ADAPTIVE VISON AIDED INTEGRATED NAVIGATION FOR
DYNAMIC UNKNOWN ENVIROMENTS

By

HEBA NEMATALLAH

A thesis submitted to the
Department of Electrical & Computer Engineering
In conformity with the requirements for
The degree of Masters of Applied Science

Queen’s University
Kingston, Ontario, Canada
(December, 2015)

Copyright ©Heba NematAllah, 2016
Abstract

In this research, a novel method for visual odometry (VO) and the integration with multi-sensors navigation systems for vehicular platforms is proposed. The proposed method partitions the field of single camera view into regions of interests where each region likely contains different types of visual features. By applying computer vision processing techniques, ambiguous pose estimation is calculated up to a scale factor. The proposed method uses aiding measurements from vehicle’s odometer to adaptively resolve the scale factor ambiguity problem in monocular camera systems. Unlike some state-of-art approaches, this work does not depend on offline pre-processing or predefined landmarks or visual maps. In addition, this work addresses unknown uncontrolled environments where moving objects likely exist. Innovative odometer-aided Local Bundle Adjustment (LBA) along with a fuzzy C-mean clustering mechanism is proposed to reject outliers corresponding to moving objects. A Gaussian Mixture approach is also applied to detect visual background regions during stationary periods which enables further rejection of moving objects. Finally, an empirical scoring method is applied to calculate a matching score of the different visual features and to use this score in a Kalman filter as measurement covariance noise to integrate VO-estimated pose changes within a larger multi-sensors integrated navigation system. Experimental work was performed with a physical vehicular platform equipped by MEMS inertial sensors, GPS, speed measurements and GPS-enabled camera. The experimental work includes three testing vehicular trajectories in downtown Toronto and the surrounding areas. The experimental work showed significant navigation improvements during long GPS outages where only VO is fused with inertial sensors and the vehicle’s speed measurements.
Acknowledgements

Though only my name appears on the cover of this thesis, a great many people have contributed to its production. I owe my gratitude to all those people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

My deepest gratitude is to Dr. Geoffrey Chan. I have been amazingly fortunate to have an advisor who gave me the guidance to recover when my steps faltered. His support helped me overcome crisis situation and finish this dissertation.

The graduate program Assistance, Debra Fraser, has been always there to listen and give advice. I am deeply grateful to her support.

I would like also to thank my colleagues at the Navigation and Instrumentation (NavINST) research group led by Dr. Noureldin: Tashfeen Karamat, Shiefi Liu and Mohamed Atia. They were really helpful and cooperative. They did not hesitate to share knowledge and experience with me.

Many friends have helped me stay sane through these difficult years. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me.

My Husband, Mohamed Atia, without his infinite source of energy, hope, and love, I would not be able to write this thesis by any mean. He was always there cheering me up and stood by me through the good times and bad.

Most importantly, none of this would have been possible without the love and patience of my parents. My parents have been a constant source of love, concern, support and strength all these years. My kids Abdelrahman, Abdallah and Fatima were always source of happiness and motivation.

I am very grateful for my supervisors Dr. Noureldin and Dr. Korenberg for giving me the opportunity to pursue this thesis and to enjoy this research.

