns made in each set of data (i.e., one at each corner of the table around which participants are walking). Timings of these turns can be determined from the video recording.

We can assume that the magnitude of the turns should be approximately 90 degrees (since that is the corner angle participant has to turn to navigate the path). Each of the heading tracking methods will produce an estimate of the magnitude of turn detected and this will then be compared to the expected turn magnitude making it possible to quantify accuracy of turn estimation by different sensors.

In addition to turns, for every set of data participant traversed two of the longer sides of the table where their heading remained static and they moved forwards. Because walking introduces noise into the sensor data (be it from acceleration or natural movement of human body through gait phases), the stability of the heading tracking can be evaluated by looking at the deviation of the sensor data during these straight segments (again identifiable from the video recording). Deviation of data will permit quantifying and comparing the heading tracking stability as another measure of performance of different tracking methods.

4.7. Absolute Reference Annotation

To accurately evaluate the performance of various step and turn tracking methods, accurate absolute reference must be available. Such a reference was obtained by annotating video recording of the study using ELAN multimedia annotation framework (Brugman et al., 2004; Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, 2012). The sensor recorder emits a high pitched sound at the beginning of its operation that is synchronised with the video recording where the pitch is present (camera is set to record both video and audio for this very reason). Activities on the video (such as steps, turns, straight lines and complete circles) are annotated with their beginning and end times marked (where applicable).Due to synchronicity achieved via audio signal matching to the beginning of the sensor recording, timestamps produced by ELAN are thus corresponding to the timestamps of the sensor recordings.

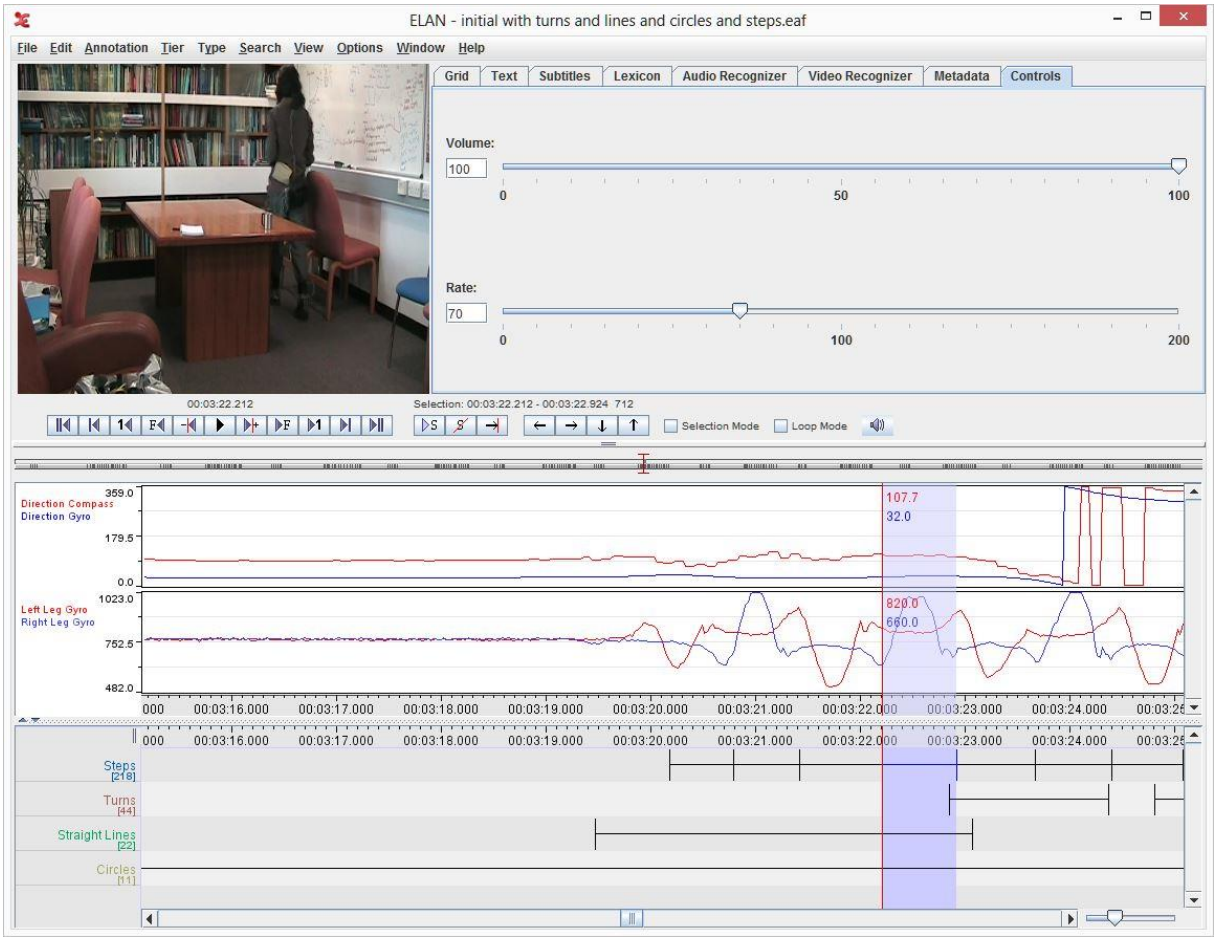


FIGURE 6. VIDEO ANNOTATION AND SENSOR SYNCHRONISATION WITH ELAN

Activities in ELAN are exported to tab delimited data file that can be imported into the analysis software and used as absolute reference for comparison purposes.

4.8. STUDY RESULTS

Before progressing to the results of the study themselves, it is important to properly calibrate the algorithms used for the relevant parts of the processing. These include threshold values for the step counting with gyroscopes and thresholds for the step counting with both normalised and Z-axis accelerometer.

4.8.1. THRESHOLD CALIBRATION

As thresholds will directly influence the performance of relevant step detection mechanisms and their calibration is an important process. Moreover, by determining the optimal threshold

for each data set, it becomes possible to determine how variable these are between different experimental data sets.

Knowing the ideal step counting result for each data set (obtained from the annotations produced with ELAN on video recording), values of different thresholds can be evaluated to determine how close the step counting results produced by threshold-based algorithms are to the expected ones.

4.8.1.1. Gyroscope Threshold

Figure 7 shows results of threshold analysis of gyroscope data. Threshold values are represented as ADC count offsets from the accelerometer readout produced with static sensors. Each ADC count corresponds to approximately 2.4 degrees per second (considering the default rest reading for zero being at 765 ADC counts and default sensing range of ±320 degrees per second with the device ability to drive analogue output between 0 and 2.5V).

Errors at different thresholds are represented by the ordinate axis whereas thresholds themselves are plotted as the abscissa. Points for lowest error for different participants (represented by different lines) all appear fairly consistent, especially looking at the error bars representing single standard deviation spread of samples.

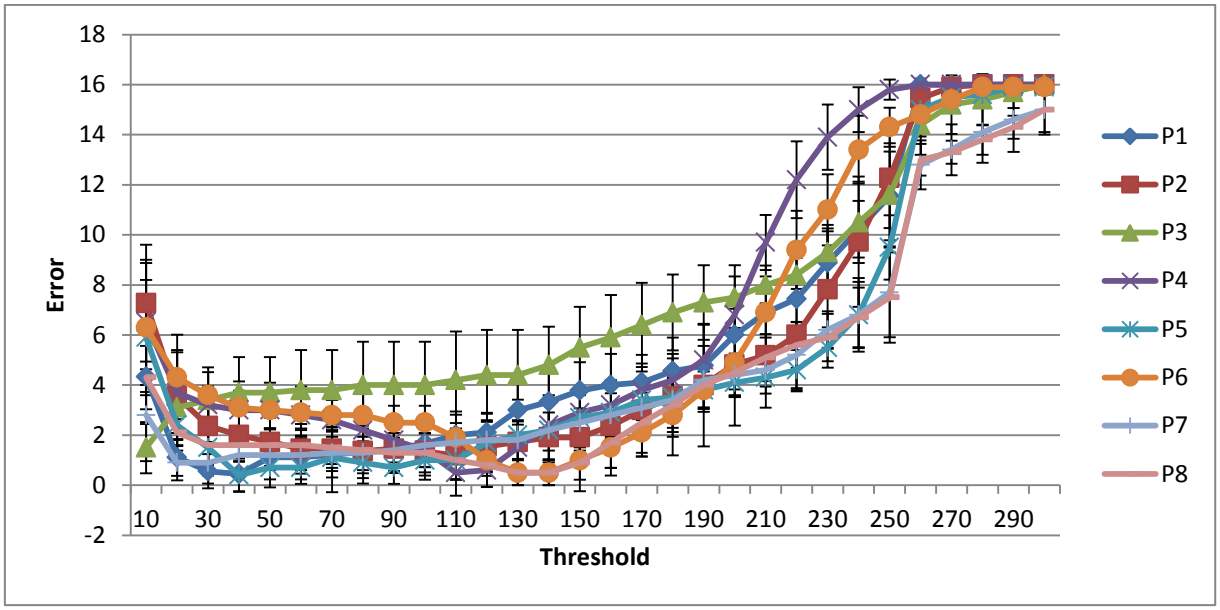


FIGURE 7. STEP DETECTION ERROR FOR DIFFERENT GYROSCOPE THRESHOLDS. ERROR BARS ARE ONE STANDARD DEVIATION

The range of potential thresholds is wide (as indicated by error bars and overall shape on Figure 7) but reasonably consistent between participants (as shown by the visually similar lengths of these error bars and the shape of the error distribution curves). The optimal threshold can best be determined from the combined results:

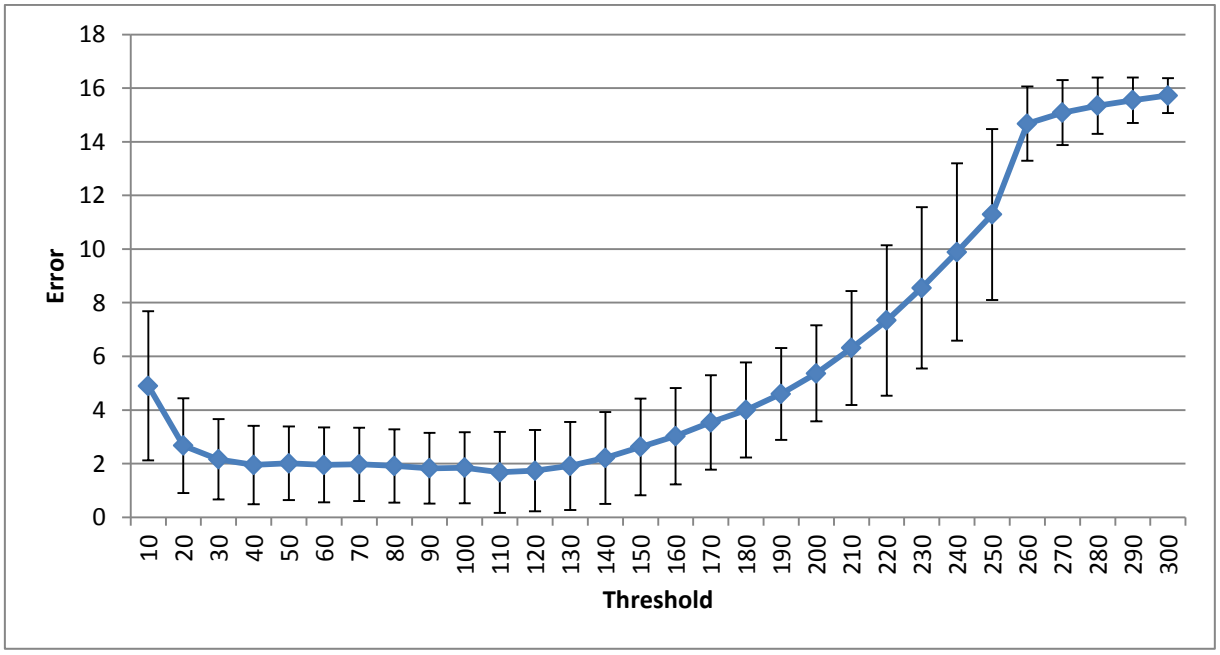


FIGURE 8. STEP DETECTION ERROR FOR DIFFERENT GYROSCOPE THRESHOLDS ACROSS ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION.

Value of 110 ADC counts provides the lowest common threshold for all participants and will be used as threshold for step detection. An alternative approach might be to define a unique threshold for each participant. However, not only would this require training and tuning for every new user of the system, it also makes comparison across participants somewhat problematic (unless one normalises the data post-recording).

4.8.1.2. Single Axis (Z) Accelerometer Threshold

Data for the single axis accelerometer thresholds differs widely between different participants. Figure 9 shows how optimal threshold for one participant corresponds to non-detection for another and potentially multiple false positives for yet another participant.

Every participant appears to have their own optimal threshold that depends on their gait. Unlike gyroscope, a common threshold is difficult to define, indicating need for individual calibration for each prospective user when relying on the single axis torso-mounted accelerometer.

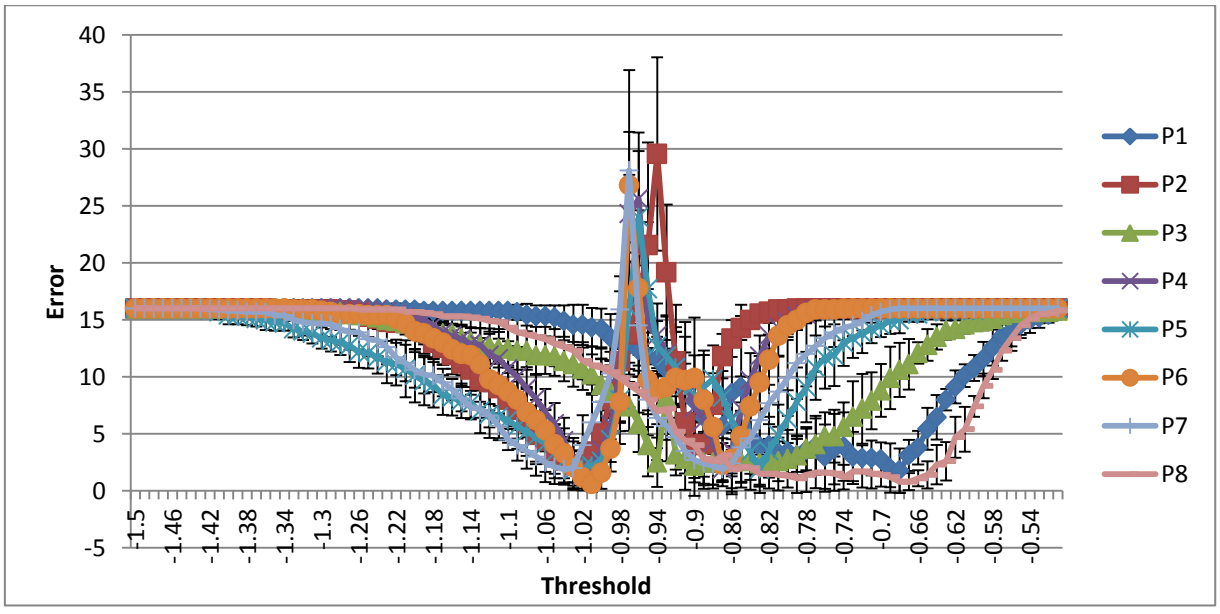


FIGURE 9. STEP DETECTION ERROR FOR DIFFERENT Z ACCELEROMETER THRESHOLDS. ERROR BARS ARE ONE STANDARD DEVIATION

There is a lot of variance in the optimal threshold values between participants, making picking a common threshold difficult.

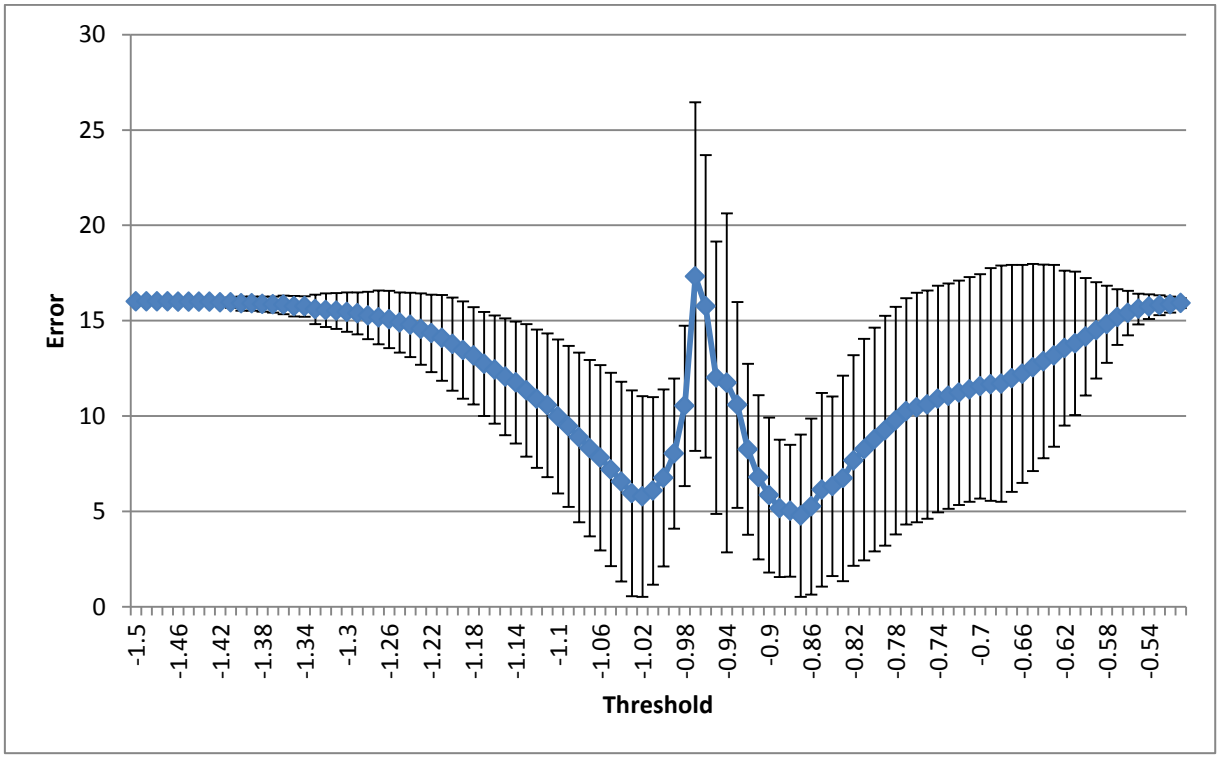


FIGURE 10. STEP DETECTION ERROR FOR DIFFERENT Z ACCELEROMETER THRESHOLDS ACROSS ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION.

Combined graph for all participants highlights this variance, having large error bars representing single standard deviation of the results. The lowest common result appears to fall on the threshold of -0.88g value and this is the value that will be used for the analysis.

4.8.1.3. Absolute (Normalised) Accelerometer Threshold

Normalised accelerometer threshold looks at the combined magnitude of all accelerations affecting the torso sensor platform.

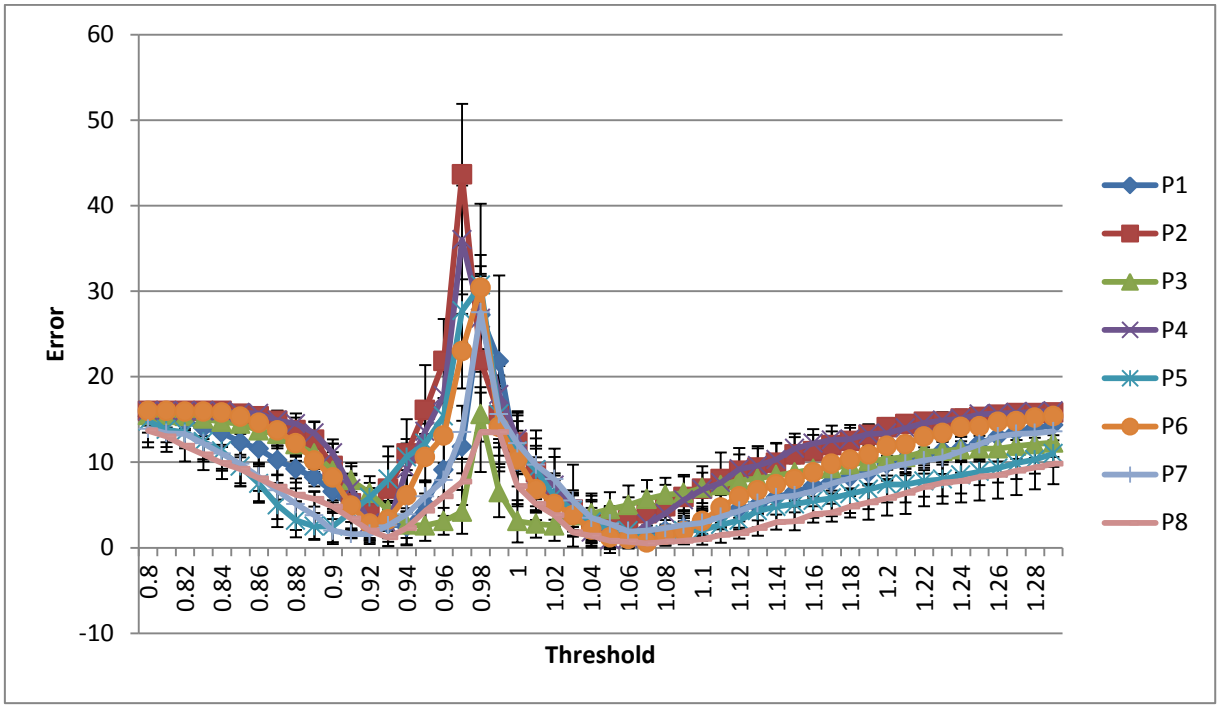


FIGURE 11. STEP DETECTION ERROR FOR DIFFERENT NORMALISED ACCELEROMETER THRESHOLDS. ERROR BARS ARE ONE STANDARD DEVIATION

The results are fairly consistent across the participants, more so than with the Z accelerometer threshold calibration. This is again highlighted by the combined graph:

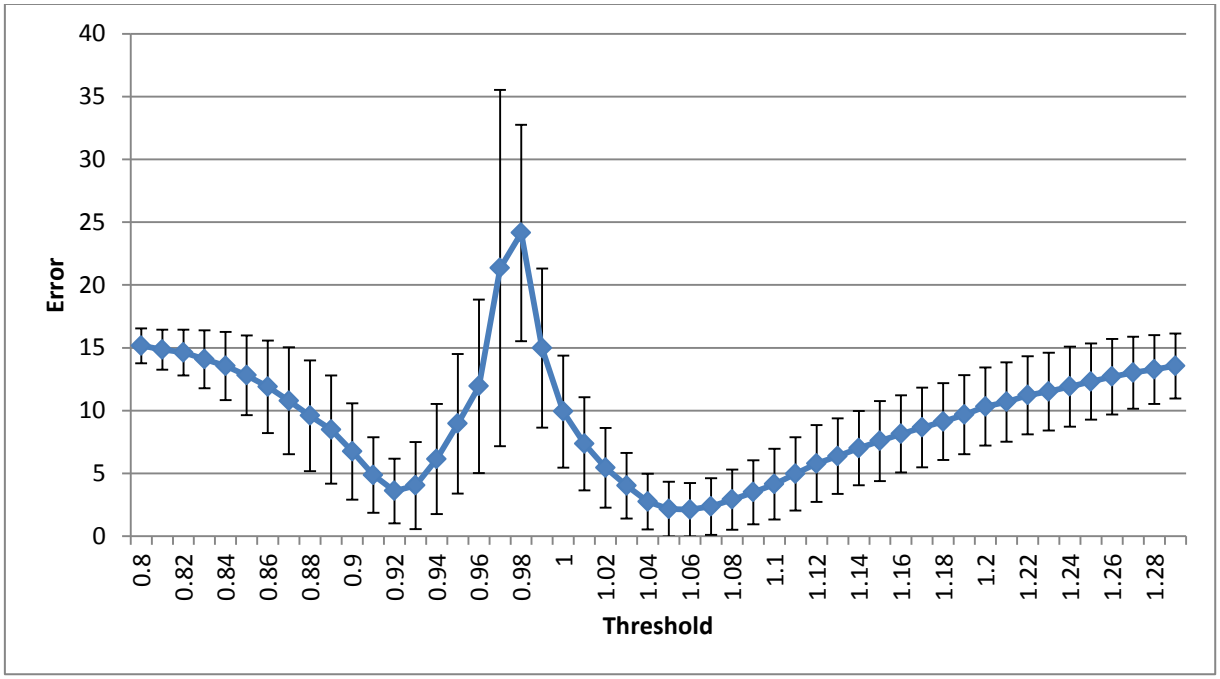


FIGURE 12. STEP DETECTION ERROR FOR DIFFERENT NORMALISED ACCELEROMETER THRESHOLDS ACROSS ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION.

This graph is similar to Figure 10 but shows much smaller variance indicating greater consistency between participants. The optimal threshold value is 1.06g.

4.8.2. STEP COUNTING RESULTS

A two-way repeated measures Analysis of Variance (ANOVA) was run on the calibrated thresholds to the step detection algorithms. The approach followed for this ANOVA is described by Gray (2012) and computer analysis performed using SPSS (version 20) statistical analysis package (IBM Corp, 2011). The two variables in the ANOVA are Sensor (absolute (normalised) accelerometer, Z-axis accelerometer and gyroscope) x Method (thresholding, zero crossing).

Within-Subjects Factors

Measure: StepError

Sensor	Method	Dependent Variable
1	1	GyroscopeThreshold
	2	GyroscopeZeroCrossing
2	1	NormalisedAccelerometerThreshold
	2	NormalisedAccelerometerZeroCrossing
3	1	ZAccelerometerThreshold
	2	ZAccelerometerZeroCrossing

Pairwise Comparisons

Measure: StepError

(I) Sensor	(J) Sensor	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	52.558*	2.219	.000	47.133	57.983
	3	56.529*	2.312	.000	50.879	62.180
2	1	-52.558*	2.219	.000	-57.983	-47.133
	3	3.972*	.354	.000	3.107	4.836
3	1	-56.529*	2.312	.000	-62.180	-50.879
	2	-3.972*	.354	.000	-4.836	-3.107

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

Pairwise Comparisons

Measure: StepError

(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-49.933*	1.743	.000	-53.407	-46.459
2	1	49.933*	1.743	.000	46.459	53.407

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

FIGURE 13. ANOVA PAIRWISE COMPARISON RESULTS FOR STEP COUNTING

Both variables show significant difference between each other (Mauchly's test showed no sphericity, so Greenhouse-Geisser adjusted degree of freedom results are used):

Tests of Within-Subjects Effects

Measure: StepError

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Sensor	317794.501	1.034	307310.341	574.767	.000	.889
Error(Sensor)	39809.535	74.456	534.669			
Method	298443.861	1.000	298443.861	821.063	.000	.919
Error(Method)	26170.897	72.000	363.485			
Sensor * Method	342574.936	1.041	329034.173	611.536	.000	.895
Error(Sensor*Method)	40333.537	74.963	538.046			

FIGURE 14. STEP DETECTION TWO WAY ANOVA RESULTS

Partial Eta Squared results provide an indication of the 'power' of the tests and show both factors (step detection method 0.919 and sensor 0.889) and their interaction to be large effects. Examining the means charts, however, highlights the erratic behaviour of the gyroscope when used with the zero crossing approach (mean step detection error of over 125 steps) and the overall poor performance of the zero crossing method (lower values of estimated marginal means of step error represent better detection):

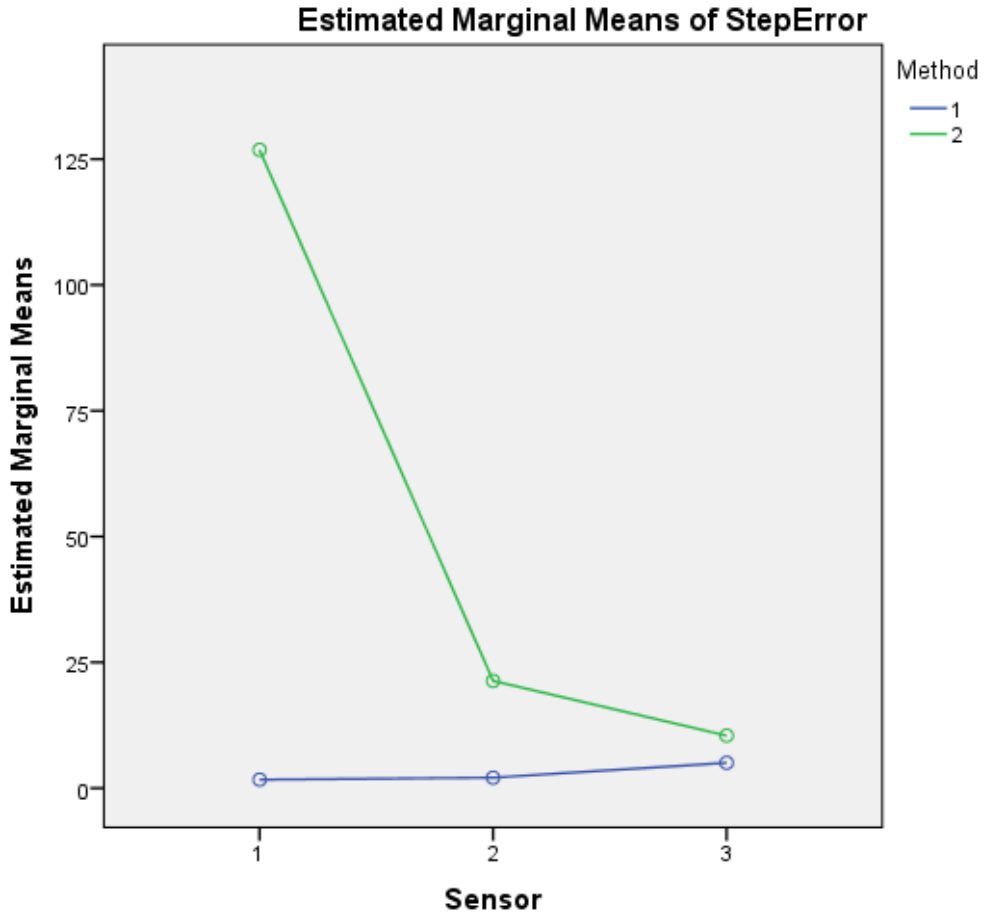


FIGURE 15. STEP ANALYSIS PERFORMANCE. METHODS ARE: 1= THRESHOLDING, 2 = ZERO CROSSING. SENSORS ARE: 1= GYROSCOPE, 2 = NORMALISED COMPASS, 3 = Z-AXIS COMPASS.

Overall, zero crossing method is outperformed by threshold-based one. Performing one way ANOVA on just the threshold method to remove the effects of the outlier behaviour of the zero crossing gyroscope (since threshold approach overall produces lesser step error as can be seen from Figure 15), shows significant difference in the performance of sensors [$F(2,239) = 42.62$, $p<0.001$].

Pairwise comparison highlights the specifics of this difference:

Pairwise Comparisons

Measure: StepError

(I) Sensor	(J) Sensor	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.411	.224	.196	-.958	.136
	3	-3.354*	.293	.000	-4.069	-2.638
2	1	.411	.224	.196	-.136	.958
	3	-2.942*	.313	.000	-3.709	-2.176
3	1	3.354*	.293	.000	2.638	4.069
	2	2.942*	.313	.000	2.176	3.709

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

FIGURE 16. ANOVA PAIRWISE COMPARISON RESULTS OF STEP DETECTION THROUGH THRESHOLDING APPROACH. SENSORS ARE: 1 = GYROSCOPE, 2 = NORMALISED ACCELEROMETER, 3 = Z-AXIS ACCELEROMETER

There is no significant difference between the performance of normalised accelerometer and gyroscope (p = 0.196). Z-axis accelerometer, however, is significantly different from the gyroscope (p < 0.001) and the normalised accelerometer (p < 0.001). Examining the data directly reinforces this conclusion:

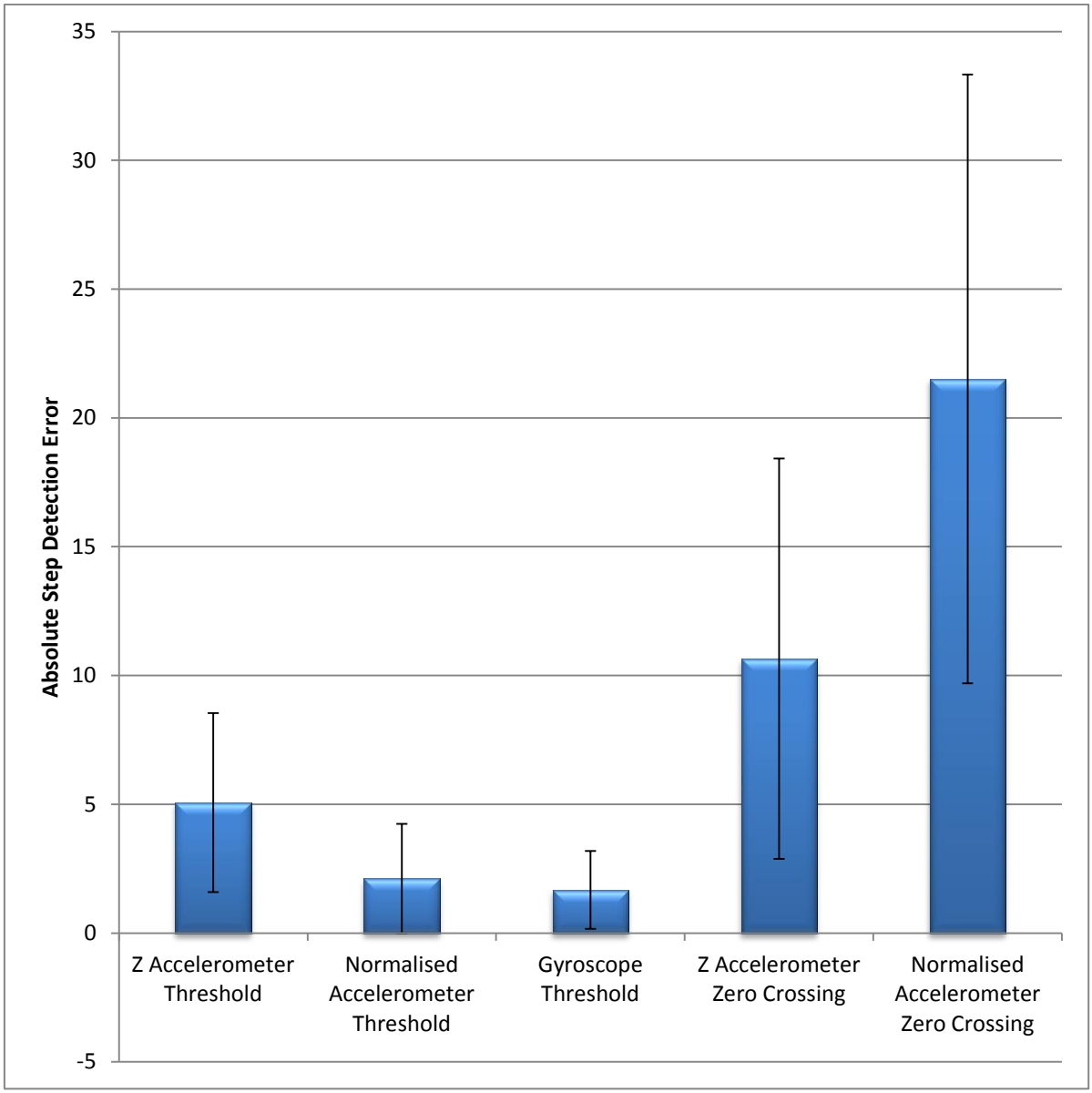


FIGURE 17. STEP DETECTION ERROR. ZERO CROSSING GYROSCOPE DETECTION REMOVED. ERROR BARS ARE ONE STANDARD DEVIATION

The threshold algorithms perform comparably with gyroscope and normalised accelerometer. Zero crossing approaches require a zero velocity update mechanism to avoid detecting steps during periods of rest (Ibarra Bonilla et al., 2011). In the current implementation zero crossing approaches show large variance and are outperformed by threshold-based approaches.

4.8.3. TURN DETECTION RESULTS

Results of turn tracking for all participants have been processed using one-way repeated measures ANOVA and produce the following (Greenhouse-Geisser adjusted because of the failed sphericity test):

Within-Subjects Factors

Measure: Turn

Sensor	Dependent Variable
1	TiltCompensatedCompass
2	IntegratedGyroscope
3	@2DCompass

Tests of Within-Subjects Effects

Measure: Turn

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Sensor	132040.902	1.365	96766.518	78.256	.000	.201
Error(Sensor)	526434.568	425.734	1236.535			

Pairwise Comparisons

Measure: Turn

(I) Sensor	(J) Sensor	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-15.649*	1.329	.000	-18.839	-12.458
	3	-28.725*	2.780	.000	-35.400	-22.051
2	1	15.649*	1.329	.000	12.458	18.839
	3	-13.077*	2.522	.000	-19.132	-7.022
3	1	28.725*	2.780	.000	22.051	35.400
	2	13.077*	2.522	.000	7.022	19.132

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

FIGURE 18. TURN DETECTION ANOVA RESULTS

There are significant differences on all pairwise comparisons between the sensors (tilt compensated compass, integrated gyroscope and 2-axis compass) according to the ANOVA

results. The sensor choice as a factor has partial eta squared value of 0.201, making it a large factor (Gray, 2012).

The actual turn detection results are shown on Figure 19. They are presented as mean difference between expected and detected turn magnitude.

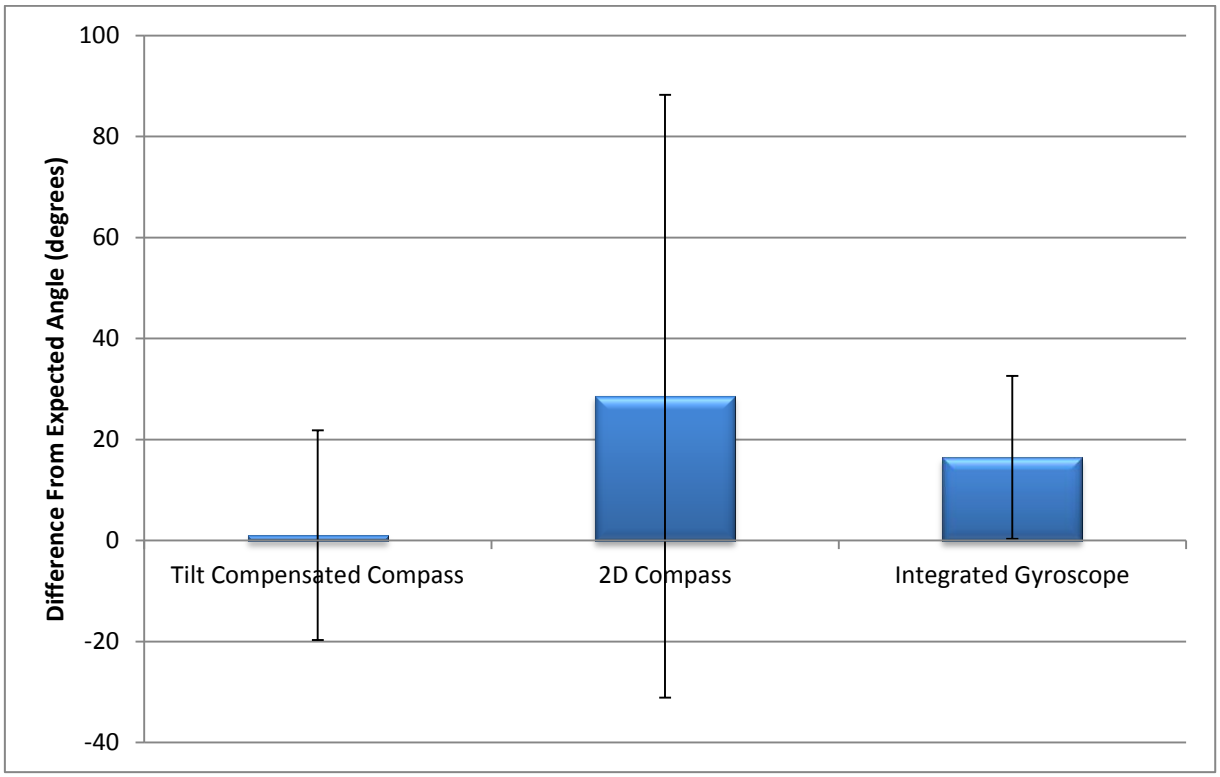


FIGURE 19. TURN DETECTION DIFFERENCE FROM EXPECTATION. ERROR BARS ARE ONE STANDARD DEVIATION

Tilt compensated compass and integrated gyroscope perform better than 2D compass due to being unaffected by platform stability issues. This is caused by tilting of the compass and high magnetic inclination angle in Birmingham, where it is approximately 67 degrees (Finlay et al., 2010). Compass results show less error but slightly greater deviation from the gyroscope results. Both sensors can be used for turn tracking but compass is significantly better overall as proven by ANOVA results.

4.8.4. HEADING STABILITY RESULTS

Heading stability is the measure of the deviation of the heading detected during waling in a straight line, where any variance in the heading readings is mainly due to sensor or processing errors. ANOVA results of heading stability analysis of sensors are as follows (Greenhouse-Geisser adjusted because of the failed sphericity test):

Within-Subjects Factors

Measure: StabilityError

Sensor	Dependent Variable
1	TiltCompensatedCompassDeviation
2	IntegratedGyroscopeDeviation
3	@2DCompassDeviation

Tests of Within-Subjects Effects

Measure: StabilityError

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Sensor	5532.940	1.225	4517.622	84.750	.000	.358
Error(Sensor)	9923.335	186.161	53.305			

Pairwise Comparisons

Measure: StabilityError

(I) Sensor	(J) Sensor	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	8.083*	.326	.000	7.295	8.871
	3	2.310*	.675	.002	.679	3.941
2	1	-8.083*	.326	.000	-8.871	-7.295
	3	-5.773*	.815	.000	-7.741	-3.805
3	1	-2.310*	.675	.002	-3.941	-.679
	2	5.773*	.815	.000	3.805	7.741

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

FIGURE 20. HEADING STABILITY ANOVA RESULTS

Once again, all the sensors differ significantly from one another, falling below standard threshold of 5% for *p*-value (significance). Partial eta squared for the sensor choice for the direction stability is 0.358, making it a large effect.

