ABSTRACT

There is a strong need for an accurate pedestrian navigation system, functional also in GNSS challenging environments, for improved safety and to enhance everyday life. Pedestrian navigation is mainly needed in urban areas and indoors, environments that are challenging for GNSS but also for other RF positioning systems and some non-RF systems such as the magnetometry used for heading due to the presence of ferrous material. Indoor and urban navigation has been an active research area for years. There is no individual system at this time that can address all needs set for pedestrian navigation in these environments, but a fused solution of different sensors can provide better accuracy, availability and continuity. Self-contained sensors, namely digital compasses for measuring heading, gyroscopes for heading changes and accelerometers for the user speed, constitute a good option for pedestrian navigation. However, their performance suffers from noise and biases that result in large position errors increasing with time. Such errors can however be mitigated using information about the user motion obtained from consecutive images taken by a camera carried by the user, provided that its position and orientation with respect to the users body are known. The motion of the features in the images may then be transformed into information about the user's motion. Due to its distinctive characteristics, this vision-aiding complements other positioning technologies in order to provide better pedestrian navigation accuracy and reliability.

This thesis discusses the concepts of a visual gyroscope that provide the relative user heading and a visual odometer that provides the translation of the user between the consecutive images. Both methods use a monocular camera carried by the user. The visual gyroscope monitors the motion of virtual features, called vanishing points, arising from parallel straight lines in the scene, and from the change of their location that resolves heading, roll and pitch. The method is applicable to the human environments as the straight lines in the structures enable the vanishing point perception. For the visual odometer, the ambiguous scale arising when using the homography

between consecutive images to observe the translation is solved using two different methods. First, the scale is computed using a special configuration intended for indoors and secondly, the scale is resolved using differenced GNSS carrier phase measurements of the camera in a method aimed for urban environments. Both methods are sufficiently tolerant for the challenges of visual perception in indoor and urban environments, namely low lighting and dynamic objects hindering the view. The heading and translation are further integrated with other positioning systems and a navigation solution is obtained. The performance of the proposed vision-aided navigation was tested in various environments, indoors and urban canyon environments to demonstrate its effectiveness. These experiments, although of limited durations, show that visual processing efficiently complements other positioning technologies in order to provide better pedestrian navigation accuracy and reliability.

TABLE OF CONTENTS

Abstract i

Table of Contents iii

List of Figures vii

List of Tables xi

Abbreviations xiii

Symbols xvii

1. Introduction 1

1.1 Research Objectives 2

1.2 Related Work 3

1.3 Author's Contribution 6

1.4 Thesis Outline 7

2. Overview of pedestrian navigation 9

2.1 Navigation Frames and Attitude 9

2.2 Absolute Positioning 10

2.2.1 Global Navigation Satellite Systems 11

2.2.2 WLAN Positioning 15

2.2.3 Other Technologies 16

2.3 Relative Positioning 17

2.3.1 Inertial Sensors 17

2.3.2 Other Self-Contained Sensors 19

2.4 Estimation 22
2.4.1 Kalman Filter 22
2.4.2 Extended Kalman Filter 24
3. Computer vision methods for navigation 25
3.1 Camera, Fundamental and Essential Matrices and Coordinate Frames 25
3.2 Feature Extraction 28
3.2.1 Filtering 29
3.2.2 SIFT-Features 29
3.2.3 Line Extraction 30
3.3 Image Matching 33
3.4 Camera Calibration 34
3.4.1 Distortion 35
4. Visual gyroscope 37
4.1 Locating the Vanishing Points 38
4.2 Attitude of the Camera 40
4.3 Error Detection 42
4.4 Performance of the Visual Gyroscope 46
4.5 Effect of Camera and Setup Characteristics on the Accuracy of the Visual Gyroscope 50
4.5.1 Experimental Results 53
4.6 Smartphone Application of Visual Gyroscope 57
5. Visual odometer 59
5.1 The Principle of the Visual Odometer 59
5.1.1 Measuring the Distance of an Object from the Camera . . . 60
5.1.2 Error Detection and Ambiguity Resolving for the Visual Odometer 62

5.1.3 Degeneracy 64
5.1.4 Performance of the Visual Odometer 65
6. Vision-aided navigation using visual gyroscope and odometer 67
6.1 Visual Gyroscope and Odometer Aided Multi-Sensor Positioning . . 67
6.1.1 Kalman Filter Used in Multi-Sensor Positioning 68
6.1.2 Test in an Indoor Office Environment 71
6.1.3 Test in Office Environment Using an Outdated WLAN Radio Map 74
6.2 Stand-Alone Visual System 76
6.2.1 Kalman Filter Used in Stand-Alone Visual Positioning . . . 76
6.2.2 Test in a Shopping Mall Environment 77
6.2.3 Test in an Urban Canyon 78
6.3 Visual Gyroscope Aided IMU Positioning 80
6.3.1 Kalman Filter Used in Visual Gyroscope Aided IMU Positioning 83
6.3.2 Equipment Setup on the Body 84
6.3.3 Equipment Setup on the Foot 89
6.4 Visual Gyroscope Implementation Using Probabilistic Hough Transform 91
7. Vision-aided carrier phase navigation 101
7.1 Ambiguity Resolution Using Differenced GNSS Carrier Phase Measurements 102
7.1.1 Ambiguous Translation Using the Fundamental Matrix . . . 104
7.1.2 Navigation Solution Incorporating the Absolute User Translation 105
7.2 Method Verification in a Sub-Urban Environment 107
7.3 Vision-Aided GNSS Navigation in an Urban Environment 109

8. Conclusions 117

8.1 Main Results 118

8.2 Future Development 121

Bibliography 123

LIST OF FIGURES

2.1 User attitude in navigation frame 10

2.2 Absolute heading error of a digital compass indoors 21

3.1 Coordinate frames in vision-aiding 27

3.2 Epipolar geometry 28

3.3 Hough transform parameters 33

4.1 Vanishing point 39

4.2 Vanishing point in an image with roll 40

4.3 Vanishing point error detection 44

4.4 Allan deviation of the visual gyroscope 48

4.5 Visual gyroscope‘s tolerance on dynamic objects 49

4.6 An image captured of the same scene with three different cameras . 52

4.7 Experiment setup for testing camera characteristics 53

5.1 Visual odometer configuration 60

5.2 Matched SIFT features between consecutive images 62

6.1 Equipment setup for testing the vision-aided multi-sensor positioning system 69

6.2 Office corridor used for experiments 72

6.3 Visual odometer speed 73

6.4 Vision-aided position solution in an office corridor 74

6.5 Vision-aided position solution in an office corridor with an outdated WLAN radio map 75

6.6 Challenging environment of Iso Omena shopping centre 78
6.7 The two-dimensional position solution in the Iso Omena shopping centre 79
6.8 Challenging environment of an urban canyon 80
6.9 Position solution in an urban canyon 81
6.10 Route for experiments on the University of Calgary campus 85
6.11 Body mounted test equipment 86
6.12 Standard deviation for different integration schemes 87
6.13 Attitude error using different integration schemes 88
6.14 Equipment setup for the foot 90
6.15 Images from one step cycle period 91
6.16 RMS position errors obtained for foot-mounted IMU 92
6.17 Line detection and vanishing point calculations using Probabilistic Hough Transform 97
6.18 Evaluation of vanishing point detection in an environment suffering from low lighting and non-orthogonal lines 98
6.19 Correcting IMU errors using a vanishing point obtained using Probabilistic Hough Transform 98
6.20 Conflict between estimated and detected vanishing 99
7.1 Setup for vision-aided carrier phase navigation 108
7.2 Position solution verification in a sub-urban environment shown in Google Earth 109
7.3 Calgary downtown 110
7.4 Number of satellites acquired in an urban canyon 111
7.5 Position solution in an urban canyon 112
7.6 Horizontal position errors in an urban canyon 113
7.7 Position solution in an urban canyon shown in Google Earth 114

7.8 Position solution using GPS only in an urban canyon shown in Google Earth 114

LIST OF TABLES

4.1 Statistics for heading change accuracy, all units degrees 47

4.2 Effect of roll error on other angle observations 50

4.3 Parameters for GoPro, Sony and Nokia cameras 52

4.4 Heading change error statistics 55

4.5 Roll and pitch error statistics 55

4.6 Processing time for different algorithms in visual gyroscope's smartphone implementation 58

5.1 Statistics of the effect of camera height for visual odometer's speed accuracy, units are in m/s 66

6.1 Positioning error statistics using different systems in an office corridor 73

6.2 Positioning error statistics using different positioning systems in an office corridor with an outdated WLAN radio map 75

6.3 Positioning error statistics for visual stand-alone and GPS position solutions 80

6.4 Attitude errors obtained for body-mounted IMU with different integration methods 87

6.5 RMS position error obtained for foot-mounted IMU with and without vision-aiding 91

6.6 Ratio of image points used for computing the Probabilistic Hough Transform presented to the image points used by Standard Hough Transform for the images processed in the experiment 97

7.1 Positioning verification error statistics using vision-aided carrier phase (VA) 108

7.2 Positioning error statistics using vision-aided carrier phase (VA) and GPS only (GPS) 113

ABBREVIATIONS

AVUPT	Absolute Visual attitude Update
BLUE	Best Linear Unbiased Estimate
C/A	Coarse/Acquisition
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
COMPASS/Beidou	Chinese Satellite Navigation System
DCM	Direction Cosine Matrix
DOP	Dilution Of Precision
E	East
ECEF	Earth Centered Earth Fixed
EKF	Extended Kalman Filter
ENU	East-North-Up
EXIF	Exchangeable Image File
Galileo	European Satellite Navigation System
GDOP	Geometric Dilution Of Precision
GLONASS	The Russian Positioning System, Global'naya Navigatsionnaya Sputknikkovaya Sistema
GNSS	Global Navigation Satellite System

GPS	Global Positioning System
HD	High-definition
HSGPS	High Sensitivity GPS
IEEE	The Institute of Electrical and Electronics Engineers
ION	Institute of Navigation
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
KF	Kalman filter
LCI	Low-coherence Interferometry
LDOP	Line Dilution Of Precision
LOS	Line Of Sight
Max	Maximum
MEMS	Micro-Electro-Mechanical
Min	Minimum
MSP	Multi Sensor Positioning
N	North
PGCP	Pseudo Ground Control Points
PDOP	Position Dilution Of Precision
PPP	Precise Point Positioning
RANSAC	RANdom SAmple Consensus

RF	Radio Frequency
RFID	Radio Frequency Identification
rms	root mean square
RSSI	Received Signal Strength Indication
SHT	Standard Hough Tranform
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous Localization And Mapping
SPAN	Synchronized Position Attitude Navigation
SVD	Singular Value Decomposition
std	standard deviation
ToA	Time of Arrival
TVUPT	Temporal Visual Attitude Update
U	Up
UAV	Unmanned Aerial Vehicle
UKF	Unscented Kalman filter
UTC	Coordinated Universal Time
UERE	User Equivalent Range Error
UWB	Ultra-Wideband
VA	Vision-aided
WiFi	Wireless network, a registered trademark of the Wi-Fi Alliance
WLAN	Wireless Local Area Network

SYMBOLS

α_i	Angle between a line i in an image and the image x-axis
β	roll
Δt	Time interval
$\Delta \mathbf{x}$	Vector offset of the user's true position and time bias from the values at the linearization point
$\delta \mathbf{x}_k$	Perturbation of the state
ϵ^-	Error of *a priori* state estimate or perturbation of the Euler angles
ϵ	Error of *a posteriori* state estimate or noise in GPS measurements or vector of errors in GNSS measurements
η_g	Noise in gyroscope or carrier phase measurement
λ	Carrier wavelength or longitude
μ	Mean
∇	Image gradient
ω	Earth turn rate
ω_{ib}^b	Body (b) turn rate with respect to the inertial (i) frame angular velocity measurement
$\tilde{\omega}_{ib}^b$	Gyroscope angular velocity measurement

$\mathbf{\Omega}$	Skew symmetrical matrix of the angular velocity vector
ϕ	pitch or latitude
$\mathbf{\Phi}$	State transition matrix
ρ	Pseudorange or the radius of a line in an image in Hough Transform
$\hat{\rho}$	Estimated pseudorange computed from the estimated user position
σ	Standard deviation
σ^2	Variance
$\sigma_C^2(t_A)$	Allan variance
θ	Heading, (azimuth)
φ	Carrier phase
b	body frame
c	Speed of light
$\mathbf{C}$	Direction cosine matrix or Convolution
$\mathbf{d}$	direction of a line in an image
d_{iono}	Ionospheric delay
d_{tropo}	Tropospheric delay
$d\rho$	Ephemeris error
dt	Satellite clock error
D_i	Distance between the starting point of line i and the vanishing point
$\mathbf{E}$	Essential matrix
f	Focal legth
$\mathbf{f}$	Specific force

$\mathbf{F}$	Fundamental matrix
$\mathbf{g}$	Mass gravitation
$\mathbf{G}$	User-satellite geometry matrix or Convolution kernel or g-sensitivity coefficient matrix
h	Height
H	Height of an image in pixels
$\mathbf{H}$	Design matrix or image homography
i	inertial frame
$\mathbf{I}$	Image matrix
k	Distortion value
$\mathbf{K}$	Kalman gain or camera calibration matrix
L1	GPS signal carrier frequency at 1575.42 MHz
$\mathbf{M}$	Image gradient magnitude matrix
N	Gaussian probability distribution or integer number of carrier waves
N^e	Inertia tensor
$\mathbf{O}$	Image gradient orientation matrix
p	pressure
$\mathbf{P}$	State error covariance or camera matrix
$\tilde{q}$	Spectral density value
$\mathbf{Q}$	Process noise covariance
r	Geometric range
r_d	Radial distance of the normalized distorted image point
$\mathbf{r}$	User position vector or Least-squares residual vector

$\mathbf{R}$	Measurement noise covariance or camera rotation matrix
R_g	Universal gas constant
$\mathbf{R}_{WLAN}$	RSSI observation vector
s	Ambiguous scale in translation observed from consecutive images
$\mathbf{s}$	Satellite coordinate vector
S	User speed
$\mathbf{S}$	Scale factor and non-orthogonality matrix
t	time
t_u	Receiver clock error
$\mathbf{t}$	User translation vector
$\mathbf{T}^i$	Satellite i's position vector
$\mathbf{T}_{rcvr}$	Receiver position vector
T_0	Temperature at the sea level
T_L	Temperature lapse rate
u	Principal point's x-coordinate
$\mathbf{u}$	User coordinate vector or the unit vector from user to satellite
u_{GC}	Satellite and user geometry change
v	Principal point's y-coordinate
$\mathbf{v}$	Kalman filter's innovation vector or user velocity vector or a vanishing point matrix
v_k	Process noise

$vfov$	Vertical field-of-view of a camera
$\mathbf{v}_x$	Vanishing point in x-axis direction in homogenous coordinates
$\mathbf{v}_y$	Vanishing point in y-axis direction in homogenous coordinates
$\mathbf{v}_z$	Vanishing point in z-axis direction in homogenous coordinates
w_i	Standardized innovation of the ith element of the innovation vector
w_k	Measurement noise
$\mathbf{x}_k$	State vector
$\hat{\mathbf{x}}_k^-$	*a priori* state estimate
$\hat{\mathbf{x}}_k$	*a posteriori* state estimate
$\overline{\mathbf{x}}_k$	Nominal value of the state
x_u	User (receiver) x-coordinate
$\mathbf{x}$	Feature coordinates in the image reference frame
$\mathbf{X}$	Object coordinates in the world reference frame or user position East component
$\dot{\mathbf{X}}$	Time derivative of $\mathbf{X}$
y_u	User (receiver) y-coordinate
$\tilde{y}(t_A)_k$	Average value of bin k in Allan variance
$\mathbf{Y}$	User position North component
$\dot{\mathbf{Y}}$	Time derivative of $\mathbf{Y}$
$\mathbf{z}_k$	Measurement vector

z_u	User (receiver) z-coordinate
Z	Depth of an object i.e. the Z-coordinate in the world reference frame

1. INTRODUCTION

In addition to commercial solutions, such as directing the user flexibly to the destination desired, pedestrian navigation is crucial in critical applications such as positioning of first responders, electronic monitoring, and military personnel. The equipment used for pedestrian navigation has to be small and light to carry, effortless to use as well as have reasonable low levels of power consumption and price. Like in all navigation systems the position information has to be accurate and available in real time. At present Global Navigation Satellite Systems (GNSS) are the superior navigation technology fulfilling all the above requirements in outdoor open-sky environments. However, instruments for pedestrian navigation are mainly needed indoors and in urban areas, where GNSS is significantly degraded or unavailable.

In these challenged GNSS environments the absolute position of the user may be obtained with other radio navigation systems like Wireless Local Area Networks (WLAN), Bluetooth, or Radio Frequency Identification (RFID). The drawbacks of these radio systems are that they need a priori prepared infrastructure and are therefore restricted to certain areas. They also have, in some environments, too low availability for the needs of pedestrian navigation, depending on the number of access points available. When the initial absolute position is known, the position of the user may be propagated using relative positioning approaches, like self-contained sensors. The propagated position may then be used to augment the position measurements obtained with GNSS or other radio sensors for more accurate and available, or even short-time stand-alone navigation.

The most commonly used self-contained sensors in pedestrian navigation are digital compasses for measuring the heading of the user, gyroscopes for heading changes, and accelerometers for the user speed. When these measurements are used as inputs to Pedestrian Dead Reckoning (PDR) algorithms or integrated with absolute position measurements using a Kalman filter, the position of the user is obtained continuously

despite the degradation of the GNSS signals. However, self-contained sensors suffer from biases and drift errors that may decrease the position accuracy substantially, especially when consumer grade Micro-Electro-Mechanical (MEMS) sensors are used.

The errors in a pedestrian indoor position solution experienced due to the above shortcomings of self-contained sensors may be mitigated using information about the user motion obtained from consecutive images. When the user is carrying a camera whose position and orientation with respect to the user's body are known, the motion of the features in the observed images may be transformed into information about the user motion. The visual motion information is not affected by the same error sources as GNSS and self-contained sensors, and is therefore a complementary information source for augmenting the positioning measurements. Visual-aiding increases the accuracy, availability, continuity and reliability of the navigation solution.

1.1 Research Objectives

The use of visual information in navigation is challenging. Motion of the features in consecutive images provides information about the change of the user's heading and translation during the time interval between two consecutive images. In order to convert this relative information into the absolute position information needed for navigation, the position and heading have to be initialized with known absolute values and then propagated using the relative measurements. The propagated position starts to drift after a while due to errors affecting the visual measurements and therefore absolute information is needed to re-initialize the trajectory from time to time.

The main error sources for the visual-aiding observable in indoor surroundings are the varying lighting conditions of the environment and the low amount of distinctive features to be detected. Urban outdoor areas do not usually suffer from low lighting during daytime. Outdoor surroundings are rich in features, but also in dynamic objects, namely humans and vehicles. When the motion of the camera, and therefore of the user, is observed by following the motion of the features in the consecutive images, the image processing method has to be able to exclude the dynamic objects in the scene from disturbing the perception of the motion.

In this research the user heading, as well as the pitch and roll of the camera, are observed by tracking the motion of vanishing points, namely features arising from pro-

jective transformations that map the three-dimensional objects into two-dimensional image points. The deficiency of the method is that it is strongly dependent of the geometry of the environment, as it requires straight parallel lines in the view of the camera, preferably in at least two orthogonal directions. In sharp turns this requirement is violated and the magnitude of the turn is often impossible to be defined using image processing alone.

Resolving translation from consecutive images using only a monocamera is a challenging research task as well. The complication in observing translation from consecutive images is that the distance between the objects seen in images from the camera contributes to the extent the image pixels move when the camera moves. When the depth is unknown the scale of translation stays unknown regardless of how many matching image points are found between the consecutive images.

The objectives of this research is to provide methods to retrieve heading and translation information using consecutive images addressing the above mentioned challenges and also to enable accurate, more reliable and available navigation solutions by augmenting other positioning systems with the information obtained. All calculations are of a sufficiently low complexity to be adopted in real time for navigation with current smartphones. The algorithms of the concept called ”visual gyroscope” providing the user heading are already developed for the Nokia Symbian environment and the feasibility of the implemented system is herewith discussed.

1.2 Related Work

The research related to visual positioning has so far mainly concentrated on navigation of vehicles and mobile robots [**?**] [**?**]. The motion of a robot or vehicle is constrained and usually only two-dimensional. The visual calculations are easier due to the fact that the location and the path of motion are to some extent known in advance. The first papers related to vision-aiding in pedestrian navigation were published in the late 90s. They used earlier prepared databases with images taken of the surroundings tagged with position information obtained using Global Positioning System (GPS), a map or a floor plan. The absolute position of the pedestrian was provided when a match was found between an image taken by the pedestrian and an image from the database [**?**]. One of the first such applications made for a smartphone was published in 2004 by Robertson and Cipolla [**?**] running the calculations on a server to which

the query image was sent. Hile and Borriello [**?**] matched features, like corners, found from the query image, into a floor plan saved in a server. The feature matching was restricted to a certain area of the floor plan using coarse position information obtained with WLAN. The database based vision-aiding applications provide accurate positioning but are restricted to a certain area and require extensive preparation.

A visual pedestrian navigation system independent of a server and of pre-existing databases needs usually integration with other positioning sensors to be functional. In such a system the relative position of the user is obtained by monitoring the motion of features in consecutive images taken by the user device and integrating the information with measurements obtained with self-contained sensors or GNSS receiver. With initial absolute position information the navigation may be performed with reduced drift and other errors. Such server independent systems have been proposed by [**?**] using visual-aided Inertial Measurement Unit (IMU) measurements. On the other hand, a Simultaneous Localization And Mapping (SLAM) system produces a map of the unknown environment while simultaneously locating the user. Traditionally the mapping has been done using inertial sensors but in recent years visual SLAM systems integrating also a camera has been developed [**?**].

Most human made environments, in indoors and urban outdoor areas consist of segments forming a Cartesian coordinate system with straight lines in three orthogonal directions. The coordinate system is called the Manhattan grid [**?**] and it provides a good basis for vision-aided navigation utilizing vanishing points. The method of integrating vanishing point based orientation information with Inertial Navigation System (INS) measurements has been implemented before for accurate indoor navigation of an unmanned aerial vehicle (UAV) [**?**] [**?**] [**?**] and for pedestrian navigation [**?**] [**?**]. The method presented in this thesis follows the mentioned vanishing points based methods but is further developed for pedestrian and especially smartphone use by developing computationally less demanding algorithms and sophisticated error detection.

The ambiguous scale in translation obtained from tracking the motion of features in consecutive images is one of the most challenging issues related to visual navigation. The magnitude of the motion of the figure in an image is dependent on the depth of the object, i.e. the z-coordinate of the distance of the object from the camera. Because the distance of the objects from the camera in the navigation environment is usually unknown, a scale problem arises and different methods for overcoming it have been

used. When the environment contains objects with known sizes, the distance may be resolved [**?**]. Also, when scale information about the environment is available, for example in the form of a floor plan [**?**], the depth of the objects may be observed. Tools aiding the resolving of the distance, like laser rangefinders, have been integrated with a camera by [**?**] and the motion of the user resolved. The requirement for special equipment reduces the applicability of the methods for pedestrian navigation at this time. When a stereo camera is used, the distance of the objects may be resolved using triangulation [**?**]. Recently some smartphones equipped with stereo cameras have been launched. In the case of stereo vision the distance between the two cameras, called the baseline, affects the accuracy of the motion obtained from images. The farther the two cameras are from each other the better the accuracy and this is due to the limitations of the resolution obtained [**?**]. Therefore a configuration using a monocular camera and images taken from two different positions provides better results for vision-aided navigation than a smartphone equipped with a stereo camera due to its very short baseline.

Certain configuration of the navigation system gives information about the distance of the objects being photographed from the camera. When the camera is pointing down to the ground the z-coordinate of the distance is constant and equals the height of the camera. The method utilizing the downward-pointing camera has been used in the applications of vehicle navigation [**?**] [**?**] and recently in pedestrian navigation [**?**]. However, one of the challenges of the visual-aiding in indoor environments is the shortage of features to be tracked. Especially floor textures are usually very homogenous and for that it is very difficult to find matching image points using a camera pointing straight to the floor. [**?**] developed an outdoor robot navigation system using a special camera configuration, namely the camera had a certain pitch towards the ground, to resolve the distance problem. Optical flow calculations for finding the camera rotation and translation were used. The method presented in this thesis follows the ideas presented in this method but is further developed for pedestrian and indoor use.

As mentioned above, GNSS is an accurate and freely accessible system for outdoor navigation widely used in smartphones. However, at least four satellites in a good geometry are needed for solving the user position, a requirement that is not always fulfilled in urban areas. When knowledge of an initial position is available, fewer satellites may be used for resolving the total change in position between two time

epochs. When the errors affecting satellite signal propagation are known, information obtained with two satellites is enough to resolve the total magnitude of translation in addition to the receiver clock error. [**?**] used the magnitude information for resolving the ambiguous scale in translation induced by motion of features in consecutive images. A method for robot navigation encompassing three cameras for visual measurements, an IMU for resolving pitch and roll of the camera and an iterative algorithm for solving the user heading was developed. In this thesis, an algorithm more suitable for pedestrian navigation is developed, utilizing less equipment and more robust vision-derived heading information.

1.3 Author's Contribution

In this thesis a novel pedestrian navigation system is presented. Two concepts are developed, namely a "visual gyroscope" providing the user heading and a "visual odometer" providing the translation. Author's contributions include also a system developed for pedestrian urban navigation, utilizing the visual gyroscope, visual odometer and signal carrier information obtained from at least two GNSS satellites.

All calculations are of a sufficiently low complexity to be adopted for navigation with current smartphones. The main contributions of the thesis are as follows:

- A visual gyroscope with lower computational requirements suitable for present smartphones. The visual gyroscope is based on observing heading, pitch and roll of the camera, using vanishing points.
- A novel error detection method which provides accurate and reliable navigation despite the unforeseeable motions of a pedestrian. The algorithm makes the visual gyroscope suitable for pedestrian navigation.
- A visual odometer, namely a method to resolve translation from images using a monocular camera. The visual odometer is suitable to be used also in indoor environments which are usually poor in features. It is feasible for seamless navigation since it leans on the visual gyroscope's orientation information and needs only the approximate height of the camera as prior information.
- A vision-aided differentiated carrier phase navigation system for pedestrians. The method is leaner than previous similar solutions. The system is independ-

ent from other sensors than the camera and the GNSS receiver because it encompasses the visual gyroscope and visual odometer providing the orientation and motion information.

The core contributions of Chapters 4-6 were first presented in [**?**], [**?**], [**?**], [**?**], [**?**], [**?**] and [**?**] in which the author of the thesis is the first author and in [**?**] in which the author of the thesis is a co-author.

1.4 Thesis Outline

In **Chapter 2**, the most prevalent systems used in pedestrian navigation - i.e. GNSS, WLAN and self-contained sensors - are presented. The computer vision principles relevant in vision-aided navigation are discussed in **Chapter 3** with an emphasis on the methods and algorithms used in the thesis. **Chapter 4** introduces the concept of a "visual gyroscope" and the novel error detection algorithm. The feasibility and challenges of the visual gyroscope are discussed as well as the effect of different camera and setup characteristics on the accuracy and applicability of the method in pedestrian navigation. In **Chapter 5** a concept of "visual odometer" is presented. The mathematics, strengths and challenges of the visual odometer and its utilization are discussed. **Chapter 6** presents results from various experiments integrating the visual gyroscope and odometer, both for indoor and urban pedestrian navigation. In **Chapter 7** the vision-aided differentiated carrier phase navigation system for pedestrians, results from experiments and its feasibility for urban pedestrian navigation are discussed. **Chapter 8** provides conclusions and recommendations for future research.

2. OVERVIEW OF PEDESTRIAN NAVIGATION

Global Navigation Satellite Systems (GNSS) are the superior navigation technology used also for pedestrian positioning. However, GNSS is significantly degraded or unavailable in environments where pedestrian positioning is mainly needed, namely in indoors and urban areas and other methods are required for augmentation or replacement in these situations. Methods other than GNSS for these indoor and urban areas may be divided into two classes based on the type of position information they provide, namely into absolute and relative positioning. Robust integration of measurements from sources providing data with different rates and perceiving observations in different reference frames is challenging. This chapter introduces the basics of GNSS based positioning, Wireless Local Area Network (WLAN) positioning and other absolute positioning methods. The relative positioning systems used in this thesis, namely Inertial Navigation System and other self-contained sensors, are also presented. Finally the Kalman filter, a set of mathematical equations used for estimating the state of a process based on a priori knowledge of the accuracy of the measurements and confidence on the model used, is discussed.

2.1 Navigation Frames and Attitude

This thesis uses five reference frames relevant for vision-aided pedestrian navigation, namely Inertial, Earth-Centered Earth-Fixed, Navigation, Body and Camera reference frames. The inertial frame has origin at the centre of the Earth and axes fixed with respect to stars, not rotating with the Earth. Earth-Centered Earth-Fixed reference frame has also origin at the centre of the Earth, but the axes rotate with the Earth with respect to the inertial frame. Both frames have their z-axis coincident with the Earth's polar axis. The navigation frame is a local geographic frame with origin defined by the initialization of the navigation setup and axes pointing at north, east and up. The body frame is a frame where the inertial navigation system is installed,

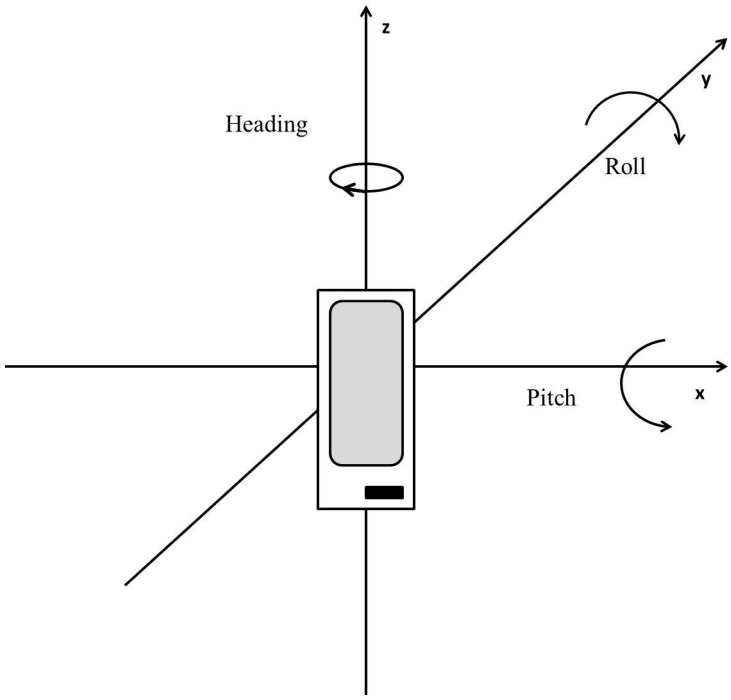

Fig. 2.1. *Heading, pitch and roll in Navigation frame.*

containing three orthogonal axis, z-axis pointing up [**?**]. In vehicle navigation the rotation around z-axis is called yaw, x-axis roll and y-axis pitch. In pedestrian navigation where the orientation of the unit is not always fixed with respect to the user the term yaw is substituted with heading and defined as the angle between the chosen body axis with respect to the Earth-fixed north axis [**?**]. The heading, pitch and roll are shown in Figure 2.1. The definition for the camera reference frame is not needed in this chapter but will be given in Chapter 3.

2.2 Absolute Positioning

Absolute positioning systems provide the actual coordinates of the user position as the relative positioning systems tell the speed (or translation) and direction of the user to be intergraded with the initial position. In reality, only GNSS provides the absolute coordinate information of the pedestrian in the Earth-Centered Earth-Fixed (ECEF) coordinate frame as the other systems provide the absolute position in some

a priori defined local reference frame, for example inside a certain building. All the absolute positioning techniques presented facilitate positioning by transmitting radio waves with different wavelengths and frequencies.