Finally, I’m thankful to Allah Almighty, the most Gracious, and the most Merciful. Allah blessed me with great people and gave me the opportunity to undertake and execute this work.
# Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements .................................................................................................................. iii
List of Figures ........................................................................................................................... iii
List of Tables ........................................................................................................................... viii
List of Abbreviations ............................................................................................................ xi
Chapter 1 Introduction ........................................................................................................... 1
  1.1 Background .................................................................................................................... 1
    1.1.1 Vision based Localization ......................................................................................... 1
    1.1.2 Monocular Visual Odometry .................................................................................... 3
  1.2 Approach ....................................................................................................................... 7
  1.3 Thesis Contribution ........................................................................................................ 8
  1.4 Outline .......................................................................................................................... 11
Chapter 2 Literature Review of Navigation in Unknown Environments .............................. 12
  2.1 Standalone Navigation Systems ...................................................................................... 12
    2.1.1 Global Navigation Systems (GPS) ........................................................................ 12
      2.1.1.1 The main concept ............................................................................................... 13
      2.1.1.2 GPS Limitations ................................................................................................. 14
    2.1.2 Inertial Sensors and Integrated Navigation Systems (INS) .................................. 15
      2.1.2.1 INS Errors ......................................................................................................... 15
      2.1.2.2 MEMS ............................................................................................................... 16
    2.1.3 Vision Sensors and Vision-Aided Navigation .......................................................... 16
  2.2 Multi Sensor Integrated Navigation ................................................................................ 17
    2.2.1 INS/GPS Integration ............................................................................................... 17
    2.2.2 Vision/GPS Integration ........................................................................................... 19
      2.2.2.1 Loosely-coupled approach .............................................................................. 20
      2.2.2.2 Tightly-coupled approach .............................................................................. 21
    2.2.3 Vision/INS Integration ............................................................................................. 21
2.2.3.1 Loosely-coupled approach .................................................. 22
2.2.3.2 Tightly-coupled approach .................................................. 22
2.2.4 Vision/INS/GPS Integration .................................................... 23
  2.2.4.1 A Two Separate Filters Integration Systems Approach .......... 23
  2.2.4.2 Two coupled Integration Filters Approach .......................... 23
  2.2.4.3 Tightly coupled and feed-forward EKF ............................ 24
2.2.5 Platform odometer/Vision Integration ...................................... 24
2.2.6 Odometer/Vision/INS Integration ........................................... 25
2.2.7 Odometer/INS /GPS Integration ............................................. 25
2.2.8 Odometer/Vision/INS/GPS Integration ..................................... 26
2.3 Conclusion .............................................................................. 26
Chapter 3 Aided Monocular Visual Odometry for Unknown Environments 27
  3.1 Monocular Visual Odometry Problem Formulation ....................... 27
    3.1.1 Mapping Physical 3-D world into 2-D Image Domain: System Model 28
    3.1.2 Modeling Camera Motion Transformation: The Essential Matrix E 29
  3.2 The Proposed Aided- Visual Odometry Approach ......................... 29
  3.3 Spatially Diverse Feature Detection ......................................... 31
    3.3.1 Region Of interest Selection .............................................. 31
    3.3.2 Hough Transform (HT) for line detection ............................. 32
      3.3.2.1 Hough Space and Lines Representation ............................ 33
      3.3.2.2 Lines Matching Challenge and the Proposed Solution .......... 34
    3.3.3 Quantization Based Road Signs Detection (RSD) ..................... 35
      3.3.3.1 RSD Detection Mechanism ........................................... 35
      3.3.3.2 RSD Processing Challenge and the Proposed Solution ........... 36
    3.3.4 Ground-Texture Detection using SIFT .................................. 36
  3.4 Feature Matching .................................................................... 37
    3.4.1 SIFT Descriptors ............................................................... 37
  3.5 Background Detection for Zero Speed and Moving Objects Rejection using Gaussian
    Mixture Models ...................................................................... 38
    3.5.1 Adaptive Gaussian Mixture Model .................................... 39
3.5.2 Zero Velocity Update and Moving Objects Rejection ........................................... 39
3.6 Outlier Removal using RANSAC ............................................................................. 40
3.7 Relative Pose Estimation ...................................................................................... 43
3.8 Odometer-Aided Local Bundle Adjustment for Automatic Scale Factor Estimation ........ 44
  3.8.1 Scale Factor Ambiguity ...................................................................................... 45
  3.8.2 Odometer based Scale Factor Estimation .......................................................... 46
  3.8.3 Local Bundle Adjustment (LBA) ........................................................................ 46
3.9 Matching Score Calculation ................................................................................... 47
3.10 Fuzzy C-mean Clustering for statistical moving objects rejection ......................... 48
3.11 Sample Results based on The Proposed Odometer-Aided Visual Odometer ............ 49
Chapter 4 Vision-Aided Integrated Navigation ............................................................ 53
  4.1 Abstract .................................................................................................................. 53
  4.2 Introduction ............................................................................................................ 53
  4.3 INS Mechanization ................................................................................................. 54
    4.3.1 Mechanization Equations: Position Vector ....................................................... 56
    4.3.2 Velocity Vector Equations ................................................................................ 57
    4.3.3 Attitude Vector Equations ................................................................................ 58
  4.4 Filter Design ........................................................................................................... 59
    4.4.1 INS-predicted Pose Change vs. VO-estimated Pose Change Updates ............... 60
    4.4.2 Linearized system error model ......................................................................... 60
    4.4.3 Measurement Model ......................................................................................... 64
      4.4.3.1 GPS measurements ..................................................................................... 64
      4.4.3.2 Visual Odometry measurements .................................................................. 65
  4.5 Conclusion .............................................................................................................. 65
Chapter 5 Experimental Work ..................................................................................... 66
  5.1 Abstract .................................................................................................................. 66
  5.2 Experimental Setup ................................................................................................. 66
  5.3 Data Synchronization ............................................................................................. 69
  5.4 Testing Trajectories ............................................................................................... 69
  5.5 Sensors Specifications ........................................................................................... 72
5.6 Data Processing and Filter Tuning ................................................................. 73
5.7 RESULTS ................................................................................................. 73
  5.7.1 Trajectory#1 .......................................................................................... 73
  5.7.2 Trajectory#2 .......................................................................................... 76
  5.7.3 Trajectory#3 .......................................................................................... 79
Chapter 6 Conclusion and Future Work .......................................................... 82
  6.1 Conclusion ............................................................................................... 82
  6.2 Future Work ............................................................................................. 83
Bibliography ..................................................................................................... 85
List of Figures