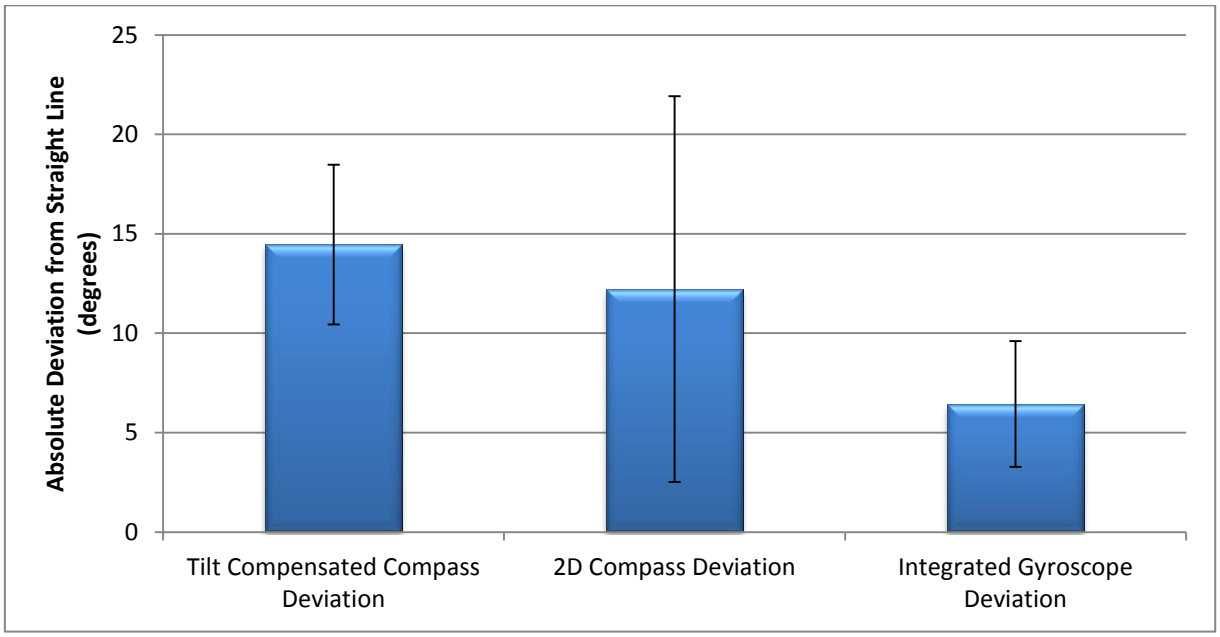


FIGURE 21. HEADING STABILITY. ERROR BARS ARE ONE STANDARD DEVIATION

Figure 21 shows heading stability results, highlighting gyroscope as sensor that maintains the greatest stability of the three heading tracking approaches compared. Whilst gyroscope normally suffers from bias drift, over relatively short straight line path it is able to hold its direction. Accelerations introduced by walking (and detected as steps in 4.8.2) are mistakenly identified as tilt in compass platform, thus causing error in heading data (Chen et al., 2010). The dual axis compass does not suffer from the same acceleration-induced errors, but its inability to compensate for tilt during walking remains a factor in its readout accuracy, giving it marginally better average result than that of the tilt compensated compass (yet with greater deviation) but leaving integrated gyroscope (especially high accuracy pre-calibrated and digitally filtered one like the one used in this research) with a clear advantage.

4.9. SENSOR CHOICE

Both the accelerometer and gyroscope perform satisfactory as pedometric sensors. In terms of algorithms, the threshold-based step detection outperforms zero crossing method. Gyroscope and normalised accelerometer perform statistically similarly and outperform single axis accelerometer with threshold-based approach. As such, either sensor is a valid choice for the step counting part of this research. Gyroscope has to be mounted on the legs and requires a pair of sensors (one on each leg) or interpolating the sensor data of one leg (doubling it) to detect all the steps. However it provides a reliable way of estimating relative step length (Tong & Granat, 1999) that is helpful for accurate dead reckoning. For purely pedometric analysis, the accelerometer would be a better choice purely due to its convenience, not requiring two sensors, nor leg mounting. If required, stride length can be inferred from accelerometer signal as well (Alvarez et al., 2012), but this requires double integration of accelerometer data (and the errors associated) and very accurate decomposition of the accelerometer signal to determine the exact vertical component. As this study has shown, relying on Z accelerometer component to be aligned with gravity normal during gait is unreliable and could further add to the error.

For heading tracking, the tilt compensated compass provides best turn tracking but suffers from stability issues when walking along straight line sections. By comparison, the gyroscope suffers from bias errors introduced by integration and accumulating hardware errors and has lower accuracy during turn tracking. It does not, however, suffer from errors introduced from acceleration due to walking and maintains high stability during straight line sections.

Dual axis compass is not affected by acceleration shifts but its inability to compensate for sensor tilt makes it least effective at both heading tracking and stability.

Because tilt compensated compass and gyroscope have strengths in different areas, especially when it comes to heading tracking during walking, a combination of the two is

going to be the better choice for pedestrian dead reckoning system than either sensor individually.

With camera tracking, however, the camera remains static for the duration of the shot (a photographer naturally tries to stabilise camera during shots) and as such does not suffer from the acceleration-introduced errors. This makes a single tilt-compensated compass adequate for the purposes of achieving heading tracking for sensor-based composition, as described in the next chapter.

5. SENSOR-BASED IMAGE COMPOSITING

This chapter describes sensor based image compositing system and its evaluation in a form of a study that aims to investigate arranging photographs taken from the same physical location. This section describes the new approach to generating the image composites that does not involve image processing; instead using only the metadata stored in the images itself, augmented by the data gathered from the self-contained sensors, as specified in section 3.4.

This chapter tries to answer the question of how images taken from the same location can be handled and evaluates the efficiency of the proposed design compared to existing alternatives. It focuses on improving the utility of composites generated while maintaining their evidential integrity.

5.1. INTRODUCTION

In recent years, photographs tagged with the compass-based direction are beginning to emerge (Casio America, Inc, 2010). The intended use-cases are not very well worked out, however, making the usefulness of compass orientation questionable. This section presents an extension on the compass-tagged photograph that enables real time composition of multiple photographs and further enhances the location tagging with the additional data and features.

Accuracy and reliability of the inertial and magnetic sensors has improved over the recent years. It is theorised that the current low cost commercial sensors can provide adequate information to compose the images producing the results comparable to those obtained by use of image processing, such as those outlined in section 3.3.

Common image processing-based compositing applications have to alter the graphical content of the source images before compositing them. This is done to project all of the

source images onto an optimal plane (Koo et al., 2009) or to compensate for inconsistent illumination and reduce the visible seams in the output panorama. Whilst this undoubtedly improves the visual appeal of the composites and makes them appear more realistic, it also renders them inadmissible for any evidential purposes (Great Britain Parliament House of Lords Select Committee on Science and Technology, 1998). Compositing the images using sensor data, on the other hand, does not alter the image content in any way or form and merely arranges the images differently. According to the definitions for admissible digital image evidence outlined in the House of Lords 5th Science and Technology report (Great Britain Parliament House of Lords Select Committee on Science and Technology, 1998), such an image collection would remain a valid evidential material, giving sensor-based compositing an additional avenue of application – evidential photography.

5.1.1. PRINCIPLE OF OPERATION

Sensor-based compositing works by obtaining the camera orientation at the point of image capture. This orientation combined with the image itself can then be composited with other images taken at the same physical location. Since the relative positions of the regions photographed are directly related to the camera orientation, the data from the sensors is adequate for a successful composition.

Assuming that the camera remains in the same physical location between shots (as is the aim of this chapter), and that its height above ground does not change (i.e. the photographer does not lean or sit to get a better shot), the changes in three angles (shown on Figure 22) need to be estimated between photographs in order to determine the relative position of the field of view covered by each photograph:

- Pitch angle, defining camera tilt up and down, its rotation around the horizontal side to side axis

- Roll angle, determining the camera tilt side to side, its rotation around the axis passing through the camera lens
- Yaw angle (camera heading direction) that determines how far the camera is rotated around vertical axis

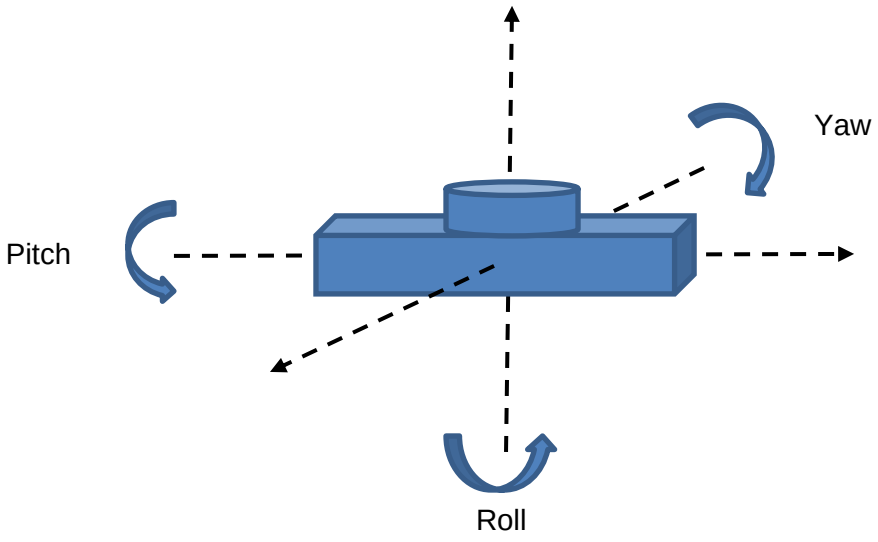


FIGURE 22. CAMERA ANGLES FOR ORIENTATION TRACKING

Pitch and roll angles can be determined using accelerometers. Accelerometers detect the acceleration normal due to gravity (as described in 4.1) and can therefore determine the angle between the sensor and the gravity normal (vertical axis). Depending on the alignment of the accelerometer, this will provide either the pitch or roll angle.

Detecting the yaw angle is harder, since it is not affected by gravity. A number of ways of determining yaw is described in chapter 2 as part of pedestrian localisation, such as use of angular rate and magnetic sensors. Both of these modules have their own strengths and weaknesses (Fang et al., 2005). A magnetic compass is reliable only when it is level with the ground plane and can suffer from the magnetic interference, whereas a gyroscope does not provide absolute readouts and suffers from temperature-dependent bias and scale errors (Ladetto & Merminod, 2002). Also, a gyroscope does not provide an absolute readout for the

heading and has to be integrated, increasing the effect of the scale errors on integrated heading value.

Camera yaw tracking is an easier task than yaw tracking for localisation, since the sensor remains in the same physical position for the duration of the task (i.e. its readings will only be necessary once the photographer is stationary and takes several photographs from the same point) thus sensor interference from external factors such as magnetic interference on the sensor are stabilised (i.e. magnetic interference could at this point be modelled as a constant effect and adjusted for). Chapter 4 showed that tilt compensated magnetic compass provides very accurate yaw information when only affected by acceleration due to gravity and thus compass is adequate for yaw tracking on camera.

The standard dual axis compass modules suffer from increasingly large errors when tilted and since the user is unlikely to maintain the level position of the camera during its normal operation, this error is likely to adversely affect the composition process.

The solution to this problem is to use a tilt compensated compass. Such a compass uses 3-axes magnetometers (instead of 2-axes employed by its non-compensated counterparts) as well as a set of accelerometers to help detect the compass orientation. Such a module can be built from the individual sensors or purchased off the shelf. Upon investigating the available options, it was decided to again use a Honeywell HMC6343 tilt compensated magnetic compass. This single unit is able to provide both the heading and the tilt data directly from the internal accelerometers and is factory calibrated. The use of HMC6343 would mean that a single module would be able to provide all the sensor reading necessary for the composition, rather than combining multiple sensors individually for the same result.

Therefore, the process of sensor based compositing becomes as follows:

1. When a photograph is being taken, read pitch and roll values from accelerometer and yaw value from compass or gyroscope

2. Estimate the field of view covered by the photograph based on the magnification factor (recorded in EXIF metadata of most commercial digital cameras (Japan Electronics and Information Technology Industries Association, 2010)) and the camera orientation
3. Rotate, scale and move the captured image in the coordinate frame of the orientation sensors according to the detected field of view

Scaling is needed to allow for magnification of different images, since images taken at greater magnification level will cover a smaller area, albeit have greater resolution when compared to their unmagnified counterparts.

The overall system hardware design is shown on Figure 23:

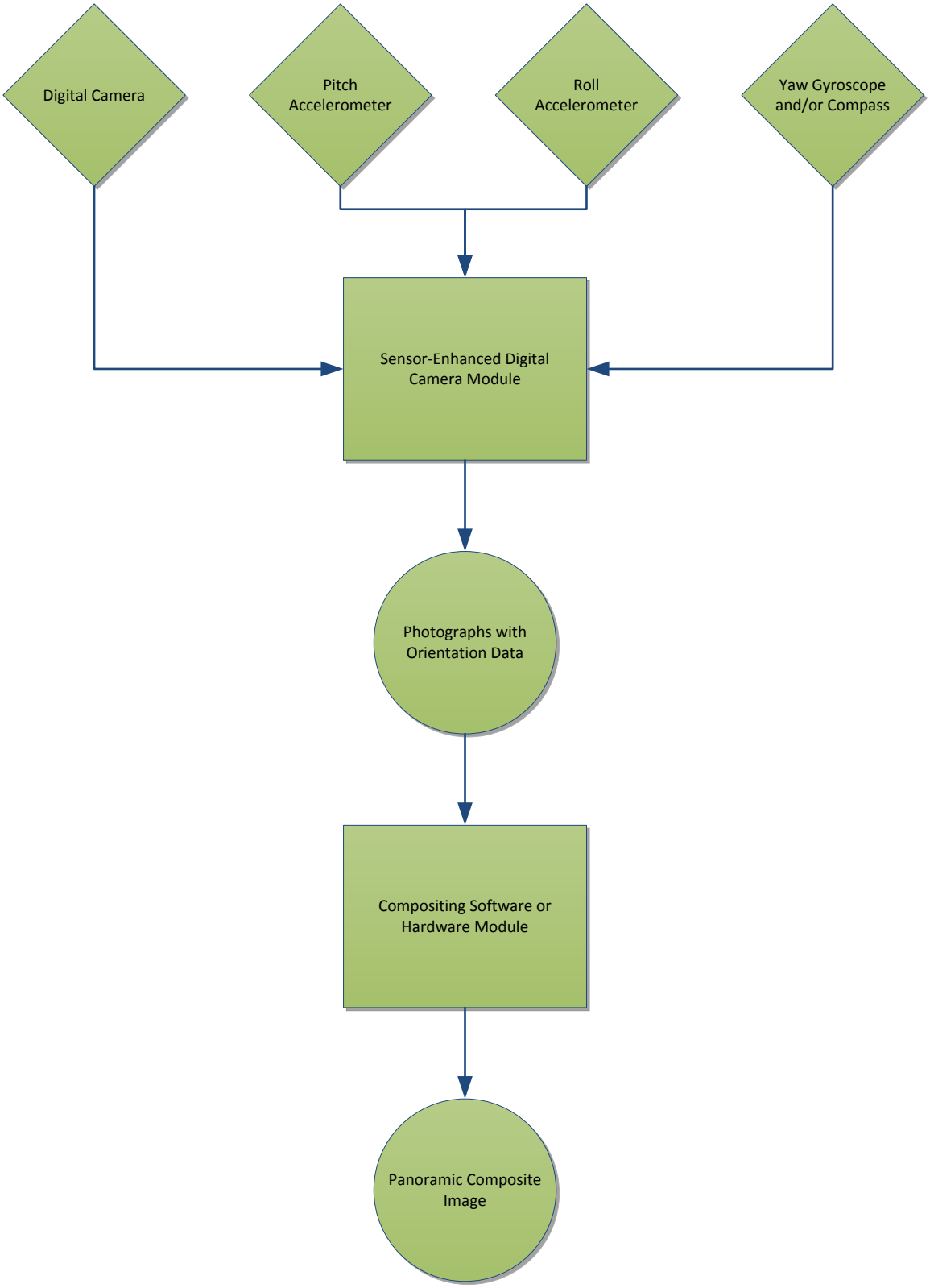


FIGURE 23. PROPOSED DESIGN OF THE SENSOR-ENHANCED IMAGE COMPOSITING SYSTEM

The proposed system stores the camera orientation data inside the metadata of the photographs themselves, similar to the way GPS (global positioning system) – based spatial tagging operates. This ensures that each photograph contains information about the camera orientation whilst it was taken inside its EXIF metadata tag (Japan Electronics and Information Technology Industries Association, 2010). During compositing, photographs will be collated by a secondary subsystem (either a software application or hardware module). This way the design becomes more versatile in its implementation and the system can be used both for normal photography and compositing, since the intermediate stage outputs (photographs with sensor readings in their metadata that do not affect the image content in any way) remain available to the system user.

The proposed compositing process itself is fairly straightforward and involves estimating the field of view of camera for each photograph in the coordinate frame of the orientation sensors and then arranging the photographs accordingly. For a more visually pleasing appearance, the photographs can then be processed by any of the image processing-based compositing approaches to handle the necessary warping (as per Koo et al (2009)) or image blending.

5.2. PORTABLE MODULE DESIGN

The design for a portable unit comprises of three parts – hardware, firmware running on said hardware (both used to obtain sensor and image data) and their software counterpart performing the actual compositing. All of these have the potential of being implemented on a single device.

5.2.1. HARDWARE DESIGN

The sensor-driven image compositing system developed for this chapter combines the high resolution digital photographs made with a commercial camera (10.2 megapixel Samsung ES15) with the sensor information collected by a custom designed hardware module.

FIGURE 24. IMAGE ACQUISITION HARDWARE
FROM LEFT TO RIGHT: A) CAMERA WITH THE COMPASS AND SONAR SENSORS AND THE CATEGORY SELECTION PAD ABOVE, B) MAIN PROCESSING MODULE, C) POWER SUPPLY AND INTEGRATED USB HUB, C) BATTERY

The digital camera is used in a largely unmodified form. The only alteration involves wiring the camera shutter button (a two state button, with the half-depressed state triggering automatic focusing) to the sensor module.

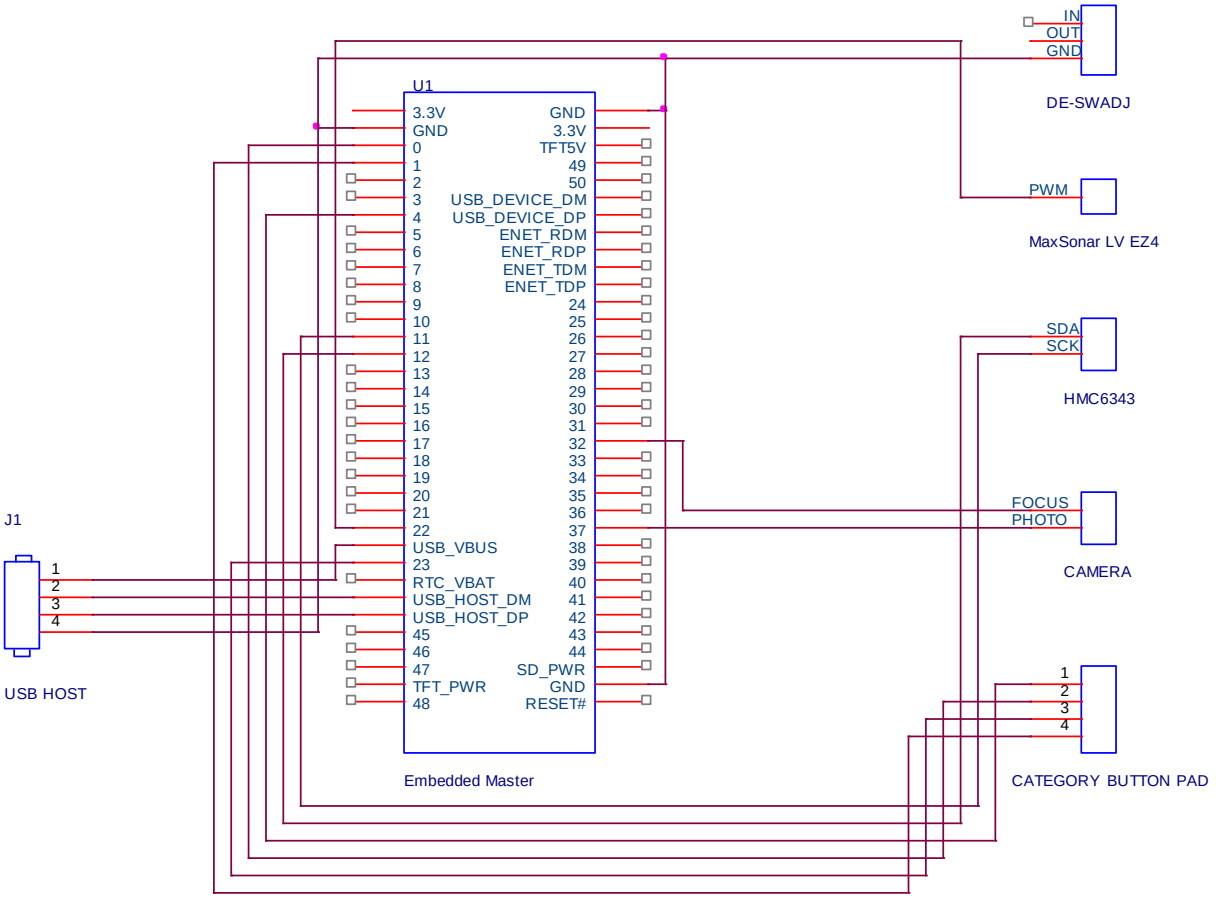


FIGURE 25. SENSOR COMPOSITING HARDWARE DESIGN SCHEMATIC. SECONDARY CONNECTIONS AND BUSES OMITTED FOR CLARITY

The sensor module is built around a GHI Electronics Embedded Master unit, running .NET Micro Framework – based firmware. The module itself is powered by NXP Semiconductor's LPC2468 72MHz ARM7 32-bit CPU, providing efficient real-time sensor processing capabilities required for the sensor-based compositing. It is connected to the camera shutter button in an input mode (effectively reacting to the shutter presses), allowing instant detection of the camera triggering the image capture. This module is of the same product family as the EMX module used for sensor recorder in chapter 4, but has less hardware resources and advanced firmware features, since it does not need to record data continuously and needs to only be able to perform the most basic processing (compositing is taken care of during post-processing on a PC). The firmware, however, is adaptable between

the modules, making it easy to merge the systems in the future. Embedded Master is also pin- and mostly firmware-compatible with EMX module, simplifying design considerations.

The device is again powered by a 12V Lithium Polymer battery via DE-SWADJ regulator that has proven its effectiveness with the sensor recording module in 0.

The module peripherals include a powered USB mini-hub (providing 4 USB ports for use by the .NET core) and a set of sensors, mounted on the camera body to track its orientation. The main sensor used is the Honeywell HMC6343 magnetic compass. This module was chosen for its robust tilt compensation and the ability to output both the geomagnetic heading and the pitch and roll information, using built-in accelerometers. It has shown good heading tracking and stability in the evaluation in 4.8

Data obtained from the compass module are augmented by a MaxSonar LV EZ4 sonar range finder, providing 6” – 254” range information with the resolution of 1”. An assumption was made during the design stages that during the normal action of photograph taking, the camera is pointed towards the subject of the photograph and hence range information from a rangefinder aligned with the camera lens will provide an approximation of the distance to subject, useful for estimating the Z-order (occlusion) of photographs, i.e. which photographs will be rendered on top of others in the final composite.

In addition to the sensor data, a keypad with 4 buttons is provided, wired directly to the sensor module. The buttons on this keypad are used to tag an image to a user-defined category. For example, the user can assign all images containing objects of interest to one category and general environment shots to another. Categorisation of images is optional but illustrates the ease of introducing additional information into the photograph metadata.

Additional sensors and user interaction controls can be added as the need arises, considering the low power and processing requirements of the firmware and the powerful CPU module used.

5.2.2. *Firmware Design*

The main focus in the design of the firmware lies in instant response to the camera shutter. The shutter button is wired to the interrupt-capable ports of the EmbeddedMaster module for near-instantaneous response to the shutter button press.

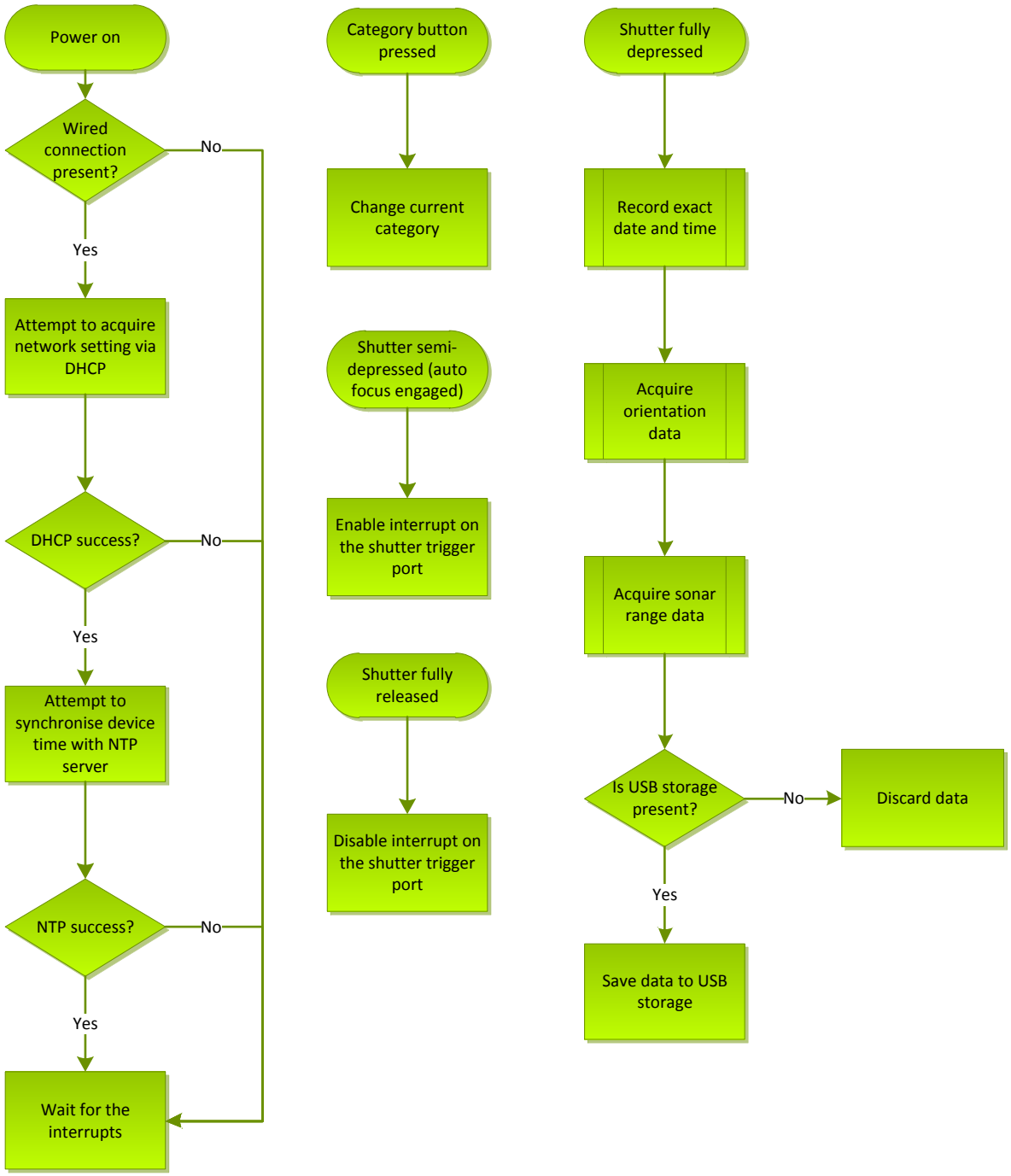


FIGURE 26. FIRMWARE DESIGN OF THE SENSOR MODULE

Upon the press of the shutter button, the time and date are recorded and then the sensor readings are all read in a single sweep, starting from the inertial and magnetic data off the Honeywell compass device, followed by the range information from the MaxSonar rangefinder. These readings are written directly to a CSV (comma separated values) file on an attached USB storage device (typically a USB flash drive) by appending them in a predefined format to the end of the file.

Several interrupt ports are wired to the category selection button pad. This way pressing a category selection button is instantly recorded and stored in the operating memory of the device. This value is then written together with the sensor and timestamp data whenever a camera snapshot occurs.

An additional feature of the firmware is its ability to synchronise time with the NTP (network time protocol) servers available on the Internet. Upon start-up, if there is a wired connection to the Ethernet port on the device, the firmware attempts to acquire the network setting from a DHCP server (thus this feature only operates on the network where such a server exists). Upon success, it attempts synchronisation of the internal clock (that is kept running using a secondary power supply in the form of a small lithium power cell) with the ntp.org freely available NTP server (NTP Project, 2011). If this succeeds, the device resumes its operation as normal and can be unplugged from the wired network. If at any stage the operation fails, device resumes operating in normal mode, awaiting the camera trigger press.

5.2.3. COMPOSITING SOFTWARE DESIGN

Due to the proprietary nature of the commercial digital camera design, as well as the legal restrictions on their modifications, the camera had to operate largely independently from the sensor module. While use of the more easily manipulated USB web camera modules was considered at the early stages of the design, the higher image quality of the typical digital camera made it preferable for the purposes of this system. Additionally, independent operation of the digital camera allows for camera to be swapped with ease for alternative

ones depending on the task requirements, e.g., moving from a conventional digital camera to one with night-vision capability.

The user can choose to press any one of the 4 buttons available on a secondary keypad before taking a snapshot. Each of these buttons corresponds to a numerical code from 1 to 4, corresponding to user-defined category (category definitions can be altered as need and situations change). The category identifier is included in the CSV file together with the sensor data. By categorizing items during photographing, the user can later, if desired, refine the compositing process by highlighting specific items.

Once image capture is completed, the images stored in the camera and the CSV file must be synchronised. This is achieved using a .NET Framework-based application running on a standard PC. The camera and the flash drive are connected to the USB ports of the PC and a custom algorithm is run to match images from the camera with the corresponding sensor readings. The result of this operation is a copy of the images with the sensor readings written to the description field of the image metadata (EXIF JPEG image data). The description field was chosen due to the fact that the information it contains for the ES15 camera is redundant and duplicated in other EXIF tags (e.g., manufacturer and camera model) and as such no useful information would be lost. This choice can be changed if required with minimal modification of the design.

When a typical digital camera takes a photograph, it adds additional information to the image, outlined in the EXIF specifications (Japan Electronics and Information Technology Industries Association, 2010), and including timestamp based on the camera user-set clock. While this clock is usually very inaccurate, being manually set by the user, it still generates timestamps that are very accurate relative to the user-set initial time. By using time-difference comparison between the camera-generated timestamps and the sensor-module recorded ones, the tagging algorithm determines which sensor readings correspond to which image.

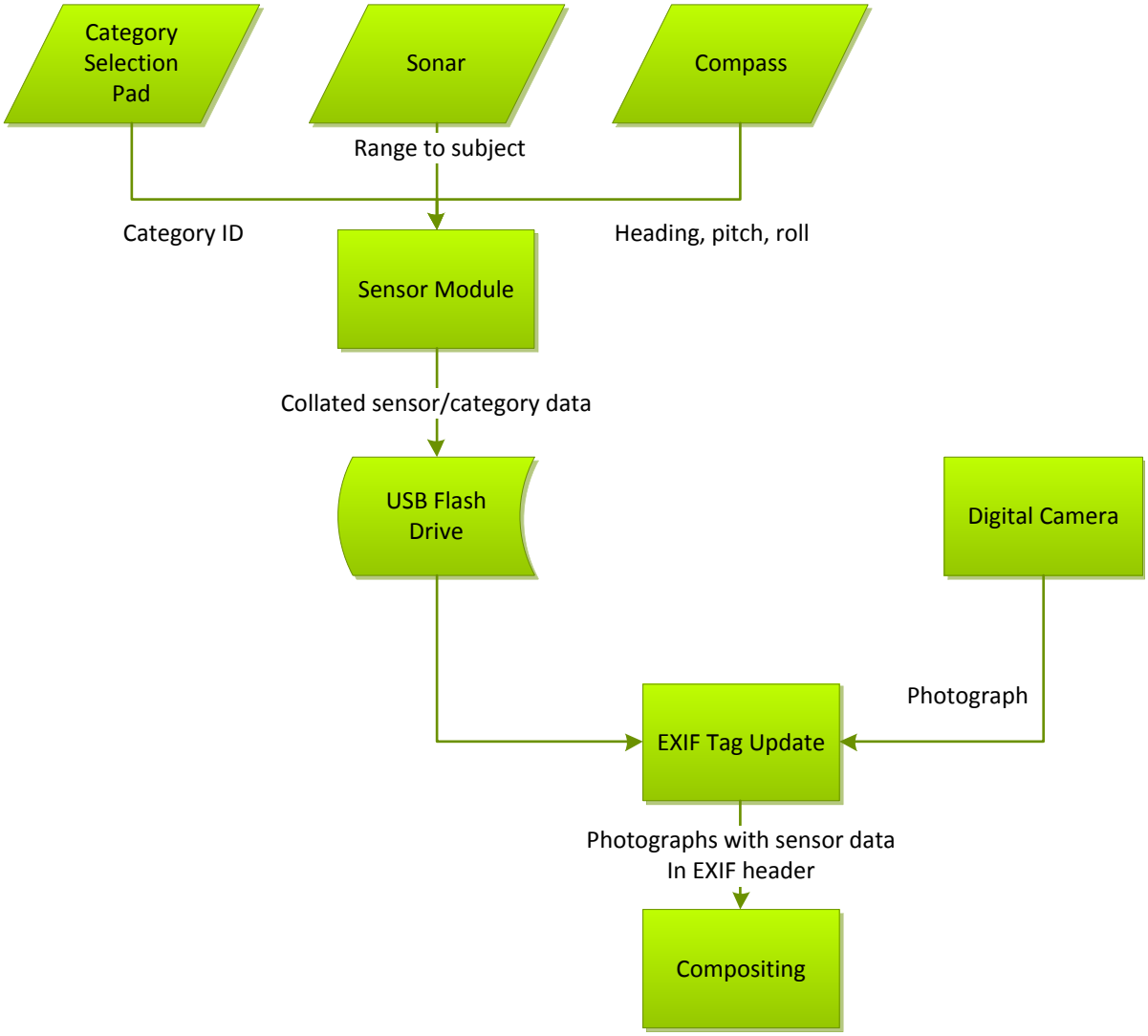


FIGURE 27. IMAGE ACQUISITION AND COMPOSITING PROCESS OVERVIEW

The compositing process uses the information embedded in the image metadata to determine an image's position in the overall composite. The data used are comprised of the EXIF metadata added to the photograph by the digital camera and the additional data recorded by the sensors.

By using the focal length of the camera and the dimensions of the camera imaging sensor (available in the camera specifications), it is possible to calculate the angle of view of the image by using the following formula:

$$\alpha = 2\, tan^{-1} \frac{d}{2\, f} \qquad \text{(Eq. 5.1)}$$ [1]

Where *d* is the dimension of the camera imaging sensor and *f* is the focal length of the camera.

Before the advent of digital cameras, focal length reading had to be estimated from the images themselves (Duffin & Barrett, 2001), but modern devices automatically include it into the image metadata (Japan Electronics and Information Technology Industries Association, 2010).

Combining this information with the 3D orientation of the camera, recorded by the sensors, it is possible to estimate the exact area of a virtual 3D sphere centred on the camera image sensor that a given image is covering.

For example, knowing that Samsung ES15 camera-created photograph taken with no zoom was taken with a focal length of 6.3mm and that the imaging sensor is 6.13 x 4.60 mm in size (known as 1/2.33'' sensor size) according to the camera specifications means that the vertical angle of view for this camera is:

$$\alpha_v = 2 \times tan^{-1} \frac{4.60}{2 \times 6.3} = 40.11° \qquad \text{(Eq. 5.2)}$$

And the horizontal angle of view is:

[1] (Wolfram Alpha LLC, 2010)

$$\alpha_h = 2 \times tan^{-1} \frac{6.13}{2 \times 6.3} = 51.89° \qquad \text{(Eq. 5.3)}$$

The calculations above do not convert the dimensions and focal lengths to SI units but this does not affect the results.

If the camera is level with the ground (has pitch and roll angles of zero) and is pointing directly north, then it covers the area approximately from -26° to +26° horizontally and from -20° to +20° vertically.

By using this information, positions of the images relative to each other can be determined and used to generate a composite. Such a composite automatically takes into account the magnification factor (zoom level) used during the camera triggering a snapshot due to relying on the camera focal length reading for the calculation.

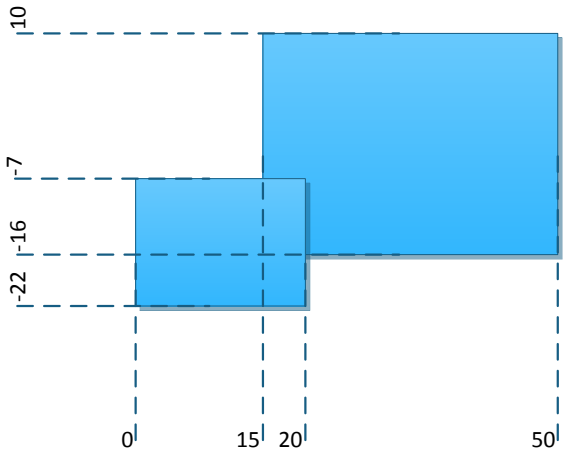


FIGURE 28. COMPOSITING IMAGES BASED ON THEIR ANGLES OF VIEW AND CAMERA ORIENTATION. SMALLER IMAGE COVERS AREA FROM 0 TO 20 DEGREES NORTH HORIZONTALLY AND -22 TO -7 DEGREES VERTICALLY. LARGER IMAGE COVERS THE AREA FROM 15 TO 50 DEGREES NORTH HORIZONTALLY AND -16 TO 10 DEGREES VERTICALLY. THE SMALLER IMAGE HAS LESS RANGE RECORDED AND OCCLUDES THE LARGER ONE

The stitching algorithms estimate the angle coverage required for the entire scene (by looking at the total horizontal and vertical coverage limits). Each image is then rotated according to its roll readout to compensate for situations where the camera is not parallel to the horizon (as is often the case since the camera is being handled manually). The intention is to ensure alignment between images in a set rather than modify the images to perform

stitching (as would be the case with image processing approaches, such as transformations described by Koo et al (2009)).

Once all the images are rotated to compensate for the roll and the overall angle of view coverage is calculated, the complete composite canvas is generated, using these total coverage data to generate a virtual grid with the horizontal angle (based on the horizontal angle of view and centred on the yaw – heading – sensor readout) being mapped to the horizontal axis of the canvas and the vertical angle (based on the vertical angle of view and centred on the pitch sensor reading) being mapped to the vertical axis of the canvas. Images are then drawn on the canvas based on their calculated position on the grid.

The order in which the images are composited is driven by the order of their proximity to the camera, as recorded by the sonar (effectively using the sonar for Z-ordering the images), with the closer images occluding the further ones. At this stage it is assumed that the object that the camera is pointed at is the intended target and as such the sonar measurement provides a reasonable estimate for use with the occlusion.

Another important assumption about the compositing process is that the camera is rotated around its centre of projection when taking photographs. This assumption is not realistic, since the camera is handheld and usually rotates much more freely but experiments show that inconsistencies and inaccuracies introduced because of this assumption do not detract from the utility of the result (as discussed in 5.4 and 5.5).

If the user selects to highlight items belonging to a certain category (for example, by choosing to assign category 1 to general images and category 2 to safety equipment), the images belonging to said category are brought forward in the Z-order, with optional highlight border rendered around them (as shown on Figure 29). Individual images can also be highlighted in the similar manner if required.

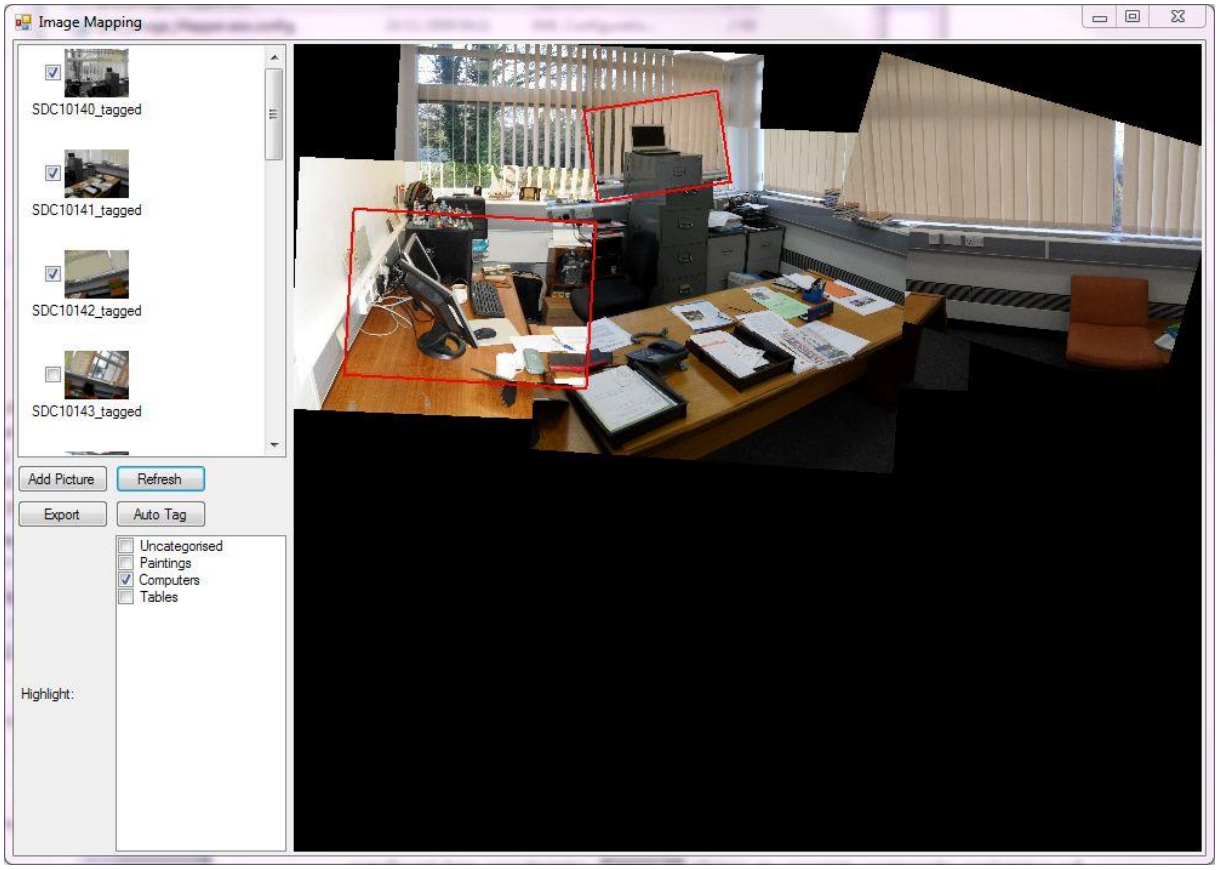


FIGURE 29. EXAMPLE OF IMAGE CATEGORY HIGHLIGHTING

The result of these operations is a composite image generated with minimal processing overheads which is able to deal with the incomplete panoramas (where images lack common overlaps), images at different magnification levels and images taken with the camera non-parallel to the horizon (such as images taken while under significant time constraints). Figure 30 shows an example composite containing roll compensated images on the left with a disconnected secondary region on the right containing a zoomed in image.