2.2.1 Global Navigation Satellite Systems

GNSS encompass the United States Global Positioning System (GPS), the Russian GLONASS, Chinese COMPASS/Beidou and the European Galileo systems. The following principles will be based on GPS because it is still the most used system due to its long existence compared to other systems mentioned above, but are true for the other systems also. In GNSS based positioning the traverse time of a signal from the satellite to the user receiver antenna is estimated. When this time is multiplied by the speed of light a geometric range between the satellite and the user is obtained. In an ideal case measurements from three satellites would provide an accurate three dimensional position of the user. In reality the measurements are erroneous, the main error source being the timing errors between the receiver clock and the satellite clock from the system time. Therefore the measured range is called the pseudorange. The satellite clocks are precise and synchronized by the ground control segment of the system. However, the clocks in the user receivers are low-cost with typically a large timing error. Therefore, it has to be estimated as a parameter in the navigation solution. Observations from at least four satellites are needed for three dimensional positioning, namely the fourth observation is used for resolving the receiver clock error. The pseudorange measurement is defined as

$$\rho_i = r^i + c(t_u - dt^i) + d\rho^i + d^i_{iono} + d^i_{tropo} + \varepsilon^i_\rho \tag{1}$$

where r^i is the geometric range between the user receiver's antenna and the satellite i $[m]$, c is the speed of light $[m/s]$, t_u is the receiver clock error $[s]$ and dt^i is the satellite clock error $[s]$ with respect to GPS time, $d\rho^i$ is the ephemeris error $[m]$, d^i_{iono} and d^i_{tropo} are the ionospheric and tropospheric delays $[m]$, respectively and ε^i_ρ encompasses noise, unmodelled errors and multipath [**?**]. Because some of the errors may be corrected using the data found in the signal and the rest may be considered negligible compared to the receiver clock error, the pseudorange measurements may be expressed as

$$\rho_i = ||\mathbf{s}_i - \mathbf{u}|| + ct_u \tag{2}$$

where $\mathbf{s}_i$ represents the coordinate vector of satellite i, ct_u is the speed of light (c) times the advance of the receiver clock t_u and $\mathbf{u}$ is the user coordinate vector (x_u, y_u, z_u) to be resolved [**?**]. These pseudorange measurements from at least four satellites may further be used for obtaining the user coordinates. Because the pseudorange equations are non-linear, the user position and clock error have to be linearized by expanding using a Taylor series as

$$\begin{aligned} f(x_u, y_u, z_u, t_u) = f(\hat{x}_u + \Delta x_u, \hat{y}_u + \Delta y_u, \hat{z}_u + \Delta z_u, \hat{t}_u + \Delta t_u) = \\ f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u) + \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{x}_u} \Delta x_u + \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{y}_u} \Delta y_u \\ + \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{z}_u} \Delta z_u + \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_U, \hat{t}_u)}{\delta \hat{t}_u} \Delta t_u + ... \end{aligned} \tag{3}$$

where $(\hat{x}, \hat{y}, \hat{z}, \hat{t})$ are approximated values of the true position and true clock error (x_u, y_u, z_u, t_u) and $(\Delta x_u, \Delta y_u, \Delta z_u, \Delta t_u)$ are the differences between the true and approximated values. The first-order partial derivatives are

$$\begin{aligned} \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{x}_u} &= -\frac{x_i - \hat{x}_u}{\hat{r}_i} \\ \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{y}_u} &= -\frac{y_i - \hat{y}_u}{\hat{r}_i} \\ \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{z}_u} &= -\frac{z_i - \hat{z}_u}{\hat{r}_i} \\ \frac{\delta f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\delta \hat{t}_u} &= c \end{aligned} \tag{4}$$

and the higher order derivatives are discarded to eliminate nonlinear terms. The variables $\hat{r}_i$ for the estimated geometric ranges are defined as

$$\hat{r}_i = \sqrt{(x_i - \hat{x}_u)^2 + (y_i - \hat{y}_u)^2 + (z_i - \hat{z}_u)^2}. \tag{5}$$

The pseudorange measurement may now be presented as

$$\rho_i = \hat{\rho}_i - \frac{x_i - \hat{x}_u}{\hat{r}_i} \Delta x_u - \frac{y_i - \hat{y}_u}{\hat{r}_i} \Delta y_u - \frac{z_i - \hat{z}_u}{\hat{r}_i} \Delta z_u + c\Delta t_u \tag{6}$$

and finally the difference between the measured pseudorange ρ_i and the pseudorange computed using the estimated user position $\hat{\rho}_i$ is

$$\Delta\rho_i = \frac{x_i - \hat{x}_u}{\hat{r}_i} \Delta x_u + \frac{y_i - \hat{y}_u}{\hat{r}_i} \Delta y_u + \frac{z_i - \hat{z}_u}{\hat{r}_i} \Delta z_u - c\Delta t_u. \tag{7}$$

The differences between the approximated and true position and clock error $\Delta\mathbf{x}$ is $\Delta\mathbf{x} = \mathbf{H}^{-1}\Delta\rho$ where the matrices for n measured satellites are

$$\Delta\rho = \begin{bmatrix} \Delta\rho_1 \\ \vdots \\ \Delta\rho_n \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} \frac{x_1-\hat{x}_u}{\hat{r}_1} & \frac{y_1-\hat{y}_u}{\hat{r}_1} & \frac{z_1-\hat{z}_u}{\hat{r}_1} & 1 \\ \vdots & \vdots & \vdots & 1 \\ \frac{x_n-\hat{x}_u}{\hat{r}_n} & \frac{y_n-\hat{y}_u}{\hat{r}_n} & \frac{z_n-\hat{z}_u}{\hat{r}_n} & 1 \end{bmatrix} \quad \Delta\mathbf{x} = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -c\Delta t_u \end{bmatrix} \tag{8}$$

and by using this information the true position and clock error may be computed from the approximated values when four satellites are observed. When more than four satellites are observed, the solution is computed using the least-squares estimation as $\Delta\mathbf{x} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}\Delta\rho$.

The user / satellite relative geometry contribute to how much the combined measurement errors, the most important being ionospheric and tropospheric delay, receiver noise and resolution and multipath, expressed using a variable called User Equivalent Range Error (UERE), will affect the resulting position error [**?**]. The more measurements the better the position solution is obtained only when the measurements are linearly independent [**?**]. When the satellites used are widely spread with respect to the user receiver, the dilution of precision (DOP) is small and the position solution much more accurate than when the satellites are close to each other. The effect of the satellite geometry for the position error is related as follows. The user-satellite geometry is denoted as $\mathbf{G} = (\mathbf{H}^T\mathbf{H})^{-1}$, where the matrix $\mathbf{H}$ is called the design matrix and is as explained above. Then the covariance matrix of the position errors on the $x-$, $y-$, and $z-$components and of the user clock bias (t_u) estimate is

$$cov(\mathbf{x}) = \sigma_{UERE}\mathbf{G}. \tag{9}$$

The variances of position and clock error components are $\sigma_x^2 = \sigma_{UERE}G_{11}; \sigma_y^2 = \sigma_{UERE}G_{22}; \sigma_z^2 = \sigma_{UERE}G_{33}; \sigma_b^2 = \sigma_{UERE}G_{44}$, where G_{ii} is the *i*th entry on the diagonal of $\mathbf{G}$. The Geometric Dilution of Precision (GDOP) encompassing 3-D position and clock bias estimation error is now

$$GDOP = \sqrt{G_{11} + G_{22} + G_{33} + G_{44}} \tag{10}$$

and Position Dilution of Precision (PDOP) the square root of the sum of the three first factors. In the case where four satellites are tracked the PDOP value is smallest, and therefore the position solution best possible, when three of the satellites are evenly

distributed in azimuth near the horizon and the fourth is perpendicularly above the user receiver (i.e. at zenith).

A more accurate satellite-to-user distance is obtained when a carrier phase observation is used. The carrier phase observation φ from satellite *i* is defined as

$$\varphi_i = r^i + c(t_u - dt^i) + d\rho^i + \lambda N - d^i_{iono} + d^i_{tropo} + \varepsilon^i_\varphi \tag{11}$$

where λ the carrier wavelength, *N* is the integer ambiguity, ε^i_φ encompasses noise, unmodelled errors and multipath and the other variables are as in the case of the pseudorange measurement.

Although carrier phase measurements provide very accurate positioning, in millimeter level in favourable conditions, they have not been widely used in pedestrian navigation. In order to obtain an accurate position solution, the integer ambiguity, namely the integer number of cycles the signal has traversed since leaving the satellite, has to be resolved. This may be done using double differenced GNSS measurements [**?**] or single differenced measurements and Precise Point Positioning (PPP) [**?**], both too complex for the equipment used for pedestrian navigation with lightness and reasonable cost requirements. The carrier phase observations are also difficult to be obtained continuously in the environments typical for pedestrian positioning namely in urban areas and indoors: the carrier phase tracking loop is more vulnerable to losing lock in attenuated signal environments than the code delay tracking loop, which produces the pseudorange measurements.

However, when the carrier phase measurements obtained in two consecutive time epochs are differentiated the integer ambiguity, which is unchangeable, disappears. The differentiated measurements are left with the error and noise terms as well as the change in geometric range between the time epochs, which may further be used for pedestrian navigation, as will be shown in Chapter 7

GNSS is the superior positioning system in open outdoor areas, but its use is very limited in urban and indoor areas. Although in these challenging environments High Sensitivity GPS (HSGPS) receivers are usually used, the performance in terms of reliability and accuracy is degraded [**?**]. As the received signal power level decreases the measurement uncertainty increases due to noise. The Effective Isotopic Radiated Power (EIRP) of a GPS (L1 C/A Code) civil signal is 26.8 dBW at the time of transmission. The power decreases mainly due to free space propagation loss (~ 184.4

dBW) while the signal travels from space to the Earth. In order to be able to find the relevant information from the signal below noise, the minimum received power at the conventional receiver has to be around -160 dBW [**?**] and at a typical HSGPS receiver -186 dBW [**?**]. The requirement of the -160 dBW received power is achieved with a Line-of-Sight (LOS) signal, but the signal degrades due to the attenuation resulting from propagation through a material (i.e. shadowing) and interference, typically multipath (i.e. fading). Type of the material the signal has to penetrate affects the amount of attenuation, a good comparison of the effect of different widely used materials may be found from [**?**]. For example, while entering a concrete and steel building the mean fading of the signal ranges from 19 to 23 dB and from 12 to 21 dB when entering a residential garage, depending on the elevation angle of the satellites tracked [**?**]. The required and received signal power levels show that the use of a HSGPS provides increased availability of the GPS positioning in most GNSS challenging environments, but the accuracy is still too poor for pedestrian navigation. Therefore augmenting and replacing GPS signals in urban and indoor areas is needed and few comprehensive methods will be discussed below.

2.2.2 WLAN Positioning

Wireless Local Area Network (WLAN) based on IEEE 802.11 standard is a wireless network used for communication between closely-spaced electronic devices (occasionally also called Wi-Fi which is a registered trademark of the Wi-Fi Alliance). Because of its broad existence means for using the technology for positioning has also been developed. In a WLAN positioning solution, the prevailing fingerprinting technique uses a database, a so called radio map, of access point signal strengths collected manually during an off-line training process. The user position is determined with the radio map and Received Signal Strength Indication (RSSI) measurements, which are the power level measurements of the received radio signal. A Bayesian theorem and the Histogram Maximum Likelihood method are used to solve the user position with the measurements [**?**], [**?**]. WLAN positioning provides typically room-level accuracy but is limited to the surroundings with existing and prepared infrastructure [**?**].

The measured RSSI samples at each reference point during the training phase are utilized to estimate the parameters of Weibull distribution [**?**] used to describe the WLAN signal strength distribution [**?**]. During positioning phase the observation

vector $\mathbf{RSSI} = \{r_1, ..., r_n\}$ is used to find the position x that maximizes the conditional probability $P(x|\mathbf{R})$ using the Bayesian theorem as

$$arg\ \max_x [P(x|\mathbf{R}_{WLAN})] = arg\ \max_x \left[\frac{P(\mathbf{R}_{WLAN}|x)P(x)}{P(\mathbf{R}_{WLAN})} \right]. \quad (12)$$

The advantages of using WLAN positioning are its large coverage, typically the range is from 50 m to 100 m, and that no line of sight is required [**?**]. A major weakness of the fingerprinting procedure is its vulnerability to the changes of the environment, causing the signal propagation patterns and thus the radio map to become obsolete and therefore offering a position solution with reduced accuracy. Also electrical equipment placed to the vicinity of the access points contorts the position solution.

2.2.3 Other Technologies

The other promising and actively researched absolute positioning technologies for pedestrians contain Radio Frequency Identification (RFID), Bluetooth and Ultra-Wideband (UWB). RFID positioning is based on attaching the user with tags that are then observed by a reader. The two most widespread methods used for resolving the user position is by just acknowledging that a user is close to the reader with known position or by using the RSSI measurements as described above. Ultra-Wideband positioning is based on a transmitter emitting radio waves occupying a large frequency bandwidth, namely more than 500 MHz. The benefit of using the Ultra-Wideband signals compared to the narrow band equivalents is their ability to penetrate many building materials such as concrete, glass and wood [**?**]. The UWB based positioning may be performed similar to the RFID positioning by attaching the user with receiver tags using RSSI methods such as in WLAN positioning or similar to Time of Arrival (ToA) methods such as GNSS based positioning. However, UWB positioning may also be used without supplying the user with special equipment, namely by using the UWB transmitter as a radar. In this manner the time elapsed before an emitted signal comes back to the transmitter after reflecting from the user is measured. When the background is known, the position of the user may be estimated. Bluetooth positioning uses the same principles as WLAN positioning; the position is mainly obtained using the RSSI methods utilizing an a priori prepared database of the access points in the area. Smartphones have been equipped with Bluetooth receivers for long already,

but unfortunately the infrastructure of the access points is not even close to be as widespread as for WLANs. The benefit of using Bluetooth for positioning is that the transmitters may be manufactured to transmit signals with strong power and therefore resulting in long range positioning [**?**].

A comprehensive presentation of various techniques used for indoor positioning, not all mentioned in this thesis, may be found from [**?**].

2.3 Relative Positioning

Self-contained sensors carried by the user are desirable equipment for pedestrian navigation providing relative position information independently of the environment. With a known initial position, the position may be propagated using the sensors for a limited period of time [**?**]. The propagation is done using standard inertial algorithms incorporating the attitude obtained by integrating the gyroscope measurements and translation obtained by double integrating the accelerometer measurements. The limitation of the self-contained sensors is the cumulative measurement errors growing fast due to the procedure integrating also the measurement noise.

An important aspect strongly affecting the development in pedestrian navigation is that the tolerable number and size of the equipment used is limited compared to e.g. robot or vehicular navigation. This forces to seek for a compromise between the accuracy and usability of the system. Micro-Electro-Mechanical System (MEMS) sensors are small in size and weight, have low power consumption and are inexpensive to produce [**?**] and therefore used widely for pedestrian navigation and especially in smartphones, however with decreased measurement performance.

2.3.1 Inertial Sensors

Accelerometers and gyroscopes are called inertial sensors. A system encompassing at least one accelerometer observing the acceleration of the body and a gyro measuring the rotation is called an Inertial Measurement Unit (IMU). Two different methods are used for processing the IMU measurements, namely Pedestrian Dead Reckoning (PDR) and inertial navigation [**?**], systems using the latter referred to as Inertial Navigation Systems (INS).

PDR has three phases; step detection, step length estimation and navigation solution update by combining the step length estimate obtained using at least one accelerometer and heading using a magnetometer or a gyro augmented with a magnetometer. As the performance of PDR is less sensitive to the quality of the sensors, especially when the distance travelled is concerned, PDR is feasible to be used with MEMS sensors. The process works with a single accelerometer, but the performance increases when more sensors are used [**?**].

Inertial navigation algorithms require a full IMU with triads of accelerometers and gyroscopes. The accelerometers observe the acceleration of a body, but to be able to transform the acceleration measurements into user position the direction of the acceleration is also needed and that is done by observing the relative rotational motion of the body with respect to the inertial reference frame using rate gyroscopes. The most common type of MEMS gyros are vibratory gyros based on Coriolis force [**?**]. The formation of a navigation solution from the accelerometer and gyroscope measurements is as follows [**?**].

The accelerometers output a measurement of specific force in the body reference frame $\mathbf{f}^b$. The measurement has to be transformed into the inertial reference frame using a Direction Cosine Matrix [**?**] $\mathbf{C}_b^i$ as $\mathbf{f}^i = \mathbf{C}_b^i \mathbf{f}^b$. Matrix $\mathbf{C}_b^i$ may be computed from the angular velocities ω_{ib}^b obtained from gyroscopes using

$$\dot{\mathbf{C}}_b^i = \mathbf{C}_b^i \mathbf{\Omega}_{ib}^b \tag{13}$$

where $\mathbf{\Omega}_{ib}^b$ is the skew symmetric matrix of the angular velocity vector. The specific force contains a measure of mass gravitation ($\mathbf{g}$) that has to be accommodated resulting in

$$\left.\frac{d^2\mathbf{r}}{dt^2}\right|_i = \mathbf{f} + \mathbf{g} \tag{14}$$

where $\mathbf{r}$ is the user position vector with respect to the reference frame origin and t is time.

By integrating the obtained value once the velocity of the user in inertial reference frame $\mathbf{v}_i$ is obtained. In pedestrian navigation the final navigation solution is needed in the Earth frame. The Coriolis theorem provides means to convert the velocity

measurement into the ECEF frame using information of the Earth turn rate $\omega_{ie} = [0\ 0\ \Omega]^T$ as

$$\mathbf{v}_e = \mathbf{v}_i - \omega_{ie} \times \mathbf{r} \tag{15}$$

where $\times$ denotes a vector cross product. By integrating again the velocity measurement the position of the user r in the Earth-Centered Earth-Fixed reference frame is obtained.

Integration of INS and GNSS fills the outages in positioning and provides more robust and reliable systems than either alone. However, both the accelerometer and gyroscope measurements suffer from various error sources, the most important ones being bias, scale factor and noise. Therefore the low-cost MEMS accelerometers used especially in smartphones are too erroneous to be used to obtain the user speed without augmentation with e.g. GNSS or using calibration and special algorithms, e.g. [**?**] [**?**]. The drift in gyroscope induced user attitude, especially heading, due to the mentioned errors seems to be still the most significant challenge in indoor and urban pedestrian navigation, although the accuracy of the position solution increases as multiple IMUs are used [**?**]. In order to achieve the accuracy and continuity of positioning needed for pedestrian navigation, means for overcoming the challenges due to the mentioned errors have to be found. Next section introduces few techniques used for augmenting the inertial sensors.

2.3.2 Other Self-Contained Sensors

A magnetic compass provides absolute angle information of the user with respect to magnetic north by measuring the intensity of Earth's magnetic field [**?**]. The Earth's magnetic field has a component parallel to the Earth's surface pointing toward magnetic north that differs in Helsinki area [**?**] by approximately 8 degrees from the geographic north (magnetic declination) and has a field intensity of about 0.52 gauss. The magnetic declination varies both from place to place and with the course of time. If the compass is totally parallel to the earth's surface, the heading (i.e. azimuth) (θ) may be computed from its horizontal measurements X_M, Y_M, neglecting the vertical component Z_M, as

$$\theta = \begin{cases} 90, & \text{if } X_M = 0, Y_M > 0 \\ 270, & \text{if } X_M = 0, Y_M < 0 \\ 180 - (arctan(Y_M/X_M)) * 180/\pi, & \text{if } X_M < 0 \\ -(arctan(Y_M/X_M)) * 180/\pi, & \text{if } X_M > 0, Y_M \leq 0 \\ 360 - (arctan(Y_M/X_M)) * 180/\pi, & \text{if } X_M > 0, Y_M > 0. \end{cases} \quad (16)$$

In practice, especially in pedestrian navigation applications, the compass is not totally parallel to the Earth's surface and the tilt has to be compensated for. If the roll (β) and pitch (ϕ) of the compass are known, the compass X_M and Y_M measurements may be transformed to the horizontal plane (X_H, Y_H) as

$$\begin{aligned} X_H &= X_M \cos(\phi) + Y_M \sin(\beta) \sin(\phi) - Z_M \cos(\beta) \sin(\phi) \\ Y_H &= Y_M \cos(\beta) + Z_M \sin(\beta). \end{aligned} \quad (17)$$

Now the heading θ may be computed using the equations in 16 by substituting the variables X_H, Y_H for X_M, Y_M.

The compass measurement suffer from errors of two different types; predictable and unpredictable. The predictable errors come from sources such as orientation of the navigation platform, soft and hard iron effects and magnetic declination. These errors may be eliminated by calibration or real-time compensation algorithms [**?**]. The unpredictable errors mainly due to environmental magnetic disturbances may be high, for example as in an office corridor experiment causing a heading mean error of around 18 degrees, when using a MEMS compass built-in a smartphone [P8] as shown in Figure 2.2, and are difficult to be removed. Therefore the compass heading in indoor environments is too poor to be used without augmentation, but as the magnetic heading is an absolute measure it is an effective measurement when integrated with e.g. a gyro. The self-contained sensors presented above are poor in estimating the user z coordinate, namely the height. Barometers measure the air pressure that can be converted into altitude information in indoor environments. The pressure (p) measured by the barometer is related to the height (h) as [**?**]

$$h = \frac{T_0}{T_L} \left(1 - \frac{p^{\frac{kR_g}{g}}}{p_0} \right) \quad (18)$$

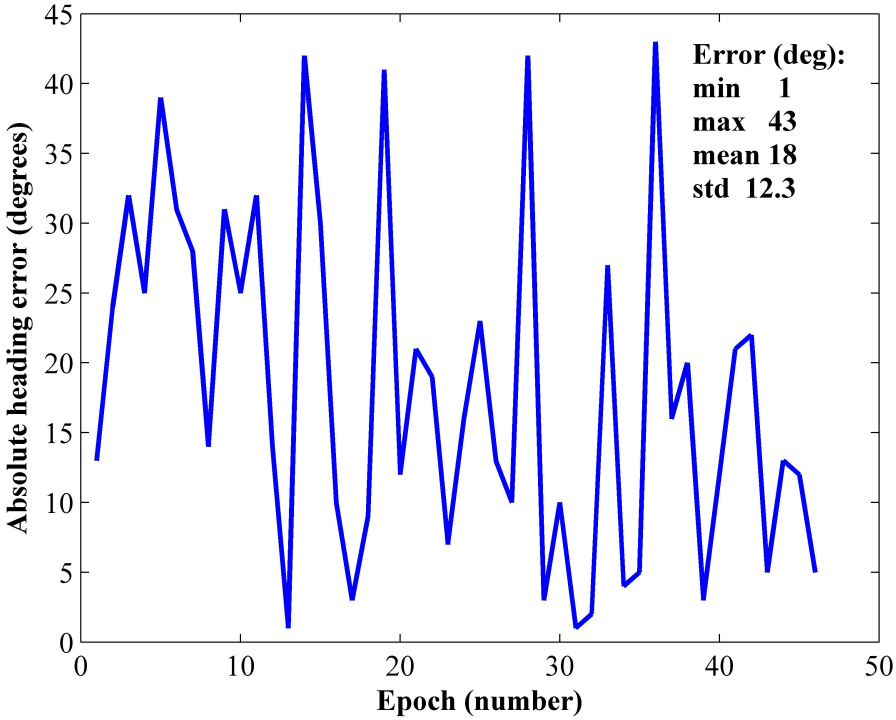

Fig. 2.2. *Absolute heading error obtained with a digital compass of a smartphone indoors.*

where R_g is the universal gas constant $8.3143(N \cdot m)/(mol \cdot K)$, p_0 is the average sea level pressure $101,325(kPa)$, T_L is the temperature lapse rate $-0.0065(K/m)$ [**?**], T_0 is the temperature at the sea level and g is the gravitation constant.

Cameras are not affected by the error sources deteriorating the measurements from gyroscopes, accelerometers and compasses and GNSS in indoor and urban areas. Observing the heading and translation from consecutive images is also free from preparing the environment a priori. Cameras are also light and small in size as well as reasonable priced. Therefore vision-aiding is a feasible method for augmenting the above mentioned systems in pedestrian navigation applications and will be discussed in the following chapters.

2.4 Estimation

Kalman filter is a set of mathematical equations used for estimating the state, e.g. position, velocity and attitude in case of pedestrian navigation, of a process recursively [**?**]. The state is estimated in a way that the mean of squared errors between the actual measurements and the expected measurements is minimized [**?**]. The recursive nature of the filter provides means for incorporating information about the past states and using the information to predict the current or even the future states. This is done by using a discrete-time stochastic system model [**?**]

$$\mathbf{x}_k = f_{k-1}(\mathbf{x}_{k-1}, v_{k-1}) \tag{19}$$

where f_{k-1} is a known linear or nonlinear function of the state $\mathbf{x}_{k-1}$ and v_{k-1} represents process noise. The measurements $\mathbf{z}_k$ are related to the state through a measurement model (h) as

$$\mathbf{z}_k = h_k(\mathbf{x}_k, w_k) \tag{20}$$

where w_k is measurement noise. In the following the fundamentals of the filter are described for the linear (Kalman filter) and nonlinear (Extended Kalman filter) cases.

2.4.1 Kalman Filter

Kalman filter estimates the state of a linear stochastic system by using measurements that are corrupted by zero-mean Gaussian noise w_k and are linear functions of the state that is also corrupted by zero-mean Gaussian noise v_{k-1} [**?**]. The discrete-time system model is [**?**]

$$\mathbf{x}_k = \mathbf{\Phi}_{k-1}\mathbf{x}_{k-1} + w_{k-1} \tag{21}$$

where k denotes the time epoch and Φ is called a state transition matrix and propagates the state from the epoch $k-1$ to k. The measurement obtained is

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + v_k. \tag{22}$$

Matrix $\mathbf{H}$ relates the state to the measurement as was the case for resolving the user position from the pseudorange measurements explained in GNSS positioning above. The state and measurement errors have Gaussian probability distributions

$$\begin{array}{rcl} p(w) & \sim & N(0, \mathbf{Q}) \\ p(v) & \sim & N(0, \mathbf{R}) \end{array} \tag{23}$$

where Q is process noise and R measurement noise covariance. Kalman filter has a prediction stage and an update stage, where the predicted state estimation is corrected using the obtained measurement. The predicted state estimation is called *a priori* estimate $\hat{\mathbf{x}}_k^-$ and the updated state estimate *a posteriori* $\hat{\mathbf{x}}_k$. The measurement $\mathbf{z}_k$ is used to update the state as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-). \tag{24}$$

The factor $(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-)$ is called the measurement innovation and it expresses the difference between the predicted $(\mathbf{H}\hat{\mathbf{x}}_k^-)$ and observed $\mathbf{z}_k$ measurement. Matrix $\mathbf{K}$ is called the Kalman gain and is computed as

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}. \tag{25}$$

The errors of a priori and a posteriori state estimates are defined as $\epsilon_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ and $\epsilon_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, respectively. The matrix $\mathbf{P}_k^-$ represents the covariance of the a priori state estimate error. The objective of the Kalman gain is to minimize the obtained a posteriori state estimate error covariance. Equation 24 shows that when the measurement error covariance is small and therefore the measurements are reliable, the gain is large and the innovation is weighted heavily as when the state estimate error covariance $\mathbf{P}_k^-$ is small, the gain is also small and a priori estimated state is trusted more.

Kalman filter is initialized by setting values for the initial state $\mathbf{x}_0$ and initial state error covariance $\mathbf{P}_0$. The measurement covariance matrix $\mathbf{R}$ is usually set a priori and kept constant as the matrix $\mathbf{H}$, process noise covariance matrix $\mathbf{Q}$ and the state transition matrix. The algorithm then recursively predicts the state as

$$\begin{aligned} \hat{\mathbf{x}}_k^- &= \mathbf{\Phi}_{k-1}\hat{\mathbf{x}}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{\Phi}_{k-1}\mathbf{P}_{k-1}\mathbf{\Phi}_{k-1} + \mathbf{Q} \end{aligned} \tag{26}$$

and updates the state estimate and state error covariance when the measurement is obtained incorporating the new Kalman gain computed using (25) as

$$\begin{aligned} \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-) \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^-. \end{aligned} \tag{27}$$

2.4.2 Extended Kalman Filter

The Kalman filter is a comprehensive method for estimating the state when both the state model f and measurement model h are linear. However this is not true for all applications including positioning and therefore other means, like an extension of the algorithm called Extended Kalman filter (EKF), should be used [**?**]. In EKF the state (19) and measurement (20) models are

$$\begin{aligned} \mathbf{x}_k &= f_k(\mathbf{x}_{k-1}) + w_{k-1} \\ \mathbf{z}_k &= h_k(\mathbf{x}_k) + v_k. \end{aligned} \tag{28}$$

In the case of the Kalman filter the state estimates are updated using the measurements as in EKF the nominal value of the state $\overline{\mathbf{x}}_k$ is updated with the perturbations $(\delta\mathbf{x}_k)$ as

$$\mathbf{x}_k = \overline{\mathbf{x}}_k + \delta\mathbf{x}_k. \tag{29}$$

EKF linearizes the measurement matrix $\mathbf{H}$ around the mean of the prior state approximating the non-linearity by a Taylor expansion. Therefore EKF is a Best Linear Unbiased Estimator (BLUE) minimizing the expectation $E(||\mathbf{x}_k - \hat{\mathbf{x}}_k||^2)$ [**?**]. The performance of EKF is however poor when the state and measurement models are highly non-linear and in such cases other estimators, e.g. Unscented Kalman Filter (UKF) [**?**] or Particle Filter (PF) [**?**] might be an alternative.

3. COMPUTER VISION METHODS FOR NAVIGATION

This chapter introduces the basics of computer vision. Herein the real life matters that are seen in the field-of-view of the camera are called objects and their two-dimensional images features. Humans inherently possess a good quality "stereo camera", namely eyes, and the human visual perception is capable of filling in missing information. Therefore it is easy for a human to understand perspectives, evaluate distances and occluded parts of the objects in the scene. In the case of the computer vision, objects in the scene are seen as sets of points of digitized brightness value functions. The form of these features, i.e. the point sets, change related to the pose of the camera and the lightness of the environment. Therefore care has to be taken while the features are extracted and matched in images. Deduction of motion information from images is also challenging. The methods used widely in vision-aided navigation research, also in the approaches presented in this thesis, are explained below.

3.1 Camera, Fundamental and Essential Matrices and Coordinate Frames

The principles explained in this section are mainly derived from [**?**]. The objects in the 3D world are mapped into 2D image features using projective transformations. These projections do not preserve the properties of shape, length, angle, distance or ratio of distances, but they do preserve the property of straightness. As a result of projective transformation the lines parallel in the scene seem to intersect in an image at a point, called the vanishing point. Therefore, to obtain a projective geometry space, the Euclidean geometry has to be augmented with a point and line in the infinity. Also, two coordinates (x, y) presenting a point in Euclidean space are replaced with a triplet $(x, y, 1)$ called homogenous coordinates in projective space.

An object point having coordinates $\mathbf{X}_N = (X, Y, Z)$ in the world (navigation) frame is transformed into the camera frame $\mathbf{X}_C$ using the rotation $(\mathbf{R})$ of the camera frame

with respect to the world frame and translation of the camera origin $(\mathbf{t})$ with respect to the world frame origin as

$$\mathbf{X}_C = \mathbf{R}\mathbf{X}_N + \mathbf{t}. \tag{30}$$

The methods presented in this thesis assume a pinhole camera model, in which an object point in the camera frame expressed in homogenous coordinates $\mathbf{X}_C = (X, Y, Z, 1)$ is mapped to the point $\mathbf{x} = (fX, fY, Z)$ in image plane. f is called the focal length and a line perpendicular to the image plane going from the camera centre, called principal axis, meets the plane at distance f in a point called principal point. World, camera and image frames as well as principal point and focal length are shown in Figure 3.1. An important note is that as opposed to the established means of coordinate frame configuration in navigation research presented in the previous chapter, in computer vision the y-axis is pointing up and z-axis along the camera's principal axis. This is an issue that has to be accommodated for while forming the navigation solution. The mapping of the world point X to the homogenous image point $\mathbf{x} = (x, y, 1)$, where $x = fX/Z$ and $y = fY/Z$ is characterized by a 3x4 camera matrix $\mathbf{P}$ as $\mathbf{x} = \mathbf{PX}$. When the calibration matrix $\mathbf{K}$ is known the world point $\mathbf{X}$ is mapped into image point $\mathbf{x}$ using camera matrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$.