Figure 1-1 Scale Factor Ambiguity ......................................................................................... 4
Figure 1-2 Flow chart of Feature-based monocular Visual Odometry .................................. 5
Figure 1-3 The partitioning of the video frame. ..................................................................... 9
Figure 2-1 loosely-coupled GPS/INS integration scheme ..................................................... 18
Figure 2-2 loosely-coupled GPS/VO integration scheme ...................................................... 20
Figure 3-1 The perspective projection of a physical point into the image plan .................. 28
Figure 3-2. Block Diagram of the Proposed System Approach ............................................ 30
Figure 3-3 the partitioning of the video image frames ........................................................ 32
Figure 3-4 Example of line mapping from image space into Hough Space ......................... 34
Figure 3-5 Examples of Road Signs that can be detected using Quantization based RSD .. 36
Figure 3-6 matched features between two sub-frames of the third region of interest (Ground texture) ........................................................................................................................................................................... 38
Figure 3-7. The flowchart of RANSAC Algorithm for Visual Odometry ......................... 41
Figure 3-8 RANSAC steps for the simple 2D line estimation example [53] ......................... 42
Figure 3-9. Flow chart of Motion Estimation Step ............................................................... 44
Figure 3-10. Scale Factor Ambiguity .................................................................................... 45
Figure 3-11 Two Different frames to estimate camera relative pose change ..................... 50
Figure 3-12 Estimated Camera Motion between the two frames shown in Figure 3-11 ...... 50
Figure 3-13. Testing Trajectory in downtown Toronto ......................................................... 51
Figure 3-14. Odometer-aided VO test results. 2D trajectory is displayed ......................... 52
Figure 4-1 The block diagram of the multi-sensor integrated navigation system ................ 59
Figure 5-1 Experimental Van ............................................................................................... 67
Figure 5-2 Novatel SPAN Unit + VTI IMU + Ublox GPS .................................................. 67
Figure 5-3. GPS-enabled MiVue 388 digital cameras ......................................................... 68
Figure 5-4. CarChip to record vehicle speed ......................................................................... 68
Figure 5-5. Testing Trajectory#1 ........................................................................................ 70
Figure 5-6. Testing Trajectory#2 ........................................................................................ 70
Figure 5-7. Testing Trajectory#3 ................................................................................................. 71
Figure 5-8. 2D Trajectory#1. .................................................................................................... 74
Figure 5-9. Errors in Trajectory#1. .......................................................................................... 75
Figure 5-10. 2D Trajectory#2. ............................................................................................... 77
Figure 5-11. Errors in Trajectory#2. ...................................................................................... 78
Figure 5-12. 2D Trajectory#3. ............................................................................................... 80
Figure 5-13. Errors in Trajectory#3. ...................................................................................... 81
List of Tables