FIGURE 30. EXAMPLE PARTIAL COMPOSITE CONTAINING TWO UNCONNECTED PARTS WITH A ZOOMED IN IMAGE IN THE SECTION ON THE RIGHT

This composite does not take into the account the fact that images require distortion in order to fit them perfectly to a virtual sphere surrounding the camera (Koo et al., 2009), instead choosing to treat the angle coverage as linear. This way the original images remain unmodified and can still be re-composited by image compositing approach if better blending is required.

5.3. STUDY DETAILS

In order to validate the performance of the new compositing approach, a study was designed, pitting it against alternative compositing approaches.

5.3.1. STUDY DESIGN

This study is aims to investigate the relative strengths and weaknesses of the sensor-based image compositing approach by comparing it to the image processing approaches employed in several commercial software packages.

The sensor-based stitching process is fully automated, requiring only the source image and sensor data to be supplied. Its results will be evaluated against the applications implementing

image processing-based stitching in a similar manner, requiring only the source images to be supplied. Therefore, any application requiring further user intervention will not be usable as a valid comparison.

The study starts with a scene being set up that would contain enough data to make recognising the imperfections of stitching easier. A set of photographs of the scene are then taken in a manner that allows compositing based on 3 criteria:

1. Images at the same magnification level which overlap and have common areas (a usual requirement for image processing algorithms)
2. Images at the same magnification level which do not have overlaps
3. Images taken at different magnification level but which overlap and have common areas

These criteria allow evaluating individual strengths and weaknesses of the compositing approaches and represent typical usage scenarios for either type of algorithm.

Images are then used by the compositing applications generating matched sets of results from each algorithm/application evaluated. The corresponding composites can be evaluated and analysed to determine their effectiveness under each condition.

5.3.2. Selection of the Software to Evaluate Against

In order to provide the quantitative evaluation of the sensor-based compositing approach, a set of alternative compositing methods has to be determined that is able to contend with the sensor-based approach.

In determining the best software applications to run the comparison, focus is mainly on the applications which have an open source license or those with known compositing algorithms used. This way the compositing algorithm is known and the evaluation can be conclusive. The applications in question must also run on Windows operating system and require

minimal user intervention into the compositing process, similar to the sensor-based compositing approach.

Each of the applications is presented with two test image sets – one containing fully overlapped images with uniform magnification and another with overlapped images at different magnification levels. Test images are generated by using Samsung ES15 digital camera and represent a subsection of the images of the scene to be used for this study, taken without use of flash. Camera is used in fully automated mode and allowed to select optimal exposure and aperture settings on its own (illustrating a typical use case for a non-professional photographer user).

The following list contains the results of the evaluation performed on the applications to help determine the best candidates for this study.

- Enblend (Mihal et al., 2010) (version 4.0) – a command line open source application. According to the developer's site (Mihal et al., 2010), the program "assume[s] that ... images are already lined up" and as such is of limited use for this evaluation. The need to pre-align the images violates the requirement for the fully automated compositing process
- Hugin (d' Angelo, 2010) (version 2010.0.0) – open source application based on the PanoramaTools algorithms by Prof. H. Dersch (Helmut Dersch, 2009). The application experienced numerous unhandled exceptions and was unusable when supplied with a normal (fully overlapped) image set. This made its evaluation impossible and its instability precludes its use for this experiment
- AutoStitch (Matthew Brown, 2010) (version 2.2 Demo) – commercial application with a published algorithm (Brown & Lowe, 2007). Successfully combines images with uniform lighting, struggles when lighting conditions are varied (lighting being determined by the camera automated exposure control). Failed to combine images with different magnification factors, despite SIFT-based approaches being theoretically able to do so

- AutoPano Pro (Kolor, 2010) (version 2.0.6 x64) – commercial application with feature-based automatic compositing reliant on the Brown/Lowe algorithm (Brown & Lowe, 2007). Handles normal panoramas and manages to compose images with different magnification factors
- PanoWizard (Magnus Egelberg, 2010) (version 1.0.27.0) – freeware graphical wrapper for PanoramaTools libraries and algorithms (Helmut Dersch, 2009). Requires manual processing and knowledge of the scene (such as angle of coverage to work). Failed to generate panoramas with the default settings because of the crash of the AutoPano component (non-commercial version, different from AutoPano Pro analysed in the previous paragraph)
- PTAssembler (Max Lyons, 2009) (version 5.0) – shareware application. Relies on a combination of Enblend and AutoPano freeware applications for its auto-stitching algorithm. AutoPano experiences a fatal crash making it impossible to produce any results, similar to PanoWizard
- PTGui (New House Internet Services BV, 2010) (version 8.3.7 Pro x64 Trial) – commercial application based in part on the PanoramaTools algorithms. Failed to generate automatic composite when supplied with the images at different magnification levels. Images under this condition require manual alignment, violating the need for fully automated compositing
- Microsoft Image Composite Editor (Microsoft Corporation, 2010a) (version 1.3.3.0 x64) – freeware application by Microsoft Research subdivision. Generates successful composites at constant magnification levels. Composites at different magnification levels can appear misaligned depending on the image set, but are nonetheless created successfully
- Microsoft Photosynth (Microsoft Corporation, 2010b) (version 2.110.317.1042) – online service based on the algorithms by Snavely et al (Snavely et al., 2006). Handles both the

images at different zoom levels and the constant magnification ones. However, does not produce 2D panoramas and does not allow exporting results into 2D format for comparison, making this application unusable for the study despite robust compositing it is capable of

Based on the analysis of the compositing software, a number of the applications are unusable for the tests. Since the evaluation consists of testing images both with the constant and varied magnification levels, applications that fail to generate an output when faced with the latter do not provide an image for comparing against. That means that the following applications can be used for the comparison:

- AutoPano Pro
- Microsoft Image Composite Editor
- Microsoft Photosynth

Microsoft Photosynth had to be removed from the list upon further evaluation. While it generates comprehensive composites, these are arranged in a 3D representation, only showing the partial composite at any given point in time with no ability to generate a complete 2D representation of the entire result (effectively “flattening” the 3D view). Since the comparison is designed to be done with the 2D images, Photosynth is not feasible for use with this study.

5.3.1. IMAGE DATA ACQUISITION

Image processing-based algorithms normally rely on finding common features in the overlaps between images and then matching the adjacent images by these characteristic features (Brown & Lowe, 2007). Therefore, an optimal scene for the image processing algorithms has to include static objects and minimal amount of motion. The outdoor scenes typically contain movement due to wind and environmental conditions that cannot be easily controlled. Such movement affects the performance of the image processing-based approaches adversely

and can bias the results towards the sensor-based one. Lighting changes due to the movement of clouds can also affect image processing-based compositing.

Therefore, an indoor scene was constructed with the uniform lighting provided by a set of the fluorescent lights. The scene contains a large number of objects, such as books, toys and appliances in order to provide multiple reference features for the imaging algorithms as well as make the analysis easier for the participants of the evaluation.

The source photographic material is obtained by taking 32 photographs of the target location divided into two subsets of 16 photographs each. The scene itself is subdivided into a 4x4 lattice will small overlaps and each photograph in the first set covers a single cell in this lattice (guideline target is shown on Figure 31 below) overlapping with all of the adjacent cells (to give image processing an overlap to work with).

FIGURE 31. TARGET PHOTOGRAPH LATTICE (EACH CELL REPRESENTS AN INDIVIDUAL PHOTOGRAPH WITH THE SMALLER CELLS REPRESENTING ZOOMED-IN VERSIONS)

The actual images and their corresponding overlaps are shown on Figure 32.

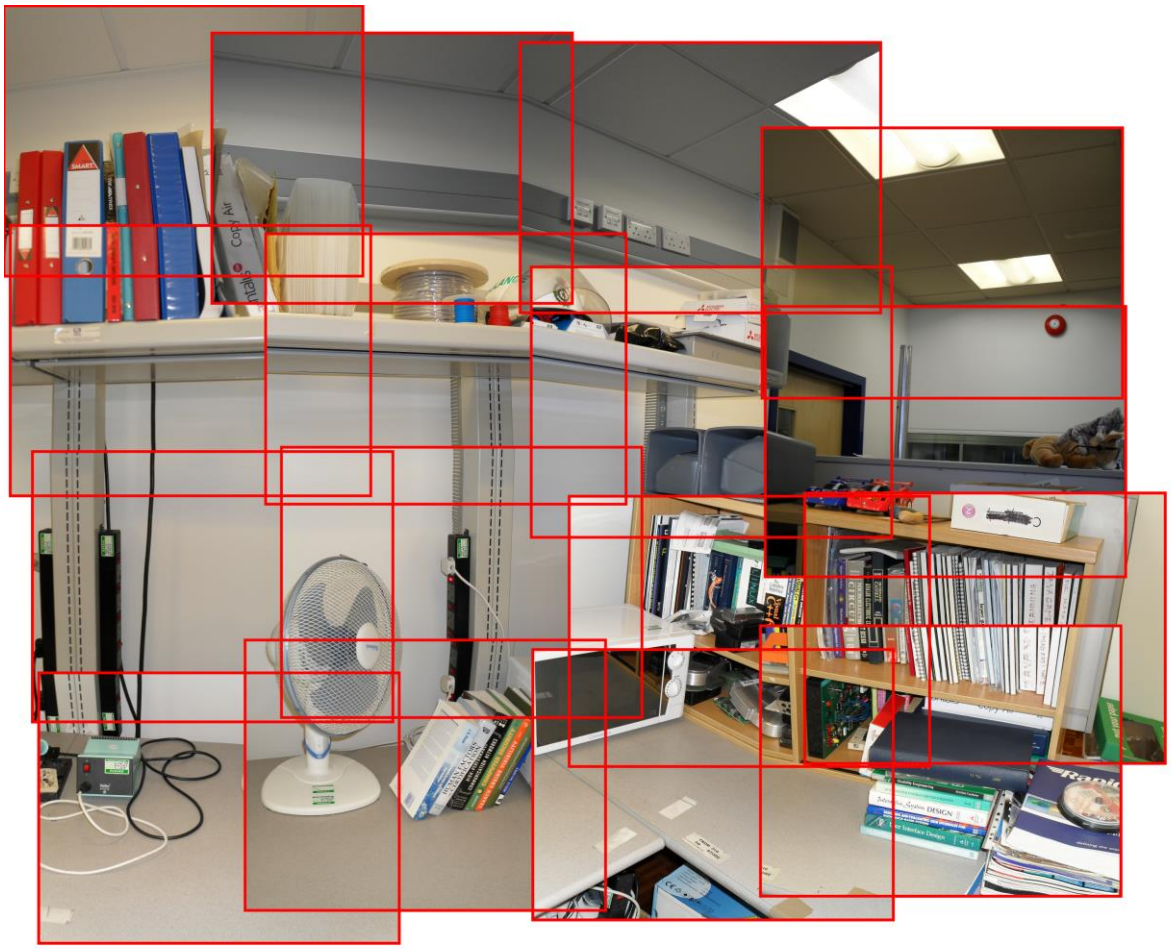


FIGURE 32. PICTURES USED FOR THE COMPOSITING EXPERIMENT WITH OVERLAPS SHOWN

The second subset contains images of the arbitrary areas of the aforementioned cells taken with the camera zoomed in on an object of interest, if available in the cell, or overlapping the adjacent cell nearest to the lattice centre otherwise. Object of interest, in this context, is defined as the most noticeable or distinctive feature in the cell. If several distinct features are present, the one closest to the lattice centre was used. Care was taken to make sure that the zoomed in version overlaps with at least a single adjacent non-zoomed in cell to provide an overlap for the image processing-based compositing software. Zoomed in pictures can be seen on Figure 33.

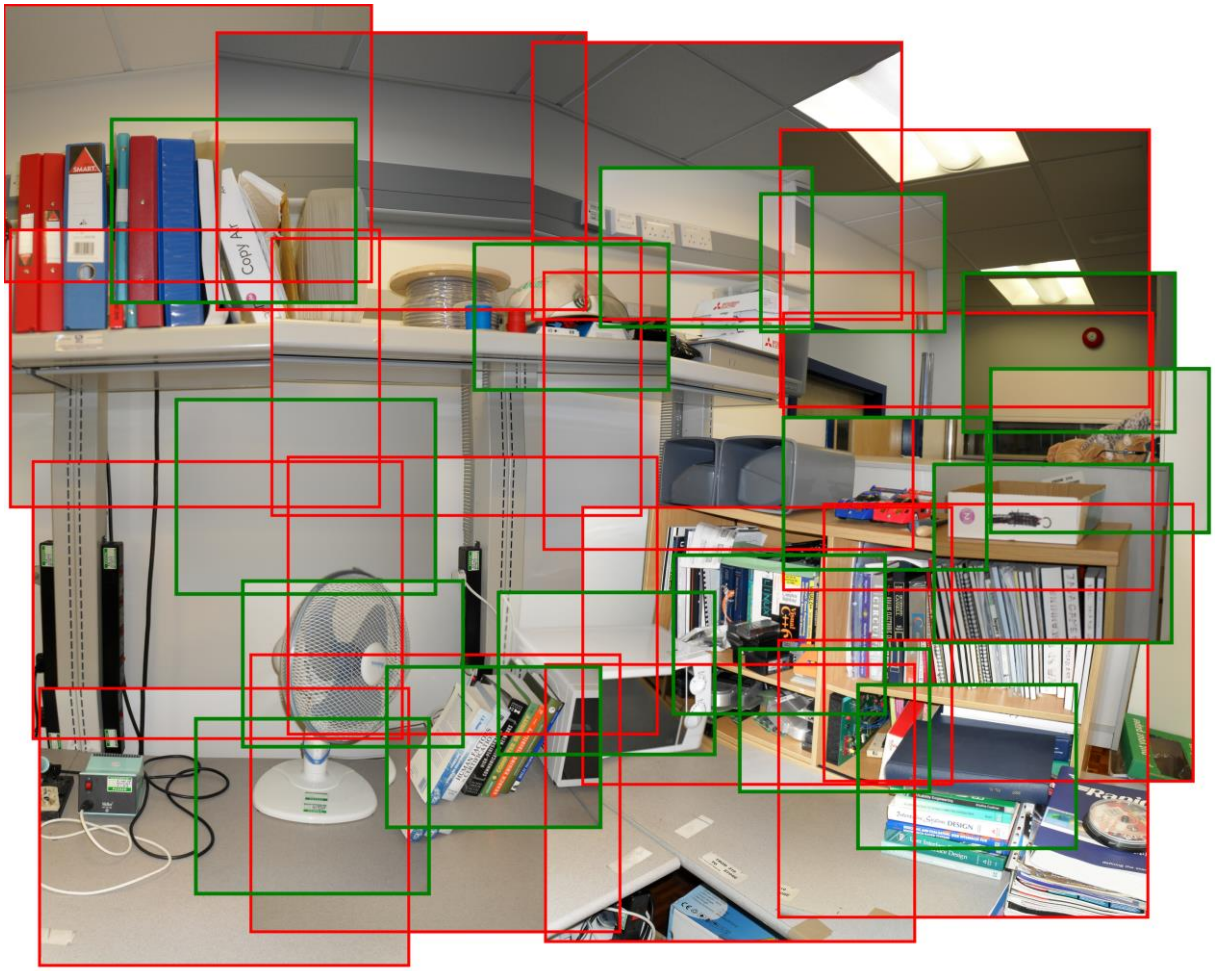


FIGURE 33. ORIGINAL (RED) AND ZOOMED-IN (GREEN) PICTURES USED FOR THE COMPOSITING EXPERIMENT

A custom software application was then used to construct randomised separate sets of images from the 32 parts available. Each of these sets was to be used for the individual compositing test.

5.3.2. Image Set Generator Software Design

In order to minimise human bias from the selection of the image sets, a random number generator-based approach was adapted for the selection of images. Each image taken was assigned a position inside one of the two 4x4 matrices – one containing normal images and the other their zoomed in counterparts. Every image in either matrix had a list of images it overlaps with (in both normal and zoomed in matrix) assigned to it.

Normal images (without magnification) were specified to overlap all of their neighbours (due to the fact that this was a precondition followed when taking the photographs, as outlined in 5.3.1). Images with magnification were manually compared to their neighbours to discover their potential overlaps. An image overlap was defined as presence of an area shared between a pair of images and a lack of overlap specified as complete absence of common regions between a pair of images.

The result of these operations was a list of 32 images each having a list of overlapping images assigned to them. A random number generator was then used to generate arbitrary sets of images which could be evaluated against several rules to discover if they satisfy them.

The following rules were applied to each set and a random image set was generated as a result:

1. No magnification is used, all images have overlapping areas with one or more other images and not a single subset of images is separated from the others.
2. No magnification is used, at least one image or image subset has no overlaps with the remaining images/image subsets.
3. At least a single image of a greater magnification factor is used and all images have overlapping areas with one or more other images. Not a single subset of images is separated from the others.

In the context of the rules above, a *subset of images* is defined as a set of overlapping images covering a continuous area of the scene.

For each of the rules above, sets of 3, 6 and 9 images were generated, for a total of 9 different sets. The varied number of source images is used to allow analysing the effectiveness of the algorithm when supplied with limited or abundant source data. Example of image sets processed by different applications can be seen on Figure 34.

5.3.3. COMPOSITE IMAGE GENERATION

Each of the 9 image sets was supplied to the compositing applications, where their automated processing mode was allowed to generate the stitched composite. Where the application managed to generate several different composites (AutoPano Pro 2 when supplied with non-overlapped images), the one with the largest number of images was used. If several resultant composites had the same largest size, the one that had a match among the other algorithms was picked (to make visual evaluation easier and fair).

Once all the composites were generated, images in the same set (for fair comparison conditions) were paired up. A random number generator was used to arrange the images in each pair (left or right) and the order of the pairs themselves. The result of this operation was a sequence of 27 image pairs (3 pairs for each of the 9 conditions formed by a combination of 3 set sizes and 3 set conditions).

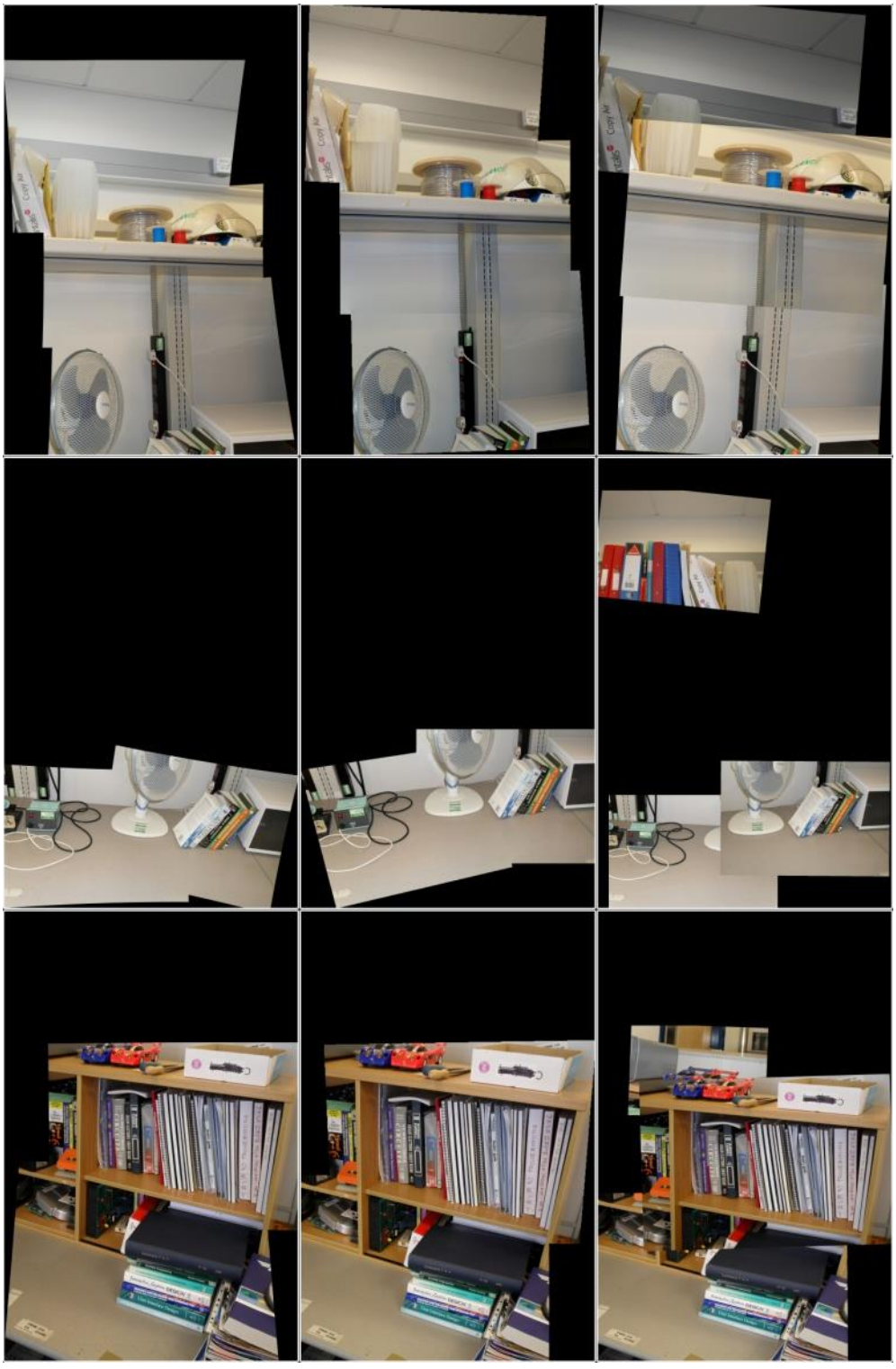


FIGURE 34. EXAMPLE IMAGES FROM THE EVALUATION
LEFT TO RIGHT: A) AUTOPANO PRO 2, B) MICROSOFT IMAGE COMPOSITE EDITOR, C) SENSOR-BASED COMPOSITING
TOP TO BOTTOM: 1) IMAGES UNDER NORMAL CONDITIONS WITH OVERLAPS, 2) IMAGES HAVING UNCONNECTED REGIONS, 3) IMAGES INCLUDING INCONSISTENT MAGNIFICATION LEVEL

Figure 34 shows examples of the images for different conditions. These images were presented in pairs in randomised order and position (left or right) to avoid biasing their decision as described in the following section.

5.3.4. WORKSHEET AND STUDY DESIGN

The images composited via different algorithms have to be presented to the study subjects in pairs, allowing them to select one better representing the scene. Paired comparison is one of the best ways to perform image quality evaluation (David, 1988; Bernas, 2002; Shen et al., 2009; Zhang & Xu, 2009). For this purpose, a set of worksheets was designed in a manner aimed to reduce the effects of biasing users towards an algorithm.

To make the worksheets more unbiased, the entire design was driven by a random number generator. Once the images were paired up, each pair's image position (left or right for side by side viewing) was decided by a random number roll.

After the image positions inside the pairs were assigned, the pairs themselves were shuffled to randomise their order. An interpretation list was also generated, showing what algorithm and rule set each image in the pair belonged to, in order to facilitate decoding each participant's choice after the study. The sequence of image pairs was combined into a Microsoft PowerPoint presentation to facilitate displa.

Each pair of images was given a sequential number. A worksheet was designed consisting of 27 pairs of checkboxes, each pair identified by a sequential number. Every checkbox was marked with "Left " or "Right", representing left or right image in a pair.

5.3.5. STUDY PROCEDURE

Study took place in a lecture theatre at the University of Birmingham and involved 26 participants (5 female and 21 male, all aged 18-35 with the average age of around 20 years) who participated in the study at the same time.

All the participants were seated at the tables in the lecture theatre, with 1 to 2 participants per table. They were each given a worksheet on which they were asked to mark their initials.

The main projector of the lecture theatre was then used to project a Microsoft Powerpoint presentation on a screen in front of the participants. The presentation initially showed the original scene and the participants were told the following:

“In front of you is a scene that you will be shown pairs of images of. We want you to mark which if the two images presented you would prefer to use if you needed to reconstruct the original scene, left or right, on the worksheets you were given”.

Image of the original scene was then removed and replaced with a pair of composites. Nobody at this stage, not even the study conductor knew which composite belonged to which software and the participants weren’t even aware of the number of approaches tested.

Once all participants have marked their answer, the next pair of images was shown, with the number of the pair clearly visible on the screen and verbally read out to the participants.

The process continued, showing one image pair at a time until all 27 were shown and selections marked on the worksheets. Each image pair was shown for 20-40 seconds (based on the time it takes every participant to mark their choice on the worksheet). Once every participant marked their choices for a specific image pair, the next pair in sequence was shown.

Worksheets were then collected and analysed as described in the following sections.

5.4. Analysis of the Perceptual Evaluation Data

Worksheets with the decisions made by the subjects of the study were entered into a spread sheet with the help of the interpretation list generated during the pair sequence creation.

The further analysis was dependent upon the choice of the algorithm to use.

5.4.1. CHOICE OF THE ANALYSIS ALGORITHM

There are a number of different image quality evaluation algorithms available (Shen et al., 2009; Bernas, 2002; Zhang & Xu, 2009). Objective algorithms look at the various aspects of the image, such as noise, gradient, colour reproduction and others in order to estimate the quality. Their results can often vary greatly and be inconsistent with the subjective quality perception (Zhang & Xu, 2009).

The traditional objective algorithms have their typical use in evaluating images that are very similar, such as differently compressed versions of the same picture. Because different compositing algorithms will alter the source images in different ways, they will generate significantly different images from the point of view of the objective evaluation techniques. Judging which image is better or worse will be difficult purely from the point of view of image quality. Thus a non-objective analysis technique is required.

In order to produce a reasonable evaluation, primarily a subjective analysis technique will be used. A common technique for evaluating image data is paired comparison technique (David, 1988). This technique is widely used wherever images differ in ways that are hard to quantify and contrast using objective techniques (Dickey-Bryant et al., 1986; Whisney et al., 1979; Kelly et al., 1999; Im, 1984; Louviere & Meyer, 1976; Kuang et al., 2007; Handley, 1081) and is one of the standard image comparison methods.

For this analysis, the images are split into pairs and the subjects asked to select preferred image out of each pair. The results are then be collated and analysed in concordance with Thurston's 5^{th} law (Thurstone, 1927), application of which is simplified greatly by using Guilford's approach to psychometrics, an efficient shortcut for Thurston's method (Guilford, 1928).

Guilford's approach is a two stage method:

1. A normal probability is determined for a certain stimulus to be selected by the subjects
2. This probability is converted to a z-score (standard score)

The probability is calculated according to the following formula:

$$p_k = \frac{N_k + 0.5 \times N_s}{N_s \times N_r} \qquad \text{(Eq. 5.4)} \quad [1]$$

Where p_k is the probability of the stimulus k being selected, N_k is the number of participants who preferred k, N_s is the total number of participants and N_r is the total number of the stimuli.

This probability can then be converted to a z-score either by looking it up in a Fechner's or Pearson's table as suggested by Guilford (1928) or by using standard z-score calculation means such as Microsoft Excel built-in statistical functions or z-score calculators.

Z-score representation of the probability is naturally on an interval scale (meaning relative distances between values on the scale are representative of how far apart these are in preference), making it easier to compare different stimuli against each other.

The accuracy of the results can be displayed on this interval scale by producing 95% confidence error bars according to the Monte Carlo simulation results (Montag, 2003). These should reflect the variance of data and how its adequacy for this study.

5.4.2. RESULTS

There are a number of criteria that could be examined to evaluate the effectiveness of the compositing algorithm. Out of them, two were selected, that provide a reflection of both subjective and objective (to a degree) performance – percentage of images consumed (objective measure) and the subjective image quality. These two measures should provide a

[1] (Guilford, 1928)

reasonable indication of how well the sensor-based compositing approach performs when compared to the conventional image processing-based ones.

5.4.2.1. Percentage of the Images Used by the Algorithms

Percentage of images used by the algorithm does not provide a good representation of the performance, since it only reflects on how much of the input the algorithm actually uses but is still an indication of the effectiveness of an approach and its comprehensiveness.

TABLE 2. IMAGE USAGE BY COMPOSITING ALGORITHMS

Criteria	Total Source Images	Images Used by AutoPano Pro 2		Images Used by Microsoft Image Composite Editor		Images Used by the Sensor-Based Compositing	
Normal Overlapped Images	3	3	100%	3	100%	3	100%
	6	4	67%	4	67%	6	100%
	9	8	89%	8	89%	9	100%
Images with Gaps	3	2	67%	2	67%	3	100%
	6	2	33%	2	33%	6	100%
	9	5	56%	7	78%	9	100%
Images with Zoom	3	2	67%	2	67%	3	100%
	6	3	50%	2	33%	6	100%
	9	5	56%	6	67%	9	100%
Total	54	34	63%	36	67%	54	100%

Table 2 shows the image usage results for the algorithms. The sensor-based approach accepts all the source images that contain the valid orientation tags. Image processing-based approaches have to rely on matching features in the overlapping region of images. When such a match cannot be made conclusively or where an overlap is not available, the

image processing algorithms have to reject a proportion of the source images that they cannot compose onto the panorama.

Even when faced with normal images having overlaps, the image processing approaches failed to use all of the source images, being unable to match features from some of the photographs.

FIGURE 35. EXAMPLE OF IMAGES LACKING DISTINCTIVE FEATURES TO BE MATCHED AND COMPOSITED

Figure 35 illustrates an example where overlapping images could not be composited by the image editing software because of lack of any distinctive features in the overlap area. This example is from a normal image set of 6 source images. The two regions of the composite are connected by the images shown, that seem easy to compose and match to a human being, but posed a problem to the image processing algorithm. The wall pattern was not distinctive enough for image processing, nor was the front and top of the microwave oven.

This shows the advantage of the sensor-based processing when photographing scenes which lack easily distinctive markings and features.

5.4.2.2. Perceptual Quality

Quality evaluation was performed using paired comparison method with the z-score calculation done using Guilford's approach, outlined in 5.4.1. Every separate calculation was

performed independently (the choices were collated and analysed independently for every separate occurrence), rather than inferred from prior calculations.

5.4.2.2.1. Overall Results

Figure 36 shows the overall results for the paired comparison test under the main conditions evaluated:

1. When images fully overlap
2. When images have gaps
3. When images include zoomed in versions

For all of the given conditions the sensor-based compositing approach (highlighted in green) was the preferred approach. It is worth noting, that the individual conditions cannot be compared to each other directly, since they all have a different zero point. Generally, zero point on an interval scale is arbitrary, so only the relative positions of each compositing method for separate conditions should be compared.

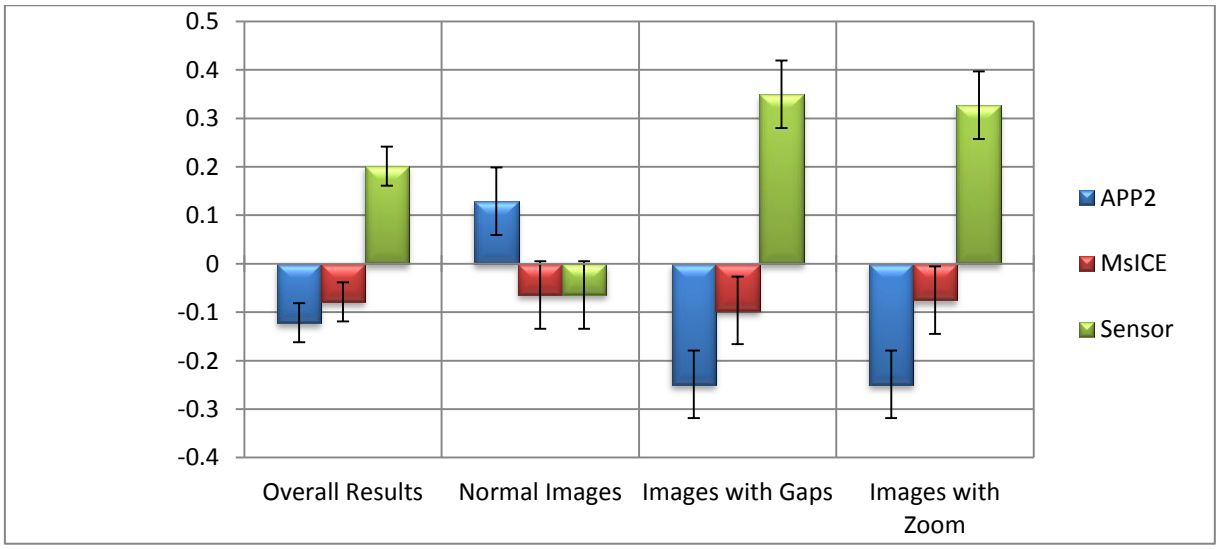


FIGURE 36. OVERALL IMAGE QUALITY RESULTS (SENSOR IS SENSOR-BASED COMPOSITING, MSICE IS MICROSOFT IMAGE COMPOSITE EDITOR AND APP2 IS AUTOPANO PRO 2). ERROR BARS SHOW 95% CONFIDENCE RANGE

Sensor-based approach was clearly preferred to the image compositing methods when faced with abnormal conditions (such as after introducing images which had incomplete overlaps

between regions, forming gaps, or when using images at different magnification factors). When faced with normal operating conditions (images having full overlaps and being at consistent magnification), the commercial software AutoPano Pro 2 took the lead.

The following sections will examine in greater detail the individual subsets of the results

5.4.2.2.2. Normal Images with Overlaps

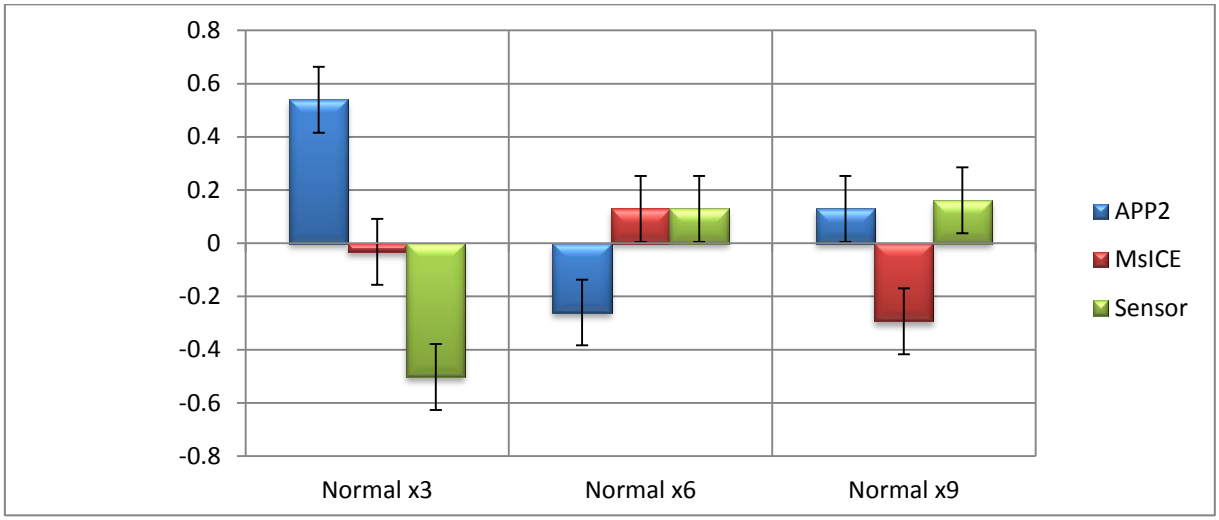


FIGURE 37. IMAGE QUALITY EVALUATION RESULTS FOR THE NORMAL CONDITIONS, WITH OVERLAPS AVIALABLE BETWEEN IMAGES (SENSOR IS SENSOR-BASED COMPOSITING, MSICE IS MICROSOFT IMAGE COMPOSITE EDITOR AND APP2 IS AUTOPANO PRO 2). ERROR BARS SHOW 95% CONFIDENCE RANGE

The results of the analysis of normal images are shown on Figure 37. Sensor-based compositing was clearly behind the image processing-based approaches when the image set was small and therefore the inconsistencies and seams caused by the composition were clearly visible. As image set grew, the scale of the overall image decreased, hiding the finer details and imperfections. As the size of the image set increased, so did the appeal of the sensor-based compositing due to the drop in finer detail perceivable by the participants because of the size of the overall composite. This becomes more apparent as the image sets with varied image count are compared to each other.

5.4.2.2.3. Image Sets with Different Image Count

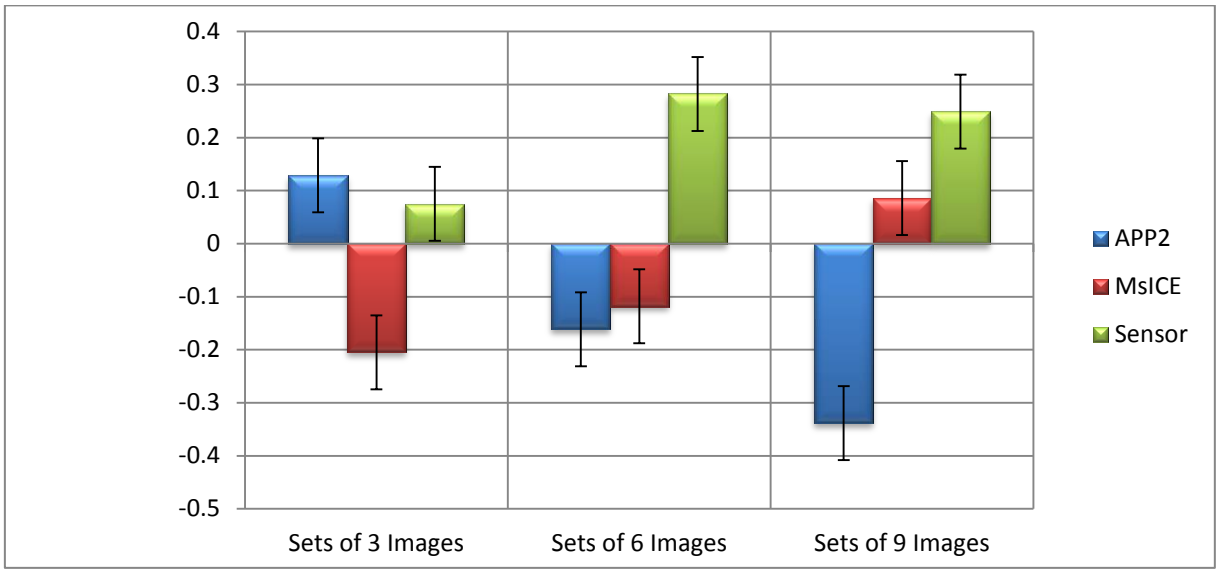


FIGURE 38. IMAGE QUALITY EVALUATION FOR THE IMAGE SETS WITH DIFFERENT IMAGE COUNT (SENSOR IS SENSOR-BASED COMPOSITING, MSICE IS MICROSOFT IMAGE COMPOSITE EDITOR AND APP2 IS AUTOPANO PRO 2). ERROR BARS SHOW 95% CONFIDENCE RANGE

Figure 38 shows comparison of image sets grouped by the image count. The conclusion outlined in 5.4.2.2.2 appears to hold true. Increasing image count increases the appeal of the sensor-based compositing. Image processing approaches seem to fair in a similar fashion, with AutoPano Pro 2 producing finer quality results when dealing with the small image count, but falling behind and being overtaken by Microsoft Image Composite Editor, that produces better looking large image count composites.

5.4.2.2.4. *Images with Inconsistent Magnification*

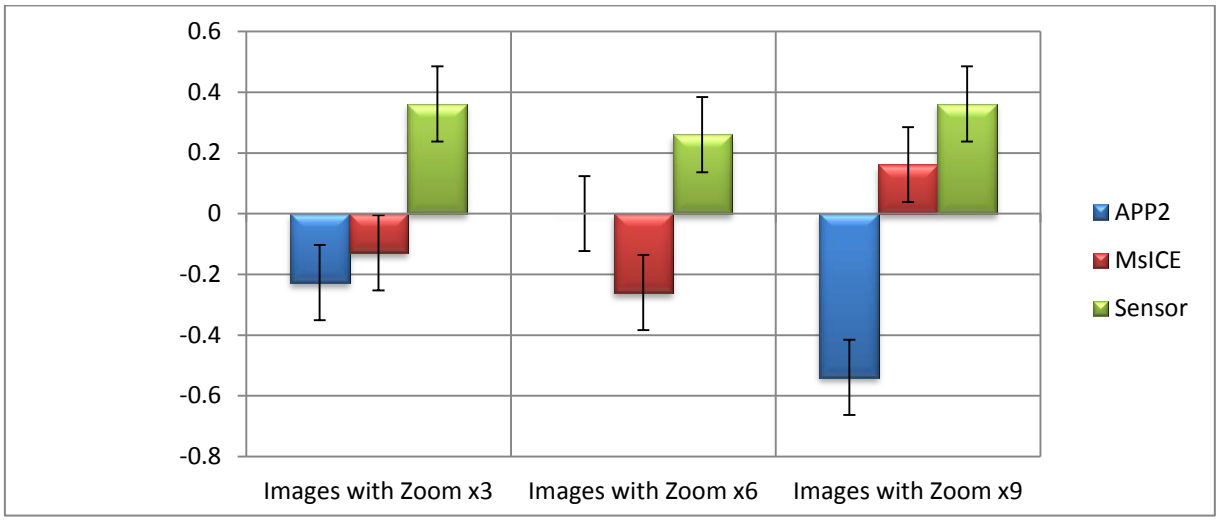


FIGURE 39. IMAGE QUALITY EVALUATION FOR THE IMAGE SETS WHICH INCLUDE ZOOMED IN IMAGES (SENSOR IS SENSOR-BASED COMPOSITING, MSICE IS MICROSOFT IMAGE COMPOSITE EDITOR AND APP2 IS AUTOPANO PRO 2). ERROR BARS SHOW 95% CONFIDENCE RANGE

When faced with the images at varying magnification, sensor-based compositing is the preferred way of handling the stitching. For the larger scale composites, containing more source images, its lead begins to drop. When examining the corresponding image consumption statistics (outlined in 5.4.2.1), it is possible to theorize that the unused photographs begin to form a smaller proportion of the overall composite as the image count increases, leading to users noticing missing elements less and paying greater attention to the quality of the overall composite. This assumption is further reinforced when examining images with gaps.