When the object points all lie on a plane, point correspondences x_i, x_i' in an image are related by a homography expressed using a 3x3 matrix $\mathbf{H}$ as $\mathbf{H}x_i = x_i'$. The matrix has three entries but is defined only up to a scale and therefore four point correspondences (each having two coordinates) are needed to resolve the ambiguous values in $\mathbf{H}$. If three of these point correspondences are collinear, the homograpy is said to be degenerate and does not have a unique solution. Degeneracy problems are addressed in more detail in Chapter 5. Image points must be normalized for the solutions of homography to be correct. The image points $\mathbf{x}$ are normalized using the camera calibration matrix $\mathbf{K}$, discussed in detail below, as $\hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}$.

When the object points do not lie on a plane but are tracked from a real 3D scene, a Fundamental matrix $(\mathbf{F})$ has to be computed. The Fundamental matrix encompasses the intrinsic projective geometry between two views, meaning that only the rotation and translation of the two camera centers (also one camera between two images) and the internal camera parameters represented by the Calibration matrix $\mathbf{K}$, affect the matrix. In other words, Fundamental matrix $(\mathbf{F})$ represents the epipolar geometry

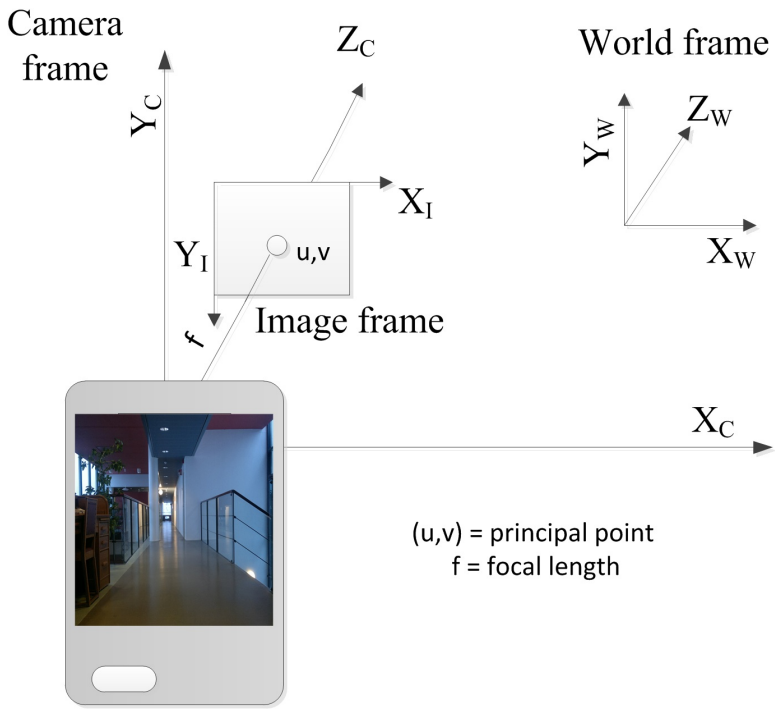

Fig. 3.1. *Camera, image and world coordinate frames.*

between the two views, visualized in Figure 3.2. The figure shows the image $\mathbf{x}, \mathbf{x}'$ of an object point $\mathbf{X}$, the epipole is an intersection of the baseline between the optical centres of the cameras and the image plane. If only the position of the first image point $\mathbf{x}$ is known, epipolar geometry restricts the location of the second image point $\mathbf{x}'$ to lie on the epipolar line, which is the line joining the image point and the epipole, and an epipolar plane is configured by the baseline and epipolar lines. The Fundamental matrix $\mathbf{F}$ is a 3x3 matrix and is defined for all corresponding points in two images $(\mathbf{x}, \mathbf{x}')$ as

$$\mathbf{x}'\mathbf{F}\mathbf{x} = 0. \tag{31}$$

At least seven corresponding points have to be matched from two images to compute the Fundamental matrix. For a general motion the rotation $\mathbf{R}$, translation $\mathbf{t}$ and cameras' internal parameters $\mathbf{K}, \mathbf{K}'$ encompassed in the Fundamental matrix relate the image points in the first and second images $(\mathbf{x}, \mathbf{x}')$, respectively, as

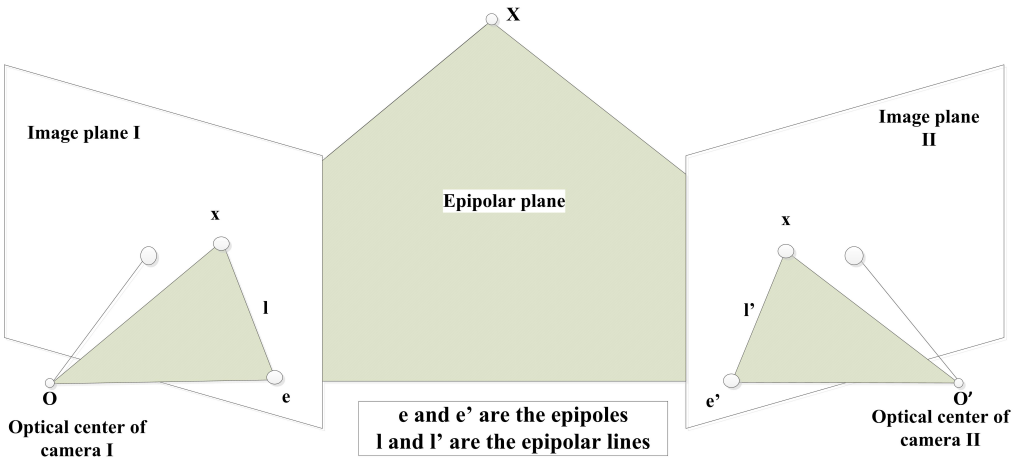

Fig. 3.2. *The epipolar geometry between two images including epipolar plane, epipoles* (e, e') *and epipolar lines* (l, l') *[?].*

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \mathbf{K}'\mathbf{t}/Z \tag{32}$$

where Z is the z-coordinate, i.e. the depth, of the object point.

When the image points are normalized as was explained above, the Essential matrix may be used instead of the Fundamental matrix, as $\hat{\mathbf{x}}'^T\mathbf{E}\hat{\mathbf{x}} = 0$, where $\mathbf{E} = \mathbf{K}'^T\mathbf{F}\mathbf{K}$.

3.2 Feature Extraction

First, the features have to be extracted for solving the motion and translation of camera between consecutive images. SIFT keys, explained below, are good features to be matched, when the environment contains many distinguishable objects. However, when the environment is poor with features, such as an office corridor, features arising from the constructions, like corners and lines, are more robust to be used. Below, first the procedure called filtering is explained, because of its use in noise reduction from images as well las edge detection. Then the extraction of two types of features used herein, SIFT keys and lines, is explained.

3.2.1 Filtering

Filtering is used for finding patterns and reducing noise from images. Filtering replaces the value of an individual pixel (x, y) with a weighted sum of its neighbors. Different weights correspond to different processes [**?**]. The pattern of weights is denoted as the kernel of the filter. Filtering is usually called convolution and is defined as

$$\mathbf{C}_{ij} = \sum_{x,y} \mathbf{G}_{i-x,j-y}\mathbf{I}_{x,y} \tag{33}$$

where the ith and jth component of the convolution result is denoted with $\mathbf{C}_{ij}$, $\mathbf{I}$ is the image and $\mathbf{G}_{i-x,j-y}$ is the kernel of the convolution.

A symmetric Gaussian kernel has the form of the probability density for a 2D Gaussian random variable and is a good candidate for a kernel for noise reduction convolution. Using a large standard deviation (σ) in convolution emphasizes the weight of the neighboring pixels and reduces the noise heavily, though causing some blurring. The Gaussian kernel is presented as

$$\mathbf{G}_{ij} = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2 + y^2}{2\sigma^2}). \tag{34}$$

3.2.2 SIFT-Features

Scale Invariant Feature Transform (SIFT) [**?**] is an approach based on transforming an image into local feature vectors; SIFT keys, describing the intensities around image points that are found as maxima or minima of a difference-of-Gaussian function. Each vector is invariant to image translation, scaling, and rotation and partially invariant to illumination changes and affine or 3D projections.

The minima and maxima of a difference of Gaussian function are computed in SIFT by building an image pyramid and resampling the data in each level. The 1D Gaussian kernel used is

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}} \tag{35}$$

and $\sigma = \sqrt{2}$. The image is convolved twice, and the difference of Gaussian is obtained by subtracting the two resulting images from each other. Then, the image resulted from the second convolution is resampled using bilinear interpolation and a pixel spacing of 1.5 resulting in an image having each pixel as a constant linear combination of the four adjacent pixels. The maxima and minima are found by comparing each pixel to its neighbours. The image resulting from the first convolution (I^1) in each level is processed for obtaining the image gradient magnitudes (M_{ij}) and orientations (O_{ij}) as

$$\begin{aligned} M_{ij} &= \sqrt{(I^1_{ij} - I^1_{i+1,j})^2 + (I^1_{ij} - I^1_{i,j+1})^2} \\ O_{ij} &= \arctan 2(I^1_{ij} - I^1_{i+1,j}, I^1_{i,j+1} - I^1_{ij}). \end{aligned} \tag{36}$$

The gradient magnitudes (M_{ij}) are given a threshold of 0.1 times the maximum possible value to provide robustness to changes in illumination. Its effect on the orientations (O_{ij}) is much lower. The orientation invariance is obtained by convolving the image using Gaussian kernel with large σ-value and by multiplying the weights with the corresponding gradient value. A histogram with 10 degree intervals is built from the convolution results and the correct orientation of the feature is the peak in the histogram. As a result the features in the image are represented with keys having a stable location, scale and orientation also invariant for changes in illumination in consecutive images. Each key is a 128-element vector.

SIFT functions well in environments full of features, such as in outdoors, but suffers from errors if the images are comprised mostly of vegetation or dynamic objects [**?**].

3.2.3 Line Extraction

Indoor and urban environments are constructed in a way that their structures constitute a three dimensional grid defining a orthogonal coordinate frame, also called Manhattan grid [**?**], containing straight parallel lines and therefore the methods based on line features are suitable for these environments otherwise poor in features [**?**]. Lines are also good features for the basis of visual positioning because they are invariant to changes in the lighting, which is crucial especially for indoor positioning, but also because straight lines remain straight in projective transformations and are

not disturbed by dynamic objects that are not blocking the view to all lines in the scene. Line extraction begins by identifying edges of all features in an image and then separating the straight lines from other features. These are explained below.

Edge Detection

Fast changes of brightness in the image indicate edges of objects. The brightness of the pixel in an image depends on the characteristics of light sources as well as the traits and orientation of the surface. The orientation of the surface is specified with surface gradients. Canny edge detector [**?**] calculates magnitudes and directions of these gradients. It is an optimal algorithm for detecting the edges requiring low error rate of the calculations, the points to be well localized, meaning that the distance between the calculated location of the edge and the real one has to be minimal, and refusing multiple responses for an edge.

In a two dimensional image the edge has a position and an orientation. The direction of the tangent to the edge contour is called edge direction. The edge is found by convolving the image using a first derivative G_n in direction n of a two-dimensional Gaussian G as a kernel and defined as

$$G_n = \frac{\delta G}{\delta n} = n \cdot \nabla G \tag{37}$$

where

$$G = exp(-\frac{x^2 + y^2}{2\sigma^2}). \tag{38}$$

Now an edge point is a local maximum when the image $\mathbf{I}$ is convolved using G_n as a kernel. The local maximum is found from an image pixel fulfilling

$$\frac{\delta^2}{\delta n^2} G \cdot \mathbf{I} = 0 \tag{39}$$

and the edge strength is calculated as

$$|G_n \cdot \mathbf{I}| = |\nabla (G \cdot \mathbf{I})|. \tag{40}$$

A pixel having a magnitude of local maximum along the gradient direction belongs to the edge and the process of finding the maximums using (39) is called non-maximum suppression. The set of possible edge pixels found by looking for local maximums

contains too many members. The pixels having weak response to an edge have to be excluded using a procedure called hysteresis. Hysteresis evaluates each pixel in the possible edge set using two thresholds. All pixels having edge strength (40)above the upper threshold are classified as part of an edge and all below the lower threshold as not belonging to the edge. Pixels between the two thresholds are evaluated based on their neighbours. If the neighbour belongs to an edge, then the pixel belongs also, otherwise not. After all pixels in the possible pixel set are deemed to belong to an edge or not, the optimal edge set is defined.

Canny edge detection is one of the most used edge detectors, but many others also exist, for example Sobel, Laplace, and Prewitt operators [**?**].

Separating the Lines from Other Edges

Canny edge detection finds all changes of brightness in an image demonstrating an edge of an object. For most computer vision applications there is still a need to find certain shapes among all edges. Hough [**?**] developed a method for identifying lines among all image pixels. His method maps all image points into a two-dimensional parameter space, the parameters being the slope and the intercept of the line. Each point is then examined and given a vote for all features possibly travelling through it. Since both the slope and intercept are unbounded, a modified form of Hough transform was developed and has been widely exploited in computer vision research [**?**]. When extended it is suitable for finding also other curves than lines. The method is based on the parameter space (ρ, θ), where ρ is the radius of a line passing through the origin and normal to the line being detected and θ is its angle with the x-axis as shown in Fig. 3.3. A straight line including the pixel *(x,y)* is then defined as a sinusoid

$$\rho = x\cos(\theta) + y\sin(\theta). \tag{41}$$

When the possible values of θ are restricted to the interval $[0, \pi]$ every line in the image plane corresponds to a unique point in the parameter space defined plane. Now the curves going through a common point in the parameter plane correspond to the image points on a specific straight line. Therefore the lines are identified by looking for the points in the (ρ, θ)-parameter space having local maximums of votes. The weakness of the otherwise sophisticated algorithm is that it is computationally heavy. As in pedestrian navigation the real-time processing of algorithms is crucial,

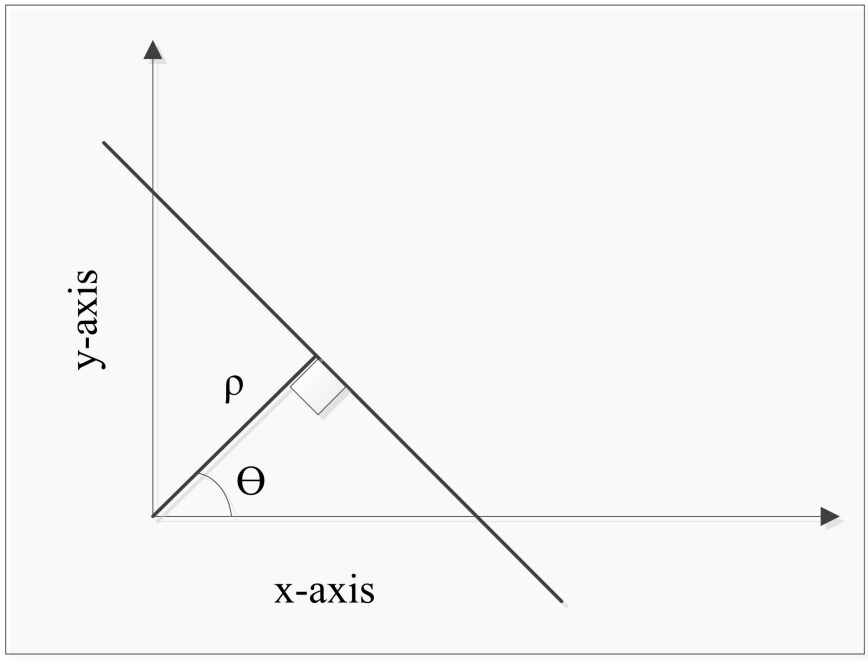

Fig. 3.3. *Formulation of parameters ρ and θ in Hough transform.*

more efficient line detection algorithm based on probabilistic Hough Transform was developed in this thesis and will be presented in Chapter 6.

3.3 Image Matching

Matching is a process of identifying the corresponding features in two images of the same scene taken at different viewpoints, different times, or by different sensors (cameras). As the SIFT keys are invariant to rotation and translation, their matching is restricted to finding the most similar keys in the two images, i.e. the keys with minimum Euclidean distance [**?**].

When more robust matching is needed from the noisy image data RANdom SAmple Consensus (RANSAC) [**?**] is used. The RANSAC algorithm enlarges the set composed by the minimum set needed for the solution with the points within some error tolerance. This is done by searching for a random sample of points that leads to a fit of the model in question for which many of the points agree. This leaves the outliers out from the data used in calculations. The algorithm is used widely in computer vision

applications such as vanishing point detection. A comprehensive explanation of the algorithm with examples is given in [**?**]. RANSAC is however computationally quite heavy and therefore not used in the methods discussed in this thesis emphasizing the computational efficiency.

3.4 Camera Calibration

As was explained earlier the operations performed for mapping the objects into images are done using projective geometry. However, navigation solutions need the information to be presented in Euclidean reconstruction. This may be done by using a calibrated camera for capturing the images. Calibration provides information about camera's intrinsic parameters and is represented using a calibration matrix $\mathbf{K}$. The camera intrinsic parameters are focal length (f_x, f_y), principal point (u, v), skew coefficient (S), aspect ratio and distortions. Focal length is defined as the distance between the centre of the cameras lens and the film while taking a focused image of an object that is infinitely far. Principal point is the intersection point of the cameras optical axis with the image plane, as was shown in Figure 3.1. Distortions blur the image due to the fact that the focal length varies in different points of the lens. The skew comes from manufacturing errors and makes the two image axes non-orthogonal, the skew coefficient defines the angle between these axes. The skew in a normal camera is usually zero, except when taking an image of an image, for example, when enlarging a negative [**?**]. A reduced form, with zero skew, of a camera matrix K is normally used for computer vision applications and is

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix}. \tag{42}$$

It is adequate for a pedestrian navigation system to calibrate the camera once and assume the parameters unchanged since. The calibration may be done by photographing a certain model image from different viewpoints and then calculating the parameters using the images and the known geometry of the model. In the methods described in the thesis calibration is done using a Matlab application [**?**]. A camera may be calibrated also with a single image using vanishing point information and the zero assumption of the skew. If the positions of three vanishing points can be

recovered, the focal length and centre of projection (the principal point) may be determined [**?**]. When the accuracy of the navigation solution may be compromised for the sake of adaptability, the focal length may be found from the images Exchangeable Image File (EXIF) data, and the principal point may be assumed to be the central point of the image.

3.4.1 Distortion

The best accuracy for vision-aided calculations is obtained when a camera with a wide angle lens offering an extended field-of-view is used, as will be shown in Chapter 4. However, the wide angle lens results in radial distortion in the images. If the distortion is not corrected, the calculation accuracy suffers. According to [**?**], the rectification of the whole image introduces aliasing effects complicating the feature detection. For optimal result, when a wide angle lens camera is used the radial distortion is corrected only for the features extracted from the images with a model presented in [**?**] and explained below.

The radial distance (r_d) of the normalized distorted image points (x_d, y_d) from the radial distortion center, which is in most cases the principal point (*u,v*), is

$$r_d = \sqrt{x_d^2 + y_d^2}. \tag{43}$$

Using the radial distance of the distorted image points, the radial distance (*r*) of the corrected image points (x_u, y_u) is obtained as

$$r = r_d(1 - k_1 r_d^2 - k_2 r_d^4). \tag{44}$$

The constants k_i are the distortion values specific to the camera and are obtained from calibration. The corrected and distorted image points are related as

$$\begin{aligned} x_d &= x_u(1 + k_1 r + k_2 r^2) \\ y_d &= y_u(1 + k_1 r + k_2 r^2). \end{aligned} \tag{45}$$

The effect of the distortion correction is shown in the case of the feasibility of the visual gyroscope in Chapter 4.

4. VISUAL GYROSCOPE

Urban scenes in indoor and downtown environments consist mainly of straight lines in three orthogonal directions. The projective transformations mapping the three dimensional scene into a two dimensional image preserves the straight lines, but not the angles, and therefore the lines parallel in the scene seem to intersect in the image. The lines in three orthogonal directions form three intersection points called vanishing points. The vanishing points arising from lines in x- and y-axis directions are called horizontal and vertical vanishing points, respectively, and from the lines in the direction of propagation (z-axis) the central vanishing point.

The central vanishing point is the intersection point $\mathbf{v}$ of a ray through the camera centre having a direction $\mathbf{d}$, and of all other lines also having direction $\mathbf{d}$, and the image plane. The vanishing point $\mathbf{v}$ is related to the direction $\mathbf{d}$ as $\mathbf{v} = \mathbf{K}\mathbf{d}$ [**?**] where $\mathbf{K}$ is the camera calibration matrix encompassing the intrinsic parameters of the camera. The directions $\mathbf{d}$ and $\mathbf{d}'$ of two vanishing points in consecutive images are related by the Rotation matrix $\mathbf{R}$ as $\mathbf{d}' = \mathbf{R}\mathbf{d}$, i.e. rotation of the camera between the two images. The change of position of the camera between the images, meaning pure translation with no rotation, has no effect on the vanishing point location. The rotation $\mathbf{R}$ of the camera may also be thought as the rotation from the initial position where the camera is aligned with the navigation frame so that the z-axis of the camera is pointing to the direction of the propagation and the x- and y-axes are orthogonal to the z-axis as shown in Figure 3.1, i.e. the orientation of the camera with respect to the navigation frame. In this initial configuration, the central vanishing point $\mathbf{v}_z$ lies at the principal point and the other two vanishing points at infinity on the x and y image axes. Then the orientation of the camera $\mathbf{R}$ is described with $\mathbf{V} = \mathbf{K}\mathbf{R}$, where $\mathbf{V}$ is the vanishing point location matrix $[\mathbf{v}_x \mathbf{v}_y \mathbf{v}_z]$, $\mathbf{K}$ is the calibration matrix containing the camera intrinsic parameters (defined in Chapter 3), and $\mathbf{R}$ is the rotation matrix of the camera [**?**].

From the change of the vanishing point locations in consecutive images the change in the camera attitude may be monitored and the camera used as a visual gyroscope. When only the central vanishing point is obtained, the visual gyroscope provides the pitch and heading of the user, if also either the horizontal or vertical vanishing point is perceived, the full three dimensional attitude is provided. First, the method for obtaining the central and vertical vanishing points is explained. Secondly, the rotation matrices for the user heading and camera pitch are given, then the configurations for full attitude. Effective error detection is crucial for not deteriorating the integrated navigation solution using erroneous visual measurements and therefore a concept of Line Dilution of Precision (LDOP) has been developed and will be presented. The influence of different camera characteristics on the performance of the visual gyroscope has been studied and will be discussed. Finally, smartphone implementation of the visual gyroscope will be represented.

4.1 Locating the Vanishing Points

In order to find the central vanishing point and furthermore the heading change and the pitch of the camera, the straight lines in the direction of the propagation must be identified in the image. Because the images are noisy, especially the ones taken with a smartphone camera and in an indoor environment, the images must be pre-processed. The images are smoothed using a Gaussian filter by replacing the image pixels with a weighted sum of their neighbour pixel values and therefore reducing the noise [**?**]. All edges in images are identified with a Canny edge detector and the straight lines separated from the set of all edges with the Hough Lines algorithm as explained in Chapter 3. All lines found are classified based on their orientation with respect to the camera frame, as going in the direction of the z-axis, totally horizontal or totally vertical and horizontal or vertical. Totally horizontal (and vertical) lines have no angle with respect to the x-axis (y-axis), i.e. the slope of the line is undefined (zero), as the ones classified as horizontal or vertical have a slope lower than a threshold with respect to the corresponding axis. The central vanishing point is found by using a voting scheme, namely each intersection of all line pairs, i.e. vanishing point candidate, is voted for by all the lines found and the point that gets most of the votes, in this case the intersection point of most of the lines, is selected as the correct one. The classification of the lines and the central vanishing point found with the method

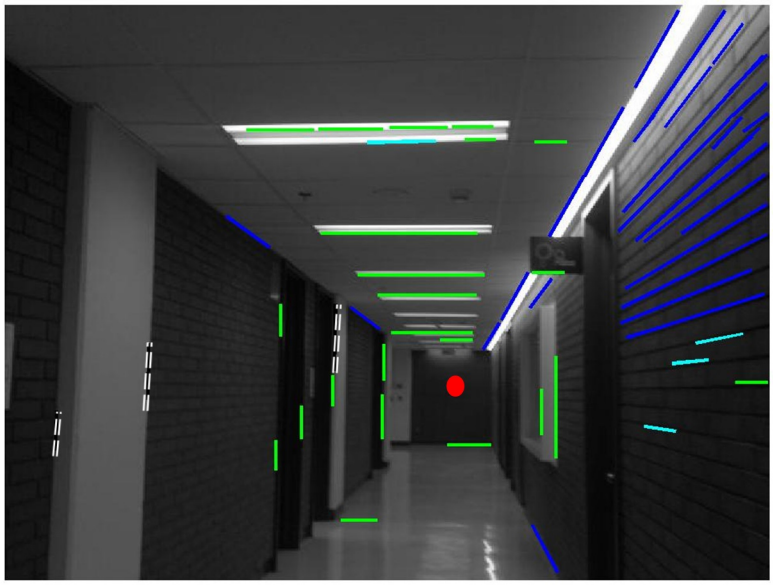

Fig. 4.1. *Lines in an image with no (or minor) roll are classified based on their slope as totally vertical or horizontal (green), vertical (white dotted), horizontal (turqoise) and along the direction of propagation (blue). Red dot is the central vanishing point.*

explained are shown in Figure 4.1.

Only the 2 degrees-of-freedom attitude may be obtained from one vanishing point location and in order to resolve also the roll at least one other vanishing point in addition to the central one has to be located. Experiments show that the horizontal lines are infrequent in urban and indoor environments and therefore in this thesis the vertical vanishing points are tracked.

When the camera is experiencing only a small roll (as in Figure 4.1), the lines in vertical directions are mainly totally vertical due to the relatively low resolution of the images, which reflects the fact that the pixels obtained with a camera experiencing small roll are overwhelmed by the noise in images. When the camera has a larger roll, most of the vertical lines have slopes deviating from zero, as is the case in Figure 4.2. The ratio of the number of vertical lines that have none-zero slopes and that of the lines that have zero slopes is calculated. If the ratio exceeds a threshold,

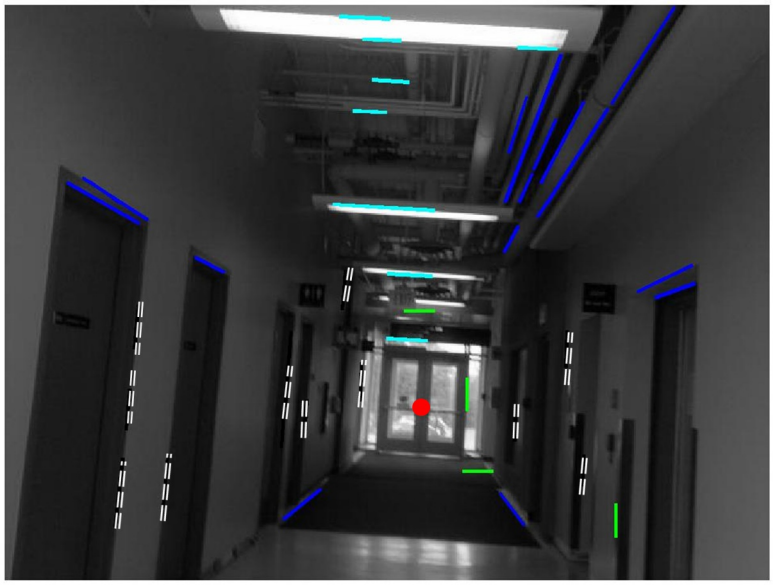

Fig. 4.2. *When the camera experiences roll the number of totally vertical lines decreases. Lines are classified based on their slope as totally vertical or horizontal (green), vertical (white dotted), horizontal (turqoise) and along the direction of propagation (blue). Red dot is the central vanishing point.*

the camera is experiencing roll, and the location of the vertical vanishing point is calculated similarly as the central one and incorporated into the rotation calculations. The locations of the central vanishing points in Figures 4.1 and 4.2 are shown using a red circle. The vertical vanishing points lie outside the image and therefore are not shown.

4.2 Attitude of the Camera

When the camera is rotated from the initial position counter clockwise with heading θ degrees the rotation matrix has the form

$$\mathbf{R} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{46}$$

and the pitch towards the floor plane ϕ degrees

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}. \tag{47}$$

When the camera experiences these two rotations the matrix $\mathbf{R}$ becomes

$$\mathbf{R} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ \sin\phi\sin\theta & \cos\phi & -\sin\phi\cos\theta \\ -\cos\phi\sin\theta & \sin\phi & \cos\phi\cos\theta \end{bmatrix}. \tag{48}$$

When the calibration and rotation matrices are as explained above, the heading change (θ) and pitch (ϕ) angles may be obtained from the location of the central vanishing point as

$$\mathbf{v}_z = \begin{bmatrix} f_x \sin\theta + u\cos\phi\cos\theta \\ -f_y \sin\phi\cos\theta + v\cos\phi\cos\theta \\ \cos\phi\cos\theta \end{bmatrix}. \tag{49}$$

The vanishing point $\mathbf{v}_z$ is presented in homogenous coordinates as $(x, y, 1)$, where the x, y are the pixel coordinates of the central vanishing point obtained using the voting scheme explained above. As the third row of the (49) equals 1, the heading change (θ) and pitch (ϕ) may be computed as

$$\begin{aligned} \theta &= \arcsin\frac{(x-u)}{f_x} \\ \phi &= \arcsin\frac{y-v}{-f_y\cos(\theta)}. \end{aligned} \tag{50}$$

An important note is that the change in heading and pitch obtained by tracking the vanishing points reverses to the definition of the navigation frame and has to be accommodated for in the integration. To be exact, when the camera rotates clockwise,

i.e. its heading increases in navigation frame, the vanishing point location moves counterclockwise and its x-coordinate decreases and therefore the obtained visual heading θ has an opposite sign compared to the heading in the navigation frame. Also in the navigation frame the pitch increases upwards as (ϕ) decreases.

Tracking the central vanishing point provides information about the user heading change and camera pitch. When also roll β is required, at least two vanishing points are needed, in this thesis the other being the vertical vanishing point as explained above. Now the rotation matrix $\mathbf{R}$ of the camera experiencing only roll has the form

$$\mathbf{R} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{51}$$

And the full rotation of the camera experiencing simultaneously change in heading, pitch and roll is

$$\mathbf{R} = \begin{bmatrix} \cos\beta\cos\theta - \sin\beta\sin\phi\sin\theta & -\sin\beta\cos\phi & \cos\beta\sin\theta + \sin\beta\sin\phi\cos\theta \\ \sin\beta\cos\theta + \cos\beta\sin\phi\sin\theta & \cos\beta\cos\phi & \sin\beta\sin\theta - \cos\beta\sin\phi\cos\theta \\ -\cos\phi\sin\theta & \sin\phi & \cos\phi\cos\theta \end{bmatrix} \tag{52}$$

and all three angles may be resolved using the two vanishing point locations.

4.3 Error Detection

In the case when the location of the vanishing point is known a priori to some extent, the intersection points deviating remarkably from the estimate may be discarded and an accurate orientation measurement obtained [**?**]. The method, used for UAVs (unmanned aerial vehicles), determines the possible vanishing point locations based on the known potential attitude of the camera. The error detection based on the estimated vanishing point location is suitable for robot and vehicle navigation, where the motion is to some extent foreseeable and stable. However, no limitations may be imposed for the possible vanishing point locations in pedestrian navigation, especially if the vision-aiding is done using a smartphone camera, because the motion of the pedestrian is much more unpredictable. Therefore a method evaluating the vanishing

point accuracy based on the geometry of lines used to compute it is developed in this thesis.