Table 1 Comparison between Corner detectors and blob detectors ............................................. 6
Table 2. Frames of the Navigation Processing .............................................................................. 55
Table 3. Testing Trajectories Distances Travelled ...................................................................... 71
Table 4. VTI Data Logger Sensors Specifications .......................................................................... 72
Table 5. Error Values for the testing trajectories ......................................................................... 79
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKF</td>
<td>Extended Kalman Filter</td>
</tr>
<tr>
<td>FOV</td>
<td>Field Of View</td>
</tr>
<tr>
<td>FCM</td>
<td>Fuzzy C-Mean</td>
</tr>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Navigation Systems</td>
</tr>
<tr>
<td>HT</td>
<td>Hough Transform</td>
</tr>
<tr>
<td>IMU</td>
<td>Inertial Module Unit</td>
</tr>
<tr>
<td>INS</td>
<td>Inertial Navigation Systems</td>
</tr>
<tr>
<td>LBA</td>
<td>Local Bundle Adjustment</td>
</tr>
<tr>
<td>MEMS</td>
<td>Micro-Electromechanical Systems</td>
</tr>
<tr>
<td>MVO</td>
<td>Monocular Visual Odometry</td>
</tr>
<tr>
<td>Odo</td>
<td>Odometer</td>
</tr>
<tr>
<td>RANSAC</td>
<td>RANdom SAmple Consensus</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RSD</td>
<td>Road Signs Detection</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale Invariant Feature Tracking</td>
</tr>
<tr>
<td>SLAM</td>
<td>Simultaneous Localization And Mapping</td>
</tr>
<tr>
<td>SINS</td>
<td>Strap-down Inertial Navigation System</td>
</tr>
<tr>
<td>VO</td>
<td>Visual Odometry</td>
</tr>
<tr>
<td>ZUPT</td>
<td>Zero Velocity Update</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

1.1 Background.

Global Navigation Satellite Systems (GNSS) [1] provide consistent accurate positioning in open-sky and if reliable satellites visibility is available. However, the accuracy is not yet sufficient for everywhere emerging applications in challenging environments such as urban areas and tunnels where GNSS signals are mostly blocked. Although integration with inertial navigation systems (INS) [2] enhances navigation accuracy by bridging GNSS interruption periods, the performance needs further improvements in dense urban areas where GNSS signals suffer from severe multipath that distracts the INS/GNSS fusion filter by feeding the filter significantly erroneous GNSS measurements updates. To overcome this problem, more sensors are integrated to increase the consistency and accuracy of the dead-reckoning navigation during GNSS outages. One of the new emerging sensors to assist INS/GNSS navigation systems is to incorporate vision sensors from digital cameras mounted in the vehicle platform [3] [4].

1.1.1 Vision based Localization

Localization is the process of estimating an object pose (position and attitude) relative to a reference frame, [2] based on sensor inputs. Visual Odometry (VO) is the process of estimating the egomotion of a platform (e.g., vehicle, human, or robot) using only the input of a single or multiple cameras attached to it [7]. The idea of VO is to determine the position and the orientation of a camera by analyzing the motion of fixed objects in the field of view (FOV) of the camera. In this area of
research, there are set of assumptions need to be consider. First, the sufficient illumination in the environment for better complex feature detection and tracking. Second, scene with enough number of static objects to allow apparent motion to be extracted. Third, sufficient scene overlap between the consecutive frames. Actually, due to progress in computer vision hardware and software, it became possible to consider visual odometry systems as promising navigation sensors that can complement traditional navigation sensors such as GNSSs and INS [5].

Vision based localization techniques can be classified according to the knowledge of the environment in which it will be applied [6]. The first category is vision based localization in known environments, where some prior description of the environment is available (e.g., map, landmarks or interest points to track). In the second category, vision based localization techniques in unknown environments, are applied where there is no prior information about the environment. The latter category is more challenging since real-life scenarios are mostly uncontrolled dynamic environments. Therefore, sufficient environment illumination or a known number of static objects are not valid assumptions in most of cases.

In addition to the pervious classification, vision based localization techniques can be categorized according to the number of cameras the system utilizes [7]. Commonly there are single (monocular), two (binocular or stereo), or three (trinocular) cameras, and other types of camera sensing such as omnidirectional and optical scanning. In monocular vision based localization, arises the ambiguity of scale factor of the motion parameters resulted from the transformation of the visual data from 2D image domain to 3D world domain (scale factor will be described in detail in section 1.1.2). On the other hand, the stereo camera based techniques can solve this problem using
triangulation. However, the depth information of objects that are too distant is hard to recover with the stereo cameras.

In another classification, the vision based techniques can be categorized according to the mounting of the camera which will provide different types of the visual information [6]. For example, a forward-looking camera will provide more visual information of landmarks. Conversely, a downward-looking camera will provide visual information of ground texture. Other classifications depend on the type of features that the VO method tracks. The three main categories are feature based methods, appearance based methods, and hybrid methods. Feature-based methods are based on salient and repeatable features that are tracked over the frames. Appearance-based methods use the intensity information of all the pixels in the image or sub-regions of it. Hybrid methods use a combination of the formerly mentioned two methods.