5.4.2.2.5. *Images with Incomplete Overlaps Causing Gaps*

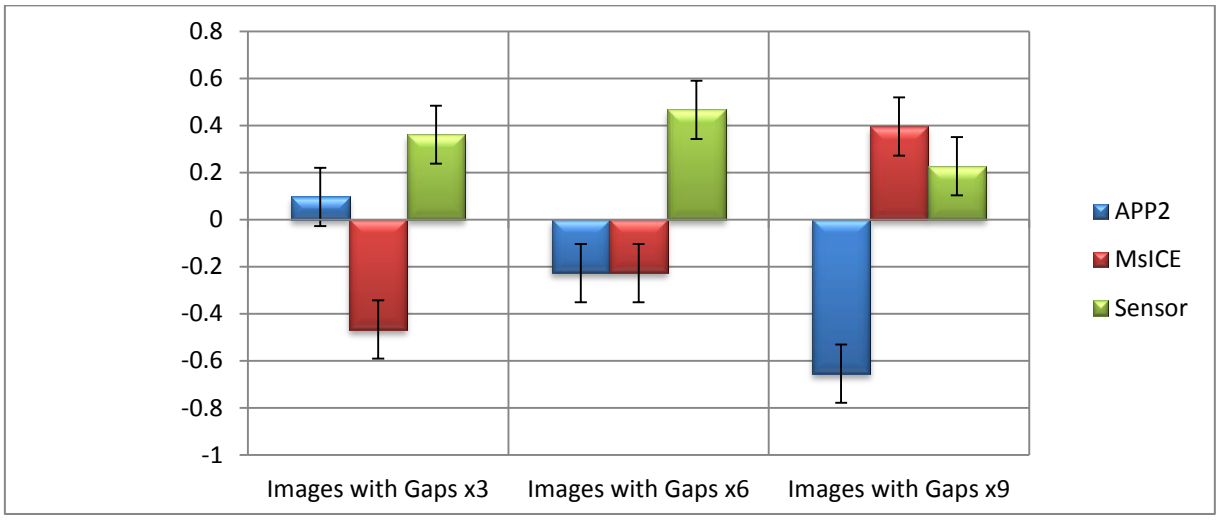


FIGURE 40. IMAGE QUALITY EVALUATION RESULTS FOR IMAGE SETS CONTAINING GAPS WITHOUT OVERLAPS (SENSOR IS SENSOR-BASED COMPOSITING, MSICE IS MICROSOFT IMAGE COMPOSITE EDITOR AND APP2 IS AUTOPANO PRO 2). ERROR BARS SHOW 95% CONFIDENCE RANGE

Figure 40 shows surprising results. Despite the sensor-based approach using more images and generating more complete composites, when faced with the maximum amount of images the MsICE (Microsoft Image Composite Editor) approach takes the lead. This can be explained by either the reduction in the percentage of the rejected images (i.e. if 1 out of 9 images is not included in the composite, it is less noticeable than 1 out of 3). It could also be explained by the peculiarities of the specific image set randomly generated in a manner that allowed Image Composite Editor (ICE) to process it despite its incomplete overlap nature.

FIGURE 41. IMAGE SET CAUSING ABNORMAL RESULTS. MICROSOFT IMAGE COMPOSITE EDITOR ON THE LEFT AND SENSOR-BASED COMPOSITING ON THE RIGHT

Figure 41 reveals the reason for the apparently abnormal results of the previous analysis. The image set generated by the random set generator includes a very wide panorama that is not easy to map to a 2D plane without distorting it. And the Distortions performed by the image processing software make it more appealing to the observer despite the sensor-based approach including additional information about several areas of the target scene.

5.5. Analysis of the Object Capture Capability of Composites

In order to evaluate the utility of the composites generated during this study, it is essential to examine not just their perceptual quality (done using paired comparison in 5.4) but also the utility of the images. A degree of utility was evaluated already during the paired comparison test since the participants were tasked with selecting a picture better suited for reconstruction of the original scene. However, by examining the objects captured on each composite, a more objective analysis can be performed.

Looking at the original scene on Figure 32, objects present in the entire image can be listed. These can be constrained to features whose dimensions are less than those of a single non-zoomed photograph (to avoid counting table which is present in the majority of photographs, for example). The complete list of the objects (moving from left to right and then top to bottom through individual subsections, such as table, shelf, etc.) would then be:

1. Folder
2. Folder
3. Folder
4. Book
5. Book
6. Folder
7. Folder
8. Book
9. Paper stack
10. Transparent folder
11. Thick wire roll
12. Blue wire roll
13. Red wire roll
14. Ethernet socket
15. Ethernet socket
16. Power socket
17. Power socket
18. Helmet
19. Small white box
20. Small white box
21. Ceiling light
22. Fire alarm
23. Black power socket
24. Black power socket
25. Soldering iron
26. Soldering base unit
27. Desk fan
28. Black wall socket
29. Book
30. Book
31. Book
32. Book
33. Book
34. Book
35. Book
36. Microwave oven
37. Small bin
38. Small bin
39. Blue toy car
40. Red toy car
41. Manual tool
42. Manual tool
43. White box
44. Brown toy animal
45. Grey toy animal
46. Shelf of books (no exact count)
47. Black box in packaging
48. Black box
49. Black box
50. Hole puncher
51. Shelf of books
52. Aluminium enclosure
53. Black box
54. Aluminium enclosure
55. Black box
56. Black box in packaging
57. Black box in packaging
58. Prototyping board
59. White and red box
60. Shelf of books
61. Book
62. Book
63. Book
64. Book
65. Book
66. Book
67. Book
68. Book
69. CD media
70. Stack of journals (no exact count)
71. Paper disposal bin

There are many different methods to identify objects. In this case, wherever it was possible to split a set of objects into individual ones, it was done. Where the angle of view or density of objects didn't allow them, these were counted as a single larger object (such as book shelves and journal stack). Using these objects that were identified, it is now possible to examine composites from the object content point. All of the objects thus detected can either be present/absent or duplicated (because of composition misalignment). An object is considered duplicated if it is present disjointed on several distinct photographs and not composed to form a singular entity.

TABLE 3. OBJECT COUNT IN COMPOSITES

Image Count	Condition	AutoPano Pro		Microsoft ICE		Sensor Compositing	
		Objects	+Duplicates	Objects	+Duplicates	Objects	+Duplicates
3	Normal	16	0	16	0	16	0
6	Normal	35	0	35	0	45	10
9	Normal	29	0	29	0	50	9
3	With Gaps	14	0	14	0	24	1
6	With Gaps	10	0	10	0	51	5
9	With Gaps	24	0	42	0	52	16
3	With Zoom	17	0	17	0	26	7
6	With Zoom	21	0	17	0	42	8
9	With Zoom	28	0	31	0	48	10

Examining the object count in composites shows that two image compositing-based methods show similar results under normal conditions. When introduced to images at varying magnifications and images without complete overlaps, image processing methods form different composites, including different objects. What remains common between them though is the fact that in all composites the object edges are well composited, leading to no duplicate object creation under all tested conditions.

Sensor-based imaging, on the other hand, creates duplicates in all but simplest cases. Whilst the amount of objects captured typically exceeds those captured on image processing-based composites, the error rate of 19% is very high. Human beings can recognise and

compensate for the inadequate composition to a degree due to the way object recognition in images is handled by human vision (Biederman, 1987). The large number of duplicates generated by alignment errors (mostly due to sensor inaccuracy and the assumption that the camera rotation occurs around its centre of projection) mean that a hybrid approach combining the composition strength of image processing with the ability of sensor-based method to compose abnormally overlapped images is likely to provide the best result. Such a hybrid approach could produce the combination of high object capture property of the sensor approach with the low duplicate count property of the image processing-based ones and is illustrated in the following section.

5.6. Future Developments

Sensor compositing approach provides several useable enhancements to existing panoramic stitching mechanisms. While the results of paired comparison test could justify the use of the sensor-based system in its current state, both the object capture analysis and empirical observation show that compositing quality of the image processing methods is far superior aesthetically (as can be seen by visually comparing Figure 42 and Figure 43).

FIGURE 42. RESULT OF PURELY SENSOR-BASED COMPOSITING

FIGURE 43. RESULT OF PURELY IMAGE COMPOSITING-BASED STITCHING

Therefore a hybrid method, combining strengths of different approaches, is probably the best choice if this system is to be used commercially outside of the environment where evidential integrity of the images is paramount.

Such a hybrid system would make use of image stitching for overlapped images and use links created by sensor compositing to connect disjoined areas. Such an example link, generated from sensor compositing panorama is shown on Figure 44:

FIGURE 44. LINK FOR IMAGE PROCESSING COMPOSITING GENERATED VIA SENSOR-BASED ONE

The result of a composite generated using the link image above is shown on Figure 45

FIGURE 45. EXAMPLE COMPOSITE GENERATED USING A HYBRID APPROACH

This composite is generated using Microsoft Image Composite Editor with the sensor link image used to allow image processing to connect disjoined regions on the right and left of picture. It provides both the object retention of sensor compositing and lack of duplicates of

image processing. By comparison, purely sensor-based and image processing techniques are less effective even to a casual observation.

Another useful feature of the sensor-based compositing is its low processing and resource requirements, allowing for very large panoramas to be generated in real time. Modern mobile devices come with a variety of applications, including those that permit panoramic compositing, in some cases making specific implementations of such compositing the unique selling point (USP) of such a device (Apple Inc., 2013). The issue with such a panoramic render is that its maximum size is often limited by the memory available to the mobile device, requiring keeping composites loaded while they are being extended by new photographs.

Sensor based compositing can help provide an alternative way of generating large scale composites easily. By using sensor composites to guide the photographer in orientation of the camera for optimal capture, a set of photographs of arbitrary size, mixing magnification levels (which frequently poses an issue for panoramic compositing application on mobile devices) can be generated with ease and then post-processed using the most effective image processing-based method.

As an example of such compositing tool, an application was created on Android mobile phone. This application uses built in phone sensors to track phone camera orientation and provide guidance to the user of the achieved overlap between images taken with the phone camera, with instant feedback showing relative overlap of the area currently pointed to by the camera and all of the photographs taken to date. Unlike existing image processing-based approaches that cannot handle zoom or panoramas consisting of large number of photograph, the sensor based approach experienced no slowdowns when tracking over 50 photographs taken at different magnification levels.

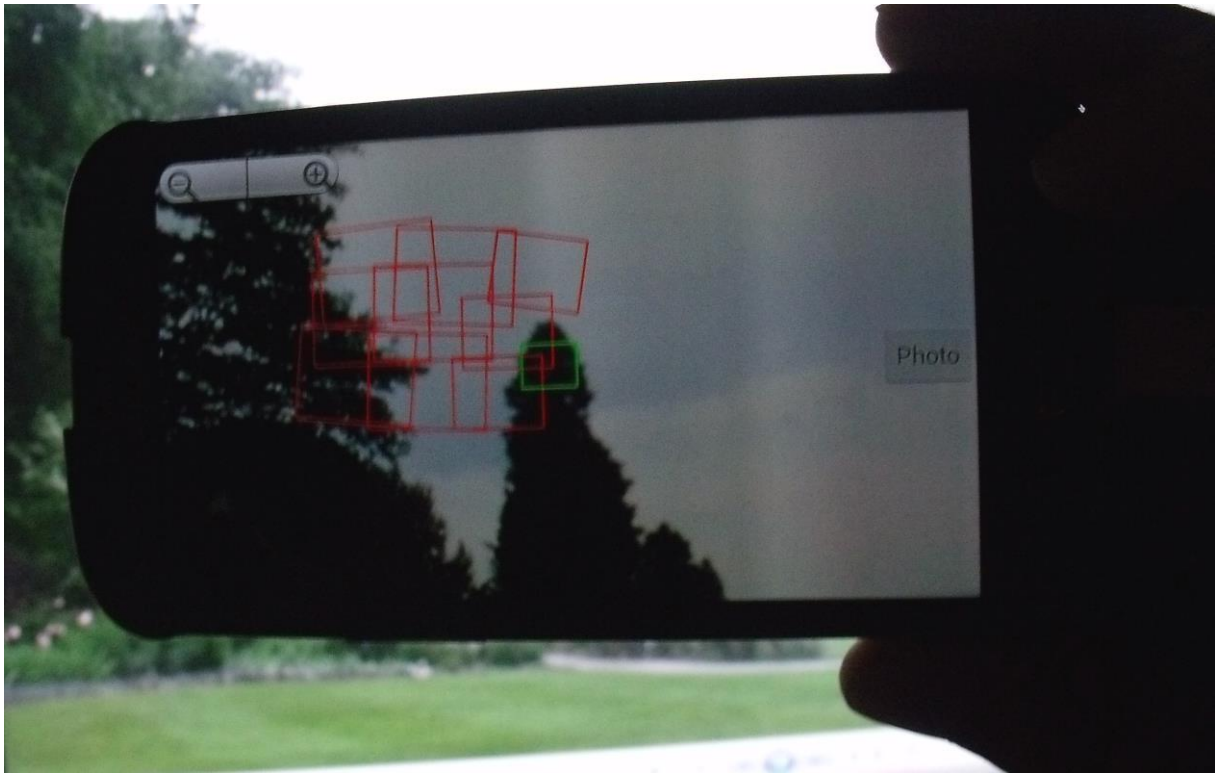


FIGURE 46. EXAMPLE ANDROID COMPOSITING ASSISTANT APP. RED RECTANGLES REPRESENT IMAGES TAKEN SO FAR. GREEN RECTANGLE REPRESENTS CURRENT IMAGE (VISIBLE CAMERA AREA ON SCREEN) IN RELATION TO THE REST, FACILITATING COMPOSITING IMAGE SET CREATION.

Whilst the example Android software does not do compositing itself, photographs created with it can then be supplied to standard image processing suite, such as the ones evaluated in 5.3.2 knowing that all the images are overlapping and will not experience issues during composition..

5.7. CONCLUSIONS

A sensor-based image compositing approach was designed and implemented. The compositing relies on tracking the camera orientation around 3 axes and stitches photographs based on their angle of view. The resultant panoramas contain unmodified original images rearranged to reflect their relative positions in the scene being photographed. All of the sensor data is provided by two sensors – a tilt compensated magnetic compass and an ultrasonic rangefinder. Rangefinder use is secondary and it can be potentially replaced by image processing to determine optimal occlusion (Zelnik-Manor & Perona,

2007). Cameras with digital compass already exist (Casio America, Inc, 2010), as do mobile phones (HTC Corporation, 2011), making the choice of the main sensor justifiable commercially. The processing algorithms are very light and require negligible resources, treating the image content of the picture as blank canvas for the purposes of compositing. This makes it feasible to implement this approach in both mobile phones and in digital cameras as an added feature. As can be seen from combined images (on Figure 32 and Figure 33), taking a photograph without visual feedback for a panorama is difficult. Sensor approach would enable seeing coverage and composition overview without image compositing that requires significantly more processing to perform in real-time mode.

The system is capable of storing additional information about the subject of the photograph, currently implemented as a user-definable category, but this can be expanded to be plain text entry or description entered using a mobile phone keyboard. The images can then be filtered by this category and if needed highlighted. This allows more robust management of scene information and its self-contained storage directly inside the metadata of the relevant images.

Sensor-based imaging arranges the images in the same way relative to each other independent on the amount of images supplied. Image processing-based approaches, on the other hand, alter their results dependent on the image set supplied, performing different degrees of warping and colour correction for different image sets. This leads to non-consistent relative image position information. Whereas with the sensor-based approach all the information about the image orientation is contained inside the image itself and doesn't alter based on context, the image processing approaches alter their results dependent on the context and surrounding images.

The sensor-based compositing approach was evaluated against two image processing-based ones – a commercial product AutoPano Pro 2 and a freeware Microsoft Image Composite Editor using a method of paired comparison. A set of 26 experiment participants

graded their preference of image in each pair and the results were arranged on an interval scale, illustrating the relative preference of the compositing approaches under a variety of conditions.

Evaluation of the compositing results showed that it performs on par with image processing algorithms when images are overlapped, being more efficient than them when the overlapped area does not contain features useable for the compositing feature recognition to match. When faced with abnormal situations, such as when the images do not overlap or are at different magnification levels, the sensor-based approach outperforms both of the image processing-based approaches. The non-sensor approaches, however, produce better results for wide panoramas, where the distortion and warping introduced by the image processing improves the appearance of straight lines and makes the results more realistic.

Analysis of object capture capability of different approaches, however, highlighted a weakness in sensor-based compositing. Without image processing to compensate for misalignment between camera sensor and centre of rotation, the resultant composites introduce a number of false duplicates of objects.

As a result of these evaluations, two enhancements to the sensor compositing system were proposed – a hybrid approach that combines sensor and image processing compositing and a real time feedback system based on sensor compositing for use on photo capture devices. These enhancements offer viable commercial applications of sensor-based compositing.

6. SENSOR-BASED IMAGE SPATIAL MANAGEMENT SYSTEM DESIGN

Chapters 4 and 5 investigated sensors usable for pedestrian dead reckoning and composition using sensor-based approaches. This chapter combines their results into a single complete solution to spatial image management. It describes the design and features of the spatial image management system (SIMS). The following chapter then focuses on the evaluation of this system and quantifies its performance.

6.1. OVERVIEW

SIMS combines sensor-based compositing with pedestrian dead reckoning using sensor decisions made as a result of sensor comparison study in chapter 4. Both hardware and software design builds up on the results of previous chapters, with features added to further benefit the potential user. These features are described using an example of a fire inspection task in 6.4.

6.2. WEARABLE MODULE DESIGN

The system requires changes to be made from the way the previous prototypes operated to accomplish the mapping task. The camera button-triggered sensor recording of the compositing sub-system's image acquisition is replaced by the continuous recording of the sensor recorder from 0. Additionally, the results of the previous experiments are factored into the design, affecting the sensor selection and arrangement.

6.2.1. HARDWARE DESIGN

The hardware for the enhanced compositing prototype represents the combination of the sensor recording unit used in the sensor study and the compositing hardware developed for the sensor-driven image compositing study.

Considering the different strengths and weaknesses of the two heading tracking methods identified as best performing – gyroscope integration and tilt compensated compass, a fusion of the two is preferable to using either sensor alone, thus both tilt compensated compass and a digital gyroscope were left on the torso sensor unit.

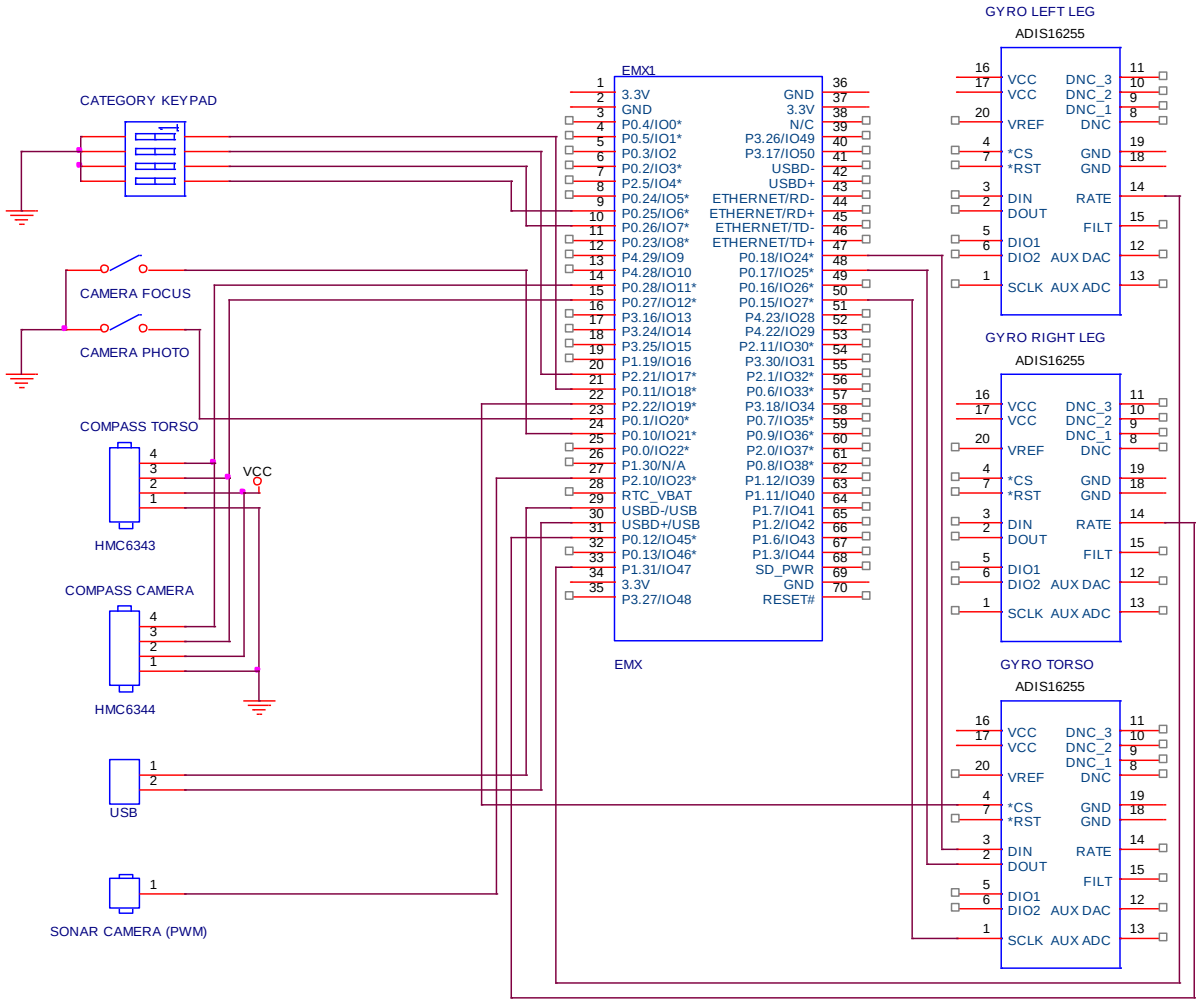


FIGURE 47. OUTLINE HARDWARE SCHEMATIC OF THE SPATIAL IMAGE MANAGEMENT SYSTEM. ONLY THE MAIN CONNECTIONS ARE SHOWN.

Figure 47 shows the hardware design outline schematic for the recording module. It is a fusion of the hardware designs from 0 and 5.2.1, providing both the sensor recording and camera orientation detection during snapshot.

6.2.1.1. Processing Unit

The design for the enhanced compositing module is again based on the .NET Micro Framework EMX processing unit. This module was used in the form of FEZ Cobra development system and is based on the LPC2478 microcontroller by NXP Semiconductors, the same microcontroller as used by the EmbeddedMaster module in the original sensor-based compositing design, but offers enhanced framework features and support.

FEZ Cobra system is housed in an enclosure with a 3.5'' 16-bit TFT display (shown on Figure 48), providing both user feedback and input through its touch screen capability. It is powered by a 12V Lithium Polymer battery via a switched mode regulator.

FIGURE 48. IMAGE MAPPING MODULE HARDWARE. BOTTOM LEFT - TORSO UNIT HOUSING GYROSCOPE AND COMPASS. TOP LEFT - DIGITAL CAMERA WITH SONAR, COMPASS AND NUMERIC PAD. TOP RIGHT - MAIN ENCLOSURE WITH FEZ COBRA SYSTEM AND TOUCH SCREEN. BOTTOM RIGHT - LEG GYROSCOPES

6.2.1.2. Sensors

The sensors used by the spatial image management system can be divided into two groups:

1. Sensors used for the compositing
2. Sensors used for the dead reckoning

Compositing sensors underwent no changes from the way they were arranged in the original compositing prototype. Evaluation showed that that sensor arrangement is adequate and produces results that are acceptable by users. Therefore the digital camera is still augmented with the Honeywell HMC6343 tilt-compensated magnetic compass, MaxSonar LV EZ4 narrow beam sonar and a category selection button pad.

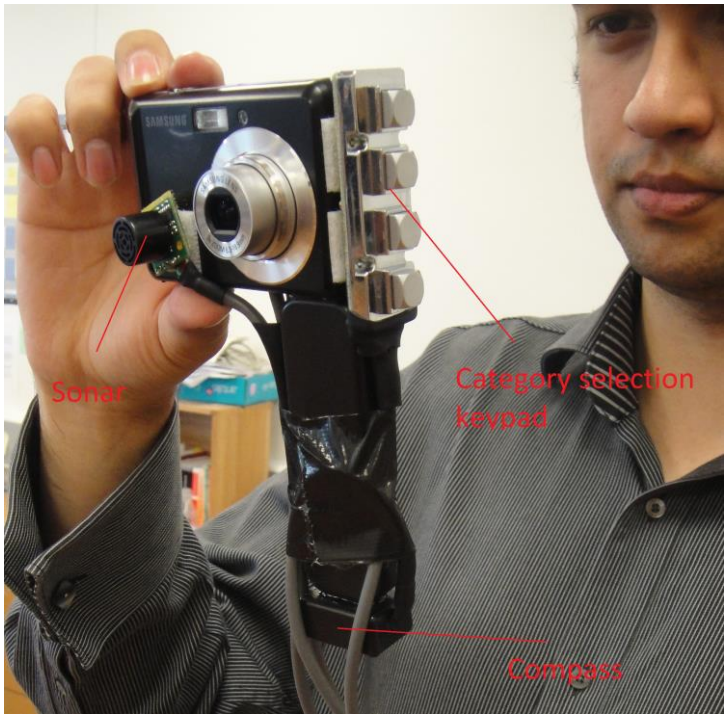


FIGURE 49. COMPOSITING SENSORS

The dead reckoning relies on two gyroscopes strapped to the legs of the user for the step detection as well as a combination of a gyroscope and a compass for the heading measurement, mounted on the torso (as described in 0). The sensors used are again Honeywell HMC6343 tilt-compensated compass as well as three AnalogDevices ADIS16255 digital gyroscopes (as shown on Figure 47). This part of the design underwent no changes from its implementation in chapter 4.

FIGURE 50. WEARABLE MODULE IN USE

It was decided to make the two groups of sensors work completely independently of each other, facilitating development and debugging of the system.

6.2.1.3. Camera

Due to the availability restrictions, Samsung ES15 camera is replaced with ES17, offering the improved image quality and resolution. As mentioned in 5.2, the replacement of the image acquisition unit is made trivial by the nature of the design.

Initial tests with the new camera revealed that it produces a strong magnetic interference that affects the compass readouts in certain quadrants. This is caused by the change in camera design and compass sensor positioning. Initial compositing system used compass mounted near the camera trigger button (as shown on Figure 24). This minimised the effects of the magnetic interference from the camera circuitry and lens magnification motor. Such a mounting, however, affected the ease of use of camera, making operation of camera trigger more awkward than without the added sensors.

It was therefore decided to reposition the magnetic compass to the bottom of the camera. Tests with this new positioning, however, showed that the magnetic interference effects had a significant impact on the compass sensitivity in certain compass quadrants (leading to a hypothesis that a soft magnetic interference is affecting one or more magnetometers in the compass module). HMC6343 has hard iron compensation circuitry and can compensate for the presence of nearby ferromagnetic components. Soft iron compensation, however, is more difficult to implement and calibrate (Ye et al., 2009) and is not available in this module (Honeywell International Inc, 2011).

To investigate the hypothesis that the errors in the compass heading are caused by soft iron interference, the outputs of the HMC6343 magnetometers were recorded while the module attached to the camera was slowly rotated around several of its axes and plotted on a 2D scatter plot. The results are shown on Figure 51 (only the relevant axes that had module fully rotated around are shown):

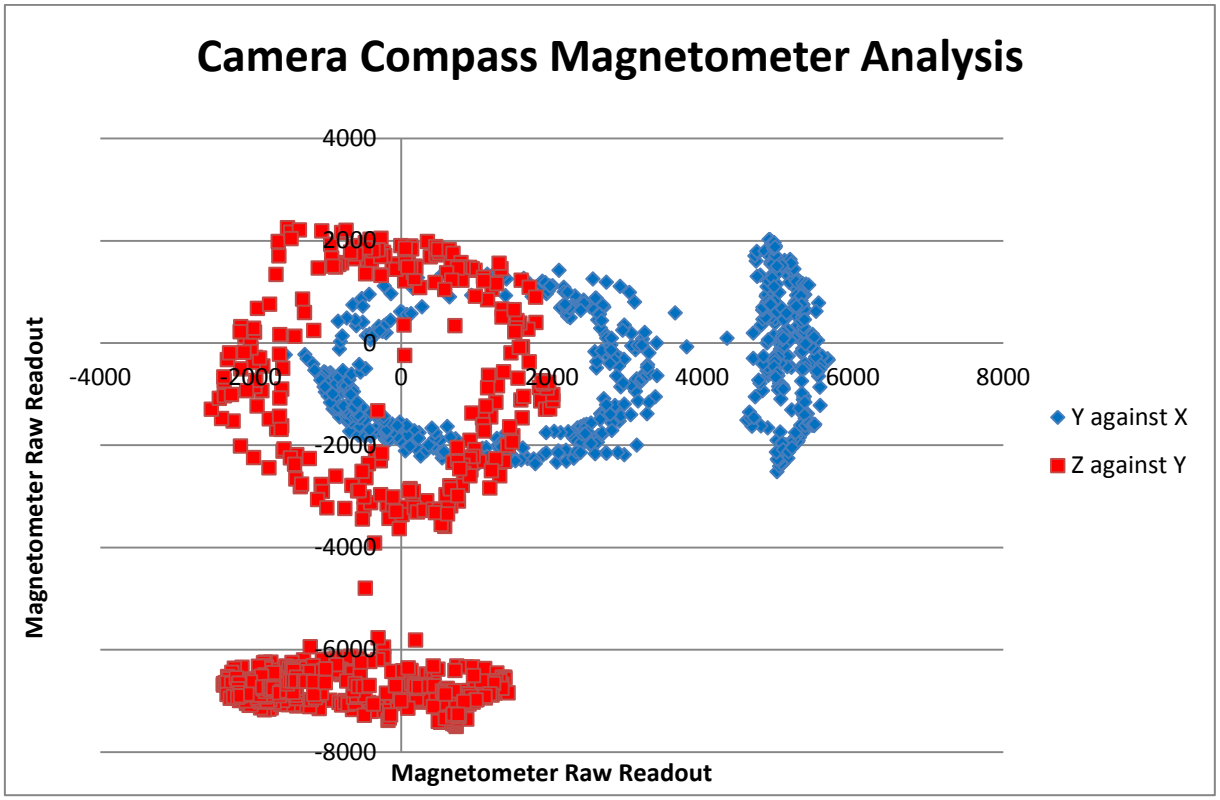


FIGURE 51. MAGNETOMETER ELLIPSES OF CAMERA COMPASS SHOWING THE EFFECTS OF SOFT AND HARD IRON DISTORTIONS

HMC6343 module has internal algorithms for pre-processing magnetometers, causing the vertically positioned magnetometer to have a large offset. Without knowing the exact processing applied to these magnetometers and the scaling factors that are not specified in the datasheet, it is very difficult to perform compass heading determination and tilt compensation manually based on these readings. What these reading do show, however, is the hard and soft iron distortions of the ellipses formed by the magnetometer scatter plot. The centres of the ellipses are offset from (0, 0) point, indicating hard iron distortions that HMC6343 can compensate for through calibration. The ellipse for Y axis magnetometer against X axis (effectively, camera horizontal plane magnetometers when camera is held in its normal landscape operating mode) is distorted, not forming a circle. This is indicative of a soft iron distortion. Such a distortion is not compensated for by HMC6343 (Honeywell International Inc, 2011) and cannot be manually compensated for (by re-implementing tilt compensation and heading calculation with distortions allowed for) without reverse

engineering magnetometer pre-processing and scaling done by HMC6343 unit that is not described in the module datasheet.

With the firmware (software) solution unavailable with the current module, it was decided to resolve the issue in hardware. This could be accomplished either by introducing the magnetic shielding to the camera or by altering the camera compass positioning. It was decided to reposition the compass and to do this a handle was introduced to the camera, with rigidly mounted compass attached to its opposite end from camera (this can be seen on Figure 49).

Retesting the magnetometers with the new compass positioning showed significant reduction in both soft and hard iron distortions (as shown on Figure 52):

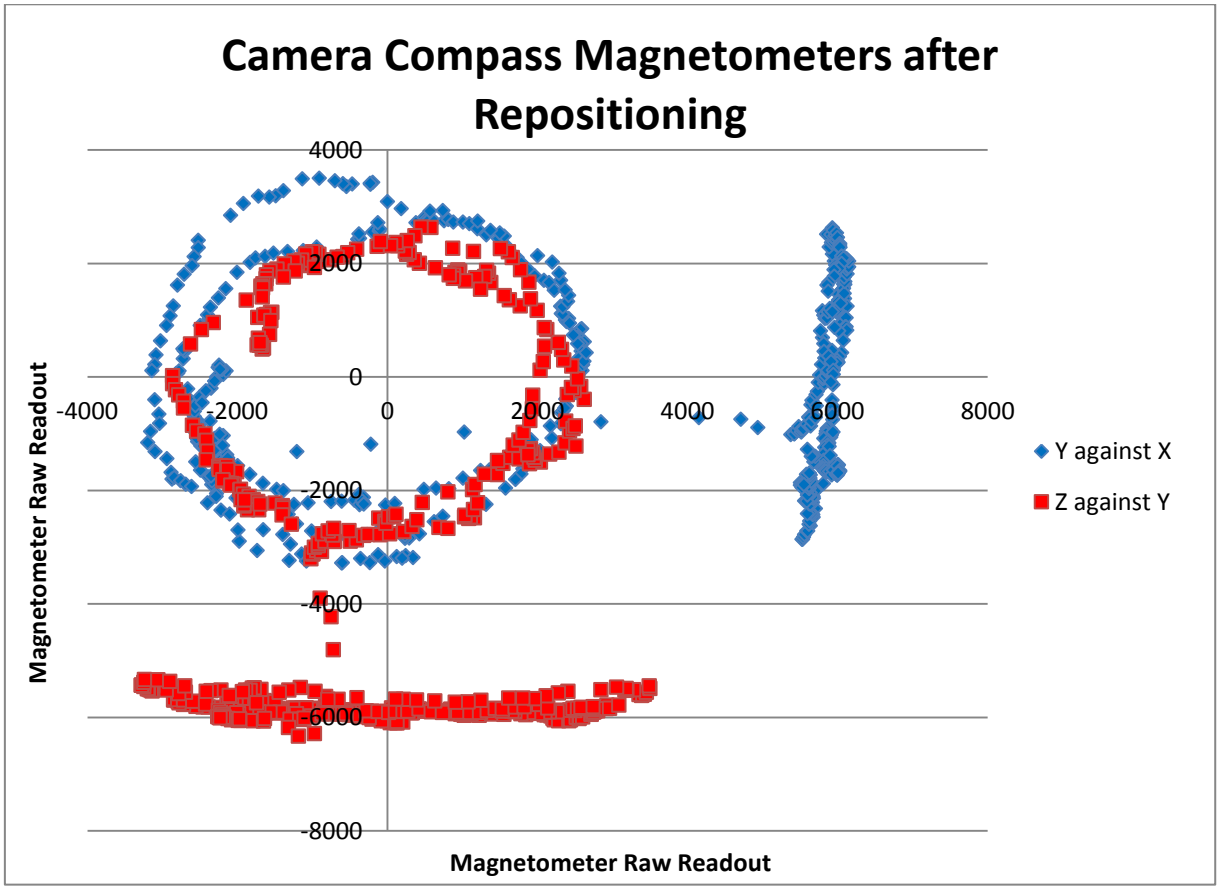


FIGURE 52. MAGNETOMETER ELLIPSES OF CAMERA COMPASS AFTER REPOSITIONING. SOFT AND HARD IRON DISTORTIONS ARE SIGNIFICANTLY REDUCED

Testing the compass heading output shows improved stability when tilted and accurate results in all compass quadrants (determined by comparing the compass output against an analogue magnetic compass).

6.2.1.4. Storage

Work with the previous prototype of the sensor-based image compositing module as well as the sensor recording module showed the feasibility of using low cost USB storage media as the persistent storage for the recording of the sensor data. Its fast seek and write speeds make such a media optimal for the real time recording of the sensor data. At the same time, low costs of such a device make its eventual deterioration and failure due to multiple write cycles easy to manage.

For ease of use, a single USB port built into the FEZ Cobra module was used (unlike a complete USB hub used in the initial compositing system described in 5.2.1). A low cost 2GB USB memory drive provided more than enough space for a walkthrough recording (since a set of 10 walks for sensor analysis study in 4 normally produces less than 1 megabyte of sensor data).

6.2.2. FIRMWARE DESIGN

The firmware running on the wearable module represents an extension to the firmware design outlined in 5.2.2 that incorporates the continuous sensor recording of firmware in 0. The design of the firmware is summarised on Figure 53:

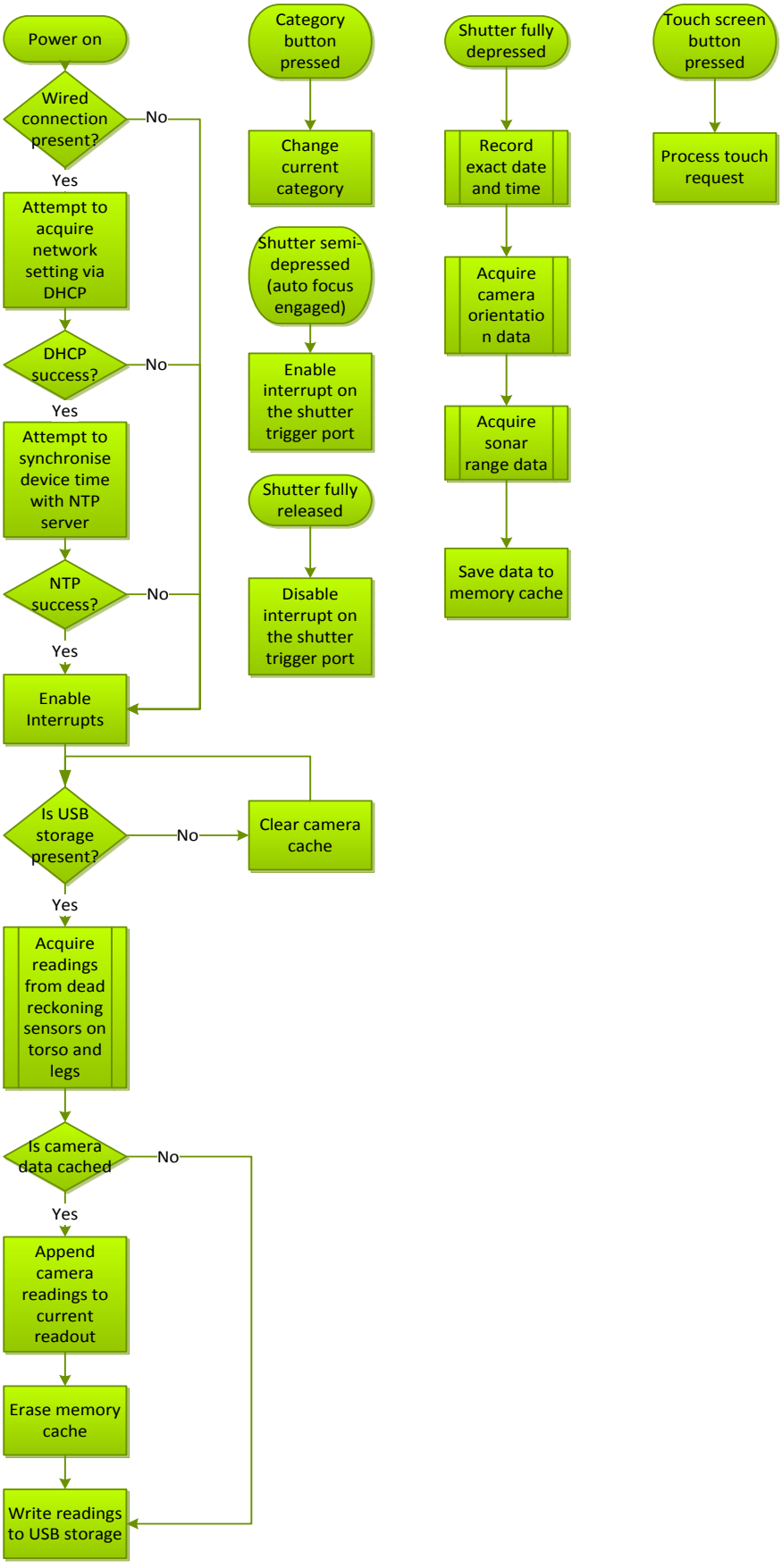


FIGURE 53. SPATIAL IMAGE MANAGEMENT SYSTEM FIRMWARE DESIGN

Upon start-up, the system attempts time synchronization with an NTP (Network Time Protocol) server in order to update its internal clock (which is being maintained in-between activations by a secondary power source), similar to 5.2.2.

Once the time check is completed (if available), the system initiates its hardware interrupts (used to detect camera shutter presses and category assignment button presses) and then begins to record the dead reckoning readings continuously, writing them to the USB storage if present. This recording happens at average rate of 30HZ, similar to 0.

Upon a camera snapshot, the interrupt subroutine records currently selected category identification, camera orientation and sonar readout and stores it in a shared memory location, available to all of the firmware threads (independently executing processing units) along with the exact timestamp of the snapshot.

The main sensor polling thread checks the shared memory location for camera readings on every polling cycle. If such data are available, then the sensor readouts are adjusted to have the timestamp of the photograph (since for dead reckoning, millisecond accuracy in timestamps is not required as opposed to the camera readings, where timestamps are used to match sensor readings to photographs as per 5.2.3), resulting in the camera sensor readouts being written to the USB storage.

Category identification works in the same way as it did with the compositing module described in 5.2.2. Upon a button press, category identification number is changed. This category is added to every snapshot (camera) data, facilitating managing and categorizing multiple photographs. The new addition to the category assignment process is the ability to alter the category by using touch screen on the FEZ Cobra module. Four buttons are available on touch screen, pressing either of which results in category assignment change.

The touch screen is also used for debugging purposes to provide additional information from the sensors to the user. Its design is described in 6.2.4.

6.2.3. *Data Recording Format*

Since continuous sensor recording suffers from string processing overheads (that are acceptable for button-triggered sensor recording of the sensor compositing prototype) outlined in 0, the firmware again opts to use binary format as the way of optimising recording speed.