The concept of Dilution of Precision (DOP), originally specifying the geometry of satellites used for obtaining a position solution with GNSS [**?**], was introduced into vision-based navigation by [**?**]. Their DOP value presented the orientation and position of pseudo ground control points (PGCP) needed in navigation using a camera and 3D maps constructed of the environment. Now the concept of an LDOP, a dilution of precision value demonstrating the geometry of the lines used for calculating the position of the vanishing point is developed. The method is based on dividing the scene in an image into four quarters around the estimated vanishing point.

If lines intersecting at the vanishing point are found from all four sections, the estimated vanishing point is correct with high probability, and it is given a minimum LDOP value, namely $\sqrt{2}$. The situation is visualized in Figure 4.3a. The justification for the minimum LDOP value selection is given below, where the situation of reduced line geometry is explained. If the lines intersecting at the estimated vanishing point are from three of the sections, the line geometry is still determined sufficiently accurate and a low LDOP value is assigned to the estimated vanishing point. When the geometry of lines is reduced, namely the lines are found only from two sections, shown in Figures 4.3b and 4.3c, or especially only from one, as in Figure 4.3d, more evaluation of the geometry must be done. In the case where lines are found from two sections, the accuracy of the estimated vanishing point is strongly dependent of the orientation of the lines found. If the lines are from opposite quarters, as in Figure 4.3c, namely the angle between the lines is close to or larger than 180 degrees, the intersection is in an incorrect location with a higher probability and a higher LDOP value is assigned than for the case where the lines are from adjacent quarters, namely the angle between them is less than 180 degrees as in Figure 4.3b. This reasoning is derived from the fact that often the lines from opposite sections are actually parts of the same line split by the line detection algorithm due to changes of brightness in the image and therefore there is in reality no intersection point.

When the line geometry is reduced into a set of lines found only from one section, as shown in Figure 4.3d, the LDOP evaluation is based on the mutual alignment of the lines using a method proposed in [**?**]. The angle between all lines in the set and the x-axis of the image is calculated and the pair with the largest difference in the angle magnitude is selected. The angle between the first line of the pair and the image

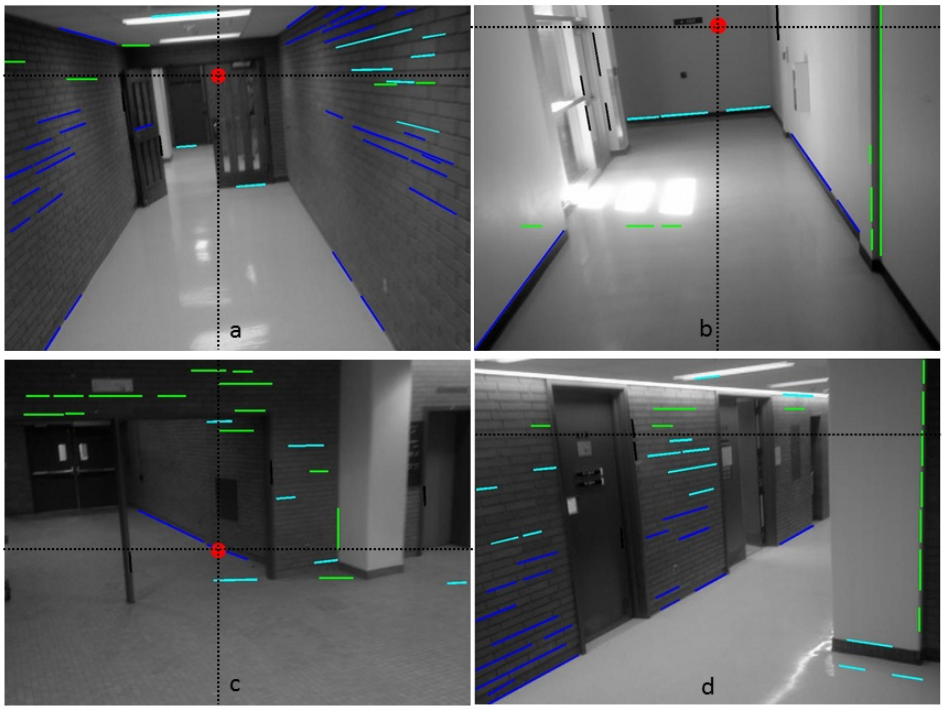

Fig. 4.3. *Four images resulting from the vanishing point calculation. The image is divided into four sections around the estimated vanishing point (section borders shown with black dotted lines) and its reliability is evaluated based on the geometry of the blue lines used for calculations. The vanishing point is found correctly in images a, b and d (outside the image). The black continuous lines are used to calculate the vertical vanishing point.*

x-axis (α_1) and between the second line of the pair and the image x-axis (α_2) is obtained using the estimated central vanishing point location (x_{v_z}, y_{v_z}), the starting point (x_i, y_i) of line *i* (*i*=1,2) and distance (D_i) of the estimated vanishing point from the starting point of line *i* as

$$\begin{aligned}\cos(\alpha_1) &= \frac{x_{v_z} - x_1}{D_1} \\ \sin(\alpha_1) &= \frac{y_{v_z} - y_1}{D_1} \\ \cos(\alpha_2 &= \frac{x_{v_z} - x_2}{D_2} \\ \sin(\alpha_2) &= \frac{y_{v_z} - y_2}{D_2}.\end{aligned} \tag{53}$$

The matrix $\mathbf{H}$, characterizing the line geometry, is

$$\mathbf{H} = \begin{pmatrix} \cos(\alpha_1) & \sin(\alpha_1) \\ \cos(\alpha_2) & \sin(\alpha_2) \end{pmatrix}. \tag{54}$$

The matrix $\mathbf{G}$ is formed from the geometry matrix $\mathbf{H}$ using $\mathbf{G} = (\mathbf{H}^T \times \mathbf{H})$ and $\mathbf{G}^{-1}$ is

$$\mathbf{G}^{-1} = \frac{1}{|\mathbf{G}|} \begin{pmatrix} \sin^2(\alpha_1) + \sin^2(\alpha_2) & -\cos(\alpha_1)\sin(\alpha_1) + \cos(\alpha_2)\sin(\alpha_2) \\ -\cos(\alpha_1)\sin(\alpha_1) + \cos(\alpha_2)\sin(\alpha_2) & \cos^2(\alpha_1) + \cos^2(\alpha_2) \end{pmatrix} \tag{55}$$

where $|\mathbf{G}|$ is the determinant of $\mathbf{G}$ and is defined as $|\mathbf{G}| = sin^2(\alpha_1 - \alpha_2)$. The GDOP in GNSS positioning applications is calculated using the diagonal values of the $\mathbf{G}^{-1}$ matrix as explained in Chapter 2 and the result may be transformed for the case of two satellites [**?**], and further used to calculate the LDOP for the lines as

$$LDOP = \sqrt{\frac{1}{|\mathbf{G}|}(\cos^2(\alpha_1) + \cos^2(\alpha_2) + \sin^2(\alpha_1) + \sin^2(\alpha_2))}. \tag{56}$$

For any two angles (α_1, α_2) (56) may be now written as $LDOP = \sqrt{\frac{2}{|\mathbf{G}|}}$. The smallest possible LDOP value is $\sqrt{2}$ and arises from the maximum angle between two lines lying in the same quarter section, namely 90 degrees. When the magnitude of the angle between the lines is more than 10 degrees, the accuracy of the estimated vanishing point location is still sufficient, but decreases rapidly when the angle decreases.

Evaluation of the accuracy of the estimated vertical vanishing point cannot be done using the line geometry, but is based on monitoring the roll obtained. A camera can be rolled over 15 degrees for the purpose of obtaining images with special viewpoints [**?**], but it is seldom done unintentionally and such large roll is not convenient for vision-aided navigation. Because the calculation of an accurate vertical vanishing point is not always possible (due to noise in the images and shortage of lines) the roll‘s magnitude must be monitored. If the roll‘s magnitude exceeds 15 degrees, the vertical vanishing point is excluded from the calculations. In these situations the roll is evaluated to be zero and the heading and pitch are calculated more accurately using only the central vanishing point. If the camera is actually experiencing roll when the calculations fail and the roll is estimated to be zero, errors appear also in the heading and pitch. The effect of this error into the user attitude observations is discussed in detail below.

4.4 Performance of the Visual Gyroscope

The visual gyroscope is a comprehensive method providing a heading change measurement but it does not provide any absolute value of the heading and must therefore be integrated with measurements from other sources. Accurate initialization of the visual gyroscope's heading is crucial using absolute heading information. The visual gyroscope is based on the vanishing point observations calculated using lines found from the image of the environment and as a result cannot be used during sharp turns, when the visibility to building boundaries forming lines is lost due to the camera being too close to the wall. If however the view to the lines is maintained during a turn, the change of the world frame may be perceived only when the image rate is high. Therefore the visual gyroscope's heading need to be augmented with heading measurements obtained using another system, e.g. a rate gyroscope, a magnetometer or a floor plan also occasionally during navigation.

Identification of the correct vanishing point location is dependent on the number and geometry of the lines found in the image. Low lighting of the navigation environment reduces the number of lines found, possibly resulting in erroneous vanishing point location. In addition to low lighting, the problem of obtaining an erroneous vanishing point location may arise from the selection criteria of the Hough Line algorithm parameters. The parameters used in this thesis are adjusted so that lines shorter than

Table 4.1. *Statistics for heading change accuracy, all units degrees*

Statistics	min error	max error	mean error	std of error
Heading change	0	18.4	0.8	0.6
Pitch	0	10.7	0.3	0.3

a threshold are left out of the computation to reduce the number of nonparallel lines disturbing the vanishing point calculation. An optimal parameter for indoor environments was found by through experimentation. When the scene consists of a plane, there are no lines in the image and the vanishing point cannot be calculated. Sometimes, despite the parameter selection, the set of lines found from the scene consists of a number of nonparallel lines resulting in erroneous vanishing point location.

The accuracy of the heading change as well as the pitch estimates were evaluated with a test containing 7555 images taken with a static camera in an office environment. The test environment had changing light as well as dynamic objects in the scene of the camera sometimes encompassing the view totally. Table 4.1 shows the statistics for the heading change and pitch obtained in the 2.5 hour time span of the test, the mean errors being 0.8 and 0.3 degrees, and standard deviations 0.6 and 0.3 degrees, respectively. Additional test results with a moving camera will be presented later in Chapter 6.

The most significant source affecting the gyroscope accuracy, especially a MEMS gyro, is the drift. The Allan variance analysis method [**?**] was originally developed for the study of oscillator stability, but has since been applied widely to gyro drift analysis. Because the Allan variance method is suitable for the noise study of any instrument, it is applied here to evaluate the noise level of the visual gyroscope. The Allan variance $\sigma_C^2(t_A)$ [**?**], modified here for the camera gyroscope, for the averaging time t_A is

$$\sigma_C^2(t_A) = \frac{1}{2(N-1)} \sum_{k=1}^{N-1} (\tilde{y}(t_A)_{k+1} - \tilde{y}(t_A)_k)^2 \tag{57}$$

where $\tilde{y}(t_A)_k$ is the average value of a bin k containing the heading change and pitch values. The averaging time t_A is the length of a bin and N is the number of bins formed of the data for the corresponding averaging time. The plotted Allan variance may be used for finding different error types for the sensors [**?**].

The Allan deviation plot for the 7555 images taken is shown in Figure 4.4. The figure

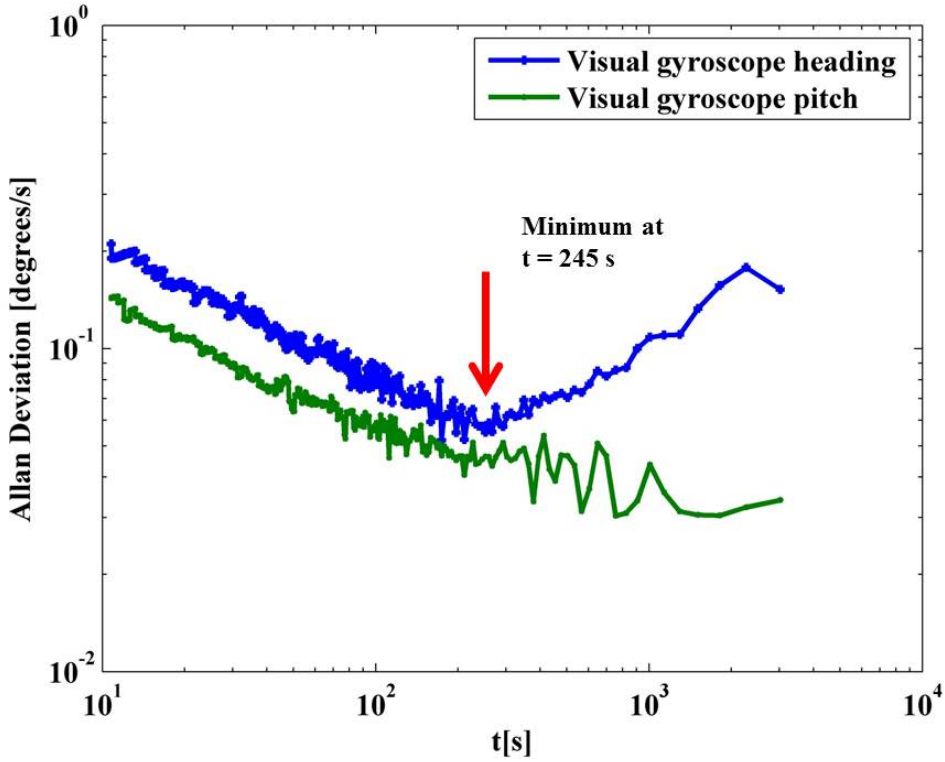

Fig. 4.4. *Allan deviation plot showing the noise in the visual gyroscope.*

shows large deviations due to the uncorrelated noise affecting the visual gyroscope stability for the short integration times. After the deviation has reached a minimum value, the rate random walk starts to increase the deviation again. The bias instability measure may be found from the minimum value, and is 0.058 degrees/second for the heading, and 0.045 degrees/second for the pitch, at the integration time of 245 seconds. The obtained measure is lower than the values obtained for a typical MEMS gyroscope tested in [**?**].

The test showed also the tolerance of the method to dynamic objects, which may be seen from Figure 4.5. Heading angle and pitch errors due to dynamic objects obscuring the scene were very infrequent. The errors in the visual gyroscope are mostly introduced by environmental factors such as lighting, construction of the environment and objects with lines not parallel to the direction of propagation.

As discussed before the roll observations may be significantly erroneous due to the

Fig. 4.5. *Calculation of the central vanishing point is largely tolerant to dynamic objects in the scene.*

restricted amount of vertical lines in the scene and therefore angle measurements over 15 degrees are not trusted but the roll is set to be zero. Evidently the roll is not totally zero in most cases and setting the roll to be zero introduces errors also to heading and pitch measurements. Fortunately, in most cases these errors are small. In most common cases when the heading changes and roll between two images are less or equal to five degrees, the errors in estimated heading and pitch are 0.6 and 0.1 degrees, respectively. In an extreme case when the camera is simultaneously experiencing roll and heading changes of 15 degrees between two consecutive images, the errors in heading and pitch are 6.1 and 2.8 degrees, respectively. When the camera is otherwise static (i.e. the heading change and pitch are around one degree between consecutive images) even a large roll causes small errors to the observed heading and pitch, namely 0.3 and 0.01 degrees, respectively. Table 4.2 summarizes some errors arising from camera motions between two consecutive images when the roll is erroneously estimated to be zero due to vertical vanishing point calculation failures.

Table 4.2. *Effect of roll error on other angle observations*

Real camera rotation (degrees)			Errors in observation when roll estimated to be zero (degrees)	
Heading	Pitch	Roll	Heading	Pitch
1	1	-15	0.3	0.01
5	5	-5	0.6	0.1
15	-15	15	6.1	2.1
-15	-15	15	5.2	2.8

4.5 Effect of Camera and Setup Characteristics on the Accuracy of the Visual Gyroscope

Image quality, camera rate and mounting location affect the visual motion perception. In this chapter these aspects are considered from the point of view of the visual gyroscope‘s performance.

Many factors affect the accuracy of measurements calculated from images. As discussed before, the vanishing points based rotation observation is dependent on the amount of straight lines found from the image and is therefore reliant on the scene. The failure of the method in unsuitable scenes sets requirements for the image rate; when the image rate is low, the probability of obtaining accurate heading change observation is reduced. Image quality and parameters for algorithms used in visual computations affect the amount and correctness of detected line features. Therefore the characteristics of the camera used as a visual gyroscope is significant for the success of the method. The quality and features of the image sensor are crucial in terms of the amount of noise present in the images. The aperture is the lens diaphragm opening inside the camera lens. It regulates the amount of light passing to the film. The aperture size is indicated by an f-number. Smaller f-number indicates more light is let in and the higher the image quality in low-light situations. The focal length of the camera, discussed in the previous chapter, influences the sharpness of the image. Images taken using a camera with a wide-angle lens, namely a lens with a short focal length, are sharper than the ones taken using a standard lens.

When the pedestrian is navigating and holding the camera in hand, the heading change of the camera may be transformed into the heading change of the user, if

the configuration of the camera with respect to the body of the user is carefully considered. The roll and pitch of the camera may be obtained from the locations of the vertical and central vanishing points, but if the camera is not aligned with the user body, they may not be considered as the orientation of the user. Likewise, if the hand is in motion in the heading direction that is not equal to the motion of the body, the heading change provided by the visual gyroscope is not accurate.

Three different cameras and two different setups were tested in an experiment done mainly indoors in a challenging environment to address all the factors affecting the quality of visual gyroscope's roll, pitch and heading change observations. The three cameras used for the experiments are a GoPro HD Hero helmet camera [**?**] directed for first responders, a Sony HD video camera aimed for extreme sports [**?**] and a Nokia N8 smartphone's camera [**?**]. Table 4.3 summarizes their most important parameters.

The GoPro Hero is a helmet camera developed for first responders and recreational users. Its wide-angle lens, providing tall HD video stream, gives extended field-of-view both in horizontal and vertical directions. The wide-angle lens increases the number of lines found in addition to providing sharper images. The camera has a fixed lens and captures video with a speed of 30 frames per second. The video was converted into still images having an image rate of 10 Hz and resolution of 1280 x 960 pixels for the camera characteristics experiment. The f-number of the lens is 2.8, resulting in increased performance in low-light indoor environments. Besides providing sharp and extensive images, the wide-angle lens produces distortion. The distortion has to be corrected for in order to obtain more accurate vision-based calculations using the method introduced in the previous chapter.

The Sony HXR-MC1 is a camera for recording video during extreme sports or in other high dynamic situations. It has a standard lens with an f-number of 3.2. The camera captures video with a speed of 30 frames per second. The video was converted into still images having an image rate of 10 Hz and resolution of 1440 x 1080 pixels. The images taken with the Sony camera are darker and blurred compared to the images captured using the other two cameras, and the view is more restricted.

The Nokia N8 smartphone camera has a wide-angle lens and an f-number value of 2.8, increasing the performance of the camera in low-light indoor environments. The camera was programmed to capture still images with a 0.8 Hz rate and resolution of

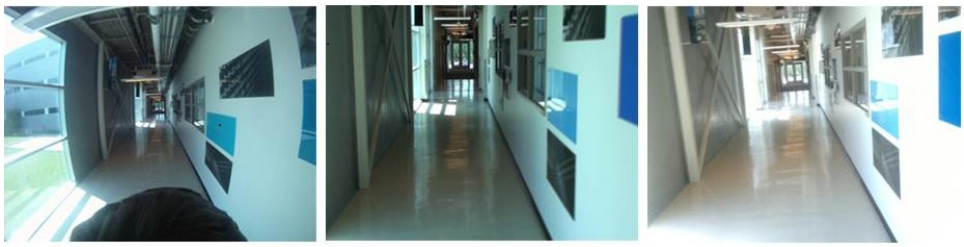

Fig. 4.6. *An image captured of the same scene with three different cameras showing the effect of different camera characteristics on the image quality.*

Table 4.3. *Parameters for GoPro, Sony and Nokia cameras*

Parameters	Camera		
	GoPro Hero	Sony HXR-MC1	Nokia N8
Focal length (35 mm equivalent)	8 mm	79.5 mm	28 mm
f-number	f/2.8	f/3.2	f/2.8
Image rate	10 Hz	10 Hz	0.8 Hz

640 x 480 pixels. The images taken with the Nokia camera have more light than the ones taken using the Sony camera, but they are not as sharp as the ones using the GoPro unit.

Figure 4.6 shows the same scene taken with the GoPro (left), the Sony (middle) and the Nokia (right) cameras. The images show the effect of the different camera parameters discussed above. The effects of the above factors, focal length value and low-light tolerance, may be seen in the figure. The images on left and right are taken with cameras having wide-angle lenses and small f-numbers and therefore the images are sharp, bright and wide, while the image in the middle is taken with a camera having a standard lens and larger f-number, in which case the image quality is poorer.

The camera's sensor is the component that converts the light in the lens projected image into electrical signal to be then digitized. Complementary metal oxide semiconductor (CMOS) and charge coupled device (CCD) chips are the two most used sensor types. The main difference between the two types of sensors is the way they read the information [**?**]. Each sensor consists of millions of light-sensitive photosites which correspond to pixels in an image. In a CMOS sensor information is read from

Fig. 4.7. *Experiment setup for testing the effect of different camera characteristics on the heading change accuracy.*

each photosite individually whereas in CCD a line of photosites is read at once. This makes the CCD sensors simple to design whereas the CMOS sensors are power efficient and therefore widely used in low-cost cameras as in smartphones but with the cost of introducing more distortion into images. All cameras used in the experiments presented in the thesis contain a CMOS sensor.

4.5.1 Experimental Results

The NovAtel SPAN-SE GPS/GLONASS receiver with Northrop Grumman's tactical grade LCI-IMU was used as a reference for both experiments and was carried in a backpack. The first round of experiments was completed using the GoPro and Sony cameras attached to the upper part of the backpack and with the Nokia N8 smartphone in the hand, as is shown in Figure 4.7. The reference was used to evaluate the accuracy of the rotation angles provided by the visual gyroscope using GoPro's and Sony's cameras and the heading change using the Nokia N8 smartphone camera. The second setup consisted of the reference system and all three cameras attached to the upper part of the backpack, enabling the evaluation of the effect of the camera

site to the heading change accuracy. Both setups were tested in experiments with duration of almost 30 minutes each. As the GoPro and Sony video streams were sampled at 10 Hz rate the data collection resulted in 14264 (first round) and 14162 (second round) images using GoPro camera, 14279 (first round) and 14158 (second round) using Sony and 1162 using Nokia N8 smartphone's camera for both rounds. The different number of images for Sony and GoPro was due to the time stamping method, in which a handheld GPS clock was shown to the camera and the first image used was the first image where the time was seen clearly.

The test was conducted on the University of Calgary campus, mainly indoors. The environment was very challenging for the vanishing point based navigation, because it consisted of many turns, doors (i.e. planes) and spacious cafeteria and hall areas (i.e. all lines were not orthogonal and view to lines was restricted in some parts). The pitch, roll and heading change between two consecutive images was computed using the vanishing point based visual gyroscope measuring the full 3D orientation. Because the heading change between two images was evaluated, when the visual gyroscope measurement failed, the heading change was again computed using the subsequent two successful consecutive images.

Statistics of heading change errors for all cameras and two rounds (test 1 and test 2) are shown in Table 4.4. The success rate of the images, based on the error detection algorithm, is also shown in the table. Before computing the statistics the measurements evaluated as erroneous by the error detection were discarded. The mean error in heading change from the visual gyroscope using the GoPro and Sony cameras was around 2.5 degrees, and that using the Nokia N8 was around 4.5 degrees. The percentage of successful images was between 70 and 80 percent for the GoPro camera, around 50 for the Nokia N8 and only around 30 for the Sony camera. While the failed images contained the ones taken during sharp turns, in situations where the light was insufficient and in spacious areas containing non-parallel lines, common for all cameras, the differences in the success rate are explained by the different characteristics of the cameras and are discussed below. Table 4.5 summarizes the roll and pitch error statistics. The roll and pitch accuracy was evaluated only for test 2, because of the lack of a reference system for the rotations of the handheld Nokia N8 in test 1. The mean roll and pitch errors were 0.5 and 2.0 degrees for the Sony, 2.1 and 2.5 degrees for the GoPro and 1.3 and 3.8 degrees for the Nokia N8. The results are elaborated for image quality, image rate, and the location of the camera on the user.

Table 4.4. *Heading change error statistics*

Camera		Statistics of heading change				
		Mean (degrees)	Std (degrees)	Min (degrees)	Max (degrees)	% of images success
GoPro Hero	Test 1	2.5	2.7	0	17	82
	Test 2	2.8	3.0	0	19	72
Sony HXR-MC1	Test 1	2.3	3.3	0	18.6	35
	Test 2	2.6	3.3	0	25	30
Nokia N8	Test 1	4.6	3.7	0	15.5	57
	Test 2	4.4	3.7	0	16	45

Table 4.5. *Roll and pitch error statistics*

Camera		Statistics of pitch and roll			
		Mean error (degrees)	Std (degrees)	Min (degrees)	Max (degrees)
GoPro Hero	roll	2.1	3.1	0	26
	pitch	2.5	4.0	0	59
Sony HXR-MC1	roll	0.5	0.9	0	15.4
	pitch	2.0	3.0	0	22
Nokia N8	roll	1.3	1.9	0	14.4
	pitch	3.8	4.9	0	43

Image Quality

Image quality is an important factor for the success and accuracy of the vanishing point based calculations. As explained before, smaller aperture size and focal length relate to sharper images, especially in low-light situations. When the images are sharp, more lines are found for the vanishing point calculations and they are less noisy. The GoPro camera surpasses significantly the other two cameras in image quality. The images converted from the Sony video stream were grainy and dark and therefore the amount of lines found was reduced. The limited number of lines disturbs the vanishing point calculations, as may be seen from the low success rate of heading change calculations using the images taken with this camera, namely 30 % and 35 %. While the success rate of calculations performed using the GoPro images was high at 72 % and 82 %, the heading change accuracy was slightly worse than when using the Sony camera (mean errors of 2.5 and 2.8 degrees compared to 2.3 and 2.6 degrees). The worse accuracy with better quality images is due to the distortion of the wide angle lens. Though the distortion was corrected before the vanishing point calculations, some effect still remains.

Image Rate

The image rate becomes an important factor when the environment appears challenging. Previous tests employing the Nokia N8 using the same method gave a mean error of 2.5 degrees for heading change observations [P7]. When the image rate is low, namely 0.8 Hz, the environment, consisting of many turns and spacious areas, with reduced amount of lines in view, leaves the method with few good observations. The effect is seen from the reduced accuracy of the visual gyroscope when using the Nokia N8 in which case the mean heading change error is 4.5 degrees as compared to 2.5 degrees with Sony's camera's lower quality images.

Camera Configuration

Surprisingly the camera configuration did not have a significant effect on the accuracy of the visual gyroscope heading change. The heading change of the camera is assumed equal to that of the user when the visual gyroscope is used for pedestrian navigation. When the camera is held in a hand, the change of the hand's posture in

the horizontal direction introduces a heading change in the camera that is inconsistent with the heading change of the user. The roll and pitch magnitudes are also larger compared to those of the configuration where the camera is tied to the user. The heading change accuracy was evaluated with two tests; in the first round the camera was held in the hand and in the second round, tied to the backpack. The difference in the mean errors was only 0.2 degrees between the two tests. The configuration of the camera affects also the success rate of the images. When the camera is held in a hand, it sometimes rotates too much, thereby reducing the sight to the lines and decreasing the success rate of visual gyroscope calculations.

Pitch and Roll

The mean pitch error values were consistent with heading change errors for all cameras. However, the roll error was over two degrees less than for the heading change using the Sony and Nokia N8, namely 0.5 and 1.3 degrees. The method described in this paper calculates the roll based on the ratio between the totally vertical lines having a slope value 0 and the vertical lines with slope deviating from zero, as explained previously. The approach decreases the effect of vertical lines non-parallel to the camera y-axis. When the camera is not experiencing any roll, the number of totally vertical lines surpasses the value of other vertical lines, also non-parallel. These lines, otherwise causing errors to vanishing point calculations, were therefore discarded. The vertical line classification however fails when the distortion is corrected. The correction changes pixel values of all lines' end points and therefore there are no lines left with zero slope value and the ratio based method is no longer valid. The effect was seen from the larger roll error when using the GoPro camera (2.1 degrees).

4.6 Smartphone Application of Visual Gyroscope

The visual gyroscope has been implemented herein in a Nokia N8 Symbian smartphone. The implementation was done using Nokia's C++ based development environment QT [**?**]. The basic visual algorithms (i.e. filtering, edge detection, Hough transform, SIFT and matching) were obtained from OpenCV open source visual library [**?**]. The total processing time for automatically capturing an image, finding the

Table 4.6. *Processing time for different algorithms in visual gyroscope's smartphone implementation*

Algorithm	processing time (s)
Capturing a photo	1.2
Edge detection (Canny)	0.17
Line detection (Hough)	1.0
Vanishing point, heading, tilt	0.07
Total	2.7

straight lines, voting for the vanishing point and calculating the orientation of the camera is on average a bit over two and half seconds, with the specification shown in Table 4.6. The bottlenecks of the calculation are the image capturing (1.2 seconds) and line detection using the Standard Hough Transform algorithm (1 second). The slowness of the Hough Transform was also acknowledged in [**?**] discussing an alternative visual gyroscope implementation in a smartphone. The time used for extracting the lines and calculating the vanishing point has to be decreased for a real-time navigation solution. This will be realized by using a more efficient line detection algorithm as discussed in Chapter 6. Also the effect of using video images instead of the still images has to be addressed in the future in pursuance of concentrating especially on the power consumption aspect.

5. VISUAL ODOMETER

The user translation derived from the accelerometer measurements suffers from errors and therefore a method for providing the information from consecutive images, namely a concept of visual odometer, is developed herein. This chapter discusses the principle of the visual odometer and especially the challenges in observing the translation from consecutive images, i.e. the unknown depth of the objects seen in images and a scale problem arising from it. Then, the visual gyroscope's error detection and performance are discussed as well as how the problems arising from degeneracy are avoided.

5.1 The Principle of the Visual Odometer

Translation of the camera between two consecutive images is constrained with a rule called homography that encompasses the calibration of the camera as well as its rotation and translation between the images as was explained in Chapter 3. The homography equation, reproduced here, is

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \mathbf{K}'\mathbf{t}/Z. \tag{58}$$

When the image points in the first ($\mathbf{x}$) and second ($\mathbf{x}'$) image are normalized homogenous coordinates the relation reduces to

$$\hat{\mathbf{x}}' = \mathbf{R}\hat{\mathbf{x}} + \mathbf{t}/Z \tag{59}$$

where $\mathbf{R}$ is the camera rotation and $\mathbf{t} = [t_x, t_y, t_z]$ the translation between the images. Z represents the distance (depth) of the photographed object from the camera and because it is usually unknown in vision-aided navigation applications, the translation is solved only within an ambiguous scale. In navigation, the absolute magnitude of the translation has to be solved for and some solutions for obtaining the depth Z and

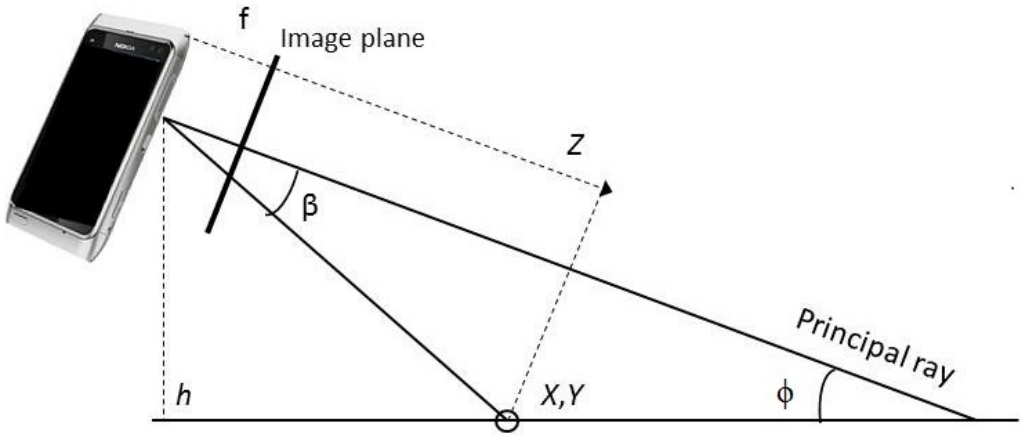

Fig. 5.1. *The special configuration of the camera for resolving the distance (Z) of the object, using the height (h) and pitch (ϕ) of the camera.*

therefore the scale were presented in Chapter 1. The visual odometer presented in this thesis is based on a special camera configuration providing means to resolve the object depth in an unknown environment. The method utilizes the camera rotation obtained using the visual gyroscope and the known height of the camera measured before starting navigation and kept sufficiently static. The definition of sufficient in this context is given later in the chapter.