1.1.2 Monocular Visual Odometry

Monocular visual odometry is the process of estimating the relative translation and rotation of a single camera by analyzing the motion of fixed objects in the camera FOV [8]. A major challenge in this approach is the scale factor ambiguity. The scale factor ambiguity results from the projection of an object at unknown depth into the camera frame. This projection could be created by an infinite number of object size and depth combinations. For example, Figure 1-1 shows two lines which have different sizes and depths but their projections on the camera frame are the same size.
Figure 1-1 Scale Factor Ambiguity.

Lines L₁, L₂ in 3-space is imaged as line l₁, l₂ by a perspective camera which its’ center is C.

In order to overcome this challenge, the triangulation technique is commonly used [7]. The general idea of triangulation is to compute the relative distance between the different combinations of the 3-D key-points at the subsequent image frames by using the following formula

\[ r = \frac{\| X_{k-1,i} - X_{k-1,j} \|}{\| X_{k,i} - X_{k,j} \|} \]  \hspace{1cm} (1.1)

Where \( r \) is the relative scale, \( X_{k,i} \) is the position of the point i at the image k.

According to [7], most VO implementations are feature based. Therefore, in this section, a quick summary of the commonly used VO main steps will be reviewed.
Figure 1-2 Flow chart of Feature-based monocular Visual Odometry

The main Steps of Visual Odometry:

1. Feature Detection Step: This step aims at identifying image correspondences. For monocular VO, the utmost feature detectors are corner and blob detectors [9]. A comparison between both detectors is shown in Table 1.
Table 1 Comparison between Corner detectors and blob detectors

<table>
<thead>
<tr>
<th>Corner detectors</th>
<th>blob detectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corner is a point at the intersection of two or more edges</td>
<td>Blob is an image pattern that differs from its immediate neighborhood in terms of intensity, color, and texture. It is not an edge, nor a corner</td>
</tr>
<tr>
<td>fast to compute but are less distinctive</td>
<td>more distinctive but slower to detect</td>
</tr>
<tr>
<td>better localized in image position</td>
<td>less localized in scale</td>
</tr>
<tr>
<td>Cannot be re-detected as often as blobs after large changes in scale and viewpoint.</td>
<td>Blobs are not always the right choice in indoor environments</td>
</tr>
</tbody>
</table>

2. Feature Description Step: In this step, the region around each detected feature is converted into a compact descriptor that can be matched against other descriptors.

3. Feature Matching Step: The feature-matching step searches for corresponding features in consecutive image frames. The set of matched corresponding to the same feature is called feature track. The Euclidean distance is commonly used as a similarity measure for feature descriptors matching.

4. Outlier Removal Step: Outliers are wrongly detected data associations. Random sample consensus (RANSAC) has been established as the standard method for model estimation in the presence of outliers. The idea behind RANSAC is to compute model hypotheses from randomly sampled sets of data points and then verify these hypotheses on different data points [7]. The hypothesis that shows the highest consensus with the other data is selected as the correct correspondence.
5. **Motion Estimation Step**: computing the relative motion between two frames. Camera pose is then computed incrementally by integrating relative motions. Integrating relative motion is known as “dead-reckoning”.

6. **Bundle Adjustment**: aims to minimize the sum of squared re-projection errors of the reconstructed 3-D point over the last \( m \) frames.

### 1.2 Approach

To overcome the problems of VO in a real-life uncontrolled unknown environment, this thesis applies common VO methods in a hybrid adaptive scheme aided by the vehicle’s odometry. In this work, the visual information of a single forward-looking camera will be processed. Different types of spatially diverse features will be detected from this visual information. The detected features can be lines, edges and connected regions. Then, a Scale Invariant Feature Tracking (SIFT) descriptor will be used for its’ proven robustness and efficiency [8]. Each detected and matched feature will have a different matching score calculated according to the platform speed, feature correspondence location, and its’ repeatability. This feature matching score will play a significant role on the motion estimation of the platform and rejection of moving objects. In addition, the RANSAC algorithm will be used for outlier removal. Finally, the bundle adjustment will be applied in an adaptive way to reduce the drift of the accumulation of the estimated trajectory. The Fuzzy C-mean Clustering (FCM) algorithm is used for the rejection of moving objects in order to further enhance the VO-estimated poses.
After estimating the pose changes using the proposed VO, it will be fused in a standard integrated Odometer/INS/GPS Extended Kalman Filter (EKF) navigation system. The integration aims at bridging GPS outages in GPS-challenging urban areas.