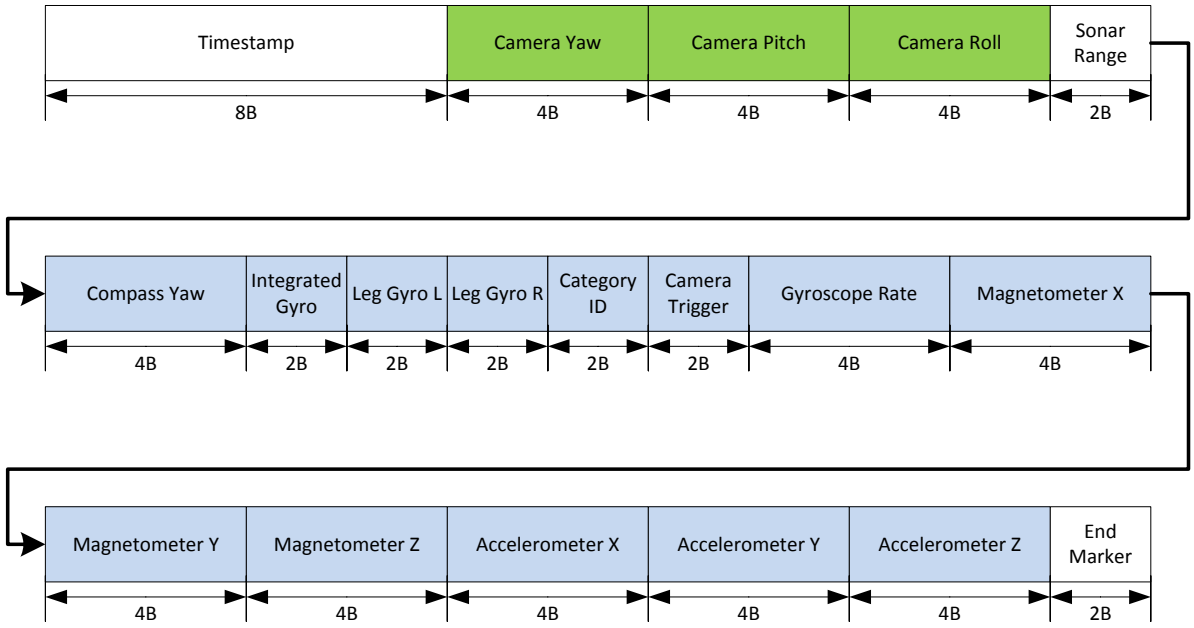


FIGURE 54. SENSOR DATA RECORDING BINARY FORMAT. DATA SIZE IN BYTES. GREEN BLOCKS REPRESENT CAMERA SENSORS. BLUE BLOCKS REPRESENT TORSO SENSORS.

Each sensor data block is 66 bytes wide. They are written sequentially and are separated by distinct end of record markers that facilitate dealing with partially recorded end blocks (when device is powered off). For convenience, when imported on PC, these are converted into standard comma separated value (CSV) files making it easier to review the data in a range of tools.

Timestamp is stored as 8 byte wide value representing count of microprocessor ticks (each one being 100 ns) elapsed since 12:00:00 midnight January 1 year 0001. This is standard convention in .NET framework, facilitating conversion of timing data into built in structures for processing. Despite excessive resolution, the overhead of using 8 bytes was deemed

acceptable because it made timing data require no pre-processing on the unit before recording (since this timing value is natively exposed by framework structures).

6.2.4. User Interface Design

The user interface design section described the touch screen interface available for the FEZ Cobra enclosure. Due to the bulky nature of the FEZ Cobra development module and its display, it cannot be easily used as the sole input method for the enhanced compositing system.

It can, however, be used to supplement button-based interaction and to provide access to the additional functionality, not commonly utilised in the normal course of operations. Such functionality includes toggling audio feedback on button presses (in the form of multi tonal beeps) and re-calibrating sensors.

The HM6343 compass module utilised for both camera and pedestrian orientation tracking is a device that includes built in hard iron calibration routines (as mentioned in 5.2.1). Thus, for example, when the compass is mounted on the camera, it has to be re-calibrated to compensate for the presence of the electro-magnetic interference from the camera (mostly from the lens motor) as described in 6.2.1.3. Similarly, compass on the torso requires re-calibration for better performance. These calibration routines only allow for hard iron adjustments but not for the magnetic distortion compensation, however.

The calibration mode is not required for every operation, only when the torso unit (compass and gyroscope) is in proximity of sources of interference that are spatially fixed relative to the compass (such as, for example, metallic object on the neck next to the torso unit). Therefore, the calibration mode represents a secondary functionality that can be invoked on a per-need basis.

Figure 55 shows the photograph of the user interface. Since FEZ Cobra unit restricts operating under emulator without physical hardware and the .NET Micro Framework does

not contain WYSIWYG (what you see is what you get)-type user interface designer, this was the only easy way of producing a UI screenshot.

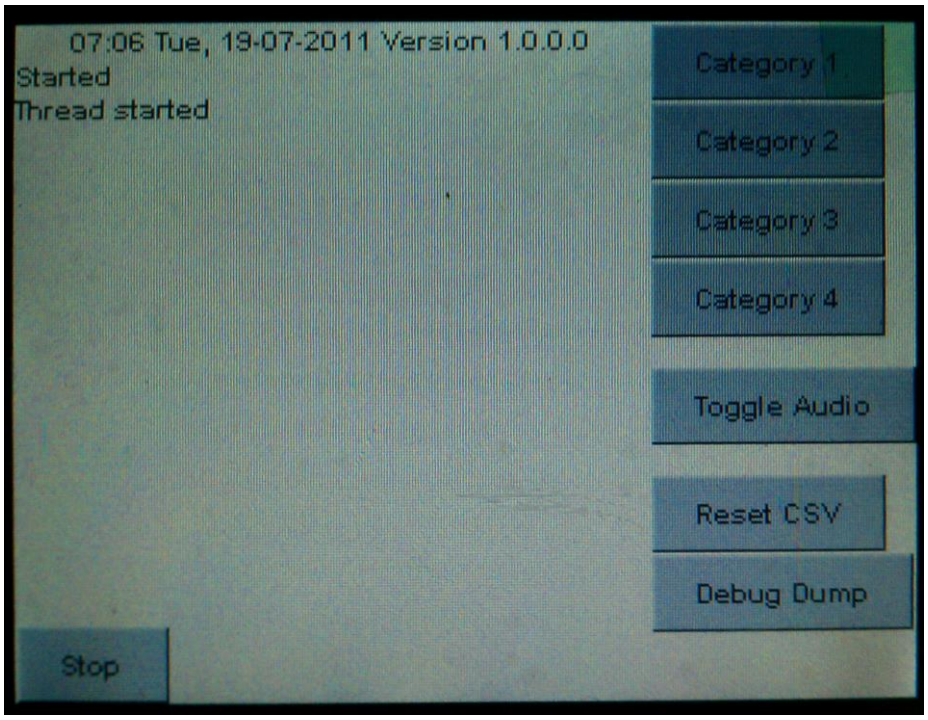


FIGURE 55. WEARABLE MODULE USER INTERFACE SCREENSHOT

Initially, the calibration subroutine access was made available from the UI directly, but it was removed in the later versions due to the adverse effects of accidentally entering calibration mode (resulting in sensor reset or incorrect calibration parameters) and now requires firmware alteration to invoke.

The start/stop button on the bottom left is used to begin or pause sensor recording (this can also be initiated by pressing number 1 on the numeric pad on the camera). Category assignment buttons on the top right are used as secondary means of category reassignment (duplicating the functionality of the keypad).

Toggle audio button toggles the audio feedback from button and shutter presses. Each category assignment key has a different tone that is played when it is pressed to confirm successful category selection. Disabling this merely removes the audio feedback without affecting the functionality.

Reset CSV button is used to erase currently recorded sensor readings to begin a new recording. Despite the use of binary formatting for the sensor recording, naming of this debugging button was left untouched to for legacy reasons.

And, finally, debug dump button outputs current sensor reading on screen in text form to facilitate sensor validation. It is used predominantly for development but the compass heading it presents is also useful for the operator of the module as a way of determining their current direction.

6.3. Sensor Fusion Filtering Design

Since there are two sensors now measuring the heading (yaw) of the user, it becomes possible to fuse their readings in order to produce a better estimate, as suggested by the research in 2.3. Gyroscope and magnetic compass suffer from different sources of error (outlined in 4.1.1) and thus should be able to complement each other's readings.

6.3.1.1. Complementary Filter

A complementary filter is typically used where the noise in the sensors can be filtered based on its frequency response (Higgins, 1975). Gyroscopes mainly suffer from low frequency noise (mostly bias drift) whereas magnetic compass errors are largely high frequency in nature. A complementary filter can therefore apply high-pass digital filter to the gyroscope signal and low-pass filter to the compass signal and then combine the results of these to produce an estimate of the true value of the heading.

Due to the fact that gyroscope measures angular rate of heading change, rather than the heading itself, its readings have to be integrated. The overall complementary filter then looks the following way:

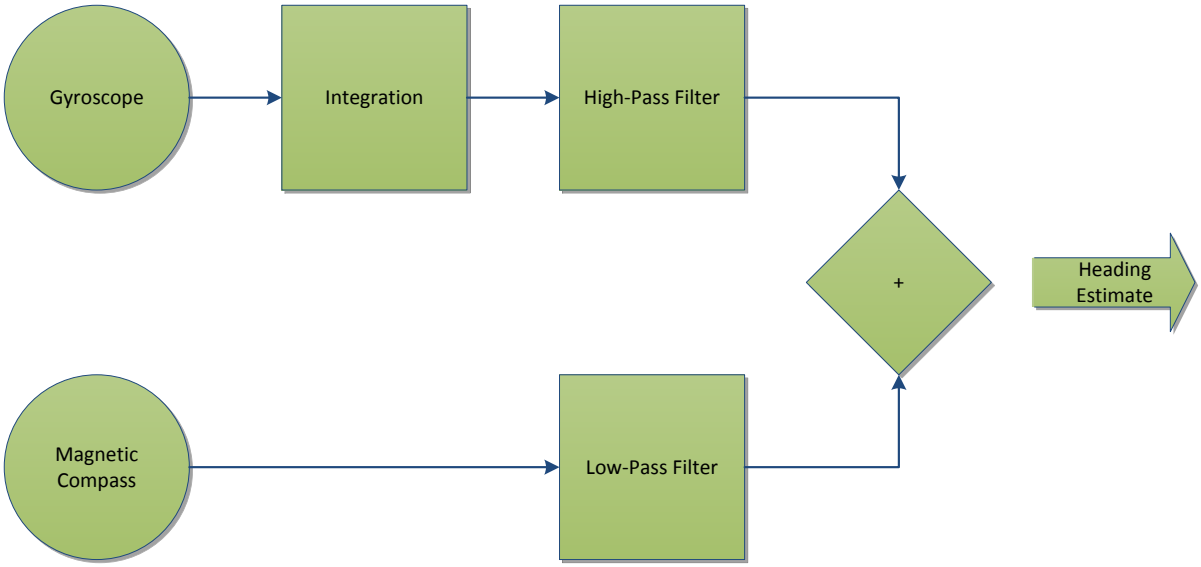


FIGURE 56. COMPLEMENTARY FILTER DESIGN

This filter can be written as:

$$\varphi = G(s) \times \alpha + (1 - G(s)) \times \beta \qquad \text{(Eq. 6.1)}$$ [1]

Where α and β are two estimates of the same value, in this case α is the integrated gyroscope heading estimate, β is the compass heading reading and *G(s)* is the high-pass digital filter transfer function, that can be written as:

$$G(s) = \frac{s\tau}{s\tau + 1} \qquad \text{(Eq. 6.2)}$$ [2]

Where τ is the filter time constant and *s* is the Laplace variable. The low pass filter transfer function will then be:

$$1 - G(s) = \frac{1}{s\tau + 1} \qquad \text{(Eq. 6.3)}$$ [3]

[1] (Higgins, 1975)

[2] (Higgins, 1975)

[3] (Higgins, 1975)

Since the signal processing (filtering) is happening in digital form, it can be implemented instead as a recursive algorithm. According to Smith (1999), such a recursive filter will be represented in the following general form:

$$y[n] = a_0x[n] + a_1x[n-1] + a_2x[n-2] + \cdots + b_1y[n-1] + b_2y[n-2] + \cdots \quad \text{(Eq. 6.4)}$$ [1]

Where *x* is the input signal and *y* is the output (filtered) signal. By varying coefficients *a* and *b* different filters can be implemented. In order to implement first order low pass filter, the following coefficients must be used:

$$a_0 = 1 - z$$
$$b_1 = z \quad \text{(Eq. 6.5)}$$ [2]

And the corresponding first order high pass filter coefficients are:

$$a_0 = \frac{1+z}{2}$$
$$a_1 = -\frac{1+z}{2} \quad \text{(Eq. 6.6)}$$ [3]
$$b_1 = z$$

Where *z* is the filter constant determined by the desired filter cut-off frequency f_c:

$$z = e^{-2\pi f_c} \quad \text{(Eq. 6.7)}$$ [4]

Using recursive first order algorithms, the complementary filter becomes addition of heading estimate generated using low-pass filtered compass signal and high-pass filtered integrated

[1] (Smith, 1999)

[2] (Smith, 1999)

[3] (Smith, 1999)

[4] (Smith, 1999)

gyroscope signal (since gyroscope doesn't measure the heading directly, rather, its rate of change).

Gyroscope integration process can be represented using the following formula (using trapezoid method for integration):

$$\alpha_k = \alpha_{k-1} + \Delta t \times \frac{\omega_k + \omega_{k-1}}{2} \quad \text{(Eq. 6.8)}$$

Where ω is the angular rate reading from the gyroscope and Δt is the time between adjacent sample readings (sample rate). The high pass filtering of the gyroscope heading α (based on the integrated angular rate ω) can then be done according to the following formula:

$$\alpha_n = \frac{1+z}{2}(\alpha_{n-1} + \omega_n \times \Delta t) - \frac{1+z}{2}(\alpha_{n-2} + \omega_{n-1} \times \Delta t) + z \times \alpha_{n-1} \quad \text{(Eq. 6.9)}$$

And the low pass filtering of the compass heading β based on the compass readings ϑ can be accomplished according to the following equation:

$$\beta_n = (1-z)\vartheta_n + z \times \beta_{n-1} \quad \text{(Eq. 6.10)}$$

Addition of these two filtered signals will produce an overall heading estimate. The processing requirements are minimal and the filter can be easily ported to an embedded platform, if necessary, satisfying the constraints of this research.

Because gyroscope produces integrated reading that has a bias error which increases the longer it is integrated and compass produces absolute heading, integrating gyroscope independently from compass will result in ever decreasing accuracy of the filtered signal as the integrated gyroscope reading deteriorates further and further. To prevent this, integration of angular rate data can happen only on the filter window signal and can use the previously filtered data as the integration basis. This will ensure greater stability of the fusion and

prevent the signals from the two sensors from deviating far from one another whilst still retaining the strongest features of each sensor.

If filtered heading (yaw) reading is γ_n at a time *n,* so that*:*

$$\gamma_n = \alpha_n + \beta_n \qquad \text{(Eq. 6.11)}$$

Then equation (Eq. 6.8) for integration of the sensor data can be rewritten as:

$$\alpha_k = \gamma_{k-1} + \omega_k \times \Delta t \qquad \text{(Eq. 6.12)}$$

And (Eq. 6.9) can be represented as:

$$\alpha_n = \frac{1+z}{2}(\gamma_{n-2} + (\omega_n + \omega_{n-1}) \times \Delta t) - \frac{1+z}{2}(\gamma_{n-2} + \omega_{n-1} \times \Delta t) + z \times \alpha_{n-1} \qquad \text{(Eq. 6.13)}$$

This ensure high pass filtering on the correct segment of the signal based on gyroscope readings only whilst also minimising the deviation between the two sensors due to accumulation of integration errors.

The value for the filter constant will be determined experimentally by tuning the filter and will depend upon the performance and noise characteristics of the gyroscope and compass sensors.

6.3.1.2. Kalman Filter Design

Review of different methods for navigational sensor fusion reveals the Kalman filter to be used extensively for this purpose, especially when combining angular rate sensors (gyroscopes) with the magnetic compass sensors.

Kalman filter design involves modelling the system (in this case the enhanced compositing module) as a combination of two main equations (Maybeck, 1979):

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad \text{(Eq. 6.14)} \quad [1]$$

$$z_k = Hx_k + v_k \quad \text{(Eq. 6.15)} \quad [2]$$

In these equations x_k is the state of the system at time instance *k*, *u* is the control signal (optional, depending on the design of the model and system), *w* is the process noise, *z* is the measurement (actual sensor readings) and *v* is the measurement noise (caused by the errors introduced by the sensors themselves). *A, B* and *H* are the constants that relate different aspects of the filter to each other.

For the purposes of this chapter and for convenience, system state can be represented as a vector:

$$x = \begin{bmatrix} Yaw \\ Angular\ Rate \end{bmatrix} \quad \text{(Eq. 6.16)}$$

Since the sensors performing the measurement are gyroscope and magnetic compass, the measurement can be represented as:

$$z = \begin{bmatrix} \alpha \\ \omega \end{bmatrix} \quad \text{(Eq. 6.17)}$$

Where α is the compass measurement angle and ω is the rate of turn as measured by the gyroscope. This simplifies constant *H* used to relate state to the observed measurement into an identity matrix:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{(Eq. 6.18)}$$

Measurement noise can then be represented as a normal distribution *N(0,R)* where:

$$R = \begin{bmatrix} Magnetic\ compass\ error & 0 \\ 0 & Gyroscope\ error \end{bmatrix} \quad \text{(Eq. 6.19)}$$

[1] (Welch & Bishop, 1995)

[2] (Welch & Bishop, 1995)

The values of the measurement errors for the sensors are available in the relevant sensor datasheet and typically do not affect each other. This error representation uses a typical Kalman filter assumption of the error being Gaussian in nature (Welch & Bishop, 1995).

Assuming that in the absence of external factors, rate of turn changes between adjacent samples are minimal (a reasonable assumption for the state transition in the absence of forces affecting the system) the transition matrix *A* becomes:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad \text{(Eq. 6.20)}$$

Where Δt is the polling interval. This matrix then assumes that the rate of turning remains static between state updates and the yaw is obtained by integrating its previous value with the angular rate reading. Such an assumption is not strictly true, since the angular rate changes between measurements, but such a change cannot be easily related to the previous state.

The process noise can then be modelled as a normal distribution *N(0,Q)* in a manner that will accommodate the fact that the angular rate is not truly static:

$$Q = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \quad \text{(Eq. 6.21)}$$

This simplified representation of the process noise assumes *x* noise for the estimation of the heading and introduces a variable *y* to represent the variation in the angular rate readings between polling intervals. For more precise results and robustness, the process noise should be estimated using filter tuning without setting the covariance matrix *Q* elements to zero, but for the purposes of this research, the simplified form above should be sufficient. It might also be necessary to inject process uncertainty into the filter by setting covariance matrix elements to non-zero elements. The need for these will be determined experimentally. This matrix assumes that noise in gyroscope does not affect the compass and vice versa.

Kalman filter is a recursive filter that operates by forming predictions and refining them using the actual measurements (Maybeck, 1979). It operates in two stages:

1. Prediction stage, when the next state of the system is estimated (also known as time update stage)
2. Correction stage when the prediction is verified using noisy measurements from the sensors (also known as measurement update stage)

During the time update (prediction) stage, the *a priori* system state $\hat{x}_k^-$ and the *a priori* estimate for the error covariance P_k^- are estimated according to the following equations:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \qquad \text{(Eq. 6.22)} \quad [1]$$

$$P_k^- = AP_{k-1}A^T + Q \qquad \text{(Eq. 6.23)} \quad [2]$$

Since the control signal is not present in the model, substituting the values for the constants, the equations (Eq. 6.22) and (Eq. 6.23) can be re-written as:

$$\hat{x}_k^- = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \hat{x}_{k-1} \qquad \text{(Eq. 6.24)}$$

$$P_k^- = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} P_{k-1} \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & y \end{bmatrix} \qquad \text{(Eq. 6.25)}$$

The measurement update stage involves computing the Kalman gain variable K^k and calculating the *a posteriori* state estimate $\hat{x}_k$ and the error covariance P_k using the measurements and the Kalman gain according to the following equations:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \qquad \text{(Eq. 6.26)} \quad [3]$$

[1] (Welch & Bishop, 1995)

[2] (Welch & Bishop, 1995)

[3] (Welch & Bishop, 1995)

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad \text{(Eq. 6.27)} \quad [1]$$

$$P_k = (I - K_k H)\, P_k^- \quad \text{(Eq. 6.28)} \quad [2]$$

Substituting into these equations the identity matrix for the value of *H*, as per (Eq. 6.18), these become:

$$K_k = P_k^-(P_k^- + R)^{-1} \quad \text{(Eq. 6.29)}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \quad \text{(Eq. 6.30)}$$

$$P_k = (I - K_k)\, P_k^- \quad \text{(Eq. 6.31)}$$

Due to the use of matrix calculations, especially calculations of matrix inverse, the computational requirements of the Kalman filter are higher than those of the complementary one. Nonetheless, matrix mathematics libraries are often available on the embedded devices. This is typically explained by the need for the matrix transforms in the drawing libraries used to render the graphics elements on the screen (where the graphical output is present on the device).

6.3.1.3. Custom Toggle Filter

Gyroscope performs better than compass during straight line segments but is less efficient at tracking turns, according to the evaluation in 4.8. This knowledge can be used to construct a custom filter that will switch between the sensors depending on the situation.

By examining the difference between two adjacent values for direction from the integrated gyroscope, if the value falls below a given threshold, gyroscope integration is used. Otherwise tilt compensated compass:

[1] (Welch & Bishop, 1995)

[2] (Welch & Bishop, 1995)

$$\gamma_k = \begin{cases} \beta_k, & \alpha_k > t \\ \gamma_{k-1} + \omega_k \times \Delta t, & \alpha_k \leq t \end{cases} \quad \text{(Eq. 6.32)}$$

Where β_k is the compass reading at time *k*, γ_k is the filtered reading, ω_k is the gyroscope angular rate reading, *t* is the calibrated threshold and α_k is the direction change according to gyroscope integration, that is found according to trapezoid rule as:

$$\alpha_k = \Delta t \times \frac{\omega_k + \omega_{k-1}}{2} \quad \text{(Eq. 6.33)}$$

This filter again requires tuning to find the optimal threshold value for selecting between the compass and gyroscope readings. It is very computationally efficient, requiring minimal processing power and directly uses results of the sensor comparison in chapter 4.

6.3.1.4. Custom Compensating Filter

Compensating filter is a custom filter that operates by compensating for errors in compass readouts by using gyroscope integration. Since a number of additional sensors are exposed in addition to the gyroscope and compass, it is possible to use these secondary sensors to determine the likelihood of existing sensors providing erroneous readings.

Gyroscope is used a secondary source of data, used only when the primary source of heading data – tilt compensated compass is deemed to be likely to produce erroneous readings.

There are two potential sources of interference that can affect the compass readings and are detectable by the additional sensors recorded by the wearable module – acceleration due to walking and magnetic interference. Acceleration due to walking can be detected when the magnitude of the acceleration affecting the compass (which can be calculated from (Eq. 4.1)) differs from the magnitude of gravity normal acceleration. This is known as the *Acceleration-Magnitude Detector* (Skog et al., 2010) for zero velocity updates.

Magnetic interference can be detected similarly to the acceleration due to walking.

$$|m| = \sqrt{m_x^2 + m_y^2 + m_z^2} \qquad \text{(Eq. 6.34)}$$

Where m_x, m_y, m_z are components of the magnetic field detected by magnetometers and $|m|$ is the absolute strength (magnitude) of the magnetic field.

Strength of the magnetic field at a given set of geographical coordinates is a known value and is equal to 48,882nT (0.489 Gauss) in Birmingham (Finlay et al., 2010; National Oceanic and Atmospheric Administration, accessed 2012). If the magnitude of the magnetic field differs from this reading, then the compass is likely to be affected by magnetic interference and its results are less reliable.

The heading estimate is then generated according to:

$$\gamma_k = (1 - w_k) \times (\gamma_{k-1} + \omega_k \times \Delta t) + w_k \times \beta_k \qquad \text{(Eq. 6.35)}$$

Where γ_k is the fused heading estimate at time k, ω_k is the angular rate detected by gyroscope, β_k is the tilt compensated compass heading, Δt is the time interval between samples and w is the confidence in the compass reading weight ranging between 0 (when compass readings are considered to be very unreliable) and 1 (when compass readings are likely to be accurate).

This weight is calculated as the minimum of the two values - estimate of compass confidence based on accelerometer readings and estimate of compass confidence based on magnetic field strength attenuated by a calibrated constant:

$$w'_k = c - \max\begin{cases} \big|1 - |a|\big| \\ \big|0.489 - |m|\big| \div 0.489 \end{cases} \qquad \text{(Eq. 6.36)}$$

Where $|a|$ is the magnitude of the acceleration affecting the compass, $|m|$ is the strength of the magnetic field in Gauss as detected by the magnetometers and c is a calibrated constant so that $c \in [0, 1]$ and can allow attenuating reliance on the compass in favour of gyroscope if needed.

The weight used by filter is then found according to:

$$w_k = \begin{cases} w'_k, & w'_k \geq 0 \\ 0, & w'_k < 0 \end{cases} \qquad \text{(Eq. 6.37)}$$

This enables limiting the reliance on tilt compensating compass and better tune filter performance.

6.4. Spatial Visualisation Software Design

In order to spatially represent the images on a common grid, the visualisation software needs to be able to do two things:

- create composites out of images taken from the same viewpoint
- arrange composites (or single photographs) based on the spatial coordinates of said viewpoint

6.4.1. Design of the Basic Functionality

Dead reckoning subsystem of the overall application is able to produce a path relative to the starting location. Path generation is based on sequentially rendering steps as interconnected unit lines with the direction specified by the current orientation (determined from the heading sensor fusion). Upon detection of the step, fused heading data (obtained via one of the algorithms described in 6.3) is used to determine direction of the step and a line segment is generated, representing this step.

The entire process uses given participant's average single step length as a base unit for relative positioning. This value remains fairly static for a given person and can be approximated as a constant for a given step frequency (Feliz et al., 2009). This value is also directly proportional to the magnitude of the inclination angle detected by the gyroscope (Tong & Granat, 1999) and therefore variance in it can be compensated for by using the gyroscope data (the greater the swing detected by gyroscope, the longer the stride):

$$stride\ length\ \approx segment\ inclination\ range \times length\ of\ the\ leg \quad \text{(Eq. 6.38)}\ [1]$$

If a more accurate representation is needed, or conversion to metric system, then knowing length of a given path segment and number of steps taken to traverse it allow approximating stride length. Alternatively, stride length can also be determined from the magnitude of vertical acceleration during step (Kim et al., 2004; Jimenez et al., 2009) or by analysing gyroscope data from the legs directly (Miyazaki, 1997; Tong & Granat, 1999). For the purposes of relative image arrangement, however, use of unit average step as range unit is deemed acceptable.

Since the sensor recording is being written sequentially, images forming a composite can be determined by grouping photographs together that do not have any step activities in between them. The viewpoint for these images will then be the last rendered path location. Knowing this it is possible to put a marker on the path for the location of every composite.

The result of these is a path containing markers for the locations of composites. This fulfils one of the software design tasks – mapping locations of composites.

Interacting with a composite marker (clicking mouse on it) renders the composite image for that location in a side panel. This fulfils the second task – creating composites themselves.

The composite rendering utilises the same algorithms as the ones described in 5.2.3. There has been only a single alteration to the way the compositing operates – instead of simply bringing all the magnified images forward in Z-Order (drawing them on top of the non-magnified ones), the effective distance of the image is reduced by the magnification factor (determined from the focal length of the camera recorded in EXIF (Japan Electronics and Information Technology Industries Association, 2010) image data) for the purposes of the

[1] (Tong & Granat, 1999)

image occlusion calculation. In other words, an image at 10m (according to the sonar readout on the camera) taken with 2x magnification will be treated as if it has been taken at 5m.

This was tested and determined to provide a more dynamic and robust image occlusion in an indoor environment, especially with composites containing a large number of component parts taken at a large range of magnification levels.

6.4.2. Design of the Enhanced Functionality

While the design so far fulfils the main requirements of spatially arranging and compositing images, this functionality can be further enhanced by adding additional features.

Both the initial compositing system and the new spatial image management version have the ability to categorise images. In the standard sensor-based compositing software this functionality was only used to render an outline around images belonging to certain categories if requested by user (effectively highlighting objects of the same category), as described in 5.2.3. The new system, however, take this a step further as described in this section.

6.4.2.1. Snapshot Marker Enhancement

Availability of spatial composite arrangement allows further enhancing the highlight by category functionality of the original compositing system. For instance, composite markers are colour coded to display what categories do their component images belong to. This also allows filtering visible composite markers by category. Composite marker is rendered as a filled circle. Depending on the amount of categories represented by its component images, the marker is split into equally sized sectors of different configurable colours (colour configuration interface is shown on Figure 59) to represent categories. For example, if a composite only contains images belonging to one category, it will be filled with a solid colour corresponding to that category. If, however, an image of another category is present among

its component parts, then the marker is split into two semi-circles with colours corresponding to these categories and so on.

FIGURE 57. ENHANCED COMPOSITE MARKERS EXAMPLE. FROM LEFT TO RIGHT, SHOWS MARKERS WITH 3, 1 AND 2 CATEGORIES

6.4.2.2. Composite Rendering Enhancement

The colour coding of the categories is further propagated into the rendering of the composites themselves. If enabled, every composite is surrounded by a border with the preconfigured colour matching that of the category assigned to it. For example, if the category 1 is assigned colour red, then all the images belonging to this category in the composite will have a red border around them.

The border is being rendered at the bottom of Z-Order (in other words, all the images are rendered on top of the border), this way images overwrite sections of the highlight and it does not interfere with the composite itself, as shown on Figure 58.

FIGURE 58. ENHANCED COMPOSITE RENDERING EXAMPLE.

The software can be altered with relative ease to force the highlights to be rendered on top of images or together with their associated component part, but this was deemed to be too intrusive and not in line with the main purpose of the software – spatial image arrangement.

6.4.2.3. Composite Filtering Enhancement

Since each composite marker has information about the categories of its component parts associated with it, the ability to filter visible composites by category is implemented. Each category has a checkbox associated with it in the rendering software as shown on Figure 59.

FIGURE 59. CATEGORY FILTERING AND COLOUR RECONFIGURATION INTERFACE

Clicking on any of the colour buttons next to a category allows selection of a colour to assign to that category. Ticking and clearing of a checkbox to the left of the category name enables category-based composite filtering. The application is designed to only display composite markers that have at least one of their categories enabled in the filtering. So, for example, if a composite marker has categories 2 and 3, it will only be visible if either or both of these categories are enabled. Otherwise it will not be shown on the route display. Figure 60 and Figure 61 show the same route with no filtering applied and filtered to only show category 3 (blue marker colour) respectively.

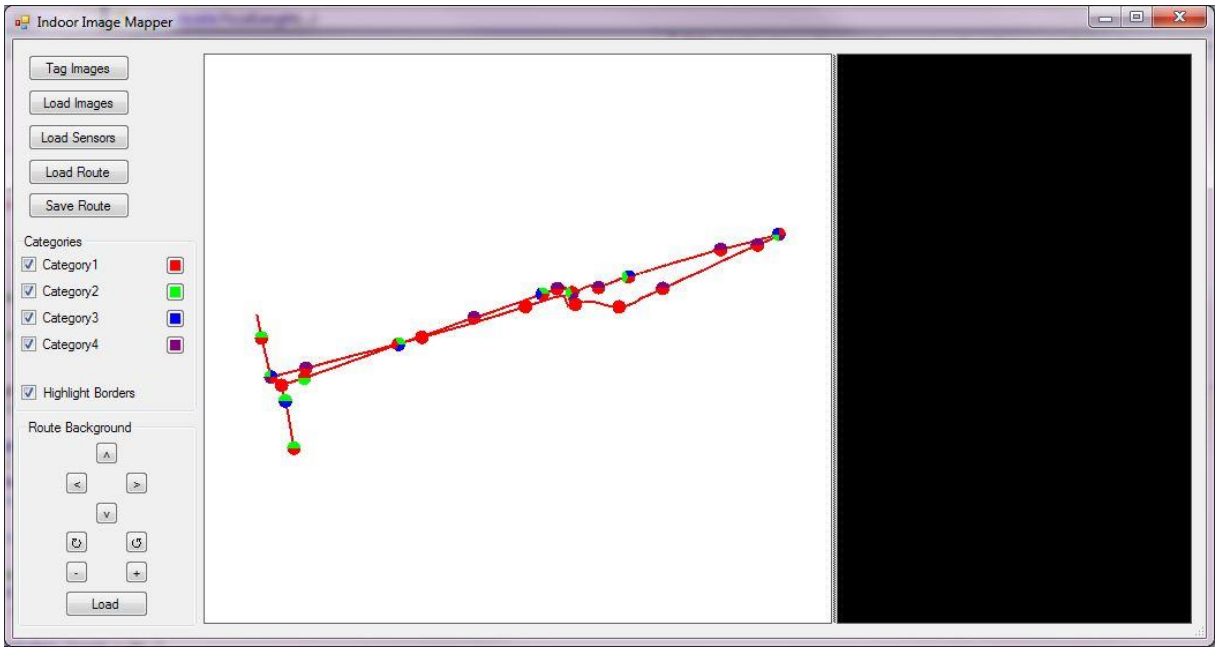


FIGURE 60. EXAMPLE OF ROUTE WITH ALL CATEGORIES ENABLED

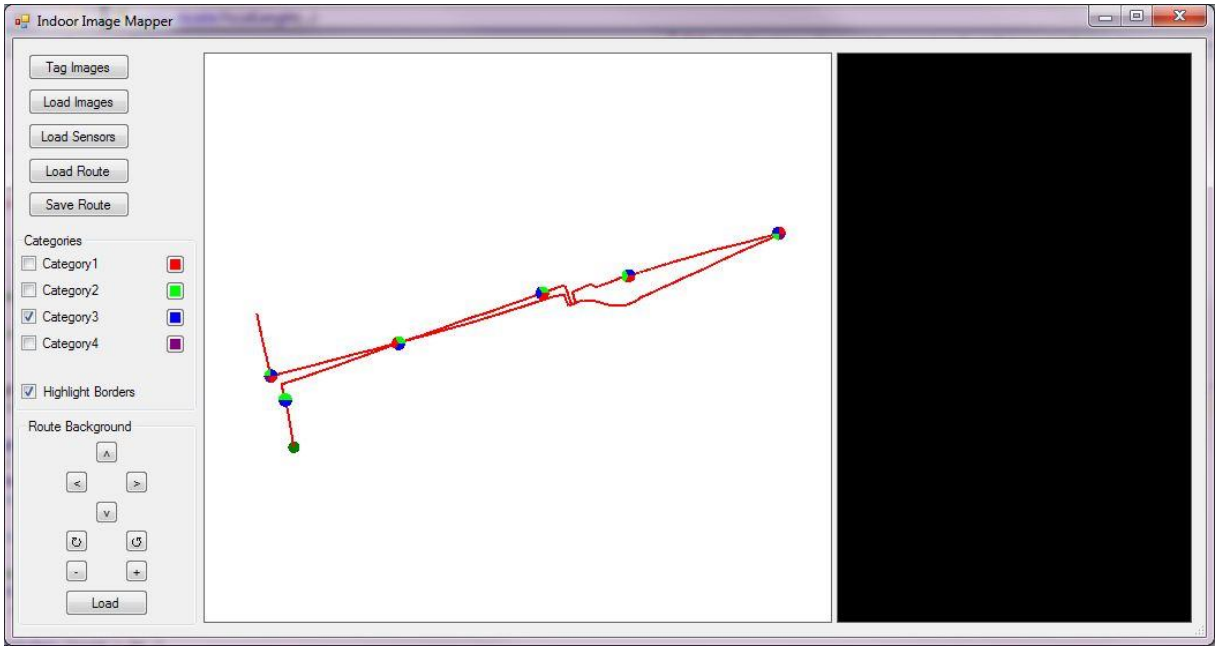


FIGURE 61. ROUTE WITH FILTERING ONLY SHOWING MARKERS CONTAINING CATEGORY 3

Such filtering allows fine control over the display of composites and facilitates the at-a-glance localising of photographs of certain category.

6.4.2.4. Composite Direction Display

When a composite marker is clicked to render the composite, an arrow is drawn originating from the marker and showing the direction of the composite horizontal middle (i.e. if the composite covers horizontal area of -10 to +20 degrees North, then the arrow will be pointed to +5 degrees North). When the mouse is moved over to the composite display and moved over it, the arrow changes its direction to indicate in which direction the part of the composite under cursor is from the capture location (represented by the marker itself).

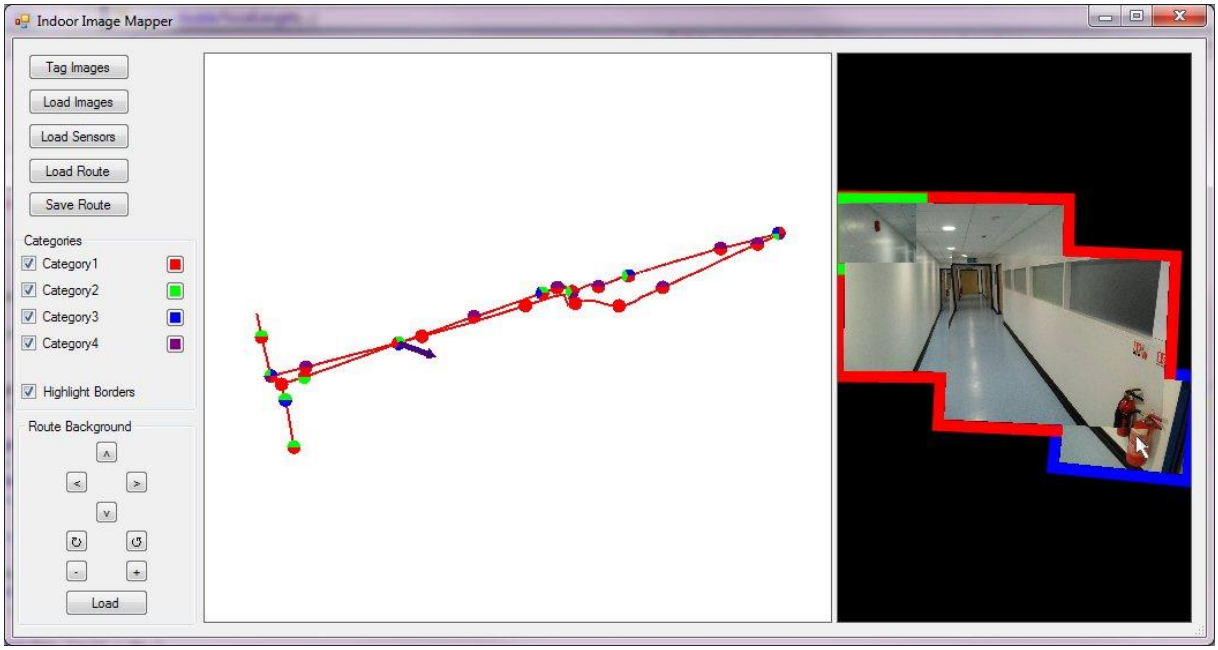


FIGURE 62. EXAMPLE OF THE COMPOSITE DIRECTION DISPLAY. ARROW ON THE ROUTE DISPLAY INDICATES THE DIRECTION OF IMAGE UNDER THE CURSOR (FIRE EXTINGUISHER)

This enhancement makes it easier to identify the direction of objects relative to the photographer's viewpoint during image generation. During the early testing of the software it was discovered that it is sometimes difficult to relate items on the photograph to their location on the map. For instance, with a photograph like the one on Figure 62 it is difficult to tell which side of the corridor is being photographed without good prior knowledge of the location – western or eastern. Placing a directionality marker resolves this issue helping relate the images to the map.

6.5. CONCLUSIONS

The spatial image management system is a combination of pedestrian dead reckoning and sensor-based image compositing that can enhance process of photography of indoor locations with additional metadata information, such as location and categorisation. It is employing a combination of sensors to spatially arrange images based on the photographer location and camera direction and is able to filter images based on their metadata.

The system is able to employ a number of algorithms to fuse sensors but their relative effectiveness remains to be determined. Comparison of system performance with different filters as well as calibration of individual filter parameters is described in the following chapter (Chapter 7) which evaluates the system performance, in particular dead reckoning accuracy, since compositing performance has already been evaluated in chapter 5.

7. SPATIAL IMAGE MANAGEMENT SYSTEM EVALUATION STUDY

The system designed and implemented in chapter 6 uses the same compositing mechanism as the system independently evaluated in chapter 5. This compositing approach underwent no changes from its original implementation, since the evaluation showed its results to be acceptable to users and comparable to state of the art image based compositing approaches when it comes to robustness and versatility, if losing when it comes to perceptual quality of the resultant composites.

The dead reckoning subsystem, however, underwent no tests beyond comparing the effectiveness of different sensors and sensor processing approaches in chapter 4. Moreover, the system can employ several different types of filtering to optimise the performance of the heading tracking, each of which requires calibration. The best sensor fusion filter has to be determined and overall pedestrian tracking performance quantified.

This raises need for another study – evaluation of the final system.

7.1. STUDY DESIGN

The study needs to evaluate the accuracy of the heading tracking mechanism employed in a quantifiable way. The most common way of doing this, as per research in chapter 2, is by using a closed loop path and determining the loop closure accuracy. If a given path starts and ends at the same coordinates, then the distance between starting and ending point will represent the magnitude of the error in the accuracy of the path tracking.

The study in chapter 4 used a closed loop to facilitate navigation. The final system evaluation study is an extension of the sensor analysis one and uses the same closed circuit path,

adding the photography task to generate images for compositing and create identifiable coordinates to analyse.

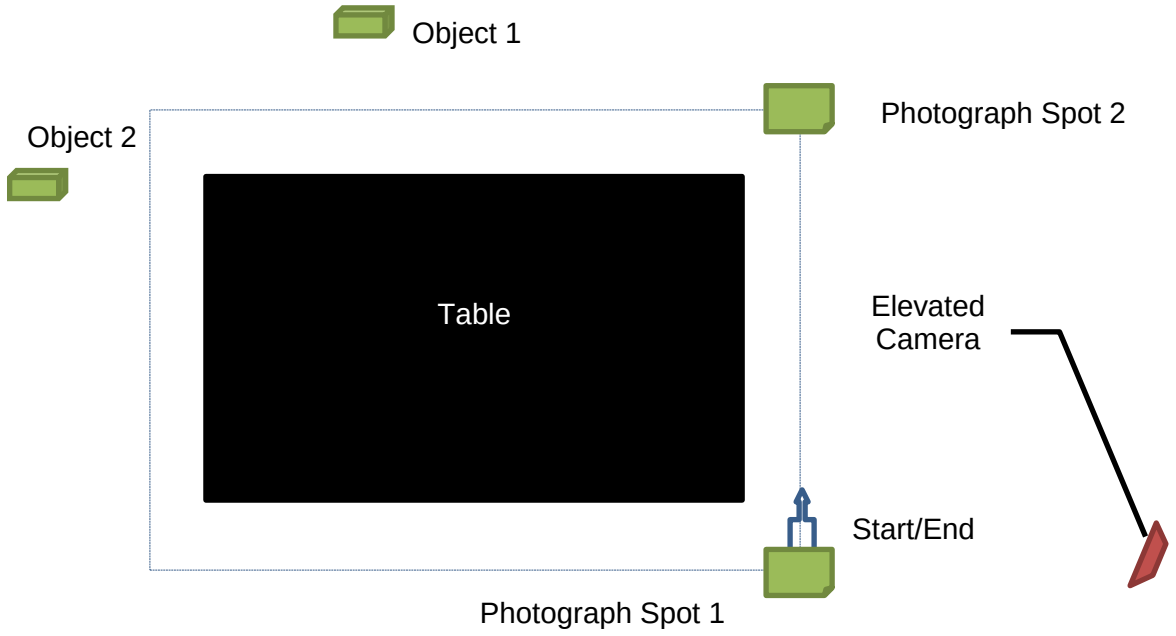


FIGURE 63. ENVIRONMENT FOR THE IMAGE LOCALISATION STUDY. TWO SEPARATE PHOTOGRAPH SPOTS ARE USED TO CREATE PHOTOGRAPHS OF THE OBJECTS ARRANGED IN A CLUSTER ON A BOARD AT APPROXIMATELY SHOULDER LEVEL. BOTH SPOTS ARE IN DIRECT VIEW OF VIDEO CAMERA.