5.1.1 Measuring the Distance of an Object from the Camera

The distance Z from the camera to the object is calculated using information of the height of the camera (h), the focal length in units of vertical pixels (f_y), and the height of the image in pixels (H) [**?**]. The height of the camera must be known but the focal length and height of the image may be obtained by camera calibration. Figure 5.1 visualizes the configuration for resolving the depth Z of an object having coordinates (*X, Y, Z*) and using the parameters listed above, camera pitch ϕ computed by the visual gyroscope and β computed as follows.

When the image height (H) and the vertical component of the focal length (f_y) are known , the vertical field-of-view (*vfov*) may be calculated as

$$vfov = \arctan\left(\frac{H}{f_y}\right) \tag{60}$$

and the angle (β) between the principal ray of the camera and the ray from the camera to the object as

$$\beta = \arctan\left(\left(\frac{2y}{H} - 1\right) tan\left(\frac{vfov}{2}\right)\right). \tag{61}$$

Finally, using β and the pitch of the camera ϕ, the distance Z is obtained as

$$Z = \frac{h\cos(\beta)}{\sin(\phi + \beta)}. \tag{62}$$

In order to be able to determine the user translation by using the special configuration, the object is required to lie in close vicinity of the camera, namely between the camera and the point where the principal ray intersects with the floor plane with the prevailing pitch angle. The vicinity requirement is rational also in the sense that the motion of the far-off objects is very small in term of pixels and may therefore be overwhelmed by noise. Also, the method for resolving the ambiguous scale using the known height of the camera requires the image points to lie on or close to the floor plane. Experiments have shown [P4] that lines used by the visual gyroscope to find the vanishing point are mainly found from the floor, namely from the junction of the floor and walls, especially with a camera having a pitch angle larger than zero towards the floor plane. If no such points are found, a coarser method is introduced, considering all points found below the vanishing point, or if the vanishing point is not found, the principal point. As the method presented in the thesis uses the rotation perceived by the visual gyroscope, the amount of matching image points needed is reduced and therefore the limitation of the region for finding suitable objects does not incur substantial limitations for the translation observation.

SIFT features are extracted from the two consecutive images and matched using Matlab algorithms [**?**] and the restrictions of the object's location described above; matching is shown in Figure 5.2. The lines join image points matched in the consecutive images (first on left and second on right), the red dot is the central vanishing point and the green point on the floor of the left image is the only matched point inside the region suitable for the visual odometer. Due to the low amount of features in indoor environments, less certain matches are accepted, so that some matches are found for most images. Now the translation of the camera between the two images may be resolved from the matched normalized homogenous image points $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$ in the first and second image, respectively, using (59). When the image points followed are projections of the objects lying on the floor, the translation matrix z- and x-component show

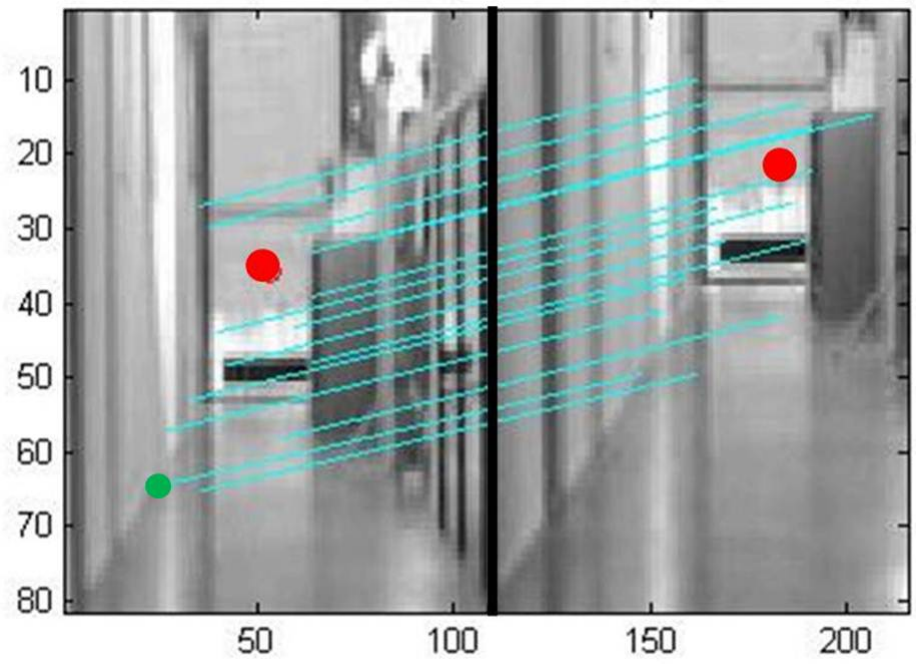

Fig. 5.2. *Matched Sift features between consecutive images.*

the horizontal translation that may further be transformed into translation in East and North directions using the visual gyroscope induced heading. The loose matching criteria, as well as the occasional use of the coarse floor plane recovery, necessitate careful error handling for the robust visual odometer measurements which will be explained next. Also the ambiguity detection will be discussed.

5.1.2 Error Detection and Ambiguity Resolving for the Visual Odometer

As was explained in Chapter 3, the image point presented with homogenous coordinates $\mathbf{x} = (x, y, 1)$, is related to the corresponding object coordinates $\mathbf{X} = (X, Y, Z, 1)$ as

$$\mathbf{x} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X} \tag{63}$$

where $\mathbf{t}$ is the translation of the camera center, $\mathbf{R}$ its orientation with respect to the world coordinate frame center and camera matrix $\mathbf{P}$ is such that $x = \mathbf{PX}$. When the configuration is done in a manner that the location and attitude of the camera while

capturing the first image is set as the world frame center, the attitude and location of the second image's points reflect the change of the location and attitude of the camera between the images, i.e. rotation and translation of the camera.

The coordinates $(\mathbf{X})$ of an object represented by two image points $(x, y, 1)$ and $(x', y', 1)$ in consecutive images are estimated using triangulation as [**?**]

$$\begin{aligned} x\mathbf{p}_3^T\mathbf{X} &= \mathbf{p}_1^T\mathbf{X} \\ y\mathbf{p}_3^T\mathbf{X} &= \mathbf{p}_2^T\mathbf{X} \\ x'\mathbf{p}_3^T\mathbf{X} &= \mathbf{p}_1^T\mathbf{X} \\ y'\mathbf{p}_3^T\mathbf{X} &= \mathbf{p}_2^T\mathbf{X} \end{aligned} \tag{64}$$

where $\mathbf{p}_i$ represents the *i-th* row of the camera matrix $\mathbf{P}$. The four equations may be expressed in a matrix form as $\mathbf{AX} = 0$ for a suitable $\mathbf{A}$. An estimation $\hat{\mathbf{X}}$ for $\mathbf{X}$ may now be obtained using the Linear-Eigen method, namely it is the unit eigenvector corresponding to the smallest eigenvector of $\mathbf{A}^T\mathbf{A}$ minimizing $|\mathbf{A}|$ and having the condition $|\mathbf{X}| = 1$. This is done using the Singular Value Decomposition (SVD). SVD is a factorization of a matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{UDV^T}$ [**?**]. Matrices $\mathbf{U}, \mathbf{V}$ are orthogonal and $\mathbf{D}$ diagonal with non-negative values. The decomposition of a matrix $\mathbf{M}^T\mathbf{M}$ is $\mathbf{M}^T\mathbf{M} = \mathbf{VD}^2\mathbf{V}^{-1}$ and the values of $\mathbf{D}^2$ are its eigenvalues and columns of $\mathbf{V}$ eigenvectors.

Because the object should be lying on the floor plane, the Y-coordinate of the estimation $\hat{\mathbf{X}}$ has to be equal to the height of the camera, and the scale ambiguity present in the image homography may be solved. The scale factor is the ratio of the camera height and the object Y-coordinate $\frac{h}{Y}$ [**?**]. When the translation obtained using the homography is multiplied by the scale factor, the real translation in the horizontal plane is obtained, and the translation in the vertical direction is assumed to be zero based on the configuration requiring the camera height to be static.

The reprojection error is a measure used for evaluating if the image points are matched correctly and is used for discarding erroneously matched points [**?**]. It is done by estimating the object point $\hat{\mathbf{X}}$ from the image point correspondences $\mathbf{x}, \mathbf{x}'$ as described above and then reprojecting the estimated object point to the matched image points. Error detection in the case of the visual odometer cannot be based on monitoring the reprojection error for the motion is mainly forward and therefore the rays from the

camera to the image point in consecutive images are nearly parallel. Herein the Y-coordinate values are monitored and the ones deviating more than a threshold from the mean of all observations are discarded. The deviation is constrained as

$$Y_i - \mu(\mathbf{Y}) < 2\sigma(\mathbf{Y}) \tag{65}$$

where Y_i is the *i-th* object's Y-coordinate, $\mu(\mathbf{Y})$ the mean of all obtained Y-coordinate values and $\sigma(\mathbf{Y})$ their standard deviation.

5.1.3 Degeneracy

Degeneracy problems arise from special situations while resolving motion of the camera from consecutive images. When the camera rotates about its centre, or the camera is static but its intrinsic parameters change between the images, a motion degeneracy arises [**?**] because the epipolar geometry between the consecutive images is not defined. Structure degeneracy arises in the case when all image points used for matching are coplanar because then the epipolar geometry between the images cannot be uniquely determined. This is due to the fact that the camera matrix $\mathbf{P}$ presented in Chapter 3 has 11 degrees of freedom and when the image points are coplanar they define a homography with only 8 degrees of freedom.

The visual odometer proposed in this thesis resolves the degeneracy problems as follows. When the depth of the object (Z) computed for matched image points as explained above is constant for two consecutive images, the translation between the images is set to zero. Thus, the homography resulting in errors due to the motion degeneracy is not computed. As the heading is computed using the visual gyroscope the heading measurement does not suffer from the degeneracy problem. The structure degeneracy arising from the image points being planar is avoided from the configuration of the translation and rotation solution. As the camera is calibrated a priori, the rotation is obtained from the visual gyroscope and the translation in the vertical direction is assumed zero, and in theory only one matched image point between the consecutive images is needed to resolve the horizontal translation. Therefore the image points may be coplanar, the missing 3 degrees of freedom needed to resolve the camera matrix are already solved using these other visual parameters.

5.1.4 Performance of the Visual Odometer

The visual odometer provides relative information about the user position, i.e. translation, and therefore the initial position has to be obtained using an absolute positioning system. If the camera used is calibrated, the visual odometer does not need additional calibration before or during navigation, however the performance of the navigation system substantially increases if the absolute position of the user is occasionally calibrated, reducing the effect of unavoidable error occurrences in visual observations. The visual odometer does not depend on any knowledge of the environment, but only the camera height must be estimated and kept sufficiently static. However, its performance is dependent on the accuracy of the visual gyroscope's measurements. The most drastic errors may be avoided by monitoring changes in pitch; if the change is considerable it is most likely due to an error in vanishing point location and in this case the previous pitch and heading values are used. The method of the visual odometer is not as tolerant to dynamic objects as the visual gyroscope, but again the error arising from monitoring the motion of a dynamic object depends on how many matching static points are found. The mean error of the user speed obtained in different navigation environments was found to be less than 0.3 m/s, with a standard deviation of 0.3 m/s. Analysis of the errors arising from different navigation environments will be discussed in the following chapter.

The camera height has to be evaluated a priori and kept sufficiently constant during navigation for the visual odometer to perform correctly. The effect of using an incorrect height estimate on the visual odometer's performance due to erroneous a priori measurement or failure in keeping the camera at a constant height, which would naturally occur when using smartphone in hand, was evaluated. The statistics from the visual odometer perceived user speed while the height of the camera was correct and kept constant were compared to a situation where the height of the camera was ± 10 cm and ± 30 cm off the correct value. In an experiment done in an office corridor and resulting in 183 images, the mean error in speed was 0.26 m/s when the correct height was used and no effect was seen when a height value with an error of -10 cm was used. However, when the height was erroneous with the same magnitude in the other direction, the mean error increased to 0.38 and 0.54 m/s when the error was 10 and 30 cm, respectively. Table 5.1 shows the statistics. The results show that a vertical motion of the camera less than or equal to 10 cm does not deteriorate the performance of the visual odometer substantially, hence the upwards the motion may

Table 5.1. *Statistics of the effect of camera height for visual odometer's speed accuracy, units are in m/s*

Statistics	min error	max error	mean error	std of error
Correct height	0	1.5	0.26	0.24
Height -10 cm	0	1.4	0.26	0.26
Height -30 cm	0	0.8	0.28	0.18
Height +10 cm	0	1.5	0.38	0.29
Height +30 cm	0	1.5	0.54	0.29

be even larger.

6. VISION-AIDED NAVIGATION USING VISUAL GYROSCOPE AND ODOMETER

This chapter discusses pedestrian indoor and urban navigation solutions utilizing a visual gyroscope and a visual odometer to obtain a vision-aided integrated system. The collected data, visual and other measurements from different sensors and radio positioning systems, are integrated using a Kalman filter. The durations of tests and test environments are varied in order to obtain an extensive understanding of the suitability of the algorithms for real life pedestrian navigation.

Due to the different start times and measurement rates of different systems, all measurements have to be time stamped carefully. Because cameras do not usually provide a time stamp to images in Coordinated Universal Time (UTC) like other sensors often do, but in relation to the camera's own clock, the time for images has to be resolved differently. In this thesis this is done using two different methods, namely by initializing the system through keeping all sensors static and then looking at the time of the start of motion from images or alternatively by showing a handheld GPS clock to the camera before starting navigation. The initialization of the world frame with respect to the navigation frame is explained separately for each implementation discussed. Calculations for all experiments are done in post-processing using Matlab. The experiment setups and results are then discussed.

6.1 Visual Gyroscope and Odometer Aided Multi-Sensor Positioning

The measurements from the visual gyroscope detecting heading and pitch changes and the translation from the visual odometer were integrated with GPS position obtained using a Fastrax IT500 high-sensitivity receiver (sensitivity being -165 dBm in navigation), WLAN fingerprinting observations from Nokia N8 smartphone, and speed and heading measurements from the MSP (multi-sensor positioning) device [**?**]

as well as Nokia 6710 accelerometers and magnetometers using the Kalman filter presented below and explained in detail in [**?**]. The GPS receiver, as well as the reference system, were initialized outdoors. The visual measurements were calculated from images taken using the Nokia N8 camera that has a resolution of 12 Mega pixels. Processing of such large images is very time consuming and therefore the resolution was reduced to 640×480 to enable real-time performance. A NovAtel SPAN (Synchronized Position Attitude Navigation) GPS/INS high-accuracy positioning system was used as reference. The equipment was placed in a cart using the setup shown in Figure 6.1. The start position was set to be the origin of the navigation frame and heading was initialized using the reference solution at the beginning of the experiment. This initial heading was used for setting up the visual gyroscope's world frame, hence at the initial point the visual gyroscope's heading was stated to be equal to the initial heading and during navigation heading changes between consecutive images were monitored to propagate the heading. As the camera was attached to a holder in a cart its roll was incremental and therefore the visual gyroscope measuring only the pitch and heading change, presented in Chapter 4 was used. The time synchronization in the experiments testing the performance of vision-aided multi-sensor positioning was done by monitoring the motion start time from the self-contained sensors and the images.

The experiments were done in an office corridor first by obtaining WLAN position measurements from a functional WLAN radio map and then using an outdated map, otherwise the setup was the same for both rounds. The relatively short test time is due to the difficulties in obtaining an accurate reference trajectory for a longer time indoors. The results are however anticipated to produce similar accuracy for longer time periods due to the possibility to obtain occasional absolute position updates. The results from the experiments show that the vision-aiding increases the accuracy, precision and availability of the navigation solution as described below.

6.1.1 Kalman Filter Used in Multi-Sensor Positioning

A constant speed model, which is often used in pedestrian navigation, was also used and is defined as

Fig. 6.1. *Equipment setup for experimenting the vision-aided multi-sensor positioning. The Nokia N8 phone acquiring the images was attached to a holder in the front of the cart. The GNSS antenna is that of the NovAtel SPAN reference system.*

$$\begin{aligned} X_{k+1} &= X_k + \dot{X}_k \Delta t + v_1 \\ Y_{k+1} &= Y_k + \dot{Y}_k \Delta t + v_2 \\ \dot{X}_{k+1} &= \dot{X}_k + v_3 \\ \dot{Y}_{k+1} &= \dot{Y}_k + v_4 \end{aligned} \tag{66}$$

where X and Y are the latitude and longitude transformed into the ENU (East, North, Up) coordinate frame, $\dot{X}$ and $\dot{Y}$ are their time derivatives, k denotes the current epoch, Δt is the time interval between two epochs, and v_i is the state uncertainty component of the element i. The state vector $(\mathbf{x})$, transition matrix (Φ) and process noise matrix $(\mathbf{Q})$ for the model are

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{X}_k \\ \mathbf{Y}_k \\ \dot{\mathbf{X}}_k \\ \dot{\mathbf{Y}}_k \end{bmatrix} \quad \Phi_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{Q}_k = \begin{bmatrix} \tilde{q}_1 \frac{\Delta t^4}{4} & 0 & \tilde{q}_1 \frac{\Delta t^3}{2} & 0 \\ 0 & \tilde{q}_2 \frac{\Delta t^4}{4} & 0 & \tilde{q}_2 \frac{\Delta t^3}{2} \\ \tilde{q}_1 \frac{\Delta t^3}{2} & 0 & \tilde{q}_1 \Delta t^2 & 0 \\ 0 & \tilde{q}_2 \frac{\Delta t^3}{2} & 0 & \tilde{q}_2 \Delta t^2 \end{bmatrix} \tag{67}$$

where $\tilde{q}_1$ is a spectral density value for the North component and $\tilde{q}_2$ for the East component, both having a value of $(2\ m)^2/s$ based on an empirical assessment of the quality of the sensors in the hardware platform. The measurements consisted of position, speed and heading from GPS, WLAN (position), Nokia phone (ACC1) and MSP (ACC2) accelerometer as well as the visual odometer (V) (speed), and Nokia phone (DC1) and MSP (DC2) digital compasses as well as the initialized visual gyroscope (V) (heading). The measurement vector $\mathbf{z}$ is

$$\mathbf{z}_k = \begin{bmatrix} X_{GPS} \\ Y_{GPS} \\ X_{WLAN} \\ Y_{WLAN} \\ S_{ACC1} \cos\theta_{DC1} \\ S_{ACC1} \sin\theta_{DC1} \\ S_{ACCV} \cos\theta_V \\ S_{ACCV} \sin\theta_V \\ S_{ACC2} \cos\theta_{DC2} \\ S_{ACC2} \sin\theta_{DC2} \end{bmatrix}. \tag{68}$$

Matrices $\mathbf{H}$ and $\mathbf{R}$ are

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{69}$$

$$R_k = diag(\sigma^2_{X_{GPS}}, \sigma^2_{Y_{GPS}}, \sigma^2_{X_{WLAN}}, \sigma^2_{Y_{WLAN}}, \sigma^2_{S_{ACC1}\cos\theta_{DC1}}, \sigma^2_{S_{ACC1}\sin\theta_{DC1}}, \sigma^2_{S_{ACCV}\cos\theta_V}, \sigma^2_{S_{ACCV}\sin\theta_V}, \sigma^2_{S_{ACC2}\cos\theta_{DC2}}, \sigma^2_{S_{ACC2}\sin\theta_{DC2}}). \quad (70)$$

The variance values $\sigma^2_{i_{DEV}}$ were chosen by testing the accuracy of the corresponding measurements (i) obtained using the device (DEV). The devices utilized had different measurement rates, i.e. 36 Hz for self-contained sensors, 1 Hz for GPS measurements, 0.8 for the visual gyroscope and odometer and 0.1 Hz for WLAN; the size of the matrices therefore varied based on the number of measurements obtained for the epoch under consideration.

6.1.2 Test in an Indoor Office Environment

The performance of the vision-aided multi-sensor positioning using the visual gyroscope and the visual odometer was tested in a typical office corridor as shown in Figure 6.2; in this case however the corridor was suffering from low lighting due to the dark North European winter. The data collection lasted three minutes and resulted in 148 images with successful visual calculations. Although the images were taken with a 0.8 Hz rate the vanishing point location calculation occasionally failed due to scenes that were too dark or corridor turns and therefore no visual gyroscope or visual odometer observations were obtained for a small percentage of the images captured. The error in cumulative distance obtained when propagating the position using the visual odometer measurements was only 1.3 m, with the total length of the route being 158 m. However, the result is over optimistic as may be seen in Figure 6.3 that shows the variation of the speed measurements. The mean error of the speed was 0.3 m/s and the standard deviation 0.3 m/s. In this experiment the user speed was not filtered as may be seen in the figure, but in the following experiments the user speed will be filtered with an upper limit of 1.5 m/s, which is considered a reasonable assumption for normal pedestrian navigation.

As the magnitude of the steep turns (i.e. 90 degrees or over) may not be observed using the visual gyroscope and no other feasible heading measurements were available in this and the following office experiments, the turns were detected as follows: In turning situations where the turns were sharp and the vanishing points not perceived, the visual odometer observed the turns as being the only regions where the

Fig. 6.2. *Office corridor used for evaluating the vision-aided multi-sensor implementation.*

matching failed completely and no corresponding image points were found. When only one matching image pair was lost, the turn was assumed to be most likely 90 degrees and when three points were lost the turn was evaluated to be 180 degrees. If some kind of map matching [**?**] were used as further augmentation, this procedure would become more feasible. The visual gyroscope and odometer induced heading changes and speed information were integrated with the other measurements using the Kalman filter described above and the user position during navigation was obtained. The position solution obtained was compared to a solution using only GPS or WLAN positions as measurements to the filter as well as to the integrated solution using all other measurements but the visual. All solutions obtained were compared to the ground truth obtained from the SPAN reference system. The vision-aided fused solution, i.e. integration of the visual gyroscope and odometer measurements, provided the best user position accuracy and precision, the mean error being 5.3 m and the standard deviation being 3.8 m. The corresponding values for the fused solution without vision aiding are 6.7 and 5.1 m, for GPS positioning 17.8 and 11.5 m and for WLAN positioning 5.9 and 4.7 m. The availability of the other positioning systems, when the solution is computed at 1Hz rate, is 100% except for the WLAN, for which the availability is only 10% due to the low update rate of the solution in

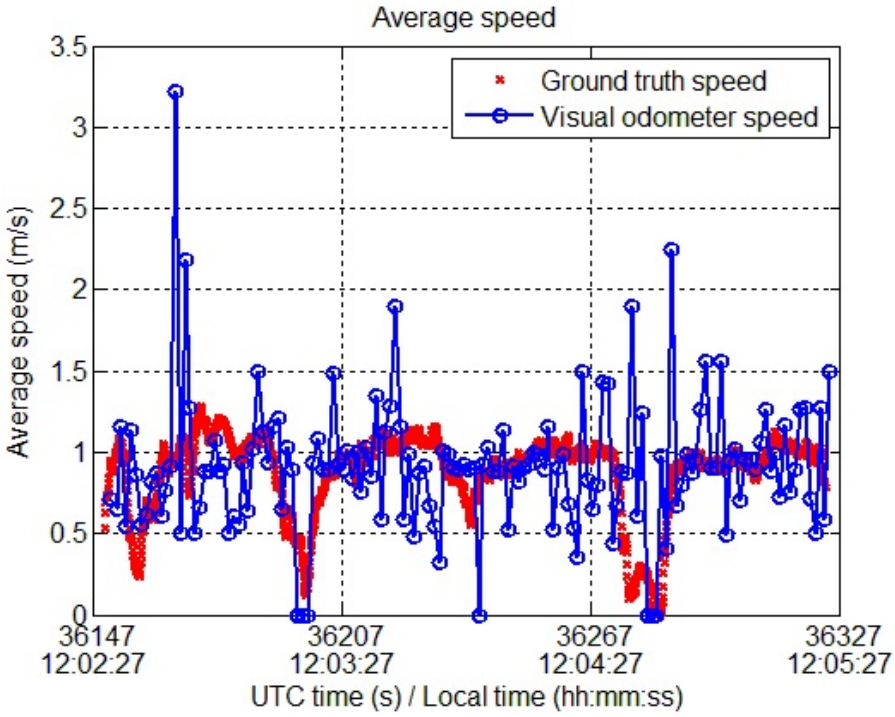

Fig. 6.3. *Average speed from ground truth (red) and visual odometer (blue).*

Table 6.1. *Positioning error statistics using different systems in an office corridor*

Statistics	min error (m)	max error (m)	mean error (m)	std of error (m)	availability (%)
WLAN	1.6	17.2	5.9	4.7	10
GPS	0.6	38.7	17.8	11.5	100
Fused	0.9	23.4	6.7	5.1	100
Vision-aided fused	0.9	19.7	5.3	3.8	100

consequence of power saving requirements of the smartphone. Table 6.1 shows the statistics. The position result is visualized in Figure 6.5 showing on the left the fused solution without vision-aiding (blue), GPS position (black), WLAN position (purple) and the ground truth (green) and on the right the vision-aided fused solution (blue), GPS position (black), WLAN position (purple) and the ground truth (green).

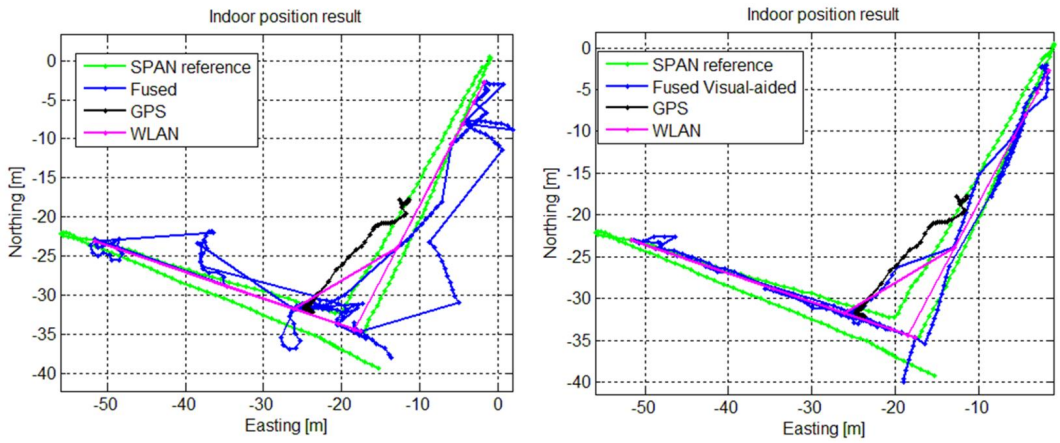

Fig. 6.4. *Indoor positioning results for a pedestrian using different sensors and fused solutions. The figure on the left has a fused solution without visual-aiding (blue) and the figure on the right a fused solution using visual-aiding from the visual gyroscope and visual odometer (blue), ground truth (green), GPS position solution (black) and WLAN position solution (purple).*

6.1.3 Test in Office Environment Using an Outdated WLAN Radio Map

An experiment using the same setup, equipment and methods as above was carried out this time using an outdated WLAN radio map and therefore deteriorating the absolute position calibration during the navigation. This test was performed to assess the impact of an incorrect WLAN map as such maps are known to change frequently due to human traffic and other changes. Two of the eight WLAN access points in the office corridor were out of order and two had changed locations after the formation of the radio map. Also, some new electrical equipment was placed to the vicinity of one access point. Thus, this altered setup caused the WLAN positioning accuracy to be reduced to 11 meters from the previous 6 m. The use of the visual odometer induced user speed filtering explained above decreased the mean error in the speed from 0.3 m/s to 0.26 m/s with a standard deviation of 0.24 m/s. Now the fused position solution without vision-aiding had a 7.8 m mean error and a standard deviation of 4.8 m while the vision-aided fused solution using the visual gyroscope and odometer measurements had a mean error of 5.8 m and a standard deviation of 3.7 m. The statistics are shown in Table 6.2. The problems in the absolute position calibration due to the outdated WLAN radio map increase the fused position solution mean error

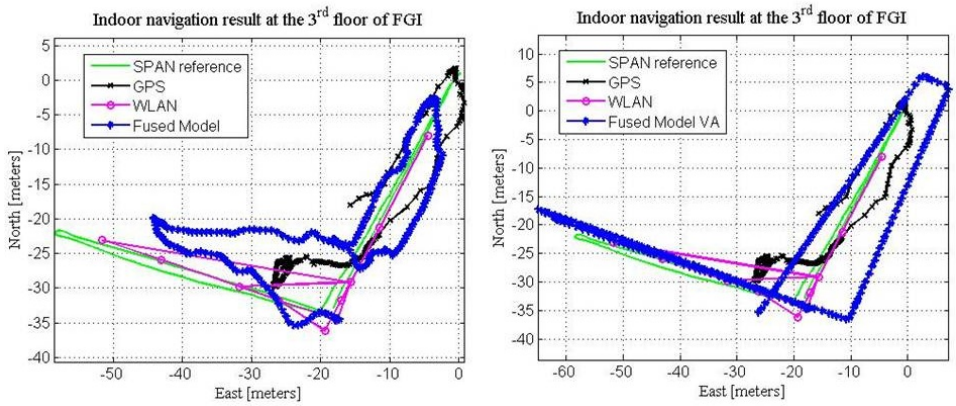

Fig. 6.5. *Indoor positioning results for a pedestrian using different sensors and fused solutions in an office corridor with an outdated WLAN radio map. The figure on the left has a fused solution without visual-aiding (blue) and the figure on the right a fused solution using visual-aiding from the visual gyroscope and visual odometer (blue), ground truth (green), GPS position solution (black) and WLAN position solution (purple).*

Table 6.2. *Positioning error statistics using different positioning systems in an office corridor with an outdated WLAN radio map*

Statistics	min error (m)	max error (m)	mean error (m)	std of error (m)	availability (%)
WLAN	0.5	36.9	11.0	10.9	10
GPS	0.3	32.6	14.9	8.4	100
Fused	0.5	16.1	7.8	4.8	100
Vision-aided fused	0.4	13.1	5.8	3.7	100

by more than 1 m compared to the fused solution using a valid WLAN radio map, but the degradation is not so significant for the vision-aided fused solution, namely only 0.5 m. The position result is visualized in Figure 6.5 showing on the left the fused solution without vision-aiding (blue), GPS position (black), WLAN position (purple) and the ground truth (green) and on the right the vision-aided fused solution (blue), GPS position (black), WLAN position (purple) and the ground truth (green).