In order to develop this system the MATLAB image processing, camera calibration, computer vision system, and fuzzy logic tool boxes are used.

1.3 Thesis Contribution

In contrast to most of the existing techniques of monocular VO in unknown environments where data are post-processed or simulated assuming pre-known static features, this work introduces innovative methods to dynamically deal with real-life scenarios. Through an extensive literature review and up to the author’s knowledge, the introduced dynamic processing in unknown environments is novel.

The main contributions of this thesis are:

- **Spatially Diverse Feature Selection:** In this work, the overall efficiency of VO is increased by ensuring selection of non-clustered features. Instead of the commonly used way of partitioning the image into fixed size grid, this work partitions the image in a different novel way. The video frame is divided into four regions as shown in Figure 1-3 and described in the following:
  - **Sky Area:** in this area, lines and corners are detected and matched.
  - **Ground Area:** in this area the ground texture features are detected and matched.
- **Right Horizon:** in this area, the road sign only is detected to take advantage of its well defined standard characteristics. Then it will be matched by its’ SIFT features.

- **The rest of Horizon:** During this work, it was observed that this area contains most of the moving objects. Therefore, it will be rejected.

![Figure 1-3 The partitioning of the video frame.](image)

A) The sky area for salient features of connected regions detection and matching  
B) is the right horizon for road sign detection.  
C) Ground Area for Ground texture Feature Detection and matching
- **Wheel Odometry-Aided scale factor ambiguity resolution:** The estimated relative translation by the odometer will be used in order to overcome the problem of scale factor ambiguity. Then, the relative scale factor of the other objects on the view is estimated.

- **Adaptive feature matching score:** Each matched feature has a different matching score according to the feature location in the image, the vehicle speed (Odometry reading) and the feature repeatability. This scoring technique is significantly important in the steps of the VO process.

- **Statistical Moving Object Rejection using Fuzzy C-Mean (FCM) clustering:** The different detected objects are clustered according to the relative motion taking into consideration their relative scale factors and matching scores. The objects that don’t belong to the cluster of the higher membership and the mean of its points matching score value are tagged as moving objects and, consequently, excluded from motion estimation.

- **Wheel Odometry-Aided Adaptive windowed bundle adjustment:** Instead of using a fixed window windowed bundle adjustment to refine the estimate of the trajectory, the processing window size will be adaptively adjusted depending on the agent speed. This technique assures fewer errors are introduced by dead-reckoning and achieves less processing as well.

- **Background Detection at Zero Velocity Update (ZUPT):** A Gaussian Mixture Model will be used to detect the background at zero velocity when the vehicle is stationary. The objects that belong to the background will have higher matching scores because of the higher possibility of being non-moving objects.
Integration of VO with multi-sensors Odometer/INS/GNSS integrated Navigation System: The VO-estimated pose changes are fused in a vehicular multi-sensors integrated navigation system that fuses measurements from INS, GPS, and odometry.

1.4 Outline

The thesis is organized as follows:

Chapter 2, Literature Review of Navigation in Unknown Environments: This chapter discusses the state of art and existing publications about localization techniques and navigation systems in unknown environments.

Chapter 3, Aided Monocular Visual Odometry (MVO) in Unknown Environments: This chapter will describe how monocular VO is used in real-life dynamic environments and the aiding by the platform odometry.

Chapter 4, Vision-Aided Integrated Navigation: This chapter discusses the integration of the visual odometry with the standard integrated Extended Kalman Filter.

Chapter 5, Experimental Results: In this chapter, the proposed integrated Extended Kalman Filter will be verified on real-road data and the results will be presented.

Chapter 6, Conclusions and future work: This chapter concludes the thesis, with a short introduction to the potential future work of the proposed approach.
Chapter 2 Literature Review of Navigation in Unknown Environments

With the recent development in communication and information systems, there is an emerging need for reliable navigation, positioning, and guidance systems that can work in all environments. The most challenging environment is an uncontrolled dynamic unknown environment. An uncontrolled environment is an environment where factors like line-of-sight visibility, pre-known land-marks, or environment structure/illumination are uncontrolled. A dynamic environment contains a variable number of randomly moving objects such as in dense urban areas. In addition, when there is no prior information provided like maps or geo-referenced signs, the environment is considered as unknown. Different navigation systems can be used in such environments. In the following section, examples of such navigation systems will be elaborated. Then different types of integration between heterogeneous navigation systems and sensors will be discussed in a latter section.