This study uses the same set of 8 participants as study in chapter 4. Each participant walks 10 complete circles taking photographs of the objects every time they are next to the photograph spot. When they reach photograph spot 1, they face object 1 and take a photograph of it. When the participants reach photograph spot 2, they face object 2 and take two photographs of it (for compositing purposes). Participants can photograph objects utilising zoom functionality to provide a close up of each individual object.

Each participant wears the hardware module described in 6.2 and produces 10 sets of data. Data is post-processed on the PC using the software algorithms described in 6.3 and 6.4. Absolute reference data is provided via annotated video through ELAN software (Brugman et al., 2004) as described in 4.7.

7.2. STUDY PROCEDURE

This study was held at the University of Birmingham and involved 8 participants producing 10 sets of data each. Each participant undertook the study separately without seeing or knowing the specifics of the study procedure beforehand.

Before the study began, a photograph was taken with the camera module of a sheet of paper containing participant identification (number and name), to facilitate determining which photographs belong to which participant.

Each participant was led into the room where the study took place one at a time. They were then told that they needed to walk around the table at their normal pace for 10 cycles counter-clockwise starting from the position indicated and stopping at the two photography spots. On one spot they needed to face the bookcase and take two photographs of the section containing theses. On the second spot, corresponding with their starting spot, they needed to face the whiteboard and take a photograph of it. The participant was then shown a typical walk performed by the study conductor who made one circle around the table starting and ending at the position closest to the camcorder, explaining his actions as he performed them. The conductor stopped at the two photography spots, pointed out and described the objects that had to be captured. If required by the participant, the explanation and/or imitation walk were repeated.

Once the participant had indicated their understanding of the study procedure, they then had the recording module strapped to them, making sure the straps were tightened comfortably and did not restrict their movement. Participant was given camera module with the category selection button pad, led to the starting position and oriented in the direction of travel.

The camcorder was then turned on and finely adjusted to make sure the entire table was visible, especially the area where participant's legs would be on the far side of the table from

the camcorder as well as the photography spots in a way that would include the handheld camera module in the video recording frame.

Participant was told to press the top button on the category selection pad when they were ready and to come to a complete stop once they've completed the last circle (which would be counted by the study conductor and verbally propagated to the participant).

At this stage, participant pressed the start button, initiating the recording and producing an audible tone recorded by camera for synchronisation purposes.

Participant began walking around the table while the conductor remained near the camera to adjust it if required. When participants started their last cycle, they were reminded to stop and hold once they were back in starting position and had taken the final photograph.

Once a participant completed the last circle and took a photograph of the whiteboard, camcorder was turned off and then the recording module was powered off and unstrapped from the participant.

The sensor recording log and the photographs taken by the participant were then transferred to the PC storage and placed in a directory marked with the participant's identification.

7.3. Camera Orientation Analysis

Before focusing on the pedestrian dead reckoning results of the study, another aspect of the sensor data can be examined – camera orientation. The repositioned sensor could have adversely affected the performance of the compass module, and therefore it is prudent to analyse the heading data recorded by the camera during snapshots taken by the participants.

Each participant takes 3 photographs in each data set – two at photography spot 2 and one at photography spot 1 (as per Figure 63). This produces 30 photographs per participant and 240 data samples in total.

Due to the orientation of the compass on the camera, photographs taken at spot 2 are directed towards the bookcase, located at 355 degrees north. The photograph taken at photograph spot 1 is taken at 60 degrees north. Examining the raw camera heading data shows that the orientation is indeed spread out among these two angles.

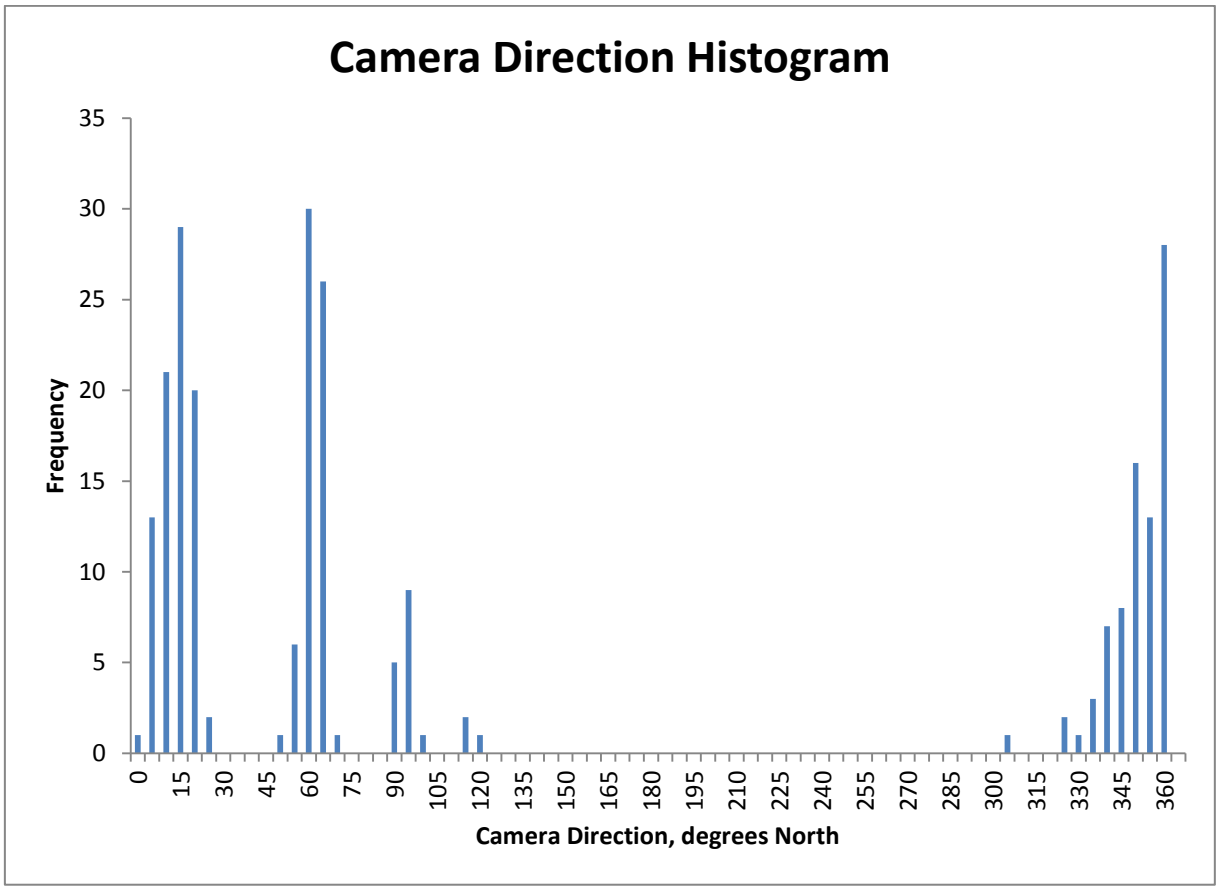


FIGURE 64. RAW CAMERA DIRECTION DATA. PHOTOGRAPHS WERE TAKEN AT 355 (-5) AND 60 DEGREES NORTH

Because the compass measures the direction in the range of 0...360 degrees, samples taken near zero degree mark are wrapped around to 360 degrees, making interpreting such a graph less convenient.

Therefore, instead of looking at the raw camera heading data, a different measure can be considered – absolute error in degrees from the expected direction:

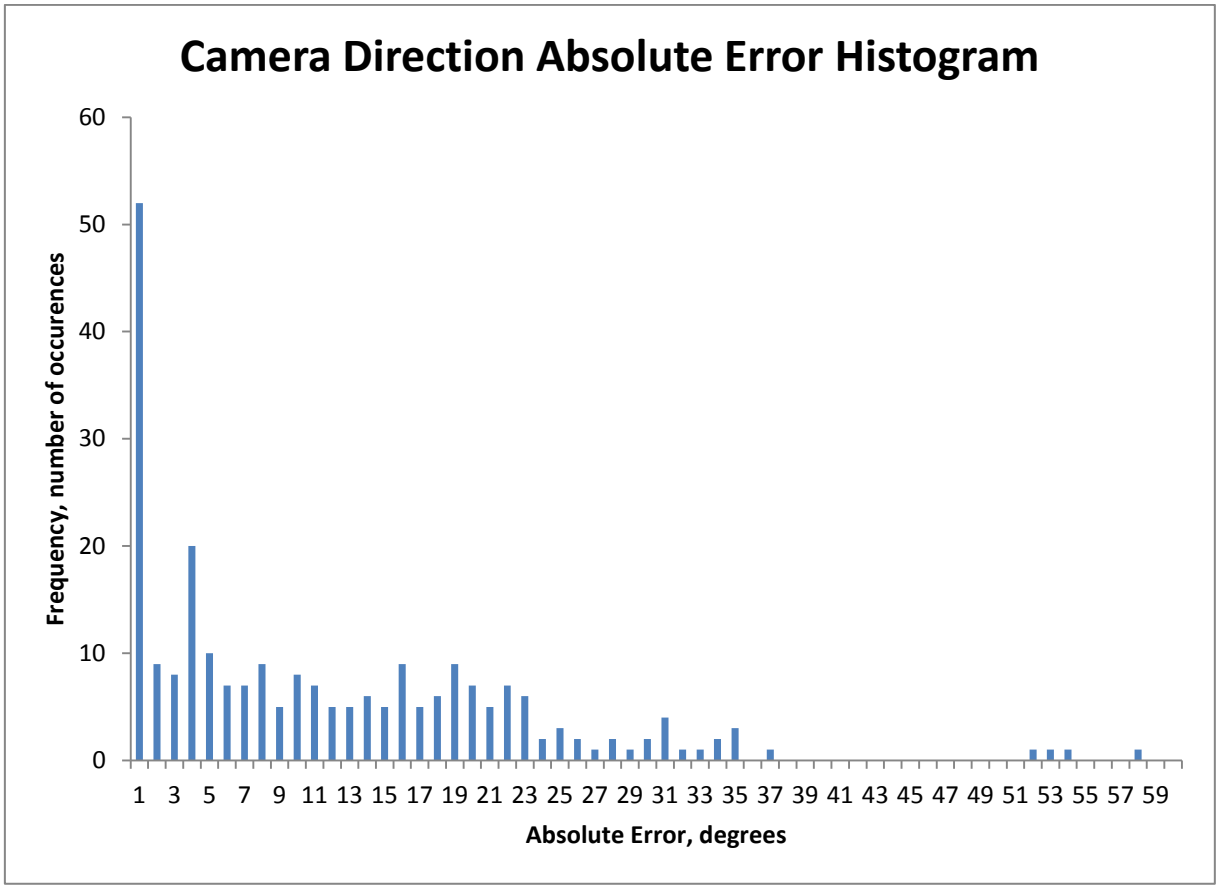


FIGURE 65. HSTOGRAM OF ABSOLUTE ERROR IN CAMERA DIRECTION

There are several outliers reaching up to 60 degrees. These have been found to be caused by a human error – one of the participants included arbitrary photographs on several of their rounds, confirmed by video recording.

The rest of the photographs remain largely centred on the directions expected, with 54% of the readings having less than 10 degree error. 95% of the readings have less than 31 degree error.

These data represent a reasonable camera orientation tracking, especially considering the fact that every participant will take a slightly different photograph of the required objects, especially closer to their final circles, explaining variance in the results.

The next stage in the analysis is validating dead reckoning performance, which requires calibrating sensors and tuning filters first.

7.4. GYROSCOPE BIAS TUNING

Before gyroscope data can be used in filters, it needs to be adjusted for bias error. Despite comprehensive filtering mechanisms built into the ADIS16255 digital gyroscope, the sensor readings still suffer from bias error that becomes significant when sensor readings are integrated (a requirement to bring gyroscope rate readout into the same domain as compass readings during filtering).

To determine the optimal gyroscope bias, each data set was analysed to determine total displacement using just the gyroscope data for different offset values. Optimal offset would correspond to the value with the smallest offset. The results of the bias calibration are shown on Figure 66:

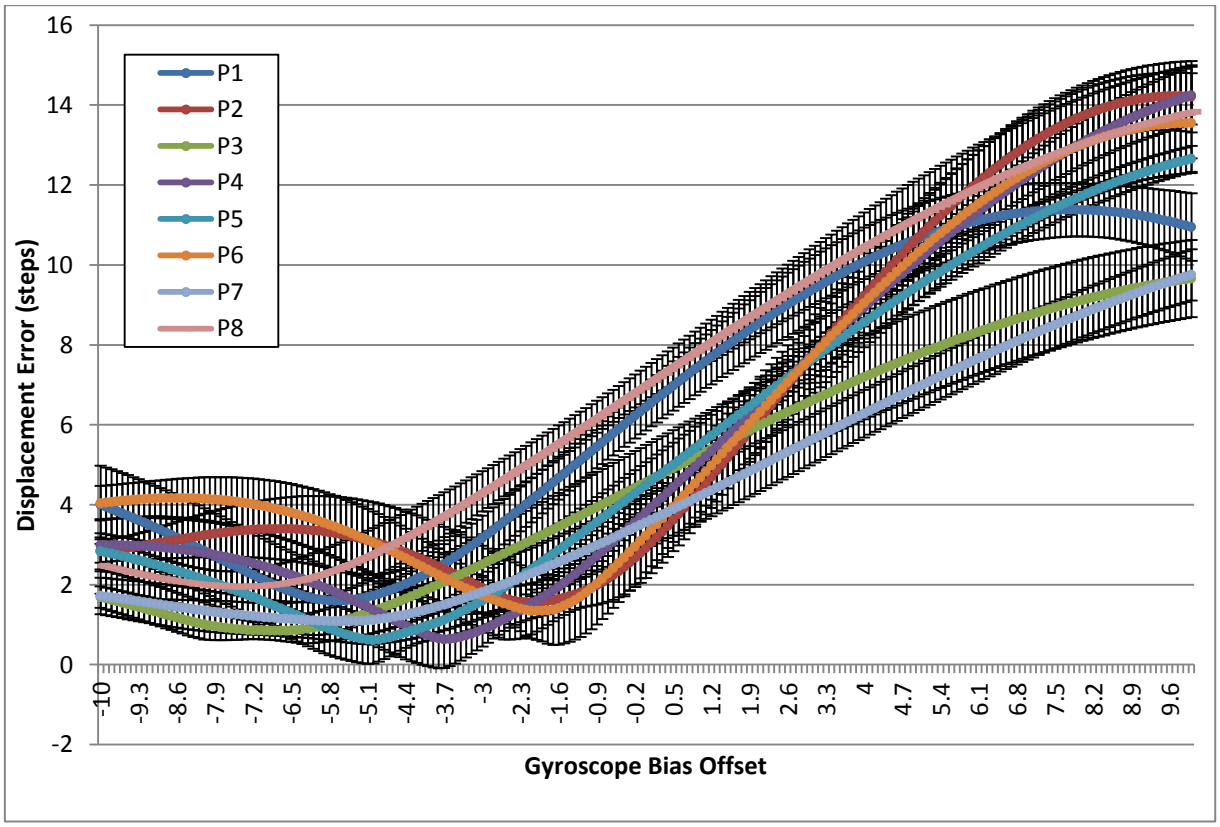


FIGURE 66. GYROSCOPE BIAS ERROR COMPENSATION CALIBRATION. ERROR BARS ARE ONE STANDARD DEVIATION

The bias offset between different participants is fairly consistent. Optimal bias offset is then determined from combined data from all the participants:

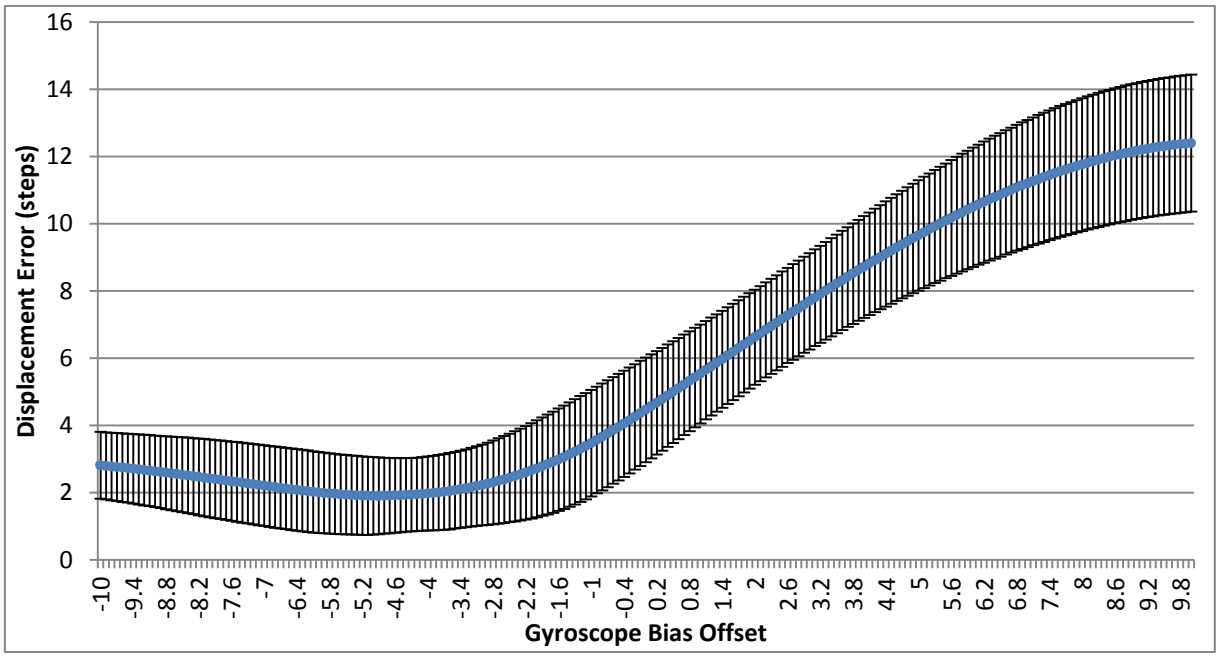


FIGURE 67. GYROSCOPE BIAS ERROR COMPENSATION CALIBRATION FOR ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION

Closer examination of the gyroscope data suggests optimal bias offset is -5.2 degrees per second. This value is therefore pre-applied to all gyroscope rate readouts before using them in filtering algorithms.

7.5. FILTER TUNING

The first step of filter processing is calibrating the parameters of the sensor fusion algorithms. Each algorithm contains one or more variables that need to be tuned for the specific filters and activity of this study. These parameters need to be tuned in a participant-independent way, obtaining a calibration set that performs the best for all the participants, if possible. This allows minimising future re-tuning, since tuning filter for individual participant indicates need of individual re-calibration and thus complicates system deployment and use by new users.

7.5.1. COMPLEMENTARY FILTER TUNING

Complementary filter has only a single parameter that requires tuning – filter constant *z* determined from the desired cut-off frequency according to (Eq. 6.7). This constant determines the frequency at which filter will start accepting predominantly gyroscope readings rather than tilt compensated compass. It can range between 0 and 1 (Smith, 1999) providing a range for calibration.

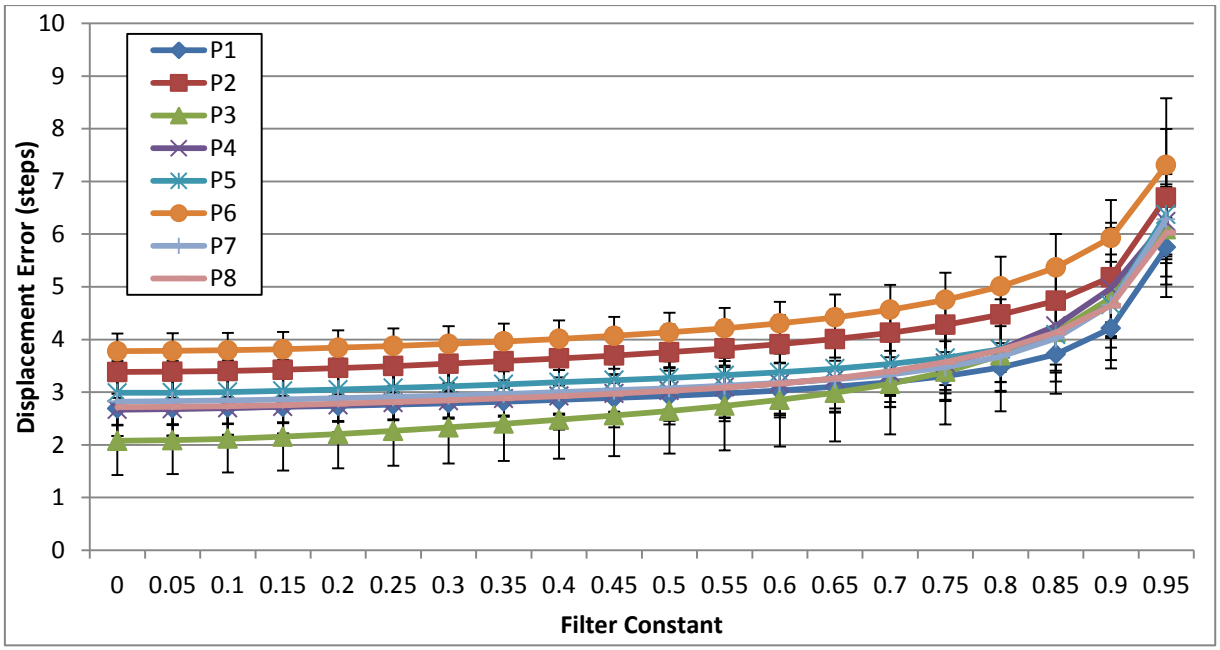


FIGURE 68. COMPLEMENTARY FILTER CALIBRATION. ERROR BARS ARE ONE STANDARD DEVIATION

The calibration data is consistent between participants as can be seen from the combined graph:

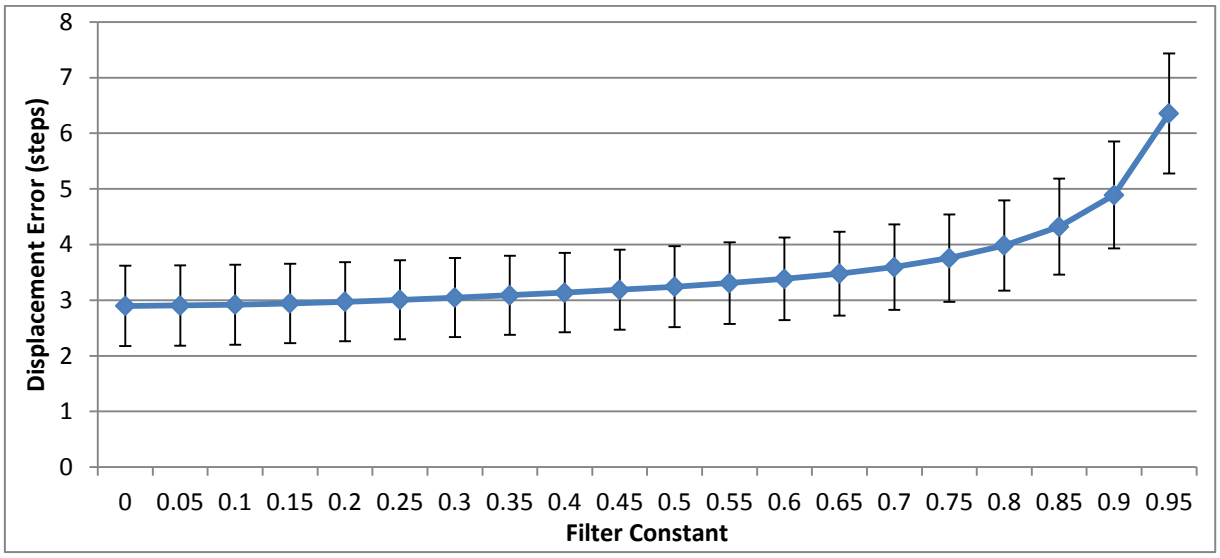


FIGURE 69. COMPLEMENTARY FILTER CALIBRATION FOR ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION

Examining the calibration data suggests that the filter performs best with low filter constant, numerically equal to 0.1 (when reviewing the underlying data).

7.5.2. KALMAN FILTER TUNING

Kalman filter has one parameter to tune – the process noise matrix described by (Eq. 6.21). This is a 2x2 matrix, but for sensor fusion of the gyroscope and compass, it can be assumed that the process noise from two sensors is not correlated and thus set elements (1, 2) and (2, 1) to zero, leaving two parameters to estimate, represented by x and y in (Eq. 6.21).

According to Ji et al. (2006), measurement noise is one tenth of the process noise for Kalman filter. Therefore, the calibration values for the filter centre around value of $\begin{bmatrix} 20 & 0 \\ 0 & 4.8 \end{bmatrix}$, since the datasheets for the sensors specify maximum noise of 2 degrees for compass and 0.48 degrees per second for the gyroscope. The range of the values for tuning was then set to start at zero (for no error) and reaching 8 times the theoretical optimal value (value selected to evaluate wider range of values).

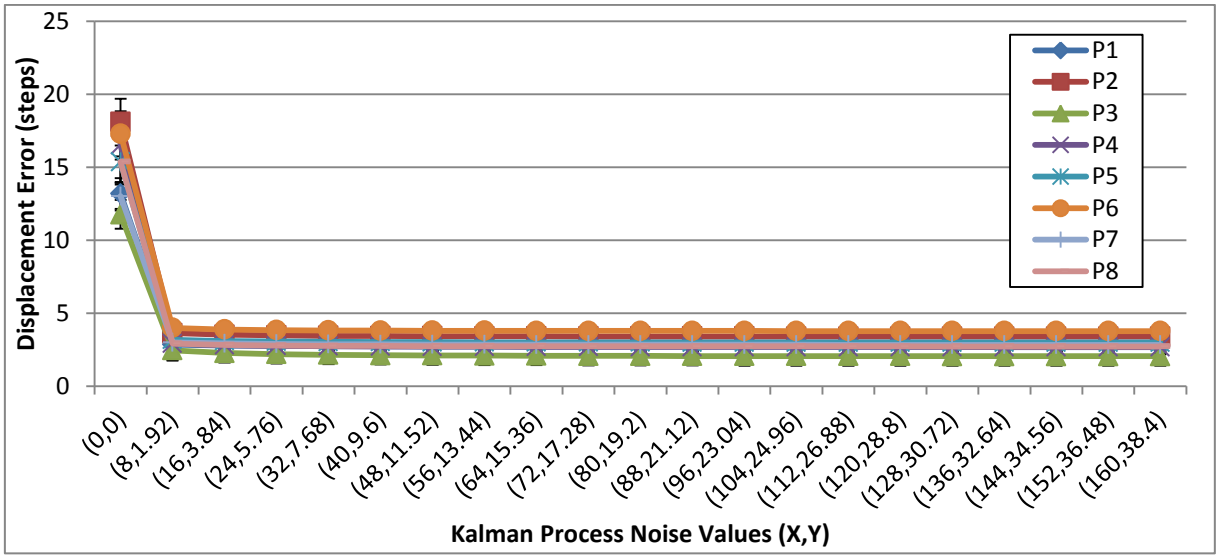


FIGURE 70. KALMAN FILTER PROCESS NOISE CALIBRATION. ERROR BARS ARE ONE STANDARD DEVIATION

The error is very large at process noise of zero. Adjusting the scale and examining the combined data for all participants shows reasonable consistency in calibration values:

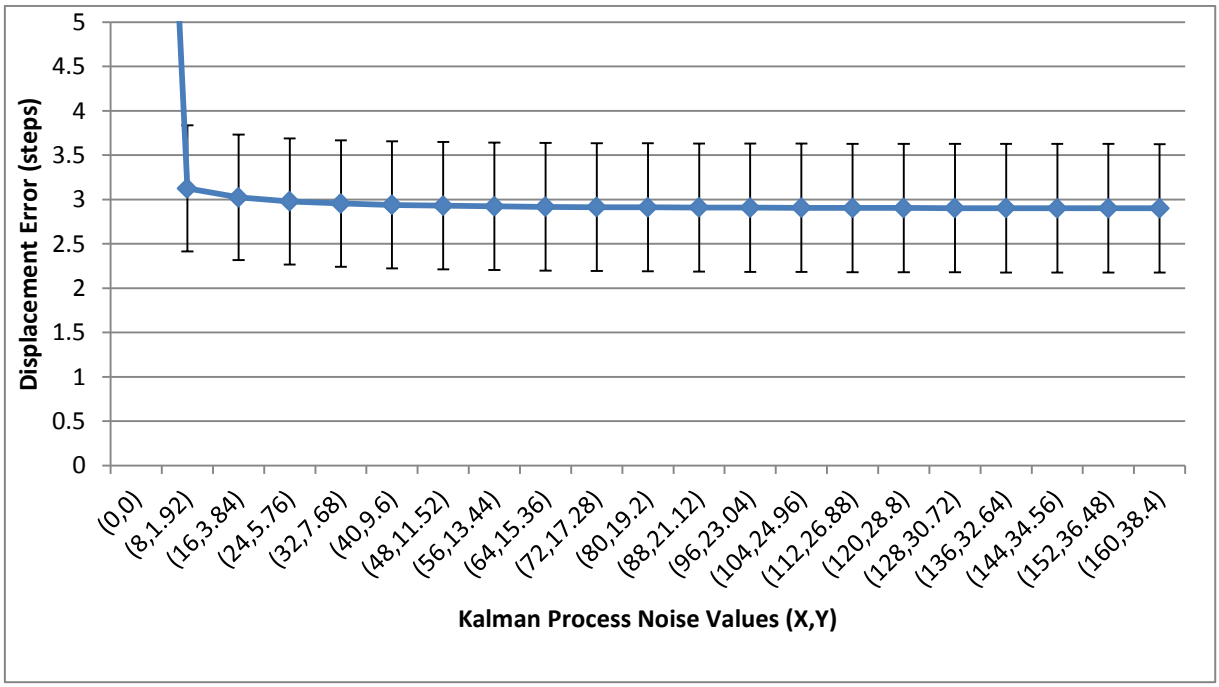


FIGURE 71. KALMAN FILTER CALIBRATION FOR ALL PARTICIPANTS ADJUSTED FOR OUTLIER VALUE. ERROR BARS ARE ONE STANDARD DEVIATION

A wide range of values can be used for the calibration, indicating the model is not very sensitive to the process noise calibration. A value of (120, 28.8), corresponding to process noise matrix of $\begin{bmatrix} 120 & 0 \\ 0 & 28.8 \end{bmatrix}$, was selected for the Kalman filter.

7.5.3. TOGGLE FILTER TUNING

Toggle filter needs to tune the threshold at which the switch between the compass and gyroscope rates will occur. Considering the normal values for the gyroscope rate readout absolute values average at around 16 degrees per second, a reasonable range to evaluate was decide to be between 0 and 30 degrees per second.

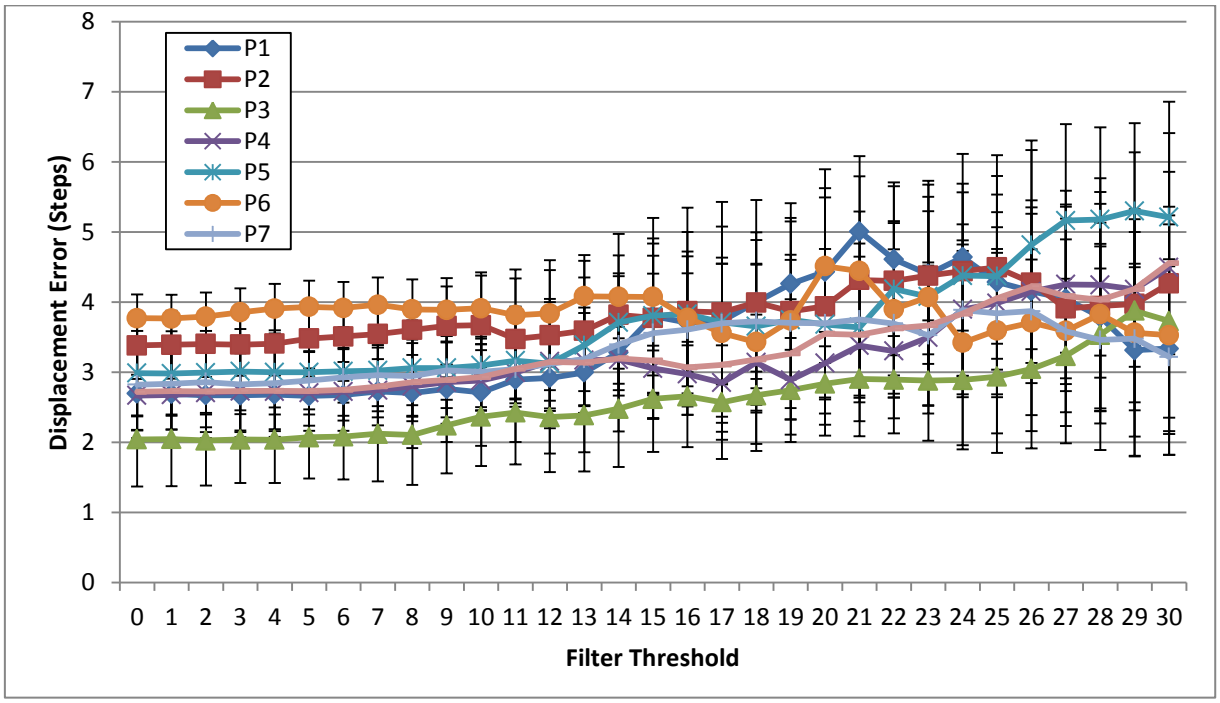


FIGURE 72. TOGGLE FILTER CALIBRATION. ERROR BARS ARE ONE STANDARD DEVIATION

Toggle filter performs best when threshold is low, as expected.

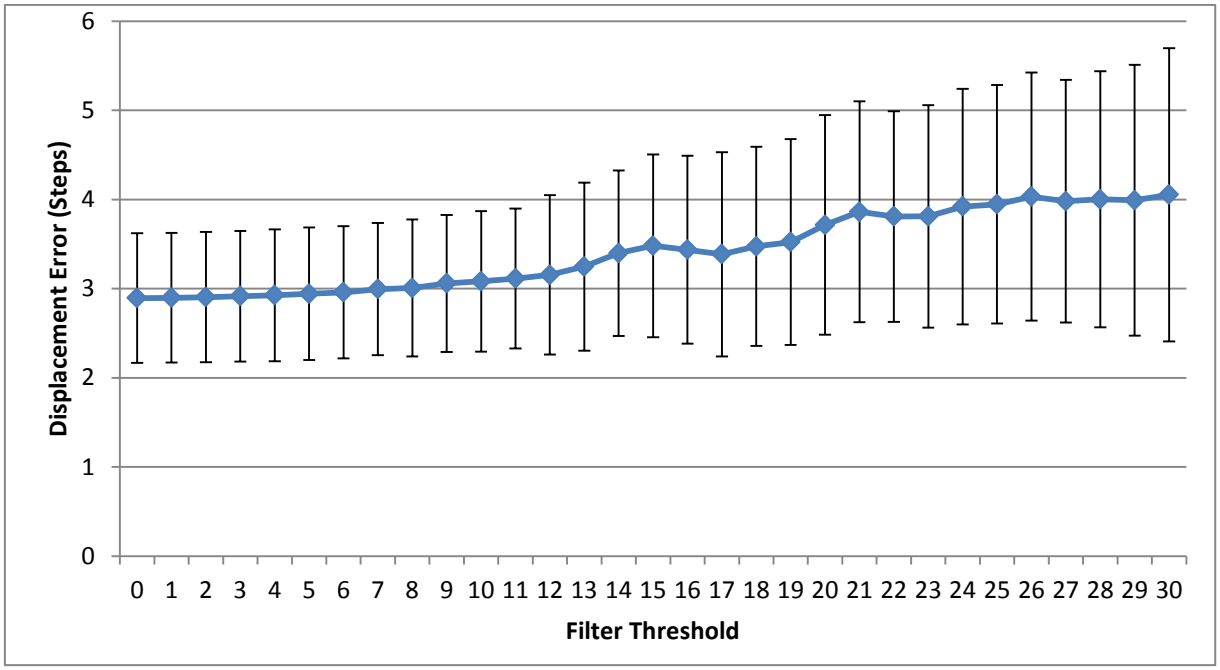


FIGURE 73. TOGGLE FILTER CALIBRATION FOR ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION

Using value of 3 degrees provides the best performance. This implies that for heading changes below 3 degrees, integrated gyroscope will be used. For changes above that value, tilt compensated compass.

7.5.4. COMPENSATING FILTER TUNING

The compensating filter has a single parameter that require calibrating – the constant *c* affecting the weighing of compass and gyroscope readings. This constant ranges from 0 to 1 by design to facilitate combining of gyroscope and compass estimates and thus specifies and definite range of values to investigate.

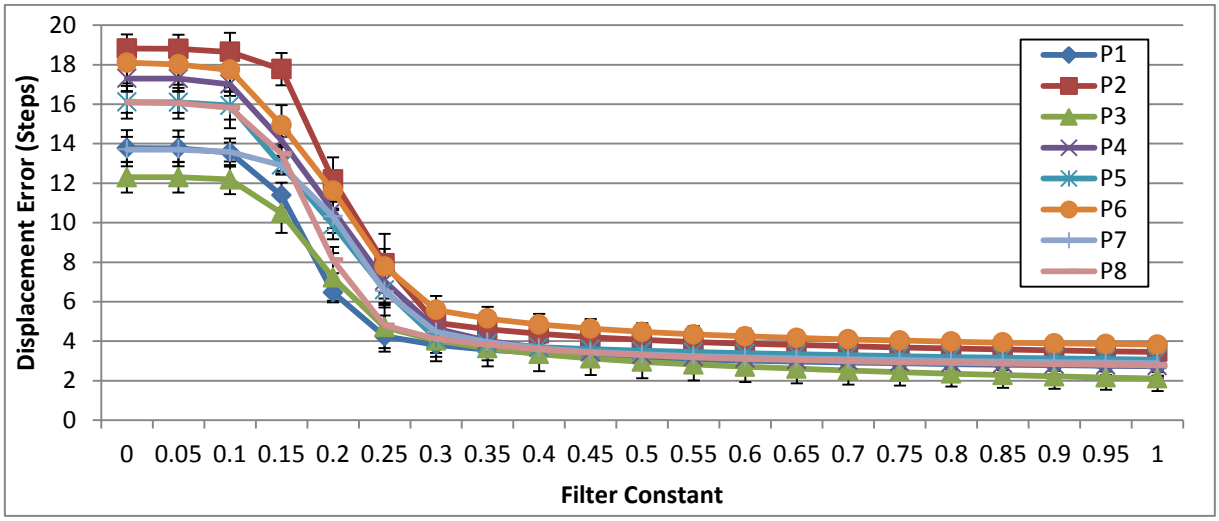


FIGURE 74. COMPENSATING FILTER CALIBRATION. ERROR BARS ARE ONE STANDARD DEVIATION

The filter performance is consistent among the participants. The actual calibration value can be determined from the combined graph:

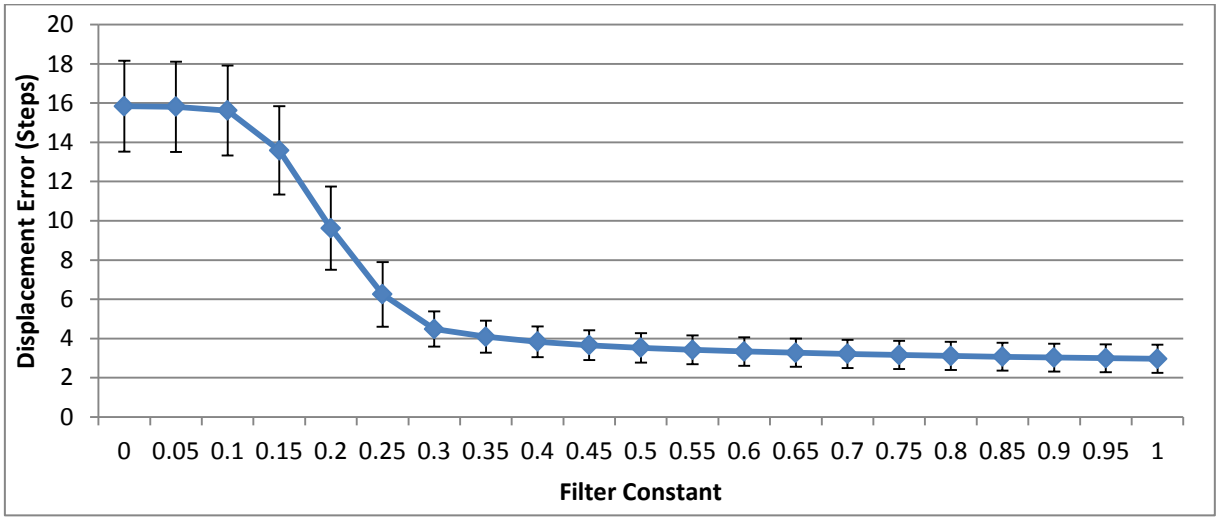


FIGURE 75. COMPENSATING FILTER CALIBRATION FOR ALL PARTICIPANTS. ERROR BARS ARE ONE STANDARD DEVIATION

Compensating filter does not require bias towards gyroscope for optimal performance. Therefore its filter constant value is set at 1 for best performance.

7.6. PEDESTRIAN DEAD RECKONING PERFORMANCE

There are different ways to evaluate numerically the performance of the dead reckoning. By design, the study ensures one of the photographic spots (the end of path photograph) corresponds with the start location. Therefore proximity of this photograph spot with the start location determines accuracy of loop closure that provides a reasonable way of determining effectiveness of tracking methods and is employed in a number of research papers (Feliz et al., 2009; Jiménez et al., 2010).