6.2 Stand-Alone Visual System

The two previous experiments assessed the performance of a visual gyroscope and odometer aided fused navigation solution in an environment most favorable for the visual methods developed herein and consisting of scenes having good line geometry and only a few dynamic objects. Because pedestrian navigation is needed in various environments, the method was tested in the most challenging environment for the visual aiding approach, namely in a shopping mall with shoppers in motion, wide corridors restricting frequently the view of their sides and therefore making the line geometry degraded, varying lighting conditions and various objects forming many non-orthogonal lines. As the method is also aimed for urban navigation, it was tested in an urban environment, frequently close to the wall of a tall building. As these environments lacked an absolute positioning system for integration, the performance of the visual gyroscope and odometer was tested as a stand-alone system, propagating the initial position and orientation using the visual heading and speed measurements. The visual measurements were calculated from images taken using the Nokia N8 camera and reduced resolution as in the previous experiments. Again, a NovAtel SPAN GPS/INS high-accuracy positioning system was used as reference. The equipment was placed into a cart as in the previous experiments. The visual measurements were propagated using a Kalman filter implemented as follows. The results obtained were compared to the ground truth and are discussed below.

6.2.1 Kalman Filter Used in Stand-Alone Visual Positioning

In this case, the navigation solution is obtained by propagating the initial position and heading using the visual gyroscope induced heading information and visual odometer speed. The propagation is done using a straightforward Kalman filter modeling the user position as

$$\begin{aligned} X_{k+1} &= X_k + S_{k+1}\Delta t \sin\theta_{k+1} + v_1 \\ Y_{k+1} &= Y_k + S_{k+1}\Delta t \cos\theta_{k+1} + v_2 \end{aligned} \tag{71}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the latitude and longitude transformed into the ENU frame, S is the speed of the user from the visual odometer, θ is the heading change obtained using the visual gyroscope, Δt is the time interval between the current epoch k and

the consecutive epoch $k+1$. v_1, v_2 are the state uncertainty components of elements $\mathbf{X}, \mathbf{Y}$ and the state vector $\mathbf{x}_k = [X\ Y]^T$. The $\mathbf{\Phi}, \mathbf{H}$ and $\mathbf{Q}$ matrices for the filter are

$$\mathbf{\Phi} = \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 2^2 & 0 \\ 0 & 2^2 \end{bmatrix}. \tag{72}$$

The variances in the measurement covariance matrix $\mathbf{R}_k = diag(\sigma_X^2, \sigma_Y^2)$ are based on the performance evaluation of the visual gyroscope and odometer and change during the processing based on the error detection results, i.e. when there is a high probability of an error occurrence the values are increased and therefore less weight is given to the measurement.

6.2.2 Test in a Shopping Mall Environment

An experiment lasting 420 seconds was done in the Iso Omena shopping mall in Espoo, Finland. The challenging environment deteriorated the visual gyroscope's performance as was already seen in Chapter 4, with the mean heading error being 4.4 degrees and the standard deviation 0.2 degrees. Figure 6.6 shows the challenges set by the environment. The incomplete line geometry due to the wide corridor and richness of objects decreases the number of straight lines found from the original image on the left and shown in the image on the right. The lines found are classified as was explained in Chapter 4, and the erroneous vanishing point found is shown as a red dot. The figure shows also a challenge set for the visual odometer, namely the shoppers in motion. However, the visual odometer induced speed did not suffer from the environment due to the increased number of objects and therefore matched image points found in the environment. The mean error of the visual odometer's speed observations was 0.25 m/s with a standard deviation of 0.2 m/s. The visual odometer propagated path was 179 meters, the total route length being 198 meters, hence yielding an agreement of 90%. Figure 6.7 shows the position solution obtained by propagating the initial position and heading using the visual gyroscope and odometer measurements (green) and the ground truth (red). As these results show the solution without any absolute positioning method calibrating the position and heading during navigation or aiding the solution in the occurrence of errors, the positioning accuracy is expected to increase substantially when the stand-alone visual

Fig. 6.6. *An image captured by the smartphone on the left and after processing using the visual gyroscope algorithm on the right. The lines found from the image are classified and colored, the ones in the direction of propagation and used for the vanishing point calculation being in blue. The central vanishing point is shown as a red dot.*

gyroscope and odometer measurements are integrated with other radio positioning and sensor measurements.

6.2.3 Test in an Urban Canyon

As pedestrian navigation is needed also in urban environments, the suitability of the developed vision-aiding method was tested in an outdoor environment, namely close to a wall of a tall building deteriorating and occasionally totally blocking line of sight GPS observations. The challenges set by the environment may be seen in Figure 6.8. On the left is the result of a successful vanishing point observation despite the presence of a dynamic object (a vehicle in this case) and bright lighting loosing edges. On the right, these challenges disturb the vanishing point location observation, the dynamic object (a human) introduces a number of non-orthogonal lines surpassing the number of lines going in the direction of propagation, a phenomena that is also partly due to the bright lighting disturbing the edge detection. The mean error of the heading obtained using the visual gyroscope in the experiment lasting for 270 seconds and resulting in 224 images was 3.3 degrees with a standard deviation of 5.1 degrees, therefore the accuracy in this urban canyon falls between the office corridor and the mall environment. The mean error in the visual odometer induced user speed was 0.2 m/s with a standard deviation of 0.23 m/s. The route length was 146 meters and the cumulative distance obtained using the visual odometer 131 meters, yielding

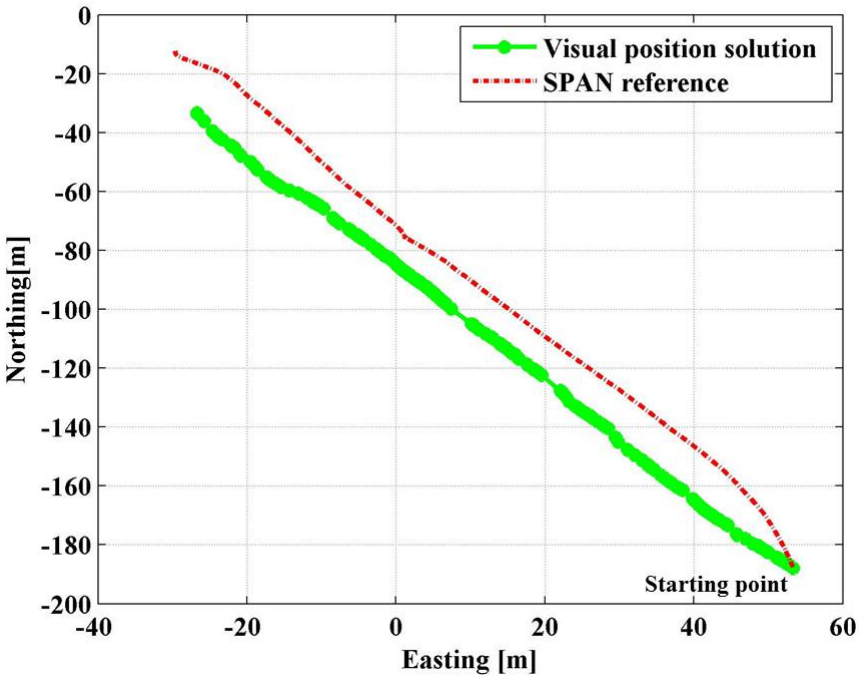

Fig. 6.7. *The two-dimensional position solution in the Iso Omena shopping centre with visual stand-alone solution (green) and SPAN reference (red).*

an agreement of 90%.

The navigation solution was again computed as a stand-alone visual system by propagating the initial position and heading using the Kalman filter explained above and the visual gyroscope and odometer measurements. The obtained position solution was compared to the one obtained using a Fastrax IT500 high-sensitivity GPS receiver, a typical consumer-grade L1 high-sensitivity GPS receiver. Figure 6.9 shows the position solution obtained using each system, with visual stand-alone system on the left and the IT500 GPS receiver on the right. The different and therefore complementary natures of the two positioning systems are seen in the figure, namely the visual position solution is immediately on the correct track but drifts slowly during navigation, while the GPS solution takes 13 seconds to converge into the correct position but is accurate thereafter after observing the necessary satellite geometry and a sufficient amount of good-quality satellite signals, as is the case throughout this experiment done in a modest urban canyon. The mean error of the visual stand-alone position

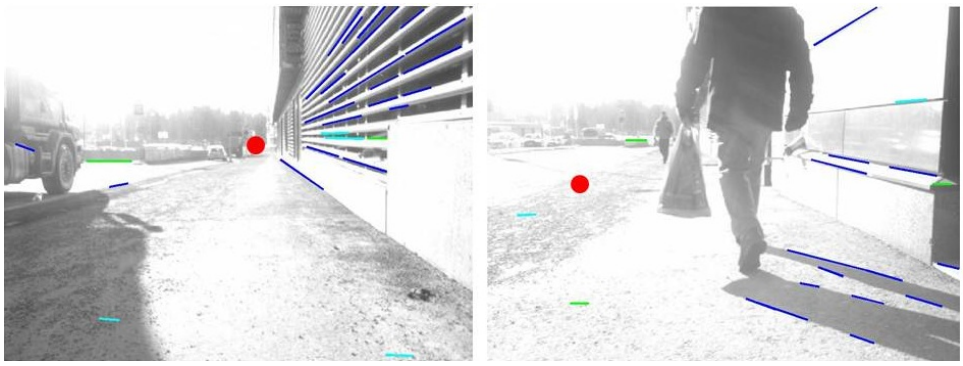

Fig. 6.8. *Challenging environment of an urban canyon having bright light results in successful vanishing point detection when the line geometry is favorable (in left) or in errors when dynamic objects create disturbances (on right).*

Table 6.3. *Positioning error statistics for visual stand-alone and GPS position solutions*

Statistics	min error (m)	max error (m)	mean error (m)	std of error (m)
Visual stand-alone system	0.3	22.5	10.3	5.8
GPS	0.4	51.0	16.7	10.1

solution was 10.3 meters with a standard deviation of 5.8 meters. The corresponding measurements for GPS solution were 16.7 and 10.1 meters, Table 6.3 shows the statistics. The experiment demonstrates how the two systems complement each other and when an integrated the solution is anticipated to improve significantly. In this section it was also shown that the performance of the stand-alone visual system is comparable to that of the other positioning systems. A harsher urban canyon situation will be discussed in Chapter 7 where the concept of vision-aided carrier phase navigation is introduced.

6.3 Visual Gyroscope Aided IMU Positioning

In the previous sections the experiments using a vision-aided multi-sensor positioning system, calibrating the position solution occasionally using absolute position information obtained from WLAN, and a visual stand-alone navigation system were described. The accuracy of the visual stand-alone system drifts in time due to the

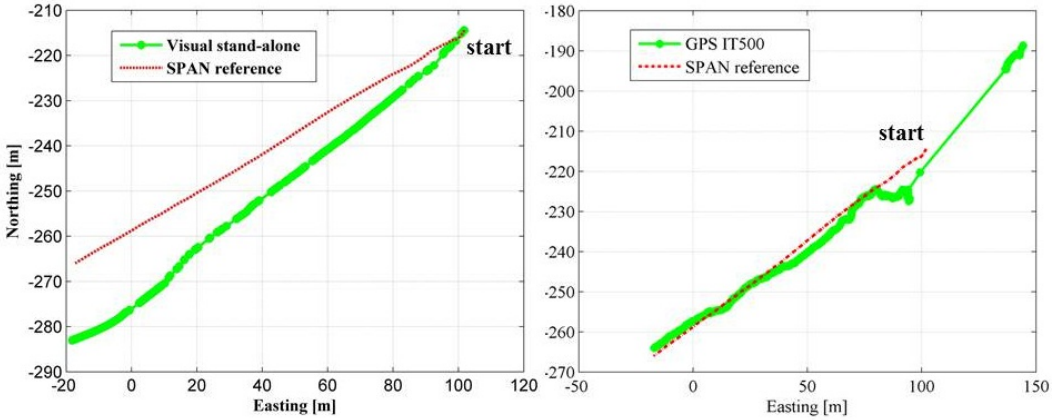

Fig. 6.9. *Position solutions obtained in an urban canyon using a visual stand-alone system (green, on left) and an IT500 GPS receiver (green, on right). The ground reference is shown in red.*

various errors discussed previously and WLAN based positioning needs a priori prepared environment, therefore other means for navigation need to be addressed. Self-contained sensors carried by the user and presented in Chapter 2 are desirable for positioning. With a known initial position, the current position may be propagated for a limited time using a triad of gyroscopes and accelerometers. The deficiency of the self-contained sensors is the cumulative measurement errors that affect the accuracy of the attitude obtained from the gyroscopes. Herein a method of updating the navigation filters attitude using vision-aiding and thereby providing accurate absolute user position for indoor navigation is presented. The method uses only equipment carried by the user, thus not requiring any additional infrastructure.

The method incorporates the visual gyroscope induced attitude as updates in a filter integrating also the GPS position and Analog Devices ADIS16488 inertial measurement unit (IMU) data. The ADI IMU is a high grade MEMS IMU, with a $12^\circ/hr$ in-run bias stability and $1620^\circ/hr$ noise level [**?**] providing measurements at 200 Hz rate. The NovAtel SPAN-SE GPS/GLONASS receiver with a Northrop Grumman's tactical grade LCI (low-coherency interferometry) IMU was used as a reference system and carried in the backpack for both experiments. The visual gyroscope measurements are obtained from images taken with a GoPro camera, discussed in more detail in Chapter 4. The filter used is a tightly coupled 21-state extended Kalman

filter (EKF) [**?**] and will be discussed below. The visual odometer measurements are not integrated in this method and therefore the speed is obtained using the IMU accelerometer alone. The visual and IMU measurements are time synchronized by showing a handheld GPS receiver clock to the camera; the IMU measurements are GPS-time tagged.

The navigation filter attitude is updated using the visual heading, pitch and roll measurements obtained from the visual gyroscope, and in the case of temporal visual attitude updates discussed below the heading change is used instead of the absolute heading. Only the measurements having an LDOP value below a specified threshold are used for the update. However, some errors arising from environments not suitable for the vanishing point based method are not identified by the error detection algorithm and have to be discarded using a fault detection algorithm. One such situation is when the user is walking along a ramp, the line geometry is good and therefore this violation of the orthogonality requirement is not perceived by the error detection and as a result all visual pitch measurements deviate from the real pitch. The fault detection is applied by accepting only the standardized visual measurement values w_i that do not surpass a pre-defined threshold value [**?**]. The standardized visual measurements are obtained from the Kalman filter's innovations of the heading, pitch and roll values v_i and their corresponding estimated standard deviations, σ_v, as

$$w_i = |\frac{v_i}{\sigma_v}|,\ i = 1 : n. \tag{73}$$

The innovation in the Kalman filter is defined to be the difference $\mathbf{v}_k$ between the measurement $\mathbf{z}_k$ and the predicted value of the state $\hat{\mathbf{x}}_\mathbf{k}$ and calculated as $\mathbf{v}_k = \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_\mathbf{k}^-$. In some cases only the visual heading measurement is found faulty and discarded, while the pitch and roll measurements are used in an update.

The relative visual gyroscope induced heading measurements are transformed into absolute heading information by observing the attitude of the camera with respect to the navigation frame and using this information in propagating the visual gyroscope's heading during navigation. The initial position and orientation of the user is obtained using the GNSS measurements at the start of the experiment. The GNSS receiver observing the position was the NovAtel SPAN-SE GPS/GLONASS receiver used also for the reference, but as the purpose of the experiment was to assess the effect of vision-aiding on gyro errors, GNSS data was only used for three minutes at the

start to provide an initial position. As the GNSS measurements were not available indoors and the gyroscope measurements are too noisy to measure the change in heading accurately and the visual gyroscope fails in sharp turns, the heading of the user has to be initialized after each sharp turn during navigation in order to obtain a robust user heading continuously. This was done by using the building layout of the navigation environment, 18 times during the experiment. All results are obtained in post-processing using.

6.3.1 Kalman Filter Used in Visual Gyroscope Aided IMU Positioning

The errors in the gyroscopes cause the attitude measurements to drift, introducing continuously increasing errors in the navigation solution. The errors consist of the gyro bias, scale factor and non-orthogonalities, and the g-dependent error and noise. The error model, discussed in more detail in [**?**], is

$$\tilde{\omega}_{ib}^{b} = \mathbf{S}_g \omega_{ib}^{b} + \mathbf{b}_g + \mathbf{G}\mathbf{f}_{ib}^{b} + \eta_g \tag{74}$$

where $\tilde{\omega}_{ib}^{b}$ is the gyroscope angular velocity measurement, $\mathbf{S}_g$ is a matrix including the scale factors and non-orthogonalities, ω_{ib}^{b} is the body (b) turn rate with respect to the inertial (i) frame measured by the gyroscope, $\mathbf{b}_g$ are the gyro biases, $\mathbf{G}$ is a 3×3 matrix of the g-sensitivity coefficients, $\mathbf{f}_{ib}^{b}$ is the specific force and η_g is the noise.

The g-dependent bias is due to high accelerations, especially affecting sensors attached to the ankle, where the acceleration may rise to a maximum of 12 g. The g-dependent bias is an error source often neglected, but significant especially in pedestrian navigation applications directed to first responders, electronic monitoring and military personnel. The g-dependent bias in the gyroscopes is a result of mass imbalances caused by the manufacturing process and can impact the MEMS gyros with an error of 100 degrees/hour/g or more when uncompensated.

The tightly coupled 21-state extended Kalman filter (EKF), developed and implemented for [**?**], consists of linear perturbations of the position, attitude, velocity, gyro and accelerometer bias, three gyro scale factor coefficients and three g-sensitivity coefficients and the model is defined as

$$
\begin{aligned}
\delta\dot{\mathbf{r}}^e &= \delta\mathbf{v}^e \\
\delta\dot{\mathbf{v}}^e &= \mathbf{N}^e\delta\mathbf{r}^e - 2\Omega_{ie}^e\delta\mathbf{v}^e - \mathbf{F}^e\varepsilon + \mathbf{R}_b^e(\mathbf{b}_a) \\
\dot{\varepsilon}^e &= -\Omega_{ie}^e\varepsilon^e + \mathbf{R}_b^e((I - \mathbf{S}_g)\omega_{ib}^b + \mathbf{b}_g + \mathbf{G}\mathbf{f}_{ib}^b) \\
\dot{\mathbf{b}}_a &= -\tau_a^{-1}\mathbf{b}_a \\
\dot{\mathbf{b}}_g &= -\tau_g^{-1}\mathbf{b}_g \\
\dot{\mathbf{S}}_g &= 0 \\
\dot{\mathbf{G}} &= 0
\end{aligned}
\tag{75}
$$

where $\mathbf{r}^e, \mathbf{v}^e$ are the position and velocity vectors in the earth centered earth fixed (ECEF) frame, ε is the perturbation of the Euler angles relating the body frame to the ECEF frame and $\mathbf{b}_a$ and $\mathbf{b}_g$ are the biases of the accelerometer and gyro. The inertia tensor is denoted with $\mathbf{N}^e$, the skew symmetric forms of the earth rotation vector Ω_{ie}^e and specific force measurement $\mathbf{F}^e$. The rotation matrix $\mathbf{R}_b^e$ rotates the specific force and angular velocity from the body to the ECEF frame.

6.3.2 Equipment Setup on the Body

The data, used for testing the feasibility of a body mounted vision-aided IMU navigation indoors, was collected through an experiment conducted on the University of Calgary campus, mainly inside buildings. The environment was again very challenging for the visual measurements, consisting of numerous sharp turns, wide regions such as cafeterias and outdoor garden areas. The experiment was conducted during office hours adding many moving humans into the images. The duration of the experiment was 48 minutes, succeeding a 10 minute walk outdoors, thereby allowing the filter to converge, the route (obtained using the reference system) is shown in Figure 6.10. GNSS data was used only for three minutes at the start of the experiment to provide an initial position. All equipment was carried in a backpack as shown in Figure 6.11. The figure shows also a close-up of the setup on the backpack, namely the camera attached to the top of the backpack and the IMU on the same plane. The mutual attitude of the IMU and camera was observed in order to be able to integrate the measurements. The results of the vision-aided IMU navigation obtained using the two different update methods mentioned before are as follows.

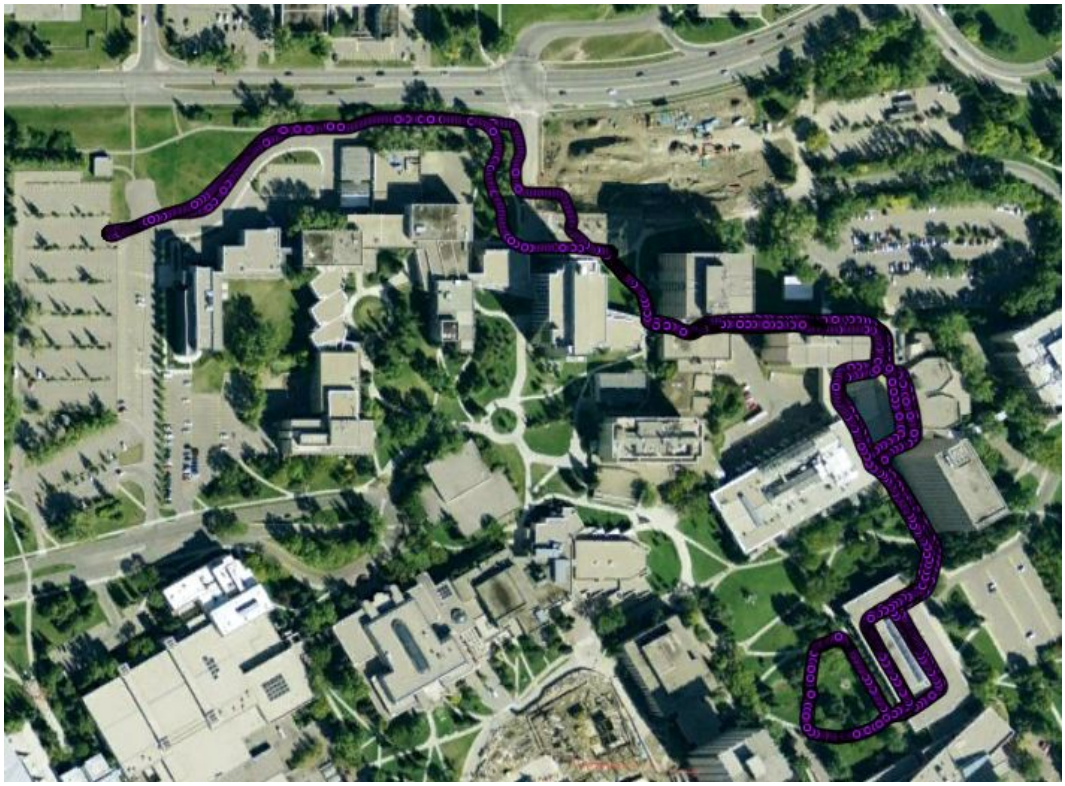

Fig. 6.10. *Route for experiments on the University of Calgary campus.*

Absolute Visual Attitude Update (AVUPT)

The absolute heading obtained by propagating the GNSS initialized heading using measurements from the visual gyroscope was used as absolute updates to the Kalman filter. The video stream obtained from the experiment was sampled into still images at 10 Hz rate and resulted in 29802 images, of which 16347 were discarded due to large LDOP values. The fault detection within the navigation filter further rejected 11% of the remaining images. Visual pitch and roll updates only, with no heading, were accepted from 38% of the images remaining from the error detection. Therefore 8337 absolute visual heading updates and 14549 absolute visual pitch and roll updates were provided to the navigation filter.

Figure 6.12 shows the standard deviations for different integration schemes. When visual updates are used, either absolute or temporal (discussed in the following sub-

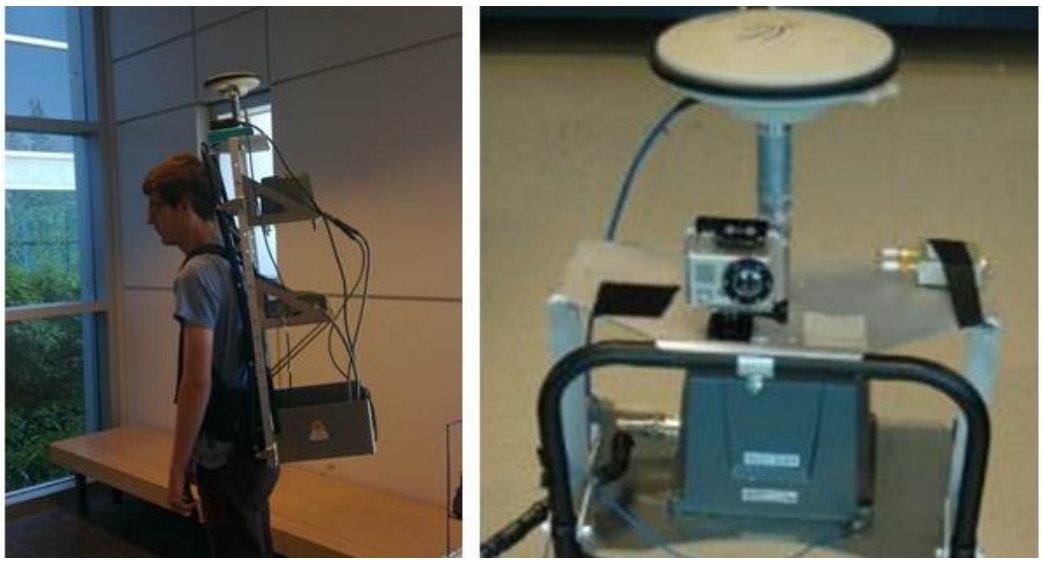

Fig. 6.11. *Equipment attached to a backpack and carried by a user in a body mounted test (on left) and a close-up of the equipment (on right).*

section), the standard deviations of roll, pitch and heading stay close to zero for the entire experiment; when no visual updates are used the standard deviations increase with time. This is mainly due to the decrease of the gyro drift growth when visual updates are used. This phenomenon should be considered with care as the update of the absolute heading using the building layout during the navigation has a significant effect on the growth of errors when the absolute update method is used. Figure 6.13 shows the effect of the visual updates on the attitude errors. The attitude errors are expressed using a measure called root mean square error (rms). The rms error is a measure expressing the spread of the values around the average. It is computed by taking the root of the averaged squared residuals as

$$rms = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{76}$$

where $\hat{y}_i$ is the predicted value of the measurement y_i, N is the total number of measurements. The vision-aiding improves the navigation solution's pitch and roll only slightly, as is seen from the figure and in Table 6.4. In the experiment, the pitch root mean square error decreases from 1.7 to 1.4 degrees and the roll from 2.0 to 1.4

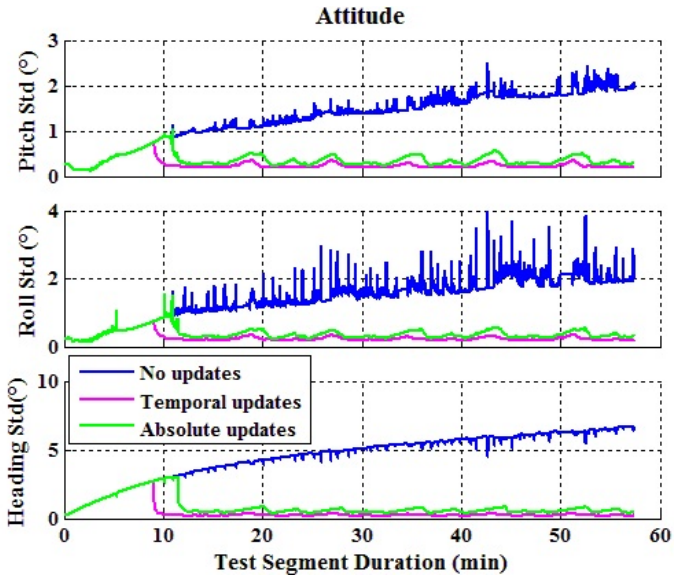

Fig. 6.12. *Standard deviation for different integration schemes, namely no visual updates used (blue), using temporal updates (purple) and absolute updates (green).*

Table 6.4. *Attitude errors obtained for body-mounted IMU with different integration methods*

Attitude errors (rms, degrees)			
	Pitch	Roll	Heading
No visual updates	1.7	2.0	29.5
Absolute vision-aided attitude updates (AVUPT)	1.4	1.4	2.1
Temporal vision-aided attitude updates(TVUPT)	1.6	1.7	17.6

degrees when the absolute vision-aided attitude updates are used. However, the heading improves significantly, namely 93% as the root mean square error decreases from 29.5 to 2.1 degrees when the navigation filter is updated with visual measurements.

Temporal Visual Attitude Update (TVUPT)

When no prior information of the environment is available, like a floor plan used in the previous experiment, the temporal attitude (i.e. the change of the attitude over a short interval) of the camera may be used. In temporal visual attitude update (TVUPT) the Kalman filter integrates the user attitude obtained from the visual gyro-

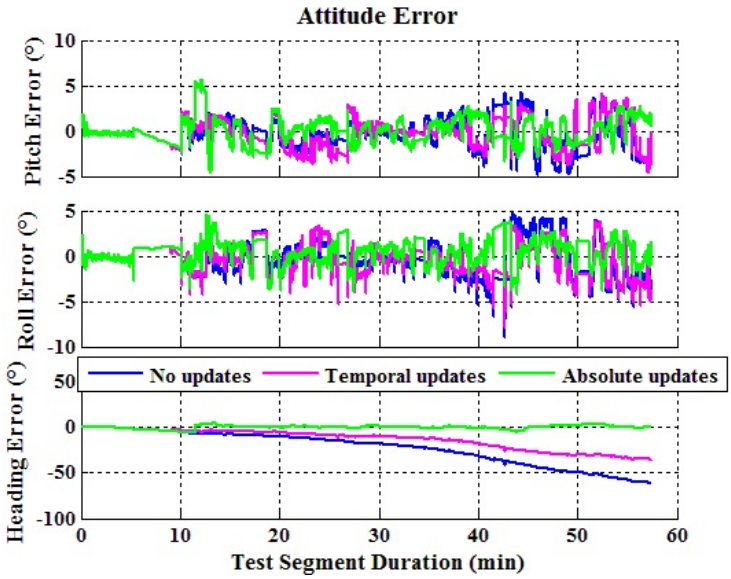

Fig. 6.13. *Attitude error using different integration schemes, namely no visual updates used (blue), using temporal updates (purple) and absolute updates (green).*

scope to estimate the errors in the IMU attitude measurements. The temporal attitude observation may be presented as

$$\begin{bmatrix} \phi \\ \beta \\ \theta \end{bmatrix}_{b_k}^{b_{k-n}} = \begin{bmatrix} \phi \\ \beta \\ \theta \end{bmatrix}_{b_k}^{e} - \begin{bmatrix} \phi \\ \beta \\ \theta \end{bmatrix}_{b_{k-n}}^{e} \tag{77}$$

where ϕ, β and θ are the pitch, roll and heading of the camera, respectively. As the filter estimates the errors in IMU attitude measurements, having a 200 Hz rate using consecutive images having a lower rate, the two consecutive images are time labeled as k and $k-n$, where n is the number of IMU epochs between two consecutive images. b represents the body frame and e the ECEF frame.

The equation may be represented in the filter's (75) rotation matrix form as $\mathbf{R}_{b_k}^{e} = \mathbf{R}_{b_{k-n}}^{e}\mathbf{R}_{b_k}^{b_{k-n}}$.

The image interval used in previous visual gyroscope aided IMU positioning experiments was 0.10 s. For the temporal visual attitude updates the interval had to be decreased, because in order to get accurate temporal visual attitude updates, the im-

age rate has to be as large as possible. Due to the computational limitations the entire data set was not sampled at the chosen 30 Hz rate, but two consecutive images were retrieved with a 0.033 s interval and then four subsequent images were discarded. The sampling resulted in 30077 images.

When the temporal visual attitude update method was used, the user heading root mean square error decreased from 29.5 degrees to 17.6 degrees, resulting in a 40% improvement. Again, the improvement in the pitch and roll accuracy was minor, namely the pitch root mean error decreased from 1.7 to 1.6 and roll from 2.0 to 1.7 degrees, as shown in Table 6.4 and Figure 6.13. The standard deviations of roll, pitch and heading stayed close to zero for the entire experiment when the temporal visual updates were used; when no visual updates are used the standard deviations increase with time, shown also in Figure 6.12. However, the visual attitude updates show an overly optimistic variance because the update is temporal rather than absolute for which the integration algorithm was originally developed and therefore this method provides variances that are not exactly indicative of the estimate.