2.1 Standalone Navigation Systems

2.1.1 Global Navigation Systems (GPS)

The Global Navigation Satellite System (GNSS) is a space-borne, radio navigation system that provides world-wide navigation and positioning coverage, availability, and high accuracy which make it one of history's most revolutionary developments. Among many GNSS operational systems, GPS is the most popular. It was developed and operated by the US Department of Defense in the 1970s [10] [11]. GPS has a constellation of 24 satellites and some extra spare satellites. The satellites are arranged in six orbits around the earth with at least four satellites per orbit. This geometry permits four to ten GPS satellites to be visible anywhere in the world, if an elevation
angle of 10° is considered [11]. Each satellite transmits a signal that contains the necessary information to enable a user receiver’s position to be calculated. This satellites constellation is called the space segment. There are also the ground and the user segments. The ground segment consists of a worldwide network of tracking stations that monitor and control satellite motion, system integrity, behavior of satellites, atomic clocks and other considerations. The user segment includes all GPS receivers. With a GPS receiver and antenna, a user can receive GPS signals and use them to calculate his/her position anywhere in the world.

2.1.1.1 The main concept

GPS utilizes the concept of time of arrival ranging [12] to determine the receiver position. The distances of a GPS receiver to the visible satellites are calculated using the time it takes the GPS radio signal to travel from each satellite to the receiver. These distances are called the *pseudo-ranges*. The location of each satellite is known to the receiver through the navigation message sent by the satellite. With these distances and satellite coordinates, the position of the GPS receiver can be computed by *Trilateration* [13]. In theory, a minimum of 3 satellites are needed to be visible to the GPS receiver to get its location. However, a 4th satellite is needed to account for the clock offset of the receiver. In addition to the pseudo-ranges, a GPS receiver may estimate the Doppler frequency of the received GPS signals, which can be used to determine the receiver’s velocity [1]. Typically, a GPS receiver can provide position and velocity information as long as it can receive a signal from at least four satellites.
2.1.1.2 GPS Limitations

Due to many factors, several forms of error in the measurement of the satellite-to-receiver range exist [10]. Some factors that affect GPS positioning accuracy are:

1. Receiver clock bias which is a time varying error in the receiver clock due to imperfection of the low-cost electronic clocks.

2. Satellite clocks drift over time. However, they can be calibrated and corrected by the control segment. The control segment monitors and estimates the satellite clock errors and sends this information with the navigation message to the receiver.

3. Tropospheric delay: The Troposphere is the lower part of the atmosphere which is from 8 to 40 km above the surface of the earth; this layer has changes in weather leading to variations of temperature, pressure, and humidity. These factors affect the speed of light, so their changes cause errors in the range measurements.

4. Ionospheric delay: The Ionosphere is the layer of atmosphere that is above 50 km. It consists of ionized air, and changes in the ionization level influence the travel time of the GPS satellite signals through the Ionosphere.

5. Multipath: Multipath errors are caused by the reflection of signals from surfaces near the receiver and causes errors in range measurements. Multipath errors are one of the most challenging errors in GPS.

6. Receiver noise: It is the error in measuring the transit time by the receiver; it is caused by many factors like electronic component nonlinearity and thermal noise.
2.1.2 Inertial Sensors and Integrated Navigation Systems (INS)

INS is self-contained navigation system that utilizes motion sensor processing to provide navigation information continuously with time [14]. The initial position and orientation is given from another external source such as a GNSS or user-defined position. The common approach to INS is the calculation of platform attitude from angular rate sensors (gyroscopes) and then project the platform accelerations (measured by accelerometers) on a local level frame and integrate once to get velocity and twice to have position. This process is also called “dead-reckoning” [5]. INS provides reliable short-term accuracy. However, the main draw-back of inertial sensors is their sensitivity to the different environmental factors such as temperature and manufacturing imperfections.

2.1.2.1 INS Errors

The INS sensors errors can be divided into biases, drifts, and scale factor error [5].

1. Biases are deterministic error values usually estimated through offline calibration.
2. Drift is an increasing error with respect to time and environmental effects. The INS drift can be stochastically modeled using techniques such as Kalman filter.
3. Scale factor is the ratio between the physical quantity and the output signal of the sensor. This error consists of a deterministic part that be determined by calibration and a stochastic part that can be modeled.
2.1.2.2 MEMS

Micro-electromechanical systems (MEMS) combine computers with tiny mechanical devices such as sensors, valves, gears, mirrors, and actuators embedded in semiconductor chips [5]. MEMS technology has made inexpensive inertial sensors, which can be employed into different applications. It is well-known in the literature that the position estimation drifts over time due to the sensors bias and noise which are amplified in double integration in the strap-down inertial navigation computing [13]. Consequently, additional redundant sensors are important to assist inertial sensors to reduce the position error.