ANOVA analysis with SPSS (IBM Corp, 2011) shows the following results (Greenhouse-Geisser adjusted because of failed sphericity test):

Within-Subjects Factors

Measure: DisplacementError

Method	Dependent Variable
1	Tilt_Compensated_Compass
2	Integrated_Gyroscope
3	D_Compass
4	Complementary
5	Kalman
6	Toggle
7	Compensating

Tests of Within-Subjects Effects

Measure: DisplacementError

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Method	25134.640	1.799	13975.032	540.374	.000	.881
Method * Subject	18702.295	14.388	1299.826	50.260	.000	.846
Error(Method)	3395.481	131.29	25.862			

FIGURE 76. ANOVA RESULTS OF DISPLACEMENT ERROR ANALYSIS

Pairwise Comparisons

Measure: DisplacementError

(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.954	.626	.949	-2.920	1.012
	3	-21.913*	.737	.000	-24.229	-19.598
	4	-.052	.027	.698	-.137	.032
	5	-.030	.012	.189	-.067	.006
	6	.015	.049	1.000	-.139	.169
	7	-.309*	.051	.000	-.469	-.149
2	1	.954	.626	.949	-1.012	2.920
	3	-20.959*	.720	.000	-23.219	-18.700
	4	.902	.631	.973	-1.080	2.883
	5	.923	.628	.963	-1.047	2.894
	6	.969	.630	.944	-1.009	2.947
	7	.645	.631	1.000	-1.337	2.628
3	1	21.913*	.737	.000	19.598	24.229
	2	20.959*	.720	.000	18.700	23.219
	4	21.861*	.740	.000	19.539	24.183
	5	21.883*	.738	.000	19.564	24.202
	6	21.928*	.733	.000	19.626	24.231
	7	21.605*	.739	.000	19.283	23.926
4	1	.052	.027	.698	-.032	.137
	2	-.902	.631	.973	-2.883	1.080
	3	-21.861*	.740	.000	-24.183	-19.539
	5	.022	.017	.990	-.031	.074
	6	.067	.039	.864	-.056	.191
	7	-.256*	.031	.000	-.355	-.158
5	1	.030	.012	.189	-.006	.067
	2	-.923	.628	.963	-2.894	1.047
	3	-21.883*	.738	.000	-24.202	-19.564
	4	-.022	.017	.990	-.074	.031
	6	.046	.045	1.000	-.095	.186
	7	-.278*	.041	.000	-.405	-.151
6	1	-.015	.049	1.000	-.169	.139
	2	-.969	.630	.944	-2.947	1.009
	3	-21.928*	.733	.000	-24.231	-19.626
	4	-.067	.039	.864	-.191	.056
	5	-.046	.045	1.000	-.186	.095
	7	-.324*	.055	.000	-.498	-.150
7	1	.309*	.051	.000	.149	.469
	2	-.645	.631	1.000	-2.628	1.337
	3	-21.605*	.739	.000	-23.926	-19.283
	4	.256*	.031	.000	.158	.355
	5	.278*	.041	.000	.151	.405
	6	.324*	.055	.000	.150	.498

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

FIGURE 77. ANOVA REPEATED PAIRWISE COMPARISON RESULTS FOR FUSION METHODS/SENSORS. SENSOR IDENTIFICATION IS ON FIGURE 76

2-Axis compass performs significantly less well than all of the other methods. Tilt compensated compass and gyroscope performance on their own do not offer statistically superior performance to the remaining sensor combinations. Similarly, the different fusion methods do not vary significantly from the performance of the gyroscope and only the Kalman filter and compensating fusion filter show significant difference from the tilt compensated compass performance.

The direct displacement values highlight these findings. For easier understanding of data, the results are presented as percentage error in the displacement of the final photograph spot compared to the length of the entire path.

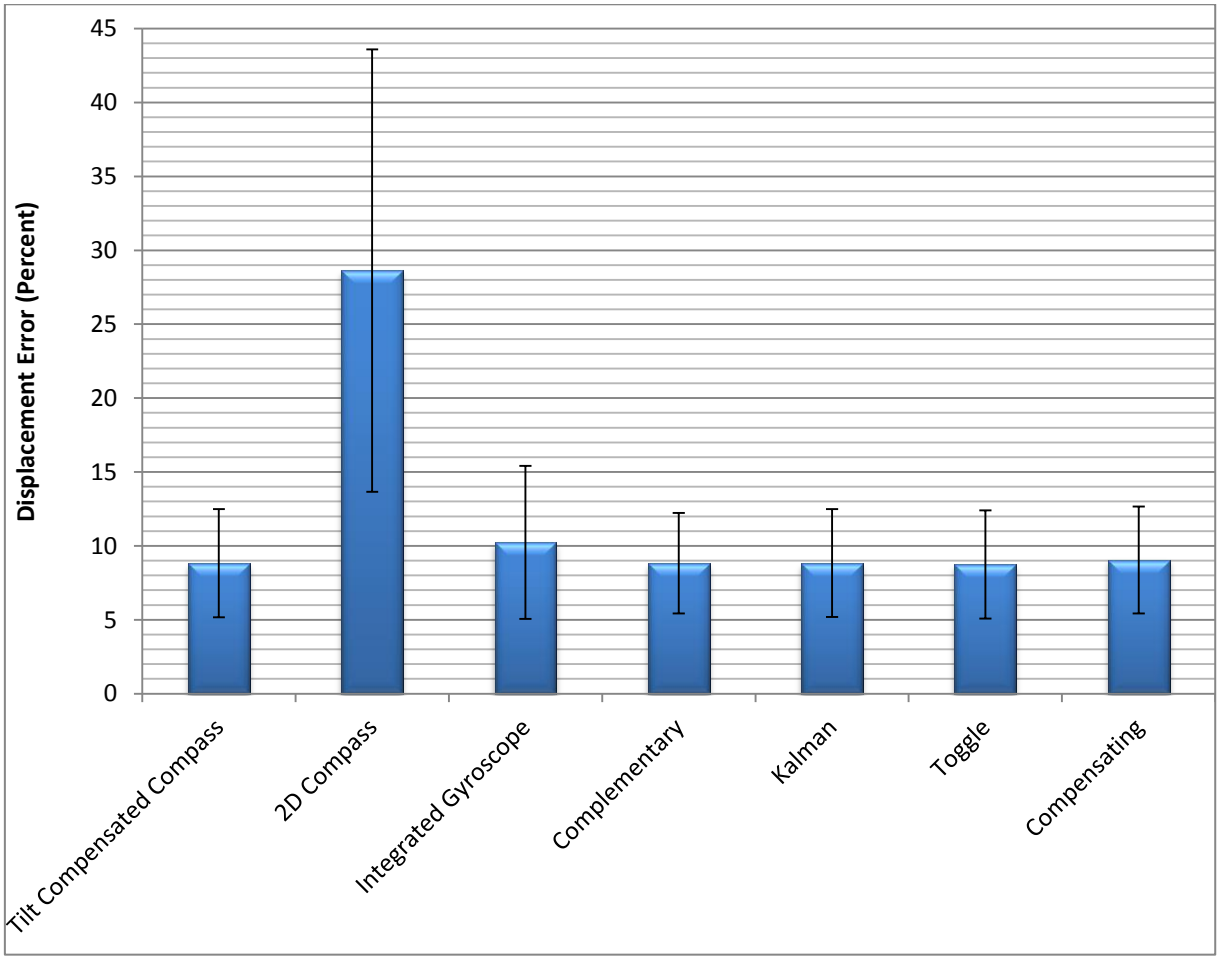


FIGURE 78. TRACKING PERFORMANCE EVALUATION. ERROR BARS ARE ONE STANDARD DEVIATION

System places the photograph spot to within 91% of the path range. Filtering algorithms offer no discernible improvement over the tilt compensated compass performance. Non-tilt compensated compass suffers from large error, averaging just 71% efficiency with large spread of readings. Gyroscope performs on par with compass when compensated for bias errors, with only 2% extra error on average falling within standard deviation, but, as shown by ANOVA analysis, this difference isn't statistically significant.

To determine if the 8 participants (producing 80 datasets) in this test are a reasonable representation of the overall population, one way ANOVA was executed on the results of different algorithms comparing both inter variance (between different participants) to the intra variance (performance of a given participant within 10 runs they did). Outlier result for the 2D fusion was removed duo to its poor overall performance.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Tilt_Compensated_Compass	Between Groups	130.490	7	18.641	1.329	.250
	Within Groups	1010.275	72	14.032		
	Total	1140.765	79			
Integrated_Gyroscope	Between Groups	2363.818	7	337.688	23.71	.000
	Within Groups	1025.272	72	14.240		
	Total	3389.090	79			
Complementary	Between Groups	122.329	7	17.476	1.235	.295
	Within Groups	1018.639	72	14.148		
	Total	1140.968	79			
Kalman	Between Groups	129.069	7	18.438	1.311	.258
	Within Groups	1012.998	72	14.069		
	Total	1142.067	79			
Toggle	Between Groups	112.740	7	16.106	1.164	.334
	Within Groups	996.575	72	13.841		
	Total	1109.315	79			
Compensating	Between Groups	110.009	7	15.716	1.119	.361
	Within Groups	1010.951	72	14.041		
	Total	1120.960	79			

FIGURE 79. ANOVA RESULTS OF PARTICIPANT VARIANCE FOR DIFFERENT ALGORITHMS/SENSORS

All but integrated gyroscope performance show significance values above the threshold of 5%, indicating that any differences between individual participants are statistically insignificant when compared to differences between individual data sets for the same participant.

Performance using the integrated gyroscope, however, varies widely enough between individual subjects to warrant post hoc analysis:

Integrated_Gyroscope

Tukey HSD[a,b]

ParticipantID	N	Subset for alpha = 0.05			
		1	2	3	4
P4 - Shaun	10	4.243339406			
P7 - Vish	10	5.272779445			
P6 - Richard	10	6.778533900			
P5 - Hossein	10	7.098047247			
P2 - Manish	11	9.317050796	9.317050796		
P1 - Aieat	9		13.68842826	13.68842826	
P3 - Ovidiu	10			17.51238708	17.51238708
P8 - Chris	10				19.70390223
Sig.		.068	.178	.328	.897

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 9.975.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

FIGURE 80. POST-HOC ANALYSIS RESULTS OF PARTICIPANT VARIANCE FOR INTEGRATED GYROSCOPE DISPLACEMENT

Post-hoc analysis split participants into 4 subsets based on their performance with the purely integrated gyroscope tracking. This large individual variance indicates that a gyroscope on its own performs differently depending on the participant and would benefit from individual tuning. As such, it cannot be recommended as the sensor for the PDR subsystem of this research.

On the other hand, the lack of statistically significant variance between participants when using tilt compensated compass or fusion algorithms indicates that the performance of these

sensors is not significantly affected by individual participant performance and does not necessarily require tuning to a specific person for optimal use.

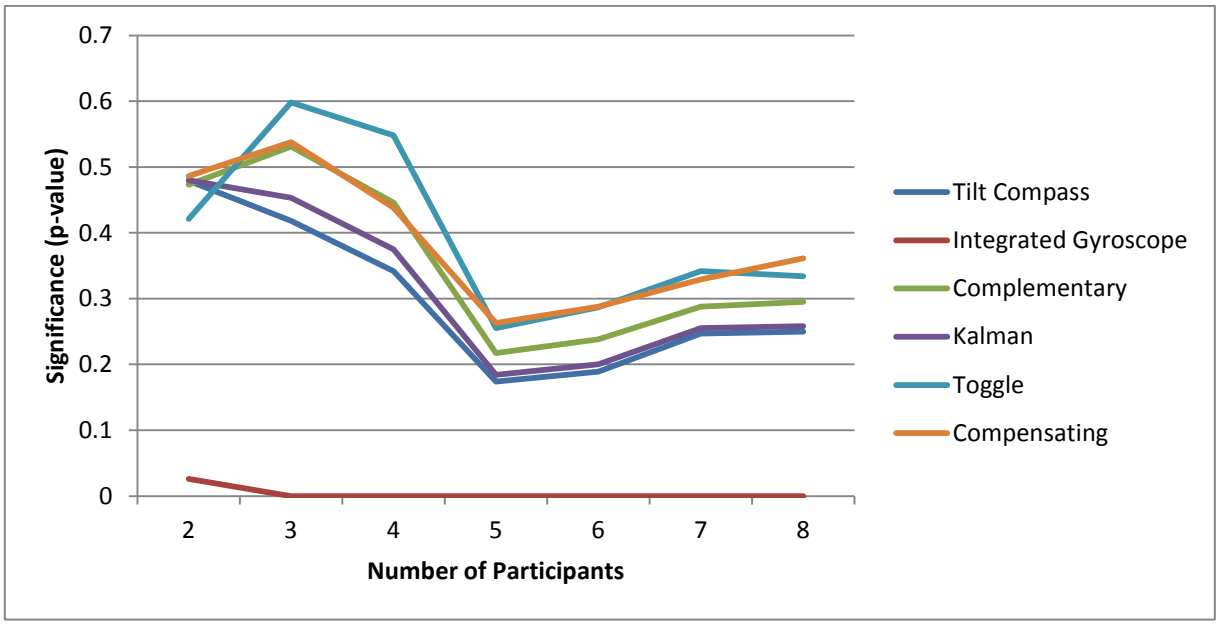


FIGURE 81. PARTICIPANT COUNT EFFECT ON ANOVA P-VALUE

Looking at the way ANOVA significance (p-value) changes with the introduction of new participants, it becomes clear that variance between participants is only significant for integrated gyroscope. For all other methods it stays above 5% threshold.

Additionally, since the integrated gyroscope is marginally outperformed by all of the other methods that show no statistically significant variance in their performance across participants, the selection of participants represents reasonable subset of the population.

7.7. CONCLUSION

The final system for spatial image management combines image compositing (that has been proven to work favourably in chapter 5) with pedestrian dead reckoning for photographer localisation. The performance of the system was tested with 80 data sets obtained from 8 participants and shown to localise the photographer to within 91% accuracy. Analysis of several distinct methods for sensor fusion including industry standard Kalman filter showed

no significant improvement in localisation when compared to the localisation using just the tilt compensated compass. Statistical analysis of the participants revealed no significant differences in the variance of the performance of the individual participants for fusion algorithms and pure tilt compensated compass-based localising, indicating that addition of more datasets is unlikely to affect the results of the study to a statistically significant amount.

The integrated gyroscope performance after bias compensation is overall comparable to that of tilt compensated compass but the variance in individual participant gait is significant enough to indicate that there is large difference in algorithm performance and it would benefit from individual calibration. Unlike gyroscope, tilt compensated compass and related fusion methods can be used without reconfiguring for new participant and is unlikely to benefit from individual tuning, making these approaches more robust and easier to deploy.

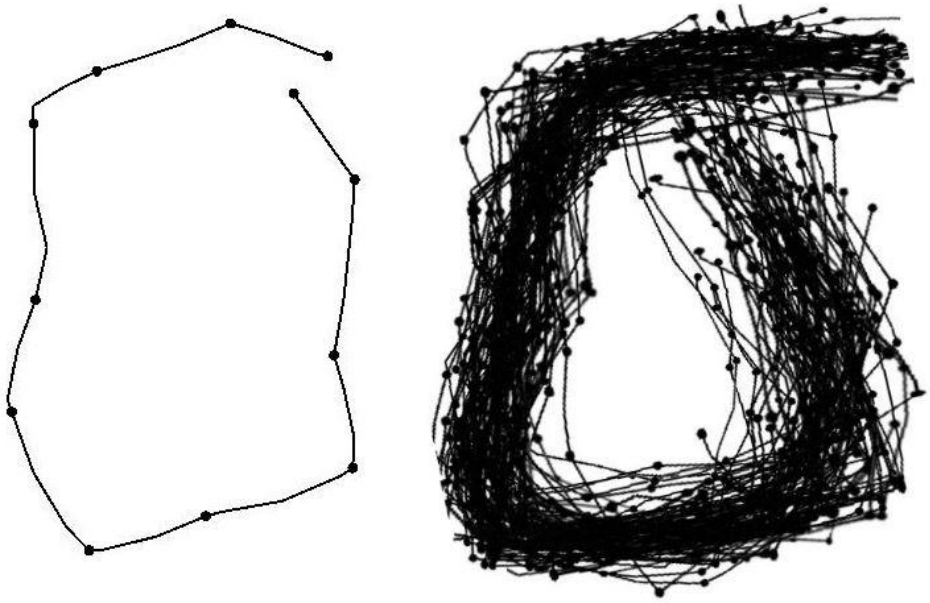


FIGURE 82. EXAMPLE RECOVERED PATH (LEFT) AND A COMPOSITE OF ALL RECOVERED PATHS (RIGHT). EACH DOT REPRESENTS AN INDIVIDUAL STEP. MOTION IS COUNTER CLOCKWISE STARTING IN TOP RIGHT CORNER

Examining the motion on the reconstructed paths (example of which is on Figure 82) as well as the ground truth video recordings shows that participants did not walk in straight lines, preferring to merge turns with one another, suggesting that using 90 degree turn assumption

for all turns is not always reasonable as it doesn’t correspond to the way people normally navigate.

8. CONCLUSIONS AND FURTHER WORK

The main goal of this research was to design and construct a prototype for an image spatial management system for photographs taken indoors and in doing so answer two main research questions:

A. How can the location at which images are captured indoors be tracked and used to enhance spatial image management

B. How well does sensor-based image compositing compare with visual methods and how can it help spatial arrangement of images

The implementation was to be guided by the constraints:

(1) Minimisation of *a priori* knowledge required in favour of increased flexibility

(2) Focus on utility of results, potentially compromising aesthetics

These goals and constraints were followed throughout the different stages of implementation and the degree to which they were achieved are summarised in this chapter.

In the course of this research, several aspects of indoor spatial image management were investigated – managing images taken from different viewpoints and those taken at the same location. A number of prototypes testing these aspects were designed and tested, all finally coalescing into a final combined system for spatial image management.

Pedestrian Dead Reckoning (PDR) was utilised for tracking the photographer in an indoor environment. Several sensors were evaluated for both compositing and PDR in a sensor analysis study with 8 participants generating 80 data sets. The study determined that both torso mounted accelerometer and leg mounted pair of gyroscopes can be used for pedestrian step counting and perform comparably. Gyroscopes were chosen for the purposes of this research because they provide direct relationship between detected data and stride size that requires no additional processing. Single integration of the gyroscope

data inside the step peak provides value that is directly proportional to the step length (Tong & Granat, 1999), allowing accounting for stride length variance with few overheads or changes to design of the sensors or firmware. Equivalent accelerometer processing would require accurate module orientation tracking and double integration of the correctly estimated vertical acceleration component (Alvarez et al., 2012) that tends to shift during walk because of natural gait.

Gyroscope and tilt compensated compass both proved reasonable sensors for heading tracking, with compass being better at turn magnitude tracking and gyroscope maintaining better stability during straight line walks because of it not being affected by acceleration errors due to walking (Chen et al., 2010). This lead to a decision to use a combination of these two sensors to try and achieve better performance than either sensor alone could provide. Non tilt compensated compass proved to be the worst performing sensor for heading tracking despite being mounted in a position to minimise platform shift during walking (Ceccato et al., 2009).

Accuracy of tilt compensated compass when it is not affected by the acceleration due to walking led to it being selected as the main sensors for sensor-based compositing. Image stitching was chosen as a way of handling multiple images taken from the same physical location. By approximating the imaging focal point to be in a similar location, sensor data could be used to compose pictures without any knowledge of the image format or data required.

An approach to image compositing reliant on a tilt compensated compass and ultrasonic rangefinder was therefore designed and implemented around a commercial digital camera. This approach requires only the most basic calculations and was tested against two commercial applications specialising in image compositing. Using paired comparison with z-score interval scale for the result display the three compositing approaches were compared to each other under different conditions. The sensor-based compositing received overall

highest score due to its ability to handle abnormal image sets, such as ones with varying lighting and magnification, limited image features and the images containing non-contiguous areas without overlapping regions. Evaluating object capture capability of the system, however, revealed that the sensor-based approach is introducing errors and duplicates into the composite, despite being able to capture the largest amount of objects compared to the image processing-based alternatives. This lead to the conclusion that a hybrid approach, reliant on image processing for stitching and sensor analysis for generating links where photographs do not overlap, would be the optimal combination of strengths of both types of compositing. On its own sensor-based approach could be used as a way of facilitating image set generation for image-processing based ones, requiring very little processing and being able to handle large image sets without degrading performance. This was illustrated with an example mobile smartphone application for Android platform.

The second half of the system, photographer localisation through pedestrian dead reckoning was tested using combined system, merging sensor compositing with step counting PDR. It was evaluated again using a further 80 datasets obtained from the same 8 participants as the sensor selection study. Final system is able to localise photographer to within 91% accuracy, comparable to systems designed by Feliz et al. (2009) for indoor performance and exceeding performance of some systems, such as PDR system by Stirling et al. (2003) with its 20% error rate, but falling short of performance of some of the alternative systems, such as one by Alvarez et al. (2012) with its 5% positioning error.

Statistical analysis of the participant performance showed no significant variance in performance of different people, indicating the solution to not require recalibration for new users, making it robust and easy to deploy. Fusion of gyroscope and tilt compensated compass for performance improvement did not produce statistically significant improvement over using just the tilt compensated compass, nor did it introduce any errors. Considering the susceptibility of compass to magnetic interference, combination of compass and gyroscope

might still be preferable in environments where such interference is likely. Under normal circumstances, however, just a torso-mounted tilt compensated compass will suffice for heading tracking.

Despite the overall success of the final design and reasonable accuracy of just 9% error, there are a number of improvements that can be introduced to the processing to further improve performance and accuracy.

Zero velocity update mechanism can help improve heading accuracy by detecting stride stances when sensors experience smallest amount of interference.

The system in its current design is unable to cope with sidestepping or moving backwards. Either of these motions is reasonable when trying to frame a photograph and either will currently adversely affect the accuracy of photographer tracking. Current step detection and navigation can be enhanced to permit determining the direction of movement with the aid of additional sensors (potentially additional or replacement accelerometers). Further refinement in the step detection to allow recognition of steps on the stairs would introduce the third dimension into the localisation, allowing tracking user through multiple floors.

Final design does not contain any way of obtaining absolute reference information, thus making the errors in heading estimation accumulate over time and gradually reduce the accuracy of the localisation. A way of dynamically re-calibrating position estimation using secondary reference sources would make the results more. Such an absolute reference could be provided by a GPS sensor, for example, if the route passes through GPS-available location, albeit the accuracy of GPS-based position estimation (Lehtinen et al., 2008; Piras & Cina, 2010; Schon & Bielenberg, 2008; Jianjun et al., 2009) might be too low for a precise re-calibration.

In closing, the goal of designing and building a spatial image management system was achieved to a reasonable degree. The end result follows all the constraints imposed during

formulation of the research questions and produces results comparable to state of the art implementations. Despite the bulky nature of the final design, its low processing requirements and lack of reliance on *a priori* knowledge makes it easy to re-implement as part of a digital imaging system. The compositing part of the design, for example, can be implemented inside of a mobile phone or a digital. Such devices already come with a variety of different sensors included (HTC Corporation, 2011; Casio America Inc, 2010; Canon U.S.A., Inc., 2011) and the low-processing compositing that can handle gaps and varying magnification levels would provide enhanced experience for users wanting to perform panoramic stitching.

The sensor-based compositing can be used as real-time feedback for a more aesthetically pleasing image processing-based one camera as illustrated by the example Android software. Thus, users will be able to take perfectly overlapping photographs guided by the instant sensor-based compositing and then supply these photographic sets to a visual panorama stitching software.

Re-implementing the complete solution on a smaller device with wireless sensors would enable its use in a variety of tasks, including but not limited to:

- Inspection tasks, such as safety inspections
- Evidential image gathering, such as crime scene photography or insurance claim investigations
- Indoor mapping of arbitrary locations
- Generating image data for tourist attraction preview media (such as brochures, web sites or maps containing images linked to photographic location views)
- Consumer level metadata-enhanced photo sharing, such as ability to share photographs enhanced with orientation and indoor location data

The results of this research show that the photographer location can be tracked using a set of inertial and magnetic sensors worn on body (that have the potential to be integrated into items of clothing once they are converted to wireless operation). Reliable tracking can be achieved by using gyroscopes and a magnetic compass without any *a priori* knowledge of the environment.

Images taken from the same viewpoint can also be arranged based purely on non-imaging sensor data provided by a tilt-compensated compass (with ability to output tilt readouts provided by accelerometers) and a rangefinder. This arrangement can be performed as image stitching (compositing) and compares favourably to image content-based method albeit a combination of the two would be a more reasonable approach. The utility of such a composite exceeds that of image processing ones when faced with abnormal conditions due to its ability to use all of the images supplied irrelevant of their overlap and image feature richness.

The entire design has the potential of being implemented inside a digital camera or a mobile phone, rather than as a standalone module, with the optional localisation support provided by external wireless sensors. Reducing the size of the image mapping module and implementing its entirety (from acquisition to composition and localisation) on an embedded microcontroller would be the final logical stage in the research and development of an indoor spatial image management system.

The research described in this thesis helps remove the boundaries of GPS-availability from the spatial image management. Introduction of information into the metadata of images ensures that each photograph contains information about itself that is not context-sensitive. This enables advanced online sharing and cataloguing that has the potential to improve the social aspect of photo sharing. The orientation and category data is not superseded by the geo-spatial satellite data of conventional GPS-based geo-tagging, but, rather, enhances and supplements it, enabling advanced devices that combine both approaches. Furthermore,

wide potential application areas mean that the techniques and technologies described in this thesis have the potential of enhancing different aspects of life, such as leisure, social and work elements.

Digital photography over the years became about more than just image capture. It became about information gathering from a number of sources (image sensor, GPS satellite positioning, etc.). This research represents the next logical step along this path.

9. BIBLIOGRAPHY

Ahn, S. & Chung, W.K., 2007. Efficient SLAM algorithm with hybrid visual map in an indoor environment. In *Control, Automation and Systems, 2007. ICCAS '07. International Conference on*. Control, Automation and Systems, 2007. ICCAS '07. International Conference on. pp. 663–667.

Alvarez, J.C., Lopez, A.M., Gonzalez, R.C. & Alvarez, D., 2012. Pedestrian dead reckoning with waist-worn inertial sensors. In *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*. Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International. pp. 24–27.

D' Angelo, P., 2010. Hugin - Panorama photo stitcher. Available at: http://hugin.sourceforge.net/ [Accessed April 5, 2010].

Aoki, H., Schiele, B. & Pentland, A., 1999. Realtime personal positioning system for a wearable computer. In *The Third International Symposium on Wearable Computers, 1999. Digest of Papers*. The Third International Symposium on Wearable Computers, 1999. Digest of Papers. pp. 37 –43.

Apple Inc., 2013. Apple (United Kingdom) - iPhone 5 - Features. Available at: http://www.apple.com/uk/iphone/features/ [Accessed May 4, 2013].

Aurnhammer, M., Hanappe, P. & Steels, L., 2006. Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval. In *Proceedings WWW2006, Collaborative Web Tagging Workshop*.

Bates, M.J., 1998. Indexing and access for digital libraries and the internet: Human, database, and domain factors. *Journal of the American Society for Information Science*, 49, p.1185–1205.

Beauregard, S., 2006. A Helmet-Mounted Pedestrian Dead Reckoning System. *2006 3rd International Forum on Applied Wearable Computing (IFAWC)*, p.1 –11.

Bernas, M., 2002. Image quality evaluation. In *International Symposium on VIPromCom Video/Image Processing and Multimedia Communications*. 4th EURASIP - IEEE Region 8 International Symposium on Video/Image Processing and Multimedia Communications. Zadar, Croatia, pp. 133–136.

Bhushan, D.B., Sowmya, V. & Soman, K.P., 2010. Super Resolution Blind Reconstruction of Low Resolution Images Using Framelets Based Fusion. In *Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on*. Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on. pp. 100–104.

Biederman, I., 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, p.115–147.

Brown, M. & Lowe, D.G., 2007. Automatic Panoramic Image Stitching using Invariant Features. *Int. J. Comput. Vision*, 74(1), p.59–73.

Brugman, H., Russel, A. & Nijmegen, X., 2004. Annotating multi-media / multimodal resources with ELAN. In *Proceedings of LREC 2004*. Fourth International Conference on Language Resources and Evaluation. pp. 2065–2068.

Bulusu, N., Heidemann, J. & Estrin, D., 2000. GPS-less low-cost outdoor localization for very small devices. *Personal Communications, IEEE*, 7(5), p.28–34.

Buse, P., 2010. The Polaroid Image as Photo-Object. *Journal of Visual Culture*, 9(2), p.189 – 207.

Camera & Imaging Products Association, 2003. What is PictBridge? [1.PictBridge overview]. Available at: http://www.cipa.jp/pictbridge/contents_e/01pictbridge1_1_e.html [Accessed January 25, 2011].

Canon U.S.A., Inc., 2011. Canon U.S.A. : Professional Imaging Products : PowerShot G12. Available at: http://www.usa.canon.com/cusa/professional/products/professional_cameras/pro_ps_digital_cameras/powershot_g12?selectedName=Features&fileName=0901e024801f6e4b_feature8.html [Accessed May 3, 2011].

Cao, F., Zhang, Z. & Zhang, Q., 2010. Seamless Image Stitching Using Optimized Boundary Matching for Gradient and Curvature. In *Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium on*. Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium on. pp. 495–498.

Casio America Inc, 2010. Casio Releases Compact EXILIM® Camera with Hybrid-GPS. *CASIO USA Press Releases*. Available at: http://www.casio.com/news/content/1B060E9E-E4B0-4211-80EA-A04348021519/ [Accessed May 16, 2011].

Casio America, Inc, 2010. Casio Exilim Press Release. Available at: http://pressroom.exilim.casio.com/press_releases/casio-releases-compact-exilim-camera-with-hybrid-gps [Accessed November 27, 2010].

Ceccato, J.-C., de Sèze, M., Azevedo, C. & Cazalets, J.-R., 2009. Comparison of Trunk Activity during Gait Initiation and Walking in Humans J. R. Ruiz, ed. *PLoS ONE*, 4(12), p.e8193.

Chang, P.C. et al., 2009. High-Frequency EM Characterization of Through-Wall Building Imaging. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(5), p.1375–1387.

Chang, T.-C., Chien, C.-A., Chang, J.-H. & Guo, J.-I., 2011. A low-complexity image stitching algorithm suitable for embedded systems. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*. Consumer Electronics (ICCE), 2011 IEEE International Conference on. pp. 197–198.

Chen, L.-J., Syu, Y.-S., Wang, B.-C. & Lee, W.-C., 2009. An analytical study of GWAP-based geospatial tagging systems. In *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on*. Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on. pp. 1–10.

Chen, W. et al., 2010. An effective Pedestrian Dead Reckoning algorithm using a unified heading error model. In *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*. Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION. pp. 340–347.

Cheng, A.-J., Lin, F.-E., Kuo, Y.-H. & Hsu, W.H., 2010. GPS, compass, or camera?: investigating effective mobile sensors for automatic search-based image annotation. In *Proceedings of the international conference on Multimedia*. Firenze, Italy: ACM, pp. 815–818.

Cheng, B.A., 2007.Automated multi-frame image capture for panorama stitching using motion sensor, U.S. Patent 20070081081, Issued 12-Apr-2007

Cheong, H., Park, S., Park, M. & Park, S.-K., 2007. Vision-based global localization in indoor environment with an object entity-based hybrid map. In *Control, Automation and Systems, 2007. ICCAS '07. International Conference on*. Control, Automation and Systems, 2007. ICCAS '07. International Conference on. pp. 218–223.

Chippendale, P., Zanin, M. & Andreatta, C., 2009. Collective Photography. In *Visual Media Production, 2009. CVMP '09. Conference for*. Visual Media Production, 2009. CVMP '09. Conference for. pp. 188–194.

Cho, W. & Hong, K.-S., 2004. Extending dynamic range of two color images under different exposures. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. pp. 853–856 Vol.4.

Clarkson, B., Mase, K. & Pentland, A., 2000. Recognizing user context via wearable sensors. In *Wearable Computers, The Fourth International Symposium on*. Wearable Computers, The Fourth International Symposium on. pp. 69–75.

Counts, S. & Fellheimer, E., 2004. Supporting social presence through lightweight photo sharing on and off the desktop. In ACM Press, pp. 599–606.

David, H., 1988. *THE METHOD OF PAIRED COMPARISONS* 2nd ed., rev., Oxford, UK: Oxford University Press.

Davis, M. et al., 2005. MMM2: mobile media metadata for media sharing. In *CHI '05 extended abstracts on Human factors in computing systems*. Conference on Human Factors in Computing Systems 2005. Portland, OR, USA: ACM, pp. 1335–1338.

Davison, A.J., Mayol, W.W. & Murray, D.W., 2003. Real-time localization and mapping with wearable active vision. In *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*. Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on. pp. 18–27.

Deissler, T. & Thielecke, J., 2009. Feature based indoor mapping using a bat-type UWB radar. In *Ultra-Wideband, 2009. ICUWB 2009. IEEE International Conference on*. Ultra-Wideband, 2009. ICUWB 2009. IEEE International Conference on. pp. 475–479.

Dickey-Bryant, L., Lautenschlager, G.J., Mendoza, J.L. & Abrahams, N., 1986. Facial attractiveness and its relation to occupational success. *Journal of Applied Psychology*, 71(1), p.16–19.

Diebel, J., 2006. *Representing Attitude: Euler Angles, Unit Quaternions, and Rotation Vectors*, Stanford University.

Van Diggelen, F., 2002. Indoor GPS theory & implementation. In *Position Location and Navigation Symposium, 2002 IEEE*. Position Location and Navigation Symposium, 2002 IEEE. pp. 240–247.

Duffin, K.L. & Barrett, W.A., 2001. Fast focal length solution in partial panoramic image stitching. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. pp. II–690 – II–695 vol.2.

Durrant-Whyte, H. & Bailey, T., 2006. Simultaneous localization and mapping: part I. *Robotics & Automation Magazine, IEEE*, 13(2), p.99–110.

Eric Tseng, H., Xu, L. & Hrovat, D., 2007. Estimation of land vehicle roll and pitch angles. *Vehicle System Dynamics*, 45(5), p.433–443.

Facebook, 2011. Welcome to Facebook. Available at: http://www.facebook.com/ [Accessed January 25, 2011].

Fang, L. et al., 2005. Design of a wireless assisted pedestrian dead reckoning system - the NavMote experience. *Instrumentation and Measurement, IEEE Transactions on*, 54(6), p.2342–2358.

Fedak, V., Veres, Z. & Nakonechny, A., 2010. Image and video super-resolution via accurate motion estimation. In *Perspective Technologies and Methods in MEMS Design (MEMSTECH), 2010 Proceedings of VIth International Conference on*. Perspective Technologies and Methods in MEMS Design (MEMSTECH), 2010 Proceedings of VIth International Conference on. pp. 223–225.

Feliz, R., Zalama, E. & Gómez, J., 2009. Pedestrian tracking using inertial sensors. *Journal of Physical Agents*, Volume 3(Issue 1).

Finlay, C.C. et al., 2010. International Geomagnetic Reference Field: the eleventh generation. *Geophysical Journal International*, 183(3), p.1216–1230.

Foxlin, E., 2005. Pedestrian tracking with shoe-mounted inertial sensors. *Computer Graphics and Applications, IEEE*, 25(6), p.38–46.

Friedl, G. & Sommer, R., 2010. Cybercasing the Joint: On the Privacy Implications of Geo-Tagging.

Fu, Q., Ren, F., Chen, L. & Xiao, Z., 2009. Multi-focus Image Fusion Algorithms Research Based on Curvelet Transform. In *Genetic and Evolutionary Computing, 2009. WGEC '09. 3rd International Conference on*. Genetic and Evolutionary Computing, 2009. WGEC '09. 3rd International Conference on. pp. 442–446.

Fu, S. & Yang, G., 2009. Uncalibrated monocular based simultaneous localization and mapping for indoor autonomous mobile robot navigation. In *Networking, Sensing and Control, 2009. ICNSC '09. International Conference on*. Networking, Sensing and Control, 2009. ICNSC '09. International Conference on. pp. 663–668.

Fu, S.-Y. et al., 2010. Compressive sensing approach based mapping and localization for mobile robot in an indoor wireless sensor network. In *Networking, Sensing and Control (ICNSC), 2010 International Conference on*. Networking, Sensing and Control (ICNSC), 2010 International Conference on. pp. 122–127.

Furukawa, Y., Curless, B., Seitz, S.M. & Szeliski, R., 2009. Reconstructing building interiors from images. In *Computer Vision, 2009 IEEE 12th International Conference on*. Computer Vision, 2009 IEEE 12th International Conference on. pp. 80–87.

Gallagher, A., Joshi, D., Yu, J. & Luo, J., 2009. Geo-location inference from image content and user tags. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. pp. 55–62.

Gallegos, G. & Rives, P., 2010. Indoor SLAM based on composite sensor mixing laser scans and omnidirectional images. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. Robotics and Automation (ICRA), 2010 IEEE International Conference on. pp. 3519–3524.

GfK Group, 2011. Emerging markets enjoy Digital Camera boom. Available at: http://www.gfk.com/group/press_information/press_releases/007119/index.en.html [Accessed August 30, 2011].

Gilman, A., Bailey, D.G. & Marsland, S., 2010. Least-squares Optimal Interpolation for Fast Image Super-resolution. In *Electronic Design, Test and Application, 2010. DELTA '10. Fifth IEEE International Symposium on*. Electronic Design, Test and Application, 2010. DELTA '10. Fifth IEEE International Symposium on. pp. 29–34.

Google Inc, 2013. Picasa. Available at: http://picasa.google.com/ [Accessed June 26, 2013].

Görmer, S. et al., 2010. Multi-exposure image acquisition for automotive high dynamic range imaging. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on. pp. 1881–1886.

Gray, C.D., 2012. *IBM SPSS statistics 19 made simple*, New York: Psychology Press.

Great Britain Parliament House of Lords Select Committee on Science and Technology, 1998. *Digital Images as Evidence : Report*, London: Her Majesty's Stationery Office.

Grzonka, S., Dijoux, F., Karwath, A. & Burgard, W., 2010. Mapping indoor environments based on human activity. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. Robotics and Automation (ICRA), 2010 IEEE International Conference on. pp. 476–481.

Guardian News and Media Limited, 2011. UK riots: every verified incident - interactive map. *the Guardian*. Available at:

http://www.guardian.co.uk/news/datablog/interactive/2011/aug/09/uk-riots-incident-map [Accessed June 13, 2012].

Guilford, J.P., 1928. The method of paired comparisons as a psychometric method. *Psychological Review*, 35(6), p.494–506.

Guivant, J. & Nebot, E., 2001. Optimization of the Simultaneous Localization and Map Building Algorithm for Real Time Implementation. *IEEE Transactions on Robotics and Automation*, 17, p.242–257.

Gunturk, B.K., 2010. Super Resolution Imaging. In R. Lukac, ed. *Computational Photography: Methods and Applications*. Taylor & Francis Group, pp. 175–202.

Guo, W. & Filer, N.P., 2006. 2.5D Indoor Mapping and Location-Sensing using an Impulse Radio Network. In *Ultra Wideband Systems, Technologies and Applications, 2006. The Institution of Engineering and Technology Seminar on*. Ultra Wideband Systems, Technologies and Applications, 2006. The Institution of Engineering and Technology Seminar on. pp. 211–215.

Gusenbauer, D., Isert, C. & Krösche, J., 2010. Self-contained indoor positioning on off-the-shelf mobile devices. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on. pp. 1–9.

Gutmann, J.-S., Burgard, W., Fox, D. & Konolige, K., 1998. An experimental comparison of localization methods. In *Intelligent Robots and Systems, 1998. Proceedings*. Intelligent Robots and Systems, 1998 IEEE/RSJ International Conference on. pp. 736–743.

Gutmann, J.-S. & Fox, D., 2002. An experimental comparison of localization methods continued. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*. Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on. pp. 454– 459 vol.1.

Gutmann, J.-S. & Konolige, K., 2000. Incremental Mapping of Large Cyclic Environments. In pp. 318–325.

Handley, J.C., 1081. Comparative Analysis of Bradley-Terry and Thurstone-Mosteller Paired Comparison Models for Image Quality Assessment. In *Proc. IS&T's Image Processing, Image Quality, Image Capture, Systems Conference*. pp. 108–112.

Hase, K. & Stein, R.B., 1999. Turning Strategies During Human Walking. *Journal of Neurophysiology*, 81(6), p.2914 –2922.

Hays, J. & Efros, A.A., 2008. IM2GPS: estimating geographic information from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8.

Helmut Dersch, 2009. Homepage H. Dersch. Available at: http://webuser.hs-furtwangen.de/~dersch/ [Accessed April 5, 2010].

Higgins, W.T., 1975. A Comparison of Complementary and Kalman Filtering. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-11(3), p.321–325.

Honeywell International Inc, 2011. 3-Axis Compass with Algorithms HMC6343. Available at: http://www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Missiles-Munitions/HMC6343.pdf [Accessed August 14, 2011].

Van House, N.A., 2007. Flickr and public image-sharing: distant closeness and photo exhibition. In ACM Press, p. 2717.

House, S. et al., 2011. Indoor localization using pedestrian dead reckoning updated with RFID-based fiducials. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011, p.7598–7601.

HTC Corporation, 2011. HTC - Products - HTC Desire HD - Specification. Available at: http://www.htc.com/www/product/desirehd/specification.html [Accessed May 17, 2011].

Hu, X., Lu, H., Zhang, L. & Serikawa, S., 2010. A New Type of Multi-focus Image Fusion Method Based on Curvelet Transforms. In *Electrical and Control Engineering (ICECE), 2010 International Conference on*. Electrical and Control Engineering (ICECE), 2010 International Conference on. pp. 172–175.

Hua Yan, Ju Liu, Jiande Sun & Xiuhua Ji, 2008. Regularization super-resolution image fusion considering inaccurate image registration and observation noise. In *Neural Networks and Signal Processing, 2008 International Conference on*. Neural Networks and Signal Processing, 2008 International Conference on. pp. 91–94.

Hubel, P.M. & Gennetten, K.D., 2005.Method and apparatus for automatically capturing a plurality of images ..., U.S. Patent 6930703, Issued 16-Aug-2005

Huber, D., Carmichael, O. & Hebert, M., 2000. 3D map reconstruction from range data. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*. Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on. pp. 891–897 vol.1.

Ibarra Bonilla, M.N., Escamilla-Ambrosio, P.J. & Ramirez Cortes, J.M., 2011. Pedestrian dead reckoning towards indoor location based applications. In *Electrical Engineering Computing Science and Automatic Control (CCE), 2011 8th International Conference on*. Electrical Engineering Computing Science and Automatic Control (CCE), 2011 8th International Conference on. pp. 1–6.

IBM Corp, 2011. *IBM SPSS Statistics for Windows*, Armonk, NY: IBM Corp.

Im, S.-B., 1984. Visual Preferences in Enclosed Urban Spaces An Exploration of a Scientific Approach to Environmental Design. *Environment and Behavior*, 16(2), p.235–262.

Ishikawa, T. et al., 2009a. In-Situ 3D Indoor Modeler with a Camera and Self-contained Sensors. In *Proceedings of the 3rd International Conference on Virtual and Mixed Reality: Held as Part of HCI International 2009*. VMR '09. Berlin, Heidelberg: Springer-Verlag, pp. 454–464.

Ishikawa, T., Kourogi, M., Okuma, T. & Kurata, T., 2009b. Economic and Synergistic Pedestrian Tracking System for Indoor Environments. In *Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of*. Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of. pp. 522–527.