6.3.3 Equipment Setup on the Foot

When the gyro is located on the ankle of the user the vertical acceleration can rise up to the maximum of 12 g causing very large g-dependent errors. The effect of correcting the errors through vision aiding of the attitude was tested by an experiment using a foot mounted system. The IMU and camera were attached rigidly to each other and located on the ankle of the user. The setup is shown in Figure 6.14. Data was collected in an experiment of 43 minutes conducted mainly indoors. Because the purpose of the research was to assess vision-aiding performance on attitude and gyro errors, GNSS data was only used for three periods of two to three minutes during the navigation in low canyons between buildings. A pedestrian navigation solution was obtained by integrating the vision-aided gyroscope attitude measurements and applying zero velocity updates to the inertial navigation filter as well as using the occasional absolute heading updates obtained from the building layout. The integration was performed using the Kalman filter described above. Due to the lack of a reference system mounted on the foot (the reference system was carried in the backpack), the attitude errors could not be evaluated but the position errors could.

The visual heading, pitch and roll measurements were used as absolute updates to the

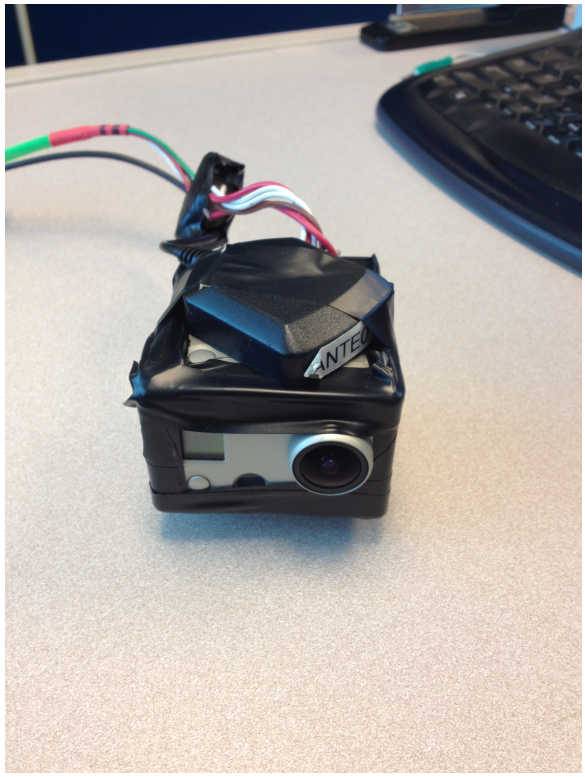

Fig. 6.14. *Equipment setup for the foot, namely the GoPro camera and IMU attached to each other.*

navigation filter attitude as explained above in the case of a body-mounted system. The calculation of visual measurements was challenging due to large camera movements at the ankle of the pedestrian. The total number of images acquired during the experiment was 25664. Only 18% received an LDOP value sufficiently good for trusting the visual measurements due to image blurring introduced by the fast motion of the foot and because the camera was pointing straight down to the floor for a short time during a step cycle period, shown in Figure 6.15. Again, fault detection was used to remove the noise from the visual measurements. The fault detection within the navigation filter further rejected 65% of the images remaining from the error detection. Visual pitch and roll updates only, with no heading, were accepted from 18% of the remaining images. This resulted into 1326 visual heading updates and 1617 visual pitch and roll updates to the navigation filter.

Table 6.5 and Figure 6.16 show the improvement of the position obtained with the

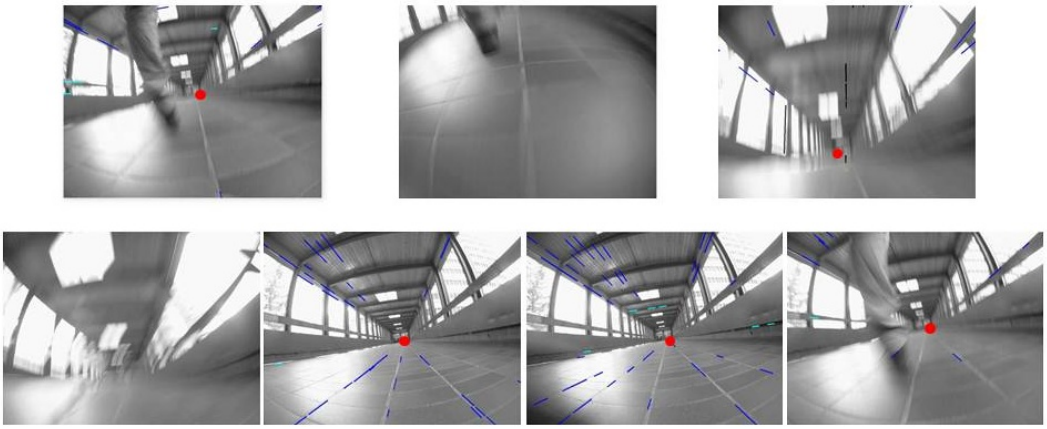

Fig. 6.15. *Images resulting from one step cycle period, three of the images are too blurred for visual gyroscope calculations and are therefore not shown, five accurate vanishing points are obtained because the rest of the images show only the floor plane or are too blurred for accurate line detection.*

Table 6.5. *RMS position error obtained for foot-mounted IMU with and without vision-aiding*

RMS Position Errors (m)		
	Horizontal	Vertical
No vision-aiding	30.9	67.5
With vision-aiding	20.0	67.7

vision-aided foot mounted navigation system. The periods when GNSS was used are shown in the figure with black squares. Vision-aiding improves the horizontal position significantly in the eperiment; the root mean square horizontal position error decreases from 30.9 m to 20 m, yielding an improvement of 34%. Vision-aiding has no effect on the vertical position error, in which case the error remains at 68 m.

6.4 Visual Gyroscope Implementation Using Probabilistic Hough Transform

The three most significant limitations of the visual gyroscope presented in Chapter 4 are its inability to monitor the heading change during sharp turns, its accuracy suffers from irregularities of the environment (namely lines violating the orthogonal-

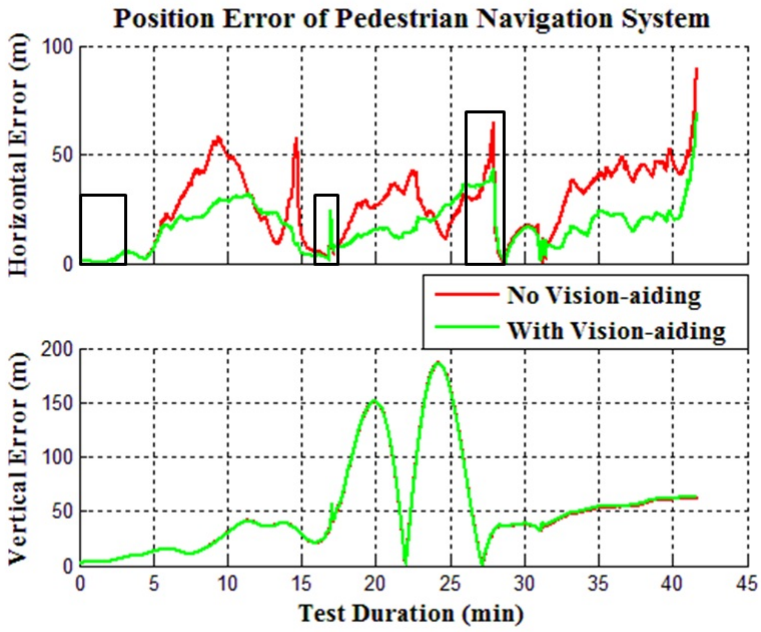

Fig. 6.16. *RMS position errors obtained for foot-mounted IMU with (green) and without vision-aiding (red).*

ity requirement) and the calculation is relatively slow for real-time implementation. The first two problems could be addressed by a tighter integration of the visual gyroscope and other positioning systems, especially a gyroscope. As the visual gyroscope presented has processed the image data independently from other positioning systems it has not had any support for exceptional situations. As was mentioned before the line detection using the Hough transform is a bottleneck in the visual gyroscope's processing. All three problems are addressed in this section by developing a method extracting the lines using a novel modification of an algorithm called Probabilistic Hough Transform [**?**] utilizing the information of the user attitude obtained from an IMU.

In [**?**] the attitude information obtained from INS was utilized to estimate the vanishing point location by calculating a probability density function for the Hough parameter space, and the method was called a predictive Hough Transform. The attitude of the user obtained from the INS may be transformed into an estimate of the vanish-

ing point using the relation

$$\tilde{\mathbf{v}} = \mathbf{K}\mathbf{C}_b^c\mathbf{C}_n^b\mathbf{R} \tag{78}$$

where $\mathbf{C}_b^c$ is a direction cosine matrix (DCM) [**?**] from the body to the camera frame and $\mathbf{C}_n^b$ from navigation frame to the body frame. The direction cosine matrix,

$$\mathbf{C}_b^n = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{11} & c_{12} & c_{13} \\ c_{11} & c_{12} & c_{13} \end{bmatrix} \tag{79}$$

where the element at the *i*th row and the *j*th column is the cosine of the angle between the i-axis of the reference frame and the j-axis of the initial frame. A vector defined in a certain axes frame may be expressed in reference axes by multiplying it by the direction cosine matrix (expressed here as transforming the vector $\mathbf{r}^b$ in body axes into the navigation frame $\mathbf{r}^n$ (the transformation may be done likewise for other transformations also) as

$$\mathbf{r}^n = \mathbf{C}_b^n\mathbf{r}^b \tag{80}$$

When the user rotates through angle ψ about the z-axis (heading), angle θ about the new y-axis (pitch) and angle ϕ about the new x-axis (roll) the transformation may be presented using the direction cosine matrix

$$\mathbf{C}_b^n = \begin{bmatrix} \cos\theta\cos\psi & -\cos\phi\sin\psi + \sin\phi\sin\theta\cos\psi & \sin\phi\sin\psi + \cos\phi\sin\theta\cos\psi \\ \cos\theta\sin\psi & \cos\phi\cos\psi + \sin\phi\sin\theta\sin\psi & -\sin\phi\cos\psi + \cos\phi\sin\theta\sin\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix}. \tag{81}$$

The reverse rotation, in this case the rotation $\mathbf{C}_n^b$ is obtained using a transpose rule $C_n^b = (C_b^n)^T$

Matrix $\mathbf{K}$ in (78) is the camera calibration matrix and $\mathbf{R}$ normalized rotation matrix of the camera this time in the navigation frame and computed using the attitude information obtained from the IMU. The expected vanishing point location $\tilde{\mathbf{v}}$ is characterised by a Gaussian density function with parameters $(\mu_\rho, \sigma_\rho^2)$ [**?**] as

$$\rho_\theta \sim N(\mu_\rho, \sigma_\rho^2) \tag{82}$$

where the distance ρ_θ related to a certain angle θ ($\theta \in [0, \pi]$) is normally distributed. The mean μ_ρ is computed for each angle θ and the corresponding line going through the estimated vanishing point $\tilde{\mathbf{v}}$. The variance σ_ρ^2 is decided based on the IMU accuracy.

In [**?**] the probability density function was used as a filter for the Standard Hough Transform (SHT) result space and provided a corrected vanishing point location. The attitude information from the accurate vanishing point was finally used for correcting the INS attitude with a Kalman filter and an improved navigation solution was obtained. The method addresses the problems arising from the erroneous vanishing point calculations due to an unsuitable environment, namely line geometry overwhelmed by non-orthogonal lines. As the line detection is done using the Standard Hough Transform, the processing time jeopardizing the real-time solution is not improved. Therefore in this thesis an accelerated line detection algorithm based on Probabilistic Hough transform is developed.

As explained in Chapter3 the Standard Hough Transform computes the parameter space (ρ, θ) for each point (x, y) found from the input image, being usually the result of the edge detection, as

$$\rho = x\cos(\theta) + y\sin(\theta). \tag{83}$$

A matrix, called accumulator, keeps count of the number of image points corresponding to a certain (ρ, θ)-pair. After examining each point in the input image the maximum value at the accumulator are identified and stated to represent lines in the image. The Probabilistic Hough Transform [**?**] is a modification of the Standard Hough Transform using only a random subset of image points for voting and deriving the number of votes needed for identifying a line using Monte Carlo evaluation theory. According to [**?**] the algorithm reduces the computation only if a priori information of the number of lines is available. As this is not usually the case herein a Progressive Probabilistic Hough Transform was developed. In the method the image points used were selected randomly and the parameter pair was selected to represent a line when the votes it had received exceeded the number that would be expected from random noise. The amount of points needed to represent a line was evaluated progressively based on the rate of the pixels examined and the pixels voting for a certain line. In this thesis a method combining the INS aided vanishing point detection

and Progressive Probabilistic Hough Transform discussed above is developed.

The attitude of the user obtained from the IMU measurements is transformed into an estimate of the vanishing point location $\tilde{\mathbf{v}}$ using (78) and the probability space (82) corresponding to the point for each possible angle θ computed. Then, a pixel is selected randomly from the set containing all pixels resulted from the edge detection. The distance ρ is calculated for all possible θ and the values in corresponding accumulator cells are increased by summing the value of the probability density function for the obtained distance and mean with the existing cell value. The Standard Hough Transform increases all accumulator cells equally because its objective is to find all straight lines present in the image, while here the objective is to find the lines supporting the vanishing point. As a result of using the probability function the closer the possible line is to the estimated vanishing point, the more the accumulator cell value is increased. When the value in the accumulator cell exceeds a predetermined threshold, a line is found. As a line is found and no more support for it is needed, all other image points belonging to the line are removed from the pixel set. Also all the votes in the accumulator arising from the line are removed for not disturbing the identification of other lines. In this way the number of image points examined and therefore the computation time needed decreases. As the points having a larger likelihood of belonging to a line going through the estimated vanishing point or a point close to it are given more weight, the lines found are likely to be in the direction of supporting the central vanishing point.

After all pairs (ρ_i, θ_i) supporting a line are identified as explained above, the correct vanishing point is found from the intersection of all lines i as follows. As discussed above, a (ρ, θ)-pair in the parameter space represents all collinear points (x_i, y_i) in the image. This is also true the other way around [**?**]; all points (ρ_i, θ_i) satisfying the equation

$$\rho_i = x\cos(\theta_i) + y\sin(\theta_i) \tag{84}$$

which represent lines going through a point (x, y). As the line detection was done by emphasizing the lines supporting the estimated vanishing point, all the lines found should intersect at the correct vanishing point which may then be found using a least-squares estimation technique for (84).

Two parameters selected for the calculation are crucial for the performance of the

visual gyroscope presented in this section, namely the threshold for deciding when a line is found and the standard deviation of the estimated vanishing point value. When the threshold for finding a line is deficient, the rate of false positives is large, and when it is too large, the computation time increases and occasionally too few lines are found from the low-light indoor environment resulting in an inaccurate vanishing point location. Also, when the standard deviation assigned for the estimated vanishing point value is too large the errors in IMU induced attitude distort the line detection by emphasizing points close to the estimated point probably not even belonging to a line. For the experiments presented below, the parameter σ was chosen to be 20 to allow the estimated vanishing point to be within ± 20 pixels from the correct vanishing point and threshold for identifying a line 0.4 through experimentation.

The method was tested using a subset of the data collected using the body mounted equipment for testing the absolute visual attitudes explained above. The subset consisted of 80 seconds of data collected indoors, resulting in 800 sampled images. Figure 6.17 shows the result of line detection and vanishing point calculations. As the images were taken using the GoPro camera with a wide-angle lens, discussed in Chapter 4, they are distorted. In order to maintain the real-time processing obtained using the developed line detection, the distortion is not corrected as described, but instead it effect is reduced by discarding the pixels close to the edges of the image. The blue lines are extracted using the Probabilistic Hough Transform. It should be noted that because the image is not distortion corrected the lines found do not agree with the lines seen in the figure but would if corrected. The green point is the vanishing point estimation based on the IMU attitude and the red point is the corrected vanishing point. The figure shows how the vanishing point is found reliably even when the IMU induced attitude and therefore the estimated vanishing point is distorted.

As stated in [**?**] the processing time of an algorithm is dependent on the computer used and algorithm and software implementation. Therefore the effect of the algorithm is shown by comparing the number of image points examined, in other words the iterations of the parameter calculation. The Standard Hough Transform examines all pixels in the input image and afterwards searches for local maxima from the accumulator to find the lines. The algorithm presented uses a fraction of the image points for extracting the lines, namely on average 45% of all image points, and therefore the computation is anticipated to be accelerated in the same proportion to the Standard Hough Transform computation time. Table 6.6 gives the test iteration statistics. It

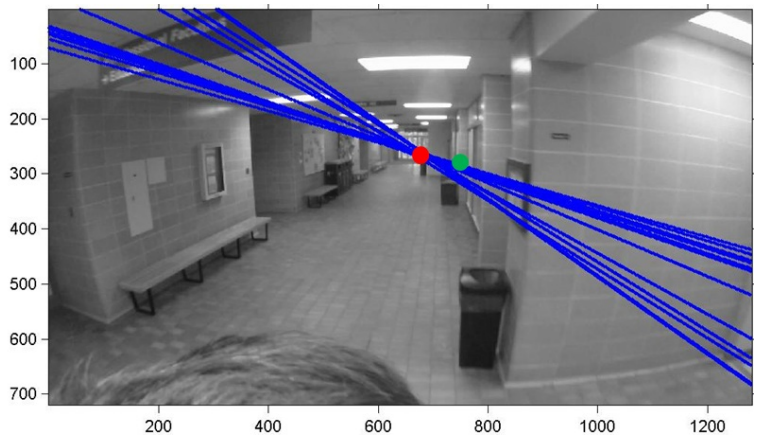

Fig. 6.17. *Line detection and vanishing point calculations using Probabilistic Hough Transform. Estimated vanishing point (green) is corrected (red) using the lines (blue) found with the Probabilistic Hough Transform.*

Table 6.6. *Ratio of image points used for computing the Probabilistic Hough Transform presented to the image points used by Standard Hough Transform for the images processed in the experiment*

Ratio of image points used compared to all image points (%)			
Min	Mean	Max	Std
27	45	67	8

should also be noted that the method already detects the lines during the point examination as well as the vanishing point, further reducing the computation time used for obtaining visual gyroscope measurements.

Restricted light conditions and lines violating the orthogonality requirement are major threats for the visual gyroscope's accuracy often resulting in errors. An example of an image suffering from both situations is show in Figure 6.18. The vanishing point computed by the visual gyroscope discussed in Chapter 4 is shown on the left while the one using the visual gyroscope with the Probabilistic Hough Transform presented in this is section on the right. The experiment shows increased tolerance for vanishing point computation because now the computation process is not dependent on visual perception only, but receives a priori information of the user attitude from the IMU.

Fig. 6.18. *Vanishing point detection in an environment suffering from low lighting and non-orthogonal lines, on the left using the visual gyroscope presented in Chapter 4, and on the right using the visual gyroscope based on Probabilistic Hough Transform.*

Although the vanishing point detection method is dependent on the IMU, the method is tolerant to large errors in the IMU measurements when the parameters of the Probabilistic Hough Transform algorithm are carefully selected. Figure 6.19 shows how an estimated vanishing point (green) resulting from large temporary errors in IMU measurements is corrected (red) through the line detection presented.

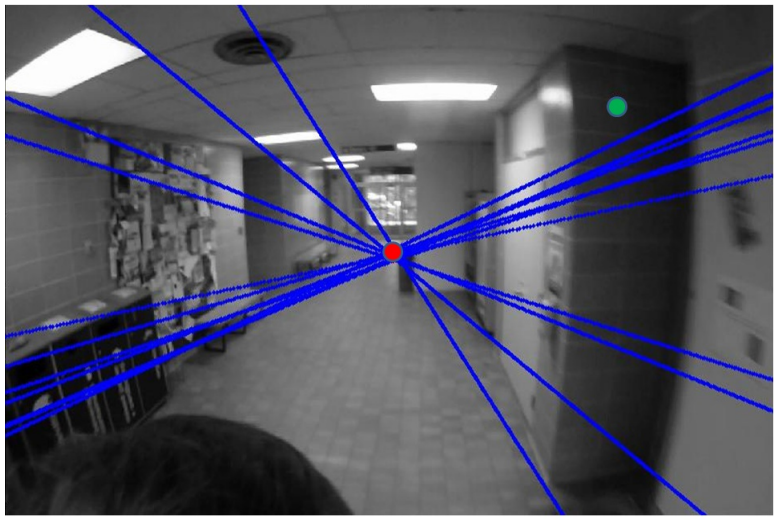

Fig. 6.19. *Detected vanishing point (red) may be used to correct large errors in IMU measurements resulting in erroneous estimated vanishing point location (green).*

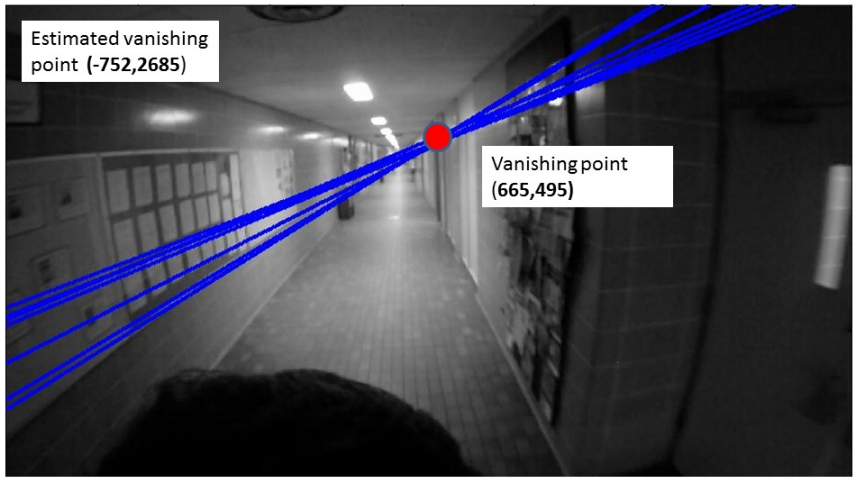

Fig. 6.20. *Conflict between estimated and detected vanishing point locations indicates the existence of a steep turn.*

The method gives also promising results for turn detection that has so far been the most significant obstacle preventing the use of vision-aided inertial sensors autonomously for navigation in unknown indoor environments. In turning situations the estimated vanishing point obtained by propagating the attitude using the method falls outside the image at the same time as the corrected vanishing point obtained from the Probabilistic Hough Transform detection is found from the other side of the image as shown in Figure 6.20. This is due to the change of the World Frame, i.e. the visual gyroscope was initialized at the beginning as having a zero heading when the camera frame is totally aligned with the world frame and now the world frame's horizontal axes are rotated 90 degrees. When this contradiction is used in integration, at least the existence of a steep turn is observed. Observing the magnitude of the turn is a future research objective.

7. VISION-AIDED CARRIER PHASE NAVIGATION

Navigation in urban areas is challenging for GNSS. Line-of-sight signals are blocked by tall buildings and therefore, the requirement for measurements from at least four satellites is not fulfilled and consequently a position solution is not obtained. Even when the solution is obtained, multipath effects deteriorate the accuracy of the position. In this chapter the visual gyroscope and a version of the visual odometer are used for aiding GNSS measurements in such areas.

The visual gyroscope is suitable also for urban environments, which consists of countless straight lines, i.e. edges of buildings and roads. The limitation of the visual gyroscope is its need for absolute heading information that has to be updated during navigation due to its deficiency to monitor the magnitude of sharp turns and due to calculation errors arising from problems in visual perception. However, the calibration need to be done only occasionally and in between the correct heading maintained by propagating the absolute heading using the visual gyroscope's measurements. Although the likelihood for observing at least four satellites needed for resolving the user position is reduced, it is still possible occasionally even in an urban canyon and therefore the visual gyroscope and GNSS positioning complement each other in these challenging environments. The translation obtained from the homography constraining consecutive images has an ambiguous scale that was resolved earlier in this thesis using a special configuration of the camera. In this chapter an alternative method is presented, namely the scale is observed using differenced carrier phase GNSS measurements. As the carrier phase measurements are differenced, the need to resolve the ambiguous integer number of the satellites' carrier phase cycles is avoided. Because only the scale, i.e. the total magnitude of translation between two time epochs is needed, using two satellites with the proper geometry is enough and therefore the method is feasible also for a dense urban environment. Below, the method is described in detail, then the verification of the method in a sub-urban environment is discussed and finally an experiment testing the method in a dense urban canyon is

presented.

7.1 Ambiguity Resolution Using Differenced GNSS Carrier Phase Measurements

Differenced satellite carrier phase measurements provide information about the magnitude of translation of the user between two time epochs which may be used for resolving the ambiguous scale in the translation obtained from images. Because only the magnitude of the translation and the receiver clock error are needed, acquisition of two satellites is enough. The idea has been approached earlier by [**?**]. The method was developed for robot navigation and visual measurements utilized from three cameras. The pitch and roll of the camera were obtained from an IMU. As the heading was not observed accurately using an IMU, it was included to the algorithm as an unknown and therefore observations from at least three satellites were needed. In the test lasting for 100 seconds centimeter level accuracy was obtained, which decreased into decimeter level when the receiver clock was calibrated and only two satellites acquired. A similar method for vehicular navigation was developed in [**?**] utilizing again an IMU for attitude, GNSS and vision-aiding. In the experiments at least three satellites were observed and meter level horizontal position accuracy was obtained. The method presented in this thesis is aimed at pedestrian navigation where the possible amount and size of equipment is limited. When the relative heading and translation information obtained from images is initialized with the absolute position and heading information is provided by GNSS, the user position may be propagated and only occasional absolute updates are needed for a functional navigation solution.

The carrier phase observation (φ^i) for the satellite i may be represented using a simplified form as

$$\varphi_i = r^i + cdt_{rcvr} + \lambda N + \eta^i + \varepsilon^i_\varphi \tag{85}$$

where r^i is the true range between the satellite and the receiver, λ is the carrier wavelength, N is the integer ambiguity, η^i is an error term incorporating ionospheric, tropospheric and satellite orbital errors and the error term ε^i_φ is the combined effect of multipath and noise. The equation assumes that the satellite clock error

is already compensated for. The carrier phase measurements obtained at two time epochs (t_1, t_2) are differentiated and the resulting measurement is

$$\Delta\varphi_i = \varphi_i(t_2) - \varphi_i(t_1) = \Delta r^i + c\Delta dt_{rcvr} + \varepsilon^i_\varphi. \tag{86}$$

The integer ambiguity term is removed by differencing the carrier phase observations over time and the change in the term encompassing the errors stays below a centimeter / second level and is therefore omitted [**?**]. The differenced range Δr^i may further be expressed [**?**] as

$$\begin{aligned} \Delta r^i &= (\mathbf{T}^i(t_2) - \mathbf{T}_{rcvr}(t_2)) \cdot \mathbf{u}(t_2) - (\mathbf{T}^i(t_1) - \mathbf{T}_{rcvr}(t_1)) \cdot \mathbf{u}(t_1) \\ &= (\mathbf{T}^i(t_2) \cdot \mathbf{u}(t_2)) - (\mathbf{T}^i(t_1) \cdot \mathbf{u}(t_1)) \\ &-(\mathbf{\Delta} T_{rcvr}(t_2)) \cdot \mathbf{u}(t_2)) - (\mathbf{T}_{rcvr}(t_1)) \cdot \mathbf{\Delta} u(t_2)) \end{aligned} \tag{87}$$

where $(\cdot)$ denotes a vector dot product, $\mathbf{T}^i$ is the satellite position vector, $\mathbf{T}_{rcvr}$ the receiver position vector and $\mathbf{u}$ the unit vector from the user to the satellite and calculated as

$$\mathbf{u} = \frac{\mathbf{T}^i - \mathbf{T}_{rcvr}}{|\mathbf{T}^i - \mathbf{T}_{rcvr}|}. \tag{88}$$

The term $(\mathbf{T}^i(t_2) \cdot \mathbf{u}(t_2)) - (\mathbf{T}^i(t_1) \cdot \mathbf{u}(t_1))$ is called the *satellite Doppler term* $(DOPP_i)$ and it arises from the motion of the satellite between the two time epochs and may be derived from the satellite observations. The term $(\mathbf{T}_{rcvr}(t_1)) \cdot \mathbf{\Delta} u(t_2)$ expresses the change in the user unit vector (the line-of-sight unit vector) and is called the *geometry change* (u_{GC}). As the *geometry* and *satellite Doppler change* terms employ the user position at the second time epoch that is not known yet but will be obtained from the calculations, an estimate of the position is used. The estimate has to be accurate only to within 100 meters [**?**]. Now the differenced carrier phase measurement (86) may be presented as

$$\Delta\varphi_i = -(\mathbf{\Delta} T_{rcvr}(t_2) \cdot \mathbf{u}(t_2)) + DOPP_i - u_{GC}. \tag{89}$$

By rearranging the terms, an equation for resolving the magnitude of the user translation $\mathbf{\Delta} T_{rcvr}(t_2)$ between the time epochs (t_1, t_2) is obtained from

$$\Delta\varphi_i^{corr} = \Delta\varphi_i - DOPP_i + u_{GC} = -(\mathbf{\Delta}T_{rcvr}(t_2) \cdot \mathbf{u}(t_2)). \tag{90}$$

The equation expressing the user translation $\mathbf{\Delta}T_{rcvr}(t_2)$ has three unknowns, namely the translation in x-, y- and z-axis directions. As the receiver clock error has to be also resolved, four satellites would be required to obtain a solution, which would not necessarily be always feasible in the dense urban environments. However, images provide information about the user translation between two time epochs, i.e. the times of capturing the two consecutive images. The translation of the user with an ambiguous scale may be obtained using the epipolarity constraint and the Fundal matrix arising from it, discussed in Chapter 3.

7.1.1 Ambiguous Translation Using the Fundamental Matrix

The fundamental matrix $\mathbf{F}$ defined by (31) may be computed, given sufficiently many matching image points $(\mathbf{x}', \mathbf{x})$, using a linear algorithm [**?**], used in the thesis; however, more robust algorithms for the computation are also presented in the reference. When the epipolar geometry is defined to be affine, i.e. the difference of an affine geometry compared to a projective geometry is that the cameras are defined to have their centres at the infinity and therefore there is a parallel projection from scene to image. The affine Fundamental matrix is

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}. \tag{91}$$

Now, each point correspondence $(\mathbf{x}', \mathbf{x})$ may be represented as

$$(x_i', y_i', x_i, y_i, 1)\mathbf{f} = 0, \text{ i.e. } \begin{bmatrix} x_i' & y_i' & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n' & y_n' & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = 0 \tag{92}$$

when n matching image points in the two consecutive images are found and $f = (a, b, c, d, e)^T$. At least four corresponding points are needed, but when there are more, as is usually the case especially in outdoor environments with favorable lighting conditions, the singular value decomposition is used. The fundamental matrix $\mathbf{F}$

may further be transformed into the Essential matrix $\mathbf{E}$, also discussed in Chapter 3, using the camera calibration matrix $\mathbf{K}$ (assumed constant between the images) and the camera motion [**?**] as

$$\mathbf{E} = \mathbf{K}^T\mathbf{F}\mathbf{K} = \Big[t\Big]_\times \mathbf{R}. \tag{93}$$

The term $\Big[t\Big]_\times$ denotes a skew symmetric matrix of the translation vector $(t_x, t_y, t_z)^T$ and is

$$\Big[t\Big]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{94}$$

The singular value decomposition of the Essential matrix $\mathbf{E}$ gives $\mathbf{E} \sim \mathbf{U}DV^T$ and the ambiguous translation obtained from $\mathbf{U}$ is $t \sim \begin{bmatrix} u_{13} & u_{23} & u_{33} \end{bmatrix}^T$, where u_{ij} denotes the element in the matrix $\mathbf{U}$ on the *i*th row and *j*th column.

7.1.2 Navigation Solution Incorporating the Absolute User Translation

The user translation $\Delta\mathbf{T}_{rcvr}(t_2)$ is perceived in the navigation frame but the visual translation is in the camera frame. In order to be able to obtain the ambiguous scale of the visual translation and turn it into a position change of the user again in the navigation frame, transformations have to be made. The visual gyroscope presented in Chapter 4 provides information about the attitude of the user with respect to the navigation frame. A Direction Cosine Matrix $\mathbf{C}_n^b$ transforming the observations from the navigation frame to the camera frame is formed using the heading, pitch and roll measurements obtained from the visual gyroscope as explained in Chapter 6. For resolving the scale, both the unit line-of-sight vector $\mathbf{u}(t_2)$ and user translation vector $(\mathbf{\Delta}T_{rcvr}(t_2))$ has to be multiplied using the direction cosine matrix $\mathbf{C}_n^b$. The user translation vector $(\Delta\mathbf{T}_{rcvr}(t_2))$ may be written using the scalar scale s of the visual translation that is still unknown and the visual translation vector $\mathbf{t}$ as $(\Delta\mathbf{T}_{rcvr}(t_2)) = \mathbf{C}_n^b s\mathbf{t}$. Now, (90) can be re-written as

$$\Delta\varphi_i^{corr} = \Delta\varphi_i - DOPP_i + u_{GC} = -(\mathbf{C}_n^b\mathbf{u}(t_2) \cdot \mathbf{C}_n^b\mathbf{t}s) \tag{95}$$

and in a matrix form

$$y = \begin{bmatrix} \Delta\varphi_1^{corr} \\ \vdots \\ \Delta\varphi_N^{corr} \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} -(\mathbf{C}_n^b \mathbf{u}_1^T(t_2)) \cdot (\mathbf{C}_n^b \mathbf{t}) & 1 \\ \vdots & 1 \\ -(\mathbf{C}_n^b \mathbf{u}_N^T(t_2)) \cdot (\mathbf{C}_n^b \mathbf{t}) & 1 \end{bmatrix} \quad \Delta\mathbf{x} = \begin{bmatrix} s \\ c\Delta dt_{rcvr} \end{bmatrix} \tag{96}$$

from which the ambiguous scale of translation s may be obtained using the least-squares equation $\mathbf{x} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$.

Occasionally, especially when only two satellites are observed, the scale computation fails. The errors are detected and discarded by monitoring the magnitude of the user speed from the absolute speed obtained, i.e. the measurement is discarded and a previous one used when the speed exceeds 3 m/s.

The translation is again transformed into the navigation frame (ENU) using a Kalman filter propagating the user position (X, Y), heading (θ) and speed (S). Speed is obtained from the translation computed as discussed above and the heading from the visual gyroscope and occasional absolute heading updates as discussed below. The Kalman filter models the user position as

$$\begin{aligned} X_{k+1} &= X_k + S\sin(\theta_k)\Delta t \\ Y_{k+1} &= Y_k + S\cos(\theta_k)\Delta t \end{aligned} \tag{97}$$

discussed in more detail in [**?**]. The state $\mathbf{x}_k$ and measurement $\mathbf{z}_k$ vectors for the model are

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{X}_k \\ \mathbf{Y}_k \\ \theta \\ S \end{bmatrix} \quad \mathbf{z}_k = \begin{bmatrix} X \\ Y \\ \theta \\ S \end{bmatrix} \tag{98}$$

and state transition matrix $\mathbf{\Phi}$ and process noise $(\mathbf{Q})$ matrices

$$\Phi_k = \begin{bmatrix} 1 & 0 & 0 & \sin(\theta_k)\Delta t \\ 0 & 1 & 0 & \cos(\theta_k)\Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{99}$$

$$\mathbf{Q}_k = \begin{bmatrix} q_1\Delta t + \frac{(a^2q_3+b^2q_4)\Delta t^3}{3} & \frac{(acq_3+bdq_4)\Delta t^3}{3} & \frac{aq_3\Delta t^2}{2} & \frac{bq_4\Delta t^2}{2} \\ \frac{(acq_3+bdq_4)\Delta t^3}{3} & q_2\Delta t + \frac{(c^2q_3+d^2q_4)\Delta t^3}{3} & \frac{cq_3\Delta t^2}{2} & \frac{dq_4\Delta t^2}{2} \\ \frac{aq_3\Delta t^2}{2} & \frac{cq_3\Delta t^2}{2} & q_3\Delta t & 0 \\ \frac{bq_4\Delta t^2}{2} & \frac{dq_4\Delta t^2}{2} & 0 & q_4\Delta t \end{bmatrix} \quad (100)$$

where q_1 is the spectral density for the position North component (X), q_2 for the East component (Y), q_3 the spectral density for the heading and q_4 for the speed. In the following section two experiments using the method discussed above are described and assessed.

7.2 Method Verification in a Sub-Urban Environment

Although the method of vision-aided carrier phase navigation is designed for urban positioning, performance verification was first performed in an easier signal environment, namely that with lower buildings blocking out only satellites having an elevation less than 30 degrees. The test setup consisted of the Novatel SPAN-SE GPS/GLONAS receiver providing carrier phase measurements and a GoPro camera for visual measurements. A Northrop Grummans tactical grade LCI-IMU and the SPAN were used for acquiring the reference solution as well as initializing the user position and heading at the beginning of the experiment and after every three minutes of navigation. The system was carried in a backpack as shown in Figure 7.1. The camera and GNSS receiver were attached to the top of the system and are indicated with a red circle in the figure. After initialization the user position was observed by propagating the heading and speed measurements using the Kalman filter explained above. The data was collected for 15 minutes and post-processed using Matlab.

As the purpose of the verification experiment was to test the feasibility of the system in extreme signal conditions when only two satellites are used, the vision-aided carrier phase navigation solution was computed using only two satellites available for the full experiment. Because the heading obtained from the visual gyroscope suffers from occasional errors in vanishing point calculation and the sharp turns cannot be observed, every three minutes the position and heading were re-initialized using the reference system. This is justified by the fact that even in a dense urban area the requirement for four satellites is fulfilled once in a while as is also seen in the real urban

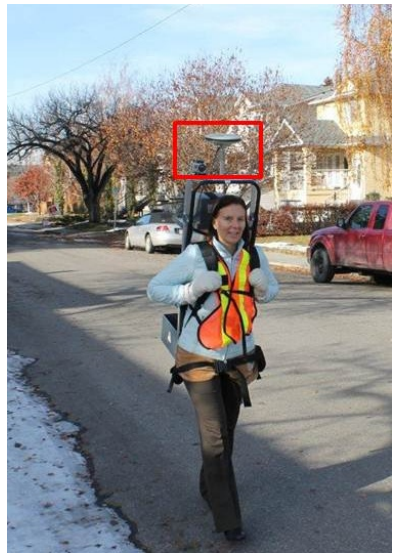

Fig. 7.1. *Setup for data collection used in verification of vision-aided carrier phase navigation.*

Table 7.1. *Positioning verification error statistics using vision-aided carrier phase (VA)*

Statistics	min error (m)	max error (m)	mean error (m)	std of error (m)
VA	0	76	24	18

experiment below. Figure 7.2 shows the path obtained using the vision-aided carrier phase navigation (red) compared to the ground truth (blue). The red light lines show the effect of the position correction after every three minutes. Table 7.1 shows the horizontal position error statistic. The fairly large mean error in position, namely 23 meters, resulted from the difficulty in obtaining an accurate heading solution using the visual gyroscope in this fairly open sub-urban environment lacking straight lines from high-rise buildings of urban canyons. This was anticipated to improve when the method is experimented in an urban canyon. The length of the obtained path agreed with the ground truth and therefore the visual odometer utilizing the carrier phase measurements was shown to provide promising results.

Fig. 7.2. *Position solution verification using vision-aided carrier phase navigation (position red dot, path red line) and compared to ground truth (blue) in a sub-urban environment in Calgary, shown in Google Earth.*

7.3 Vision-Aided GNSS Navigation in an Urban Environment

In order to show the advantages and limitations of the method in severe urban canyons using an unaided standard receiver, a test was carried out in downtown Calgary as shown in Figure 7.3. This canyon encompasses tall and reflecting buildings heavily blocking the satellite signals and/or causing multipath. The test duration was 25 minutes. The user position and heading were initialized using the reference system described in the previous section. Then, the user position was propagated using the Kalman filter described above and the following procedure. The Novatel SPAN-SE GPS/GLONASS receiver was used to acquire the pseudorange (L1), carrier phase (L1) and Doppler measurements as well as the GNSS navigation message. Only GPS measurements were used in the processing. The GoPro camera was used for capturing a video stream that was sampled at 10Hz. All equipment was attached to a backpack carried by the user as in the experiment above. When four or more satellites were acquired, the user position was computed using the pseudorange measurements and the least-squares method described in Chapter 2. As may be seen below, the GPS measurements are heavily degraded in such challenging environments and therefore the quality of the position solution was monitored by examining at the least-squares residuals. The residual $\mathbf{r}$ is computed after the least-squares final solution is obtained and it expresses the difference between the anticipated and obtained measurements. The residual vector is calculated using the pseudorange measurement vector $(\mathbf{z})$, the geometry matrix $(\mathbf{H})$ and the estimated user position vector $(\hat{\mathbf{x}})$ as

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\hat{\mathbf{x}}. \tag{101}$$

Fig. 7.3. *Calgary downtown environment used for vision-aided GNSS navigation.*

When the residual of any satellite *i* exceeded a threshold (herein 20 m, selected by experimentation), the GPS solution was discarded and a vision-aided carrier phase solution was computed instead using the position and heading from the previous epoch in the state vector for initialization. When the consecutive epochs provided successful GPS position solutions, heading between the epochs was computed. As the error in GPS-derived position using pseudoranges is commonly a few meters even in a favorable environment, the heading computed from two consecutive epochs having a time interval of only one second, would be erroneous in most cases. Therefore, the heading was computed using the longest interval of successful positions, however not exceeding 10 epochs. The heading was computed using the latitude (ϕ) and longitude (λ) of the position at the first time epoch (1) and last epoch used (n) as [**?**]

$$\theta = \begin{array}{c} mod(\arctan 2(\sin(\lambda_n - \lambda_1)\cos(\phi_n), \\ \cos(\phi_1)\sin(\phi_n) - \sin(\phi_1)\cos(\phi_n)\cos(\lambda_n - \lambda_1)), 2\pi). \end{array} \quad (102)$$

When less than four satellites were found, the translation and heading of the user were computed using the visual gyroscope and the visual odometer presented in this chapter and the position propagated using the Kalman filter. Figure 7.4 shows the number of satellites obtained for each time epoch in the experiment (blue). As the experiment was started in an open area, up to 9 GPS satellites were occasionally used. As the path proceeded into the urban canyon the number of satellites used decreased, dropping under four towards the end of the data set. The number of observed satellites does not directly however reflect how many satellites were available for position computation. The figure shows also the epochs when the obtained position accuracy was deemed unreliable based on the residual and vision-aided carrier phase solution

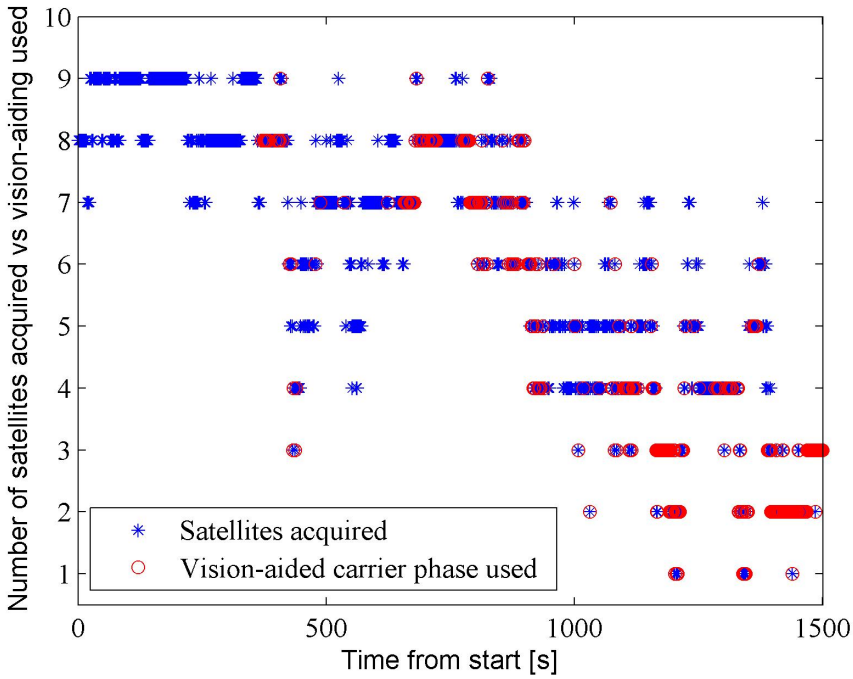

Fig. 7.4. *Number of satellites acquired in an urban canyon for each time epoch of the experiment (blue) and the epochs when vision-aided carrier phase navigation used due to too few satellites or large residuals in GNSS pseudorange position estimations (red above the blue mark).*

used instead (red above the blue mark).

Figure 7.5 shows the path obtained using vision-aided carrier phase navigation (blue), GPS-only solution (green) and the ground reference (red). As the GPS-only positions deviate strongly from the reference in the end of the data set, the figure has been zoomed showing better the vision-aided solution results. The vision-aided navigation solution provided fairly accurate results at the beginning of the experiment but deteriorated as the user entered the urban canyon and obtained poor GPS measurements, resulting in 200 meters of error in the worst case.

Figure 7.6 shows the horizontal errors of the vision-aided carrier phase navigation and GPS-only solutions. Again, at the beginning of the experiment the errors remained low but deeper inside the urban canyon they grew. The figure shows also the main reason for the error growth, namely the GPS position computation process

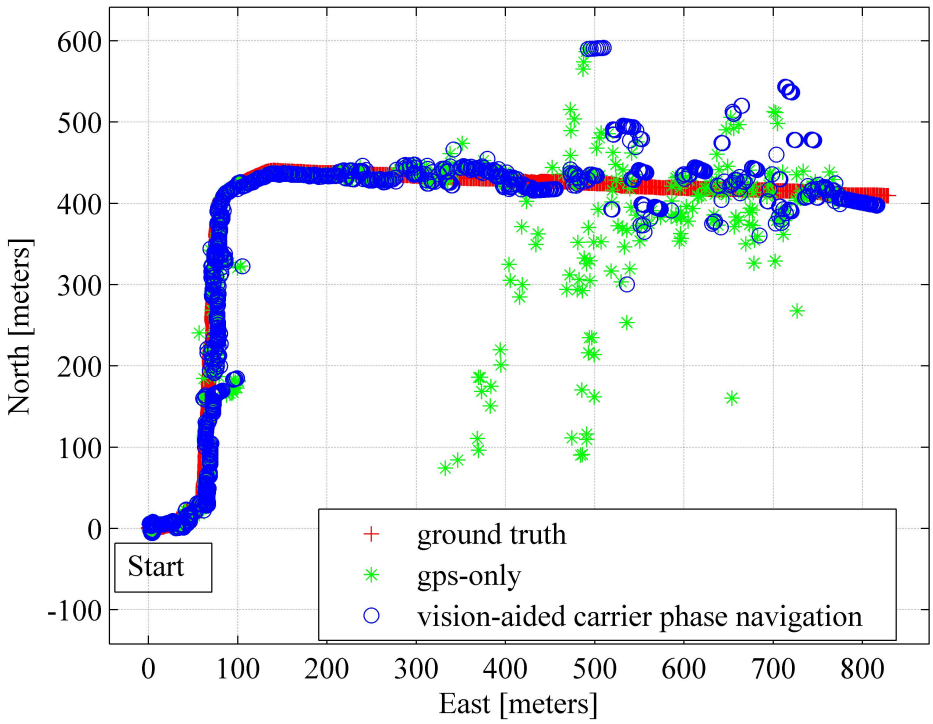

Fig. 7.5. *Position solution in an urban canyon using the ground truth (red), GPS-only(green) and vision-aided carrier phase navigation (blue).*

resulting in low range residual values and therefore used although being erroneous making the GPS position significantly deteriorated. The figure is zoomed to exclude the largest GPS-only errors enabling closer examination of the vision-aided solution errors. Heavily multipath-affected observations are difficult to discard by assessing range residuals only, especially when the measurement redundancy is low. A better filtering and observation selection for the obtained GPS position should be developed for a more accurate vision-aided navigation solution. Table 7.2 shows the statistics for the horizontal position errors, namely a mean error of 25 meters with a standard deviation of 48 meters. However, already in this simple implementation, the vision-aided carrier phase method improves the navigation solution significantly as may be seen from the table also showing the corresponding horizontal position error statistics for a GPS-only solution based on the pseudoranges (mean error 73 meters, standard deviation of 1241 meters). This positive effect of vision-aiding may also be

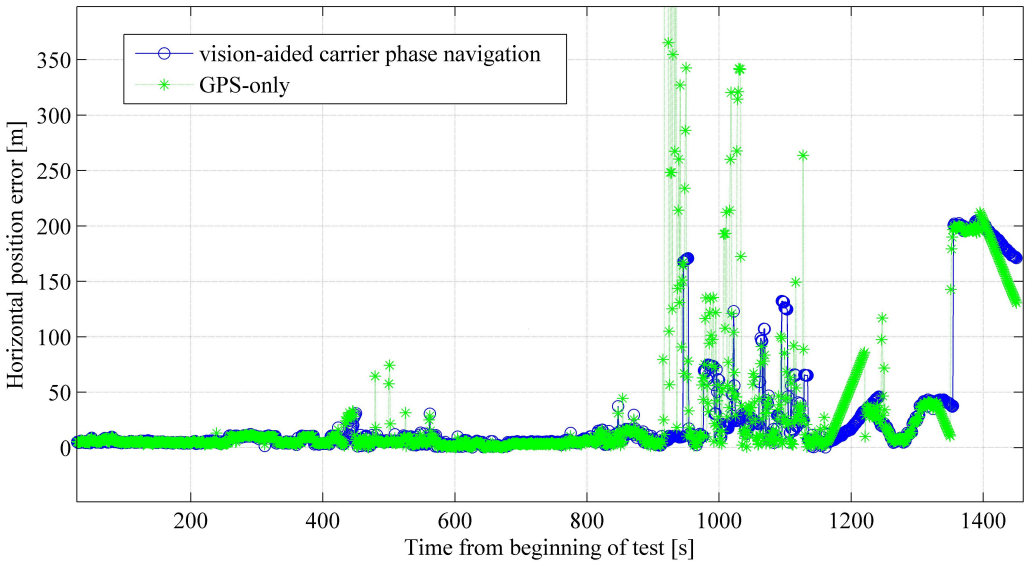

Fig. 7.6. *Horizontal position error in an urban canyon obtained using vision-aided carrier phase navigation and GPS-only.*

Table 7.2. *Positioning error statistics using vision-aided carrier phase (VA) and GPS only (GPS)*

Statistics	min error (m)	max error (m)	mean error (m)	std of error (m)
VA	0.4	200	25	48
GPS	0.1	4015	73	1241

seen by comparing the navigation paths obtained using the vision-aided carrier phase navigation shown in Figure 7.7 and using the GPS pseudorange measurements only, propagated using a simple Kalman filter but no error detection for the obtained position, shown in Figure 7.8. Both figures show the obtained position solution (green dots), its path (red line) and the ground truth (blue).

As anticipated, performance especially when using GPS alone, was significantly degraded in the urban canyon. Due to the frequent unavailability and large GPS position errors, the vision-aided solution suffered significantly. As the position and heading from GPS-only were already erroneous when the navigation solution computation switched to the vision-aided carrier phase method, the user position was still poor

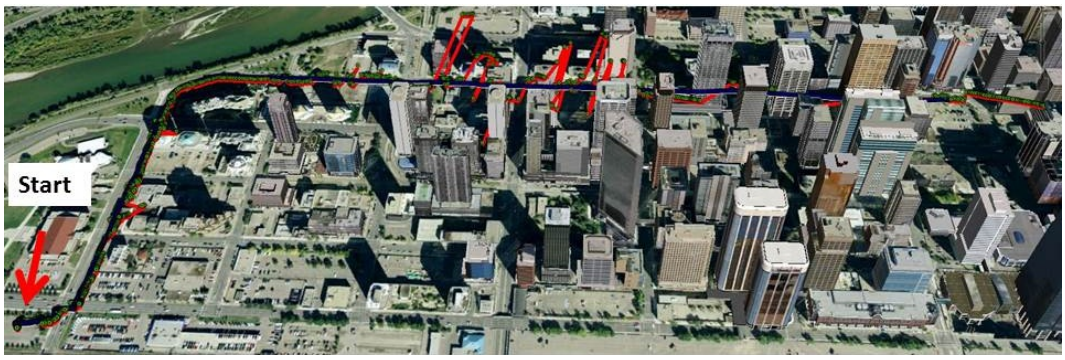

Fig. 7.7. *Position solution using vision-aided carrier phase navigation (position green dot, path red line) and compared to ground truth (blue) in an urban canyon in downtown of Calgary, shown in Google Earth.*

despite the addition of accurate visual measurements,. If other sensors were added to aid GPS, better measurement error detection for the GPS measurements would occur, in which vision-aided carrier phase navigation would provide significantly better results. The absolute heading could be obtained by using a 3D magnetometer.

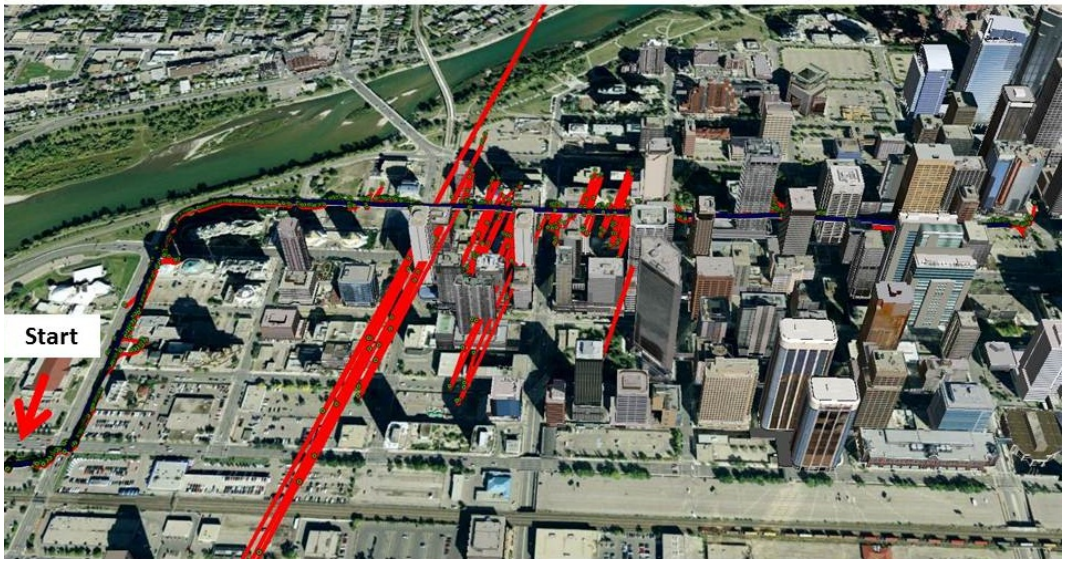

Fig. 7.8. *Position solution using GPS measurements only (position green dot, path red line) and compared to ground truth (blue) in an urban canyon in downtown of Calgary, shown in Google Earth.*

In urban canyons, magnetometer heading measurements deteriorate due to magnetic disturbances arising from nearby ferrous environments. However, techniques to mitigate these errors, especially when magnetometer measurements are combined with 3D accelerometers and rate gyros, are emerging [**?**]. Inertial sensors and barometers would also aid the error detection process, thereby eliminating large GPS measurement errors.

8. CONCLUSIONS

There is a strong need for enhanced pedestrian navigation systems for improved safety and to improve everyday life. First responders, electronic monitoring and military personnel operate in challenging situations and need a system that is available in all environments. Moreover, general users need precise indoor navigation to locate specific rooms in buildings and to use location based applications. Pedestrian navigation is mainly needed indoors and in urban environments. Although indoor and urban navigation has been an active research subject for years, no unique navigation system addressing all needs has yet been developed with a level of performance similar to that of GNSS in the outdoors. A pedestrian navigation system has to be light and small in size, have low power consumption and price, in addition to perform well in all environments. Therefore it has preferably to be independent of specific indoor infrastructures such as RF access points.

So far the most promising approach for pedestrian navigation is the fusion of many different sensors and positioning systems, the most widely used being self-contained sensors, GNSS and WLAN. The performance of GNSS is degraded indoors and in urban canyons, WLAN needs an a priori prepared infrastructure and the errors in self-contained sensors result in position solutions that drift in time and become distorted. Hence, other means for augmenting or replacing some of these methods have to be used. Vision-aiding is a feasible method in many environments because it is affected by error sources which are different from those of other navigation technologies. Consecutive images provide relative information about the attitude and translation of the camera, which can be further transformed into user heading and position information. In favorable environments and circumstances this information obtained from the images results in a much more accurate and available navigation solution when integrated with other measurements; it can also be used for stand-alone navigation for short periods of time if the absolute position and heading are known at the beginning of these periods. However, vision-aiding solution suffers from errors due to low

lighting or of scenes unsuitable for visual perception, especially those including moving objects (i.e. human, vehicles). Therefore vision-aided systems need occasional calibration from an absolute navigation system.

In this thesis, new tools for vision-aiding navigation solutions were developed and tested, namely a concept called visual gyroscope for observing the user heading and a visual odometer for translation. Both methods provide user displacement information by monitoring the motion of features in consecutive images captured by a camera carried by the user. Different methods were investigated to combine the above measurements using Kalman filtering and vision-aided navigation systems were obtained. The next section discusses the main results.

Despite the challenges in vision-aiding arising from the occasional unsuitability of the urban navigation environment for visual processing as well as difficulties related to computer vision algorithms, vision-aiding improves navigation accuracy, availability, reliability and continuity.

8.1 Main Results

The visual gyroscope tracks the motion of virtual points, called vanishing points, arising from the projective geometry mapping the parallel lines in a three dimensional scene into a two dimensional image. The change in the location of the vanishing points can further be transformed into the attitude of the user and therefore the method is called the visual gyroscope. In an ideal situation where there is enough and constant light, the environment has a favorable structure and does not contain any dynamic objects, a visual gyroscope processing images from a static camera produces almost no error compared to a reference system in the user attitude. In the situation where the lighting is not the best possible and dynamic objects disturb the vanishing point perception, a static visual gyroscope performs still better than a common MEMS gyroscope in a performance test. In a real navigation situation the conditions vary a lot and deteriorate the visual perception and therefore careful error detection is needed. An error detection algorithm suitable for pedestrian navigation, namely a method called LDOP monitoring the reliability of the visual gyroscope's attitude measurements based on the geometry of the lines used, was developed. When the visual gyroscope's measurements passing the error detection were used as updates for a Kalman filter propagating the angular velocity and acceleration provided by

an IMU, the obtained user attitude improved significantly. When the user heading was occasionally calibrated using the building layout, the attitude improved 93% and when the visual gyroscope's heading change measurements were used with no calibration, an improvement for the attitude of 40% was obtained in the experiment.

The visual gyroscope is incapable to detect the magnitude of sharp turns, namely when the turn is close to 90 degrees or more. This was addressed by developing a method using the IMU attitude measurements to aid the vanishing point detection also making the procedure more accurate. So far the detection of the sharp turns was implemented, whether it is possible to also observe the magnitude of the turns is a task for future research. Simultaneously, the algorithm of IMU-aided vanishing point detection was found to decrease the computational burden of the line detection needed in the visual gyroscope method and therefore making the real-time performance of the navigation solution possible.

The visual odometer identified the matching feature points in consecutive images and used the homography relation to observe the translation of the user i.e. the distance travelled. As the translation obtained from the images has an ambiguous scale, two different approaches to solve it were developed. A visual gyroscope feasible for indoor navigation resolved the scale by using the known, special configuration of the camera. As the height of the camera was known a priori and kept sufficiently static (vertical motion of $\pm 10cm$ was still found to provide accurate results), the pitch of the camera was obtained using the visual gyroscope, basic geometry could be used to compute the depth of the object found in the image and therefore the scale too. As the method observed the attitude of the camera using the visual gyroscope and the camera was kept static, only the horizontal translation was left to be resolved reducing the amount of matching image points needed, a profitable feature especially for indoor navigation. For the special configuration to be useful, the image points had to lie on the floor. However, again because the attitude of the camera was obtained using the visual gyroscope, the degeneracy problem arising from using image points all found from a plane, was avoided. The performance of the visual odometer was evaluated by looking at the agreement of the translation obtained using the method and the ground truth, and it was found to be over 90% in all experiments. The visual gyroscope and odometer were integrated with a multi-sensor multi-network system and tested in an office corridor with a configuration having a workable WLAN positioning solution and using an outdated WLAN radio map. The vision-aided fused solution was found

to improve the mean error of the user position from 1.5 to 2 meters. The visual gyroscope and odometer were also tested as stand-alone system in more challenging environments, namely in a shopping mall and an urban canyon, also resulting in improved position accuracy.

In urban canyons the GNSS measurements are typically available, however deteriorated. A method utilizing GNSS carrier phase observations for resolving the scale problem in translation was developed. After observing the ambiguous translation from the consecutive images, the scale was obtained by looking at the differentiated carrier phase measurements from the satellites and observed at the time epochs of capturing the images. As the carrier phase measurements were differentiated, the integer ambiguity in carrier phase measurements was cancelled. Because only the magnitude of the translation and the receiver clock error needed to be resolved, tracking two satellites was enough. By integrating the visual gyroscope's and odometer's measurements using a Kalman filter, an initial position could be propagated and a resulting navigation solution obtained. The method was first experimented in an suburban environment propagating only the visual gyroscope induced orientation and odometer's translation and solving the scale for the translation from differentiated GNSS carrier-phase measurements for two satellites. Since the errors in the visual perception deteriorate the position, the solution was calibrated using the reference system once in three minutes and a mean error of 24 meters was obtained in the 15-minute experiment. In a real urban canyon GPS was vision-aided by propagating the user heading and position using the visual gyroscope's and odometer's measurements when less than four satellites were observed or the GPS induced position was distorted based on assessing the solution's residual values. Again, as the visual gyroscope and odometer are dependent of an absolute initialization, and they were re-initialized whenever more than four acceptable satellite observations were available providing a GPS solution, the errors in the GPS position used to calibrate the visual measurements occasionally deteriorated also the vision-aided navigation result. However, the overall position accuracy improved significantly compared to the one obtained using GPS positioning only, namely the mean error decreased from 73 meters to 25 meters when the vision-aided GPS carrier-phase based processing was used.

Despite of all challenges in the urban and indoor environments the visual aiding improved the navigation solution in all different experiments discussed in the thesis. However, it should be noted that the tests were done using a limited amount of data

and therefore future research should assess the performance of the developed methods via extended duration of data collection. Due to its distinctive characteristics, visual processing complements other positioning technologies in order to provide better pedestrian navigation accuracy and reliability.

8.2 Future Development

The largest challenges in using the visual gyroscope's measurements for vision-aiding the user attitude is its incapability to observe the magnitude of the turns and therefore the need for occasional absolute calibration. This problem was preliminarily addressed by developing a visual gyroscope utilizing the attitude information obtained from an IMU to aid the vanishing point location computation. From the disagreement of the vanishing point location provided by the IMU and the visual gyroscope, the occurrence of a sharp turn was observed. Future work should address the possibility to observe the magnitude of such a turn and therefore eliminate or at least decrease the need for calibration during navigation.

As the integration of different systems is beneficial, or even mandatory, for indoor and urban area navigation, the error detection for visual measurements as well as for observation from the other systems is crucial and should be a topic for further research. Means for emphasizing the strengths of all systems involved as well as covering for their deficiencies is a subject for the development of even more functional and seamless integration means. Though this thesis addressed various challenges in the indoor and urban navigation and proposed various visual processing methods for positioning, the matter of ubiquitous pedestrian navigation is by no means yet solved by using vision-aiding, but definitely improved.

BIBLIOGRAPHY