Ito, M. & Okabe, D., 2003. Camera phones changing the definition of picture-worthy. *Japan Media Review*, 2007. Available at: http://www.ojr.org/japan/wireless/1062208524.php [Accessed September 22, 2011].

Jacobs, N. et al., 2007. Geolocating Static Cameras. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. pp. 1–6.

Jacobs, N., Roman, N. & Pless, R., 2008. Toward Fully Automatic Geo-Location and Geo-Orientation of Static Outdoor Cameras. In *Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision*. Washington, DC, USA: IEEE Computer Society, pp. 1–6.

Japan Electronics and Information Technology Industries Association, 2010. JEITA / Publications - JEITA Standards. *CP-3451A Exchangeable image file format for digital still cameras: Exif Unified Version 2.21*. Available at: http://www.jeita.or.jp/english/public_standard/ [Accessed May 3, 2010].

Ji, X.S. et al., 2006. Application of the Digital Signal Procession in the MEMS Gyroscope De-drift. In *Nano/Micro Engineered and Molecular Systems, 2006. NEMS '06. 1st IEEE International Conference on*. Nano/Micro Engineered and Molecular Systems, 2006. NEMS '06. 1st IEEE International Conference on. pp. 218–221.

Jia, J. & Tang, C.-K., 2008. Image Stitching Using Structure Deformation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(4), p.617–631.

Jianjun, Z., Lixiang, S. & Hong, Y., 2009. Assessment and research on the Self-Interference of GPS weak signal acquisition in indoor location environment. In *Information, Computing and Telecommunication, 2009. YC-ICT '09. IEEE Youth Conference on*. Information, Computing and Telecommunication, 2009. YC-ICT '09. IEEE Youth Conference on. pp. 110–113.

Jimenez, A.R., Seco, F., Prieto, C. & Guevara, J., 2009. A comparison of Pedestrian Dead-Reckoning algorithms using a low-cost MEMS IMU. In *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*. Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on. pp. 37–42.

Jiménez, A.R., Seco, F., Prieto, J.C. & Guevara, J., 2010. Indoor pedestrian navigation using an INS/EKF framework for yaw drift reduction and a foot-mounted IMU. In *2010 7th Workshop on Positioning Navigation and Communication (WPNC)*. 2010 7th Workshop on Positioning Navigation and Communication (WPNC). IEEE, pp. 135–143.

Jing, G., Shi, Y. & Lu, B., 2010. Single-Image Super-Resolution Based on Decomposition and Sparse Representation. In *Multimedia Communications (Mediacom), 2010 International Conference on*. Multimedia Communications (Mediacom), 2010 International Conference on. pp. 127–130.

Jirawimut, R. et al., 2003. A method for dead reckoning parameter correction in pedestrian navigation system. *Instrumentation and Measurement, IEEE Transactions on*, 52(1), p.209–215.

Joshi, D., Gallagher, A., Yu, J. & Luo, J., 2010. Exploring user image tags for geo-location inference. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. pp. 5598–5601.

Judd, T., 1997. A Personal Dead Reckoning Module. In *ION GPS-97: Proceedings of the 10th International Technical Meeting of the Satellite Division of the Institute of Navigation*. 10th International Technical Meeting of the Satellite Division of the Institute of Navigation. Kansas City, Missouri: Institute of Navigation.

Judd, T. & Vu, T., 2008. Use of a new pedometric dead reckoning module in GPS denied environments. In *Position, Location and Navigation Symposium, 2008 IEEE/ION*. Position, Location and Navigation Symposium, 2008 IEEE/ION. pp. 120–128.

Kang, G.-Q., 2010. Image Stitching Based on Motion Estimation. In *Information and Computing (ICIC), 2010 Third International Conference on*. Information and Computing (ICIC), 2010 Third International Conference on. pp. 82–85.

Kelly, K.M. et al., 1999. Cryogen spray cooling in combination with nonablative laser treatment of facial rhytides. *Archives of dermatology*, 135(6), p.691–694.

Khoomboon, S., Kasetkasem, T., Keinprasit, R. & Sugino, N., 2010. Increase a standalone GPS positioning accuracy by using a proximity sensor. In *Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on*. Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on. pp. 584–587.

Kim, H. & Chang, S., 2010. RFID assisted image annotation system for a portable digital camera. In *Control Automation and Systems (ICCAS), 2010 International Conference on*. Control Automation and Systems (ICCAS), 2010 International Conference on. pp. 1156–1159.

Kim, J.W., Jang, H.J., Hwang, D.-H. & Park, C., 2004. A Step, Stride and Heading Determination for the Pedestrian Navigation System. *Journal of Global Positioning Systems*, 3, p.273–279.

Kim, S., Cheong, H., Park, J.-H. & Park, S.-K., 2009a. Human augmented mapping for indoor environments using a stereo camera. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. pp. 5609–5614.

Kim, S., Park, J.-H. & Park, S.-K., 2009b. Vision-based human augmented mapping for indoor environments. In *Advanced Robotics, 2009. ICAR 2009. International Conference on*. Advanced Robotics, 2009. ICAR 2009. International Conference on. pp. 1–6.

Kleiner, A., Dornhege, C. & Sun Dali, 2007. Mapping disaster areas jointly: RFID-Coordinated SLAM by Hurnans and Robots. In *Safety, Security and Rescue*

Robotics, 2007. SSRR 2007. IEEE International Workshop on. Safety, Security and Rescue Robotics, 2007. SSRR 2007. IEEE International Workshop on. pp. 1–6.

Kodaka, K. et al., 2009. Transport services on indoor map generating system using spatially distributed RFID tags for home environment. In *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*. Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on. pp. 758–763.

Kolor, 2010. AutoPano Pro features - Panoramic photo software - panorama stitcher - photo stitch software. Available at: http://www.autopano.net/en/photo-stitching-solutions/autopano-pro.html [Accessed April 5, 2010].

Kong, J. et al., 2007. A Novel Fusion Approach of Multi-exposure Image. In *EUROCON, 2007. The International Conference on 'Computer as a Tool'*. EUROCON, 2007. The International Conference on 'Computer as a Tool'. pp. 163–169.

Konolige, K. & Agrawal, M., 2008. FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping. *Robotics, IEEE Transactions on*, 24(5), p.1066–1077.

Koo, H.I., Kim, B.S. & Cho, N.I., 2009. A new method to find an optimal warping function in image stitching. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on. pp. 1289–1292.

Kourogi, M., Ishikawa, T. & Kurata, T., 2010. A method of pedestrian dead reckoning using action recognition. In *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*. Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION. pp. 85–89.

Kourogi, M. & Kurata, T., 2003a. A method of personal positioning based on sensor data fusion of wearable camera and self-contained sensors. In *Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on*. Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on. pp. 287– 292.

Kourogi, M. & Kurata, T., 2003b. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. In *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*. Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on. pp. 103– 112.

Kourogi, M., Kurata, T., Sakaue, K. & Muraoka, Y., 2000. Improvement of panorama-based annotation overlay using omnidirectional vision and inertial sensors. In *Wearable Computers, The Fourth International Symposium on*. Wearable Computers, The Fourth International Symposium on. pp. 183–184.

Kourogi, M., Kurata, T. & Sakaue., K., 2001. A Panorama-based Method of Personal Positioning and Orientation and Its Real-time Applications for Wearable Computers. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*. ISWC '01. Washington, DC, USA: IEEE Computer Society, p. 107–.

Kourogi, M., Sakata, N., Okuma, T. & Kurata, T., 2006. Indoor/ Outdoor Pedestrian Navigation with an Embedded GPS/RFID/Self-contained Sensor System. In *Proceedings of the 16th International conference on Artificial Reality and Telexistence (ICAT 2006*. pp. 1310–1321.

Kuai, X. et al., 2010. Simultaneous localization and mapping (SLAM) for indoor autonomous mobile robot navigation in wireless sensor networks. In *Networking, Sensing and Control (ICNSC), 2010 International Conference on*. Networking, Sensing and Control (ICNSC), 2010 International Conference on. pp. 128–132.

Kuang, J. et al., 2007. Evaluating HDR rendering algorithms. *ACM Trans. Appl. Percept.*, 4(2).

Ladetto, Q. et al., 2002. Digital Magnetic Compass and Gyroscope for Dismounted Soldier Position & Navigation. In *Military Capabilities enabled by Advances in Navigation Sensors, Sensors & Electronics Technology Panel, NATO-RTO meetings, Istanbul, Turkey*.

Ladetto, Q. & Merminod, B., 2002. Digital Magnetic Compass and Gyroscope Integration for Pedestrian Navigation. In *9th Saint Petersburg International Conference on Integrated Navigation Systems, Saint Petersburg, Russia*.

Ladetto, Q., Merminod, B., Terrier, P. & Schutz, Y., 1999. On foot navigation: When GPS alone is not enough. In *Global Navigation Satellite System, GNSS, Genova, Italy*. pp. 443–449.

Lamond, B. & Watson, G., 2004. Hybrid rendering - a new integration of photogrammetry and laser scanning for image based rendering. In *Theory and Practice of Computer Graphics, 2004. Proceedings*. Theory and Practice of Computer Graphics, 2004. Proceedings. pp. 179–186.

Larsen, J., 2005. Families Seen Sightseeing. *Space and Culture*, 8(4), p.416 –434.

Le, C., Dogaru, T., Lam Nguyen & Ressler, M.A., 2009. Ultrawideband (UWB) Radar Imaging of Building Interior: Measurements and Predictions. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(5), p.1409–1420.

Lehtinen, M., Happonen, A. & Ikonen, J., 2008. Accuracy and time to first fix using consumer-grade GPS receivers. In *Software, Telecommunications and Computer Networks, 2008. SoftCOM 2008. 16th International Conference on*. Software, Telecommunications and Computer Networks, 2008. SoftCOM 2008. 16th International Conference on. pp. 334–340.

Leonard, J.J. & Durrant-Whyte, H.F., 1991. Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*. Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on. pp. 1442–1447 vol.3.

Li, S., Kanbara, T. & Hayashi, A., 1999. Making a local map of indoor environments by swiveling a camera and a sonar. In *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*. Intelligent Robots and

Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on. pp. 954–959 vol.2.

Li, Y., Wang, Y., Huang, W. & Zhang, Z., 2008. Automatic image stitching using SIFT. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*. Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on. pp. 568–571.

Lim, Y.P., Brown, I.T. & Khoo, J.C.T., 2008. An accurate and robust gyroscope-gased pedometer. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. pp. 4587–4590.

Loscos, C. & Jacobs, K., 2010. High-Dynamic Range Imaging for Dynamic Scenes. In *Computational Photography: Methods and Applications*. Taylor & Francis Group, pp. 259–279.

Louviere, J.J. & Meyer, R.J., 1976. A Model for Residential Impression Formation. *Geographical Analysis*, 8(4), p.479–486.

Lowe, D.G., 2003. Distinctive Image Features from Scale-Invariant Keypoints.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. pp. 1150 – 1157 vol.2.

Luczak, S., Oleksiuk, W. & Bodnicki, M., 2006. Sensing Tilt With MEMS Accelerometers. *Sensors Journal, IEEE*, 6(6), p.1669–1675.

Lukac, R., 2010. *Computational Photography: Methods and Applications*, Taylor & Francis Group.

Luo, C. & Yang, S.X., 2008. A Bioinspired Neural Network for Real-Time Concurrent Map Building and Complete Coverage Robot Navigation in Unknown Environments. *Neural Networks, IEEE Transactions on*, 19(7), p.1279–1298.

Luo, R.C. & Lai, C.C., 2010. Multisensor based effective indoor environment map build-up for intelligent service robot. In *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*. Intelligent Control and Automation (WCICA), 2010 8th World Congress on. pp. 1456–1461.

Magnus Egelberg, 2010. PanoWizard. Available at: http://www.egelberg.se/panowizard/ [Accessed April 5, 2010].

Malviya, A. & Bhirud, S.G., 2009a. Multi-Focus Image Fusion of Digital Images. In *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09. International Conference on*. Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09. International Conference on. pp. 887–889.

Malviya, A. & Bhirud, S.G., 2009b. Wavelet based multi-focus image fusion. In *Methods and Models in Computer Science, 2009. ICM2CS 2009. Proceeding of International*

Conference on. Methods and Models in Computer Science, 2009. ICM2CS 2009. Proceeding of International Conference on. pp. 1–6.

Matthew Brown, 2010. AutoStitch. Available at: http://cvlab.epfl.ch/~brown/autostitch/autostitch.html [Accessed April 5, 2010].

Max Lyons, 2009. PTAssembler. Available at: http://www.tawbaware.com/ptasmblr.htm [Accessed April 5, 2010].

Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, 2012. *ELAN*, Available at: http://tla.mpi.nl/tools/tla-tools/elan/.

Maybeck, P.S., 1979. *Stochastic models, estimation, and control*, Academic Press.

Mertens, T., Kautz, J. & Van Reeth, F., 2007. Exposure Fusion. In *Computer Graphics and Applications, 2007. PG '07. 15th Pacific Conference on*. Computer Graphics and Applications, 2007. PG '07. 15th Pacific Conference on. pp. 382–390.

Microsoft Corporation, 2010a. Microsoft Research Image Composite Editor (ICE). Available at: http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice/ [Accessed April 10, 2010].

Microsoft Corporation, 2010b. Photosynth - Use your camera to stitch the world. Available at: http://photosynth.net/default.aspx [Accessed April 10, 2010].

Microsoft Corporation, 2011. Windows Photo Gallery: frequently asked questions. Available at: http://windows.microsoft.com/en-US/windows-vista/Windows-Photo-Gallery-frequently-asked-questions [Accessed May 19, 2011].

Mihal, A. et al., 2010. Enblend/Enfuse - combine images with no seams. *Enblend/Enfuse - combine images with no seams*. Available at: http://enblend.sourceforge.net/ [Accessed April 5, 2010].

Miles, M.D., 1986. Measurement of Six Degree of Freedom Model Motions Using Strapdown Accelerometers. *Proceedings of the 21st American Towing Tank Conference*, p.369–376.

Miyazaki, S., 1997. Long-term unrestrained measurement of stride length and walking velocity utilizing a piezoelectric gyroscope. *Biomedical Engineering, IEEE Transactions on*, 44(8), p.753–759.

Mlgaard, C., 2004.Digital camera with integrated accelerometers, U.S. Patent 6747690, Issued 08-Jun-2004

Montag, E.D., 2003. Louis Leon Thurstone in Monte Carlo: creating error bars for the method of paired comparison. In Y. Miyake & D. R. Rasmussen, eds. *Image Quality and System Performance*. SPIE, pp. 222–230.

Naikal, N., Kua, J., Chen, G. & Zakhor, A., 2009. Image augmented laser scan matching for indoor dead reckoning. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. pp. 4134–4141.

Nakajima, K. & Tanaka, T., 2004. Study on accuracy improvement under bad condition in GPS. In *SICE 2004 Annual Conference*. SICE 2004 Annual Conference. pp. 234–238 vol. 1.

National Oceanic and Atmospheric Administration, NGDC Geomagnetic Calculators | ngdc.noaa.gov. Available at: http://www.ngdc.noaa.gov/geomag-web/#igrfwmm [Accessed April 30, 2012].

National Trust, 2011. National Trust | Chirk Castle. Available at: http://www.nationaltrust.org.uk/main/w-chirkcastle [Accessed August 10, 2011].

New House Internet Services BV, 2010. Photo stitching software 360 degree Panorama image software - PTGui. Available at: http://www.ptgui.com/ [Accessed April 10, 2010].

NTP Project, 2011. ntp.org: Home of the Network Time Protocol. Available at: http://www.ntp.org/ [Accessed September 26, 2011].

Ogaz, M., Sandoval, R. & Chacon, M., 2009. Data processing from a Laser Range Finder sensor for the construction of geometric maps of an indoor environment. In *Circuits and Systems, 2009. MWSCAS '09. 52nd IEEE International Midwest Symposium on*. Circuits and Systems, 2009. MWSCAS '09. 52nd IEEE International Midwest Symposium on. pp. 306–313.

Ojeda, L. & Borenstein, J., 2007. Non-GPS navigation with the personal dead-reckoning system. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.

Papagiannis, H., 2009. Augmented Reality (AR) joiners, a novel expanded cinematic form. In *Mixed and Augmented Reality - Arts, Media and Humanities, 2009. ISMAR-AMH 2009. IEEE International Symposium on*. Mixed and Augmented Reality - Arts, Media and Humanities, 2009. ISMAR-AMH 2009. IEEE International Symposium on. pp. 39–42.

Papliatseyeu, A., Kotilainen, N., Mayora, O. & Osmani, V., 2009. FINDR: Low-Cost Indoor Positioning Using FM Radio. In J.-M. Bonnin, C. Giannelli, & T. Magedanz, eds. *MobileWireless Middleware, Operating Systems, and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 15–26.

Piao, Y. & Xu, W., 2010. Method of auto multi-exposure for high dynamic range imaging. In *Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference on*. Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference on. pp. 93–97.

Piras, M. & Cina, A., 2010. Indoor positioning using low cost GPS receivers: Tests and statistical analyses. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on. pp. 1–7.

Pirinen, O., Foi, A. & Gotchev, A., 2010. Color High-Dynamic Range Imaging: Algorithms for Acquisition and Display. In *Computational Photography: Methods and Applications*. Taylor & Francis Group, pp. 228–256.

Prez Lorenzo, J.M. et al., 2004. A Hough-based method for concurrent mapping and localization in indoor environments. In *Robotics, Automation and Mechatronics, 2004 IEEE Conference on*. Robotics, Automation and Mechatronics, 2004 IEEE Conference on. pp. 840–845 vol.2.

Saeedi, J. & Faez, K., 2009. Fisher classifier and fuzzy logic based multi-focus image fusion. In *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*. Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on. pp. 420–425.

Saez, J.M. & Escolano, F., 2004. A global 3D map-building approach using stereo vision. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*. Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on. pp. 1197–1202 Vol.2.

Said, E.H., Homaifar, A. & Grossberg, M., 2009. Creating virtual sensors using learning based super resolution and data fusion. In *Aerospace conference, 2009 IEEE*. Aerospace conference, 2009 IEEE. pp. 1–9.

Sandnes, F., 2010a. Determining the Geographical Location of Image Scenes based on Object Shadow Lengths. *Journal of Signal Processing Systems*, p.1–13.

Sandnes, F.E., 2009a. Geo-spatial Tagging of Image Collections Using Temporal Camera Usage Dynamics. In *Pervasive Systems, Algorithms, and Networks (ISPAN), 2009 10th International Symposium on*. Pervasive Systems, Algorithms, and Networks (ISPAN), 2009 10th International Symposium on. pp. 160–165.

Sandnes, F.E., 2009b. Sorting Holiday Photos without a GPS: What Can We Expect from Contents-Based Geo-spatial Image Tagging? In P. Muneesawang et al., eds. *Advances in Multimedia Information Processing - PCM 2009*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 256–267.

Sandnes, F.E., 2010b. Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. *Multimedia Systems*, 16(4-5), p.309–318.

Schon, S. & Bielenberg, O., 2008. On the capability of high sensitivity GPS for precise indoor positioning. In *Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on*. Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on. pp. 121–127.

Shen, H. et al., 2009. Comparison of Image Quality Objective Evaluation. In *2009 International Conference on Computational Intelligence and Software Engineering*. 2009 International Conference on Computational Intelligence and Software Engineering. Wuhan, China, pp. 1–4.

Shin, H. & Cha, H., 2010. Wi-Fi Fingerprint-Based Topological Map Building for Indoor User Tracking. In *Embedded and Real-Time Computing Systems and Applications (RTCSA), 2010 IEEE 16th International Conference on*. Embedded and Real-Time Computing Systems and Applications (RTCSA), 2010 IEEE 16th International Conference on. pp. 105–113.

Siddiqui, A.B., Jaffar, M.A., Hussain, A. & Mirza, A.M., 2010. Block-Based Feature-Level Multi-Focus Image Fusion. In *Future Information Technology (FutureTech), 2010 5th International Conference on*. Future Information Technology (FutureTech), 2010 5th International Conference on. pp. 1–7.

Singh, V., Venkatesha, S. & Singh, A.K., 2010. Geo-clustering of Images with Missing GeoTags. In *Granular Computing (GrC), 2010 IEEE International Conference on*. Granular Computing (GrC), 2010 IEEE International Conference on. pp. 420–425.

Skog, I., Handel, P., Nilsson, J.O. & Rantakokko, J., 2010. Zero-Velocity Detection - An Algorithm Evaluation. *Biomedical Engineering, IEEE Transactions on*, 57(11), p.2657 –2666.

Skrba, Z. et al., 2009. Objective real-time assessment of walking and turning in elderly adults. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. pp. 807–810.

Smith, R., Self, M. & Cheeseman, P., 1990. Estimating uncertain spatial relationships in robotics. In I. J. Cox & G. T. Wilfong, eds. *Autonomous robot vehicles*. New York, NY, USA: Springer-Verlag New York, Inc., pp. 167–193.

Smith, S., 1999. *The scientist and engineer's guide to digital signal processing* 2nd edition., San Diego (Calif.): California Technical Pub.

Snavely, N., Seitz, S.M. & Szeliski, R., 2006. Photo tourism: Exploring photo collections in 3D. In *SIGGRAPH Conference Proceedings*. New York, NY, USA: ACM Press, pp. 835–846.

Spalding, J., 1989. Review: David Hockney. London, Tate Gallery. *The Burlington Magazine*, 131(1030), p.51–53.

Stirling, R., Collin, J., Fyfe, K. & Lachapelle, G., 2003. An Innovative Shoe-Mounted Pedestrian Navigation System. *CD-ROM Proceedings of GNSS*.

Subotic, N. et al., 2008. Parametric reconstruction of internal building structures via canonical scattering mechanisms. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. pp. 5189–5192.

Takezawa, S., Ishimoto, T. & Dissanayake, G., 2006. Optimal Control for Simultaneous Localisation and Mapping Problems in Indoor Environments with Stereo Vision. In *IEEE Industrial Electronics, IECON 2006 - 32nd Annual Conference on*. IEEE Industrial Electronics, IECON 2006 - 32nd Annual Conference on. pp. 4749–4754.

Tang, Y. & Jiang, H., 2009. Highly Efficient Image Stitching Based on Energy Map. In *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*. Image and Signal Processing, 2009. CISP '09. 2nd International Congress on. pp. 1–5.

Thrun, S. et al., 1998. A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots. In *Machine Learning*. pp. 29–53.

Thurstone, L.L., 1927. A Law of Comparative Judgment. *Psychological Review*, 34(4), p.273–286.

Tong, K. & Granat, M.H., 1999. A practical gait analysis system using gyroscopes. *Medical Engineering & Physics*, 21(2), p.87–94.

Tumblr, Inc., 2012. Tumblr. Available at: https://www.tumblr.com/ [Accessed December 15, 2012].

Two Pilots©, 2011. EXIF Editor, EXIF Viewer, EXIF Data. *Professional Photography Software for editing EXIF, IPTC, and XMP data*. Available at: http://exifpilot.com/ [Accessed September 18, 2011].

Venkatraman, K., Amutha, B., Karthick, K.T. & Sankar, S.R., 2010. A hybrid method for improving GPS accuracy for land vehicle navigation system. In *Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on*. Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on. pp. 74–79.

Wei Chen et al., 2009. An integrated GPS and multi-sensor pedestrian positioning system for 3D urban navigation. In *Urban Remote Sensing Event, 2009 Joint*. Urban Remote Sensing Event, 2009 Joint. pp. 1–6.

Welch, G. & Bishop, G., 1995. *An Introduction to the Kalman Filter*, Chapel Hill, NC, USA: University of North Carolina at Chapel Hill.

Wetherbie, J.O. & Smith, C.E., 2001. Large-scale feature identification for indoor topological mapping. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. Systems, Man, and Cybernetics, 2001 IEEE International Conference on. pp. 2852–2857 vol.5.

Whisney, A.J., Winakor, G. & Wolins, L., 1979. Fashion Preference: Drawings versus Photographs. *Home Economics Research Journal*, 8(2), p.138–150.

Wither, J., Coffin, C., Ventura, J. & Hollerer, T., 2008. Fast annotation and modeling with a single-point laser range finder. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*. Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on. pp. 65–68.

Wolfram Alpha LLC, 2010. Angle of view for photography - Wolfram|Alpha. *Wolfram|Alpha*. Available at: http://www.wolframalpha.com/input/?i=angle+of+view+for+photography [Accessed May 8, 2010].

Woodman, O. & Harle, R., 2008. Pedestrian localisation for indoor environments. In *Proceedings of the 10th international conference on Ubiquitous computing*. UbiComp '08. New York, NY, USA: ACM, pp. 114–123.

Wu, J. & Liu, Z., 2009. General Framework of Photo-Realistic 3D Visualization of Urban Building. In *Image and Graphics, 2009. ICIG '09. Fifth International Conference on*. Image and Graphics, 2009. ICIG '09. Fifth International Conference on. pp. 559–564.

Wulf, O., Arras, K.O., Christensen, H.I. & Wagner, B., 2004. 2D mapping of cluttered indoor environments by means of 3D perception. In *Robotics and Automation, 2004.*

Proceedings. ICRA '04. 2004 IEEE International Conference on. Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on. pp. 4204–4209 Vol.4.

Xing, J. & Miao, Z., 2007. An Improved Algorithm on Image Stitching Based on SIFT features. In *Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on*. Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on. p. 453.

Xiong, Y. & Pulli, K., 2010a. Fast image stitching and editing for panorama painting on mobile phones. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. pp. 47–52.

Xiong, Y. & Pulli, K., 2010b. Fast panorama stitching for high-quality panoramic images on mobile phones. *Consumer Electronics, IEEE Transactions on*, 56(2), p.298–306.

Xsens, 2011. Xsens MVN : Inertial Motion Capture - Xsens. Available at: http://www.xsens.com/en/general/mvn [Accessed September 22, 2011].

Xu, Z., Liu, J., Xiang, Z. & Li, H., 2002. Map building for indoor environment with laser range scanner. In *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*. Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on. pp. 136–140.

Yahoo! Inc, 2011. Welcome to Flickr - Photo Sharing. Available at: http://www.flickr.com/ [Accessed January 25, 2011].

Yang Sa, 2009. Application of Multi-Wavelet Transform in Multi-Focus Image Fusion. In *Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on*. Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on. pp. 560–563.

Ye, P., Zhai, C., Du, G. & Zhan, X., 2009. Experiment evaluation of rapid error compensation for magnetic compass in underwater vehicle. In *International Conference on Mechatronics and Automation, 2009. ICMA 2009*. International Conference on Mechatronics and Automation, 2009. ICMA 2009. IEEE, pp. 4880–4884.

Yeh, L.-W., Hsu, M.-S., Lee, Y.-F. & Tseng, Y.-C., 2009. Indoor localization: Automatically constructing today's radio map by iRobot and RFIDs. In *Sensors, 2009 IEEE*. Sensors, 2009 IEEE. pp. 1463–1466.

Yim, J., Jeong, S., Joo, J. & Park, C., 2008. Utilizing Map Information for WLAN-Based Kalman Filter Indoor Tracking. In *Future Generation Communication and Networking Symposia, 2008. FGCNS '08. Second International Conference on*. Future Generation Communication and Networking Symposia, 2008. FGCNS '08. Second International Conference on. pp. 58–62.

Yuan, J., Jiebo Luo & Wu, Y., 2010. Mining Compositional Features From GPS and Visual Cues for Event Recognition in Photo Collections. *Multimedia, IEEE Transactions on*, 12(7), p.705–716.

Zafar, I., Edirisinghe, E.A. & Bez, H.E., 2006. Multi-Exposure & Multi-Focus Image Fusion in Transform Domain. In *Visual Information Engineering, 2006. VIE 2006. IET International Conference on*. Visual Information Engineering, 2006. VIE 2006. IET International Conference on. pp. 606–611.

Zelnik-Manor, L. & Perona, P., 2007. Automating joiners. In *Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*. San Diego, California: ACM, pp. 121–131.

Zhang, F. & Xu, Y., 2009. Image quality evaluation based on human visual perception. In *2009 Chinese Control and Decision Conference*. 2009 Chinese Control and Decision Conference (CCDC). Guilin, China, pp. 1487–1490.

Zheng, Y.-T., Zha, Z.-J. & Chua, T.-S., 2010. Research and applications on georeferenced multimedia: a survey. *Multimedia Tools and Applications*, 51(1), p.77–98.

Zhu, Z. et al., 2007. Ten-fold Improvement in Visual Odometry Using Landmark Matching. In *IEEE 11th International Conference on Computer Vision, 2007 ICCV 2007 : [Rio de Janeiro, Brazil], October 14-21, 2007*. Los Alamitos, Calif.: IEEE Computer Society, pp. 1–8.

Zhuang, P., Wang, D. & Shang, Y., 2010. SMART: Simultaneous indoor localization and map construction using smartphones. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*. Neural Networks (IJCNN), The 2010 International Joint Conference on. pp. 1–8.

Zhuang, Y., Hu, X. & Wang, J., 2009. The Implement of an Image Stitching Algorithm Based on Feature Extraction. In *Education Technology and Training, 2009. ETT '09. Second International Conference on*. Education Technology and Training, 2009. ETT '09. Second International Conference on. pp. 327–330.

Zomet, A., Levin, A., Peleg, S. & Weiss, Y., 2006. Seamless image stitching by minimizing false edges. *Image Processing, IEEE Transactions on*, 15(4), p.969–977.

Zou, M. & Liu, Y., 2009. Multi-Sensor Image Fusion: Difficulties and Key Techniques. In *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*. Image and Signal Processing, 2009. CISP '09. 2nd International Congress on. pp. 1–5.

10. APPENDICES

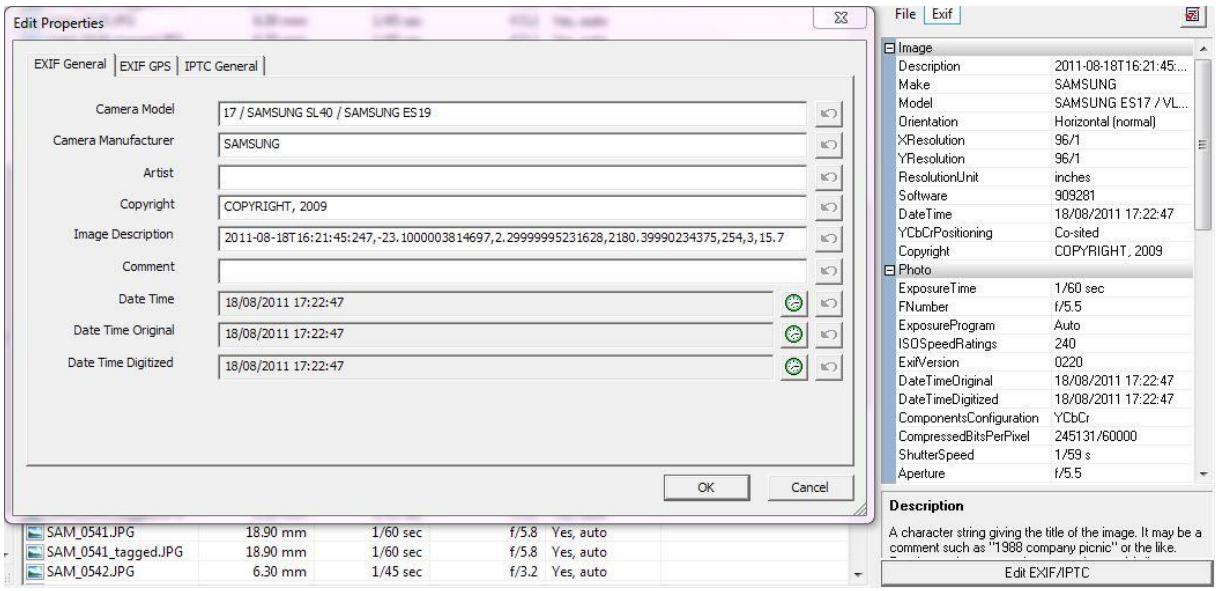
APPENDIX A: EXAMPLE OF A SENSOR COMPOSITING RECORDING

The following are the raw contents of the sensor recording file generated by the sensor-based image compositing module described in 5. File contents are presented in full. Columns are: timestamp, sonar range in inches, compass heading in degrees, pitch in degrees, roll in degrees, category ID.

```
1/1/2008 1:42:05 AM,38,293.5,20.399999618530273,3.5,1
1/1/2008 1:42:18 AM,39,293.39999389648438,0.30000001192092896,2.2000000476837158,1
1/1/2008 1:42:26 AM,63,295.29998779296875,-11.699999809265137,1,1
1/1/2008 1:42:31 AM,63,292.39999389648438,-23.600000381469727,-0.60000002384185791,1
1/1/2008 1:42:47 AM,71,333.29998779296875,25.799999237060547,3,1
1/1/2008 1:42:53 AM,70,317.70001220703125,14.699999809265137,0.5,1
1/1/2008 1:43:08 AM,65,320.10000610351562,-12.100000381469727,-1.3999999761581421,1
1/1/2008 1:43:13 AM,65,307.29998779296875,-25.299999237060547,0.5,1
1/1/2008 1:43:24 AM,116,344.70001220703125,22.899999618530273,0.30000001192092896,1
1/1/2008 1:43:31 AM,70,348.70001220703125,6.8000001907348633,0.69999998807907104,1
1/1/2008 1:43:36 AM,149,338.5,-11.699999809265137,2.5,1
1/1/2008 1:43:41 AM,56,339.60000610351562,-20.200000762939453,-0.89999997615814209,1
1/1/2008 1:43:49 AM,75,4.9000000953674316,17.700000762939453,3.5999999046325684,1
1/1/2008 1:43:56 AM,74,6.8000001907348633,7.5999999046325684,0.5,1
1/1/2008 1:44:02 AM,49,357.10000610351562,-4.4000000953674316,1.6000000238418579,1
1/1/2008 1:44:10 AM,38,4.6999998092651367,-27.200000762939453,-1,1
```

APPENDIX B: EXAMPLE OF AN EXIF TAG MODIFIED FOR COMPOSITING

The following image contains an example of an EXIF metadata tag of a photograph generated after sensor data has been merged into it. The description field has been modified to include sensor data. Screenshot is from Exif Pilot (Two Pilots©, 2011) application.



The relevant tag data is:

```
2011-08-18T16:21:45:247,-23.1000003814697,2.29999995231628,2180.39990234375,254,3,15.7
```

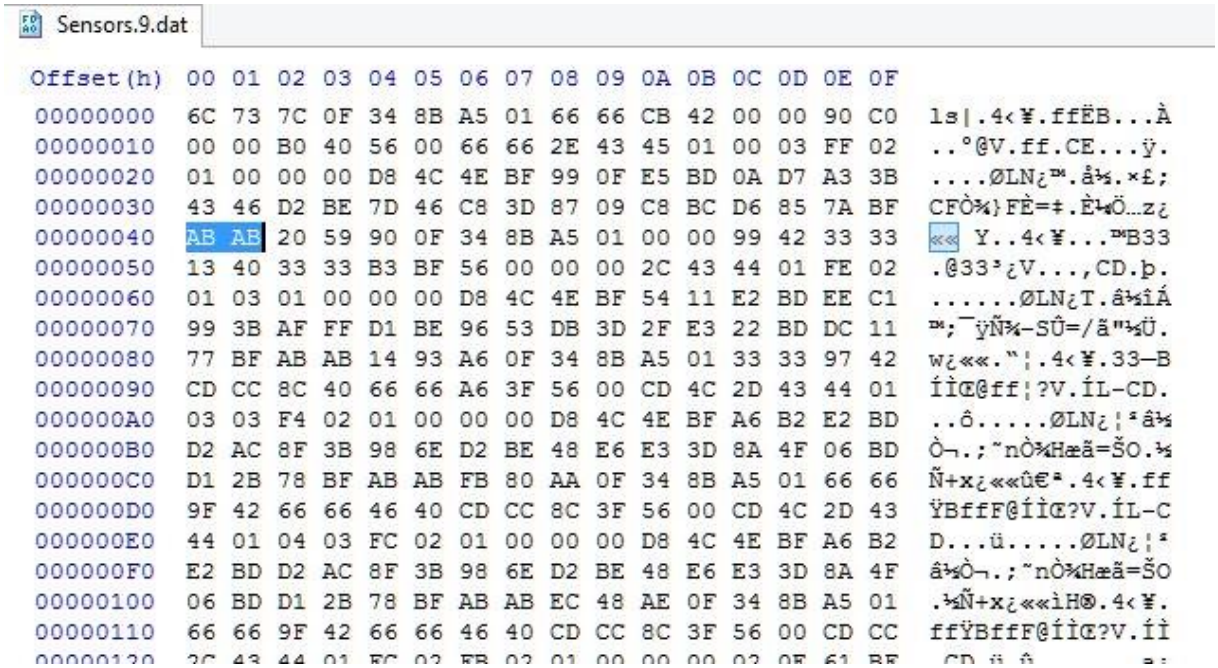
APPENDIX C: SENSOR COMPOSITING EVALUATION WORKSHEET

Initials _____
Test ID _____

1.	☐ Left	☐ Right	19.	☐ Left	☐ Right
2.	☐ Left	☐ Right	20.	☐ Left	☐ Right
3.	☐ Left	☐ Right	21.	☐ Left	☐ Right
4.	☐ Left	☐ Right	22.	☐ Left	☐ Right
5.	☐ Left	☐ Right	23.	☐ Left	☐ Right
6.	☐ Left	☐ Right	24.	☐ Left	☐ Right
7.	☐ Left	☐ Right	25.	☐ Left	☐ Right
8.	☐ Left	☐ Right	26.	☐ Left	☐ Right
9.	☐ Left	☐ Right	27.	☐ Left	☐ Right
10.	☐ Left	☐ Right			
11.	☐ Left	☐ Right			
12.	☐ Left	☐ Right			
13.	☐ Left	☐ Right			
14.	☐ Left	☐ Right			
15.	☐ Left	☐ Right			
16.	☐ Left	☐ Right			
17.	☐ Left	☐ Right			
18.	☐ Left	☐ Right			

APPENDIX D: EXAMPLE OF FINAL SYSTEM RECORDING

Final system records in binary format. An example of binary data is as follows, with end of record marker highlighted:



When decoded into CSV format for convenience and interoperability with compositing subsystem reliant on CSV data, this then becomes (columns are Timestamp (ms), Camera Yaw, Camera Pitch, Camera Roll, Sonar Range, Direction Compass Heading, Direction Gyro Heading, Left Leg Gyro, Right Leg Gyro, Category ID, Camera Snapshot, Direction Gyro Rate, Magnetometer X, Magnetometer Y, Magnetometer Z, Accelerometer X, Accelerometer Y, Accelerometer Z):

```
0,101.7,-4.5,5.5,86,174.4,35,768,767,1,0,-0.80586,-0.1118462,0.005,-
0.4106923,0.0977907,-0.02441861,-0.9786047

130,76.5,2.3,-1.4,86,172,36,766,769,1,0,-0.80586,-0.1103846,0.004692308,-
0.4101538,0.107093,-0.03976744,-0.9651163

276,75.6,4.4,1.3,86,173.3,36,771,756,1,0,-0.80586,-0.1106923,0.004384615,-
0.411,0.1112791,-0.0327907,-0.9694186

301,79.7,3.1,1.1,86,173.3,36,772,764,1,0,-0.80586,-0.1106923,0.004384615,-
0.411,0.1112791,-0.0327907,-0.9694186

326,79.7,3.1,1.1,86,172.8,36,764,763,1,0,-0.87912,-0.1106923,0.004384615,-
0.411,0.1101163,-0.03023256,-0.9683721

352,79.7,3.1,1.1,86,172.8,36,767,768,1,0,-0.87912,-0.1094615,0.006615384,-
0.4106154,0.1101163,-0.03023256,-0.9683721
```

377,79.7,3.1,1.1,86,172.8,36,768,771,1,0,-0.87912,-0.1094615,0.006615384,-0.4106154,0.1101163,-0.03023256,-0.9683721

401,85.2,2.3,0.9,86,172.8,36,769,762,1,0,-0.87912,-0.1094615,0.006615384,-0.4106154,0.1101163,-0.03023256,-0.9683721

427,85.2,2.3,0.9,86,173,36,768,768,1,0,-0.87912,-0.1094615,0.006615384,-0.4106154,0.1039535,-0.02965116,-0.9726744

452,85.2,2.3,0.9,86,173,36,768,770,1,0,-0.87912,-0.1102308,0.006153846,-0.4095384,0.1039535,-0.02965116,-0.9726744

478,85.2,2.3,0.9,86,173,36,764,769,1,0,-0.87912,-0.1102308,0.006153846,-0.4095384,0.1039535,-0.02965116,-0.9726744

504,86.1,1,1,86,173,36,766,763,1,0,-0.95238,-0.1102308,0.006153846,-0.4095384,0.1039535,-0.02965116,-0.9726744

529,86.1,1,1,86,174.3,36,769,761,1,0,-0.95238,-0.1102308,0.006153846,-0.4095384,0.1101163,-0.02325581,-0.9781395

554,86.1,1,1,86,174.3,36,768,765,1,0,-0.95238,-0.1105385,0.005692308,-0.4098462,0.1101163,-0.02325581,-0.9781395

579,86.1,1,1,86,174.3,36,769,765,1,0,-0.95238,-0.1105385,0.005692308,-0.4098462,0.1101163,-0.02325581,-0.9781395

605,85.9,1.5,2,86,174.3,36,769,765,1,0,-0.95238,-0.1105385,0.005692308,-0.4098462,0.1101163,-0.02325581,-0.9781395

631,85.9,1.5,2,86,173.9,36,760,764,1,0,-0.95238,-0.1096154,0.007384615,-0.4114615,0.1146512,-0.0227907,-1.025

655,85.9,1.5,2,86,173.9,36,761,766,1,0,-0.95238,-0.1096154,0.007384615,-0.4114615,0.1146512,-0.0227907,-1.025

681,85.9,1.5,2,86,173.9,36,768,767,1,0,-0.95238,-0.1096154,0.007384615,-0.4114615,0.1146512,-0.0227907,-1.025

706,91.9,-0.2,0.4,86,173.9,36,762,762,1,0,-0.95238,-0.1096154,0.007384615,-0.4114615,0.1146512,-0.0227907,-1.025

731,91.9,-0.2,0.4,86,173.9,36,763,771,1,0,-0.95238,-0.1085385,0.006769231,-0.4106154,0.1073256,-0.02232558,-0.9717442

756,91.9,-0.2,0.4,86,173.9,36,767,767,1,0,-0.95238,-0.1085385,0.006769231,-0.4106154,0.1073256,-0.02232558,-0.9717442

782,80.6,3.2,1.7,86,173.9,36,764,769,1,0,-0.95238,-0.1085385,0.006769231,-0.4106154,0.1073256,-0.02232558,-0.9717442

807,80.6,3.2,1.7,86,173.9,36,765,766,1,0,-0.95238,-0.1085385,0.006769231,-0.4106154,0.1073256,-0.02232558,-0.9717442