2.1.3 Vision Sensors and Vision-Aided Navigation

As explained in chapter 1, vision sensors can be used for navigation provided that recognizable features exist in the environment with sufficient illumination. As mentioned in section 1.1, the vision-based positioning techniques suffer from the accumulated error resulting from the use of the dead-reckoning principle. In addition to that, when choosing a camera configuration, it should also be taken into account that the use of a single camera induces a scale drift within time. Indeed, it is difficult to propagate the scale factor throughout the process, since with a single camera only, this scale factor is not observable during the localization process. Consequently, the scale factor is directly the subject of the accumulated errors and drifts within time, especially when many features disappear abruptly between two successive images (e.g. in sharp turns). For these reasons, the integration of vision-based systems with other sensors is necessary to improve the navigation performance.
2.2 Multi Sensor Integrated Navigation

The objective of sensor fusion and multi-sensor processing is to improve the positioning and navigation performance obtained by each sensor taken individually by combining their information to maximize the benefits and reduce or cancel complementary errors of each. In general, there are three different ways to fuse sensor measurements according to the level of depth of the information exchanged between sensors and the integration algorithm. They are as follows: 1) Loosely coupled, 2) Tightly coupled, and 3) Deeply coupled [15].

- In the loosely coupled integration, each sensor independently provides a positioning solution. Then, the final positioning solution is estimated by a filtering algorithm such as Kalman Filter.
- On the other hand, the tightly coupled integration technique integrates raw sensors information to provide a unique solution.
- In the deeply coupled integration technique, the fusion is performed inside one of the sensors processes to provide a unique solution. The most widely used filter algorithm for sensor fusion is the Extended Kalman filter (EKF).

2.2.1 INS/GPS Integration

In contrast to the INS's short-term positioning accuracy, satellite-based GPS navigation techniques offer relatively consistent accuracy if sufficient GPS signals can be tracked. However, GPS itself does not provide attitude measurements. Therefore, INS and GNSS systems are often integrated together to take advantage of their complementary strengths. Integration of INS and GNSS is often
performed to improve the performance of navigation systems in denied GNSS environment [16] [17].

There is a current trend to use low cost, or equivalently lower quality, inertial sensors with a high performance GNSS receiver where deeply-coupled integration systems are utilized [18]. This deeply-coupled integration system is an ideal navigation system for platforms that needs low-cost and high performance light-weight navigation systems. Platforms such as unmanned aerial vehicles (UAVs) which have only a limited payload capacity and also requires high manoeuvrability can benefit from this integration scheme to meet specific missions as well as precise navigation for autonomous operation. Figure 2-1 shows an example of a loosely coupled GPS/INS system. PVA stands for Position, Velocity, and Attitude.

![Figure 2-1 loosely-coupled GPS/INS integration scheme](image)
Under high dynamic conditions, an ultra-tightly coupled INS/GNSS navigation system is ideally adopted to improve GNSS signals tracking performance [19] [17]. This architecture integrates the inertial motion unit (IMU) in the receiver for improving tracking loops and range estimation accuracy.

The main drawback in the cost-effective GPS/INS system is that the integrated system becomes more dependent on the availability and quality of GNSS measurements [20]. However GNSS information which relies on external satellite signals can be easily blocked or jammed by intentional/unintentional interference. Even a short duration of satellite signal blockage can cause significant deviation in the navigation solution [21].

2.2.2 Vision/GPS Integration

A complete GPS-based navigation solution is generally not feasible in challenging signal environments such as urban canyons. However, even in these difficult environments, a partial set of GPS signal measurements may still be available. For instance, one or two satellites may generally be still visible even in dense urban canyons [10]. This limited GPS information is insufficient for a complete 3D positioning and timing. However, it can be exploited to improve the efficiency of alternative navigation aids such as vision-based navigation. Such integrated system initializes the vision system in the presence of GPS.

In [22], a method combining limited GPS carrier phase measurements with features that are extracted from images of a monocular video camera is introduced. An integrated GPS/vision solution estimates position and orientation (pose) changes of the camera’s body-frame; and,
initializes ranges to vision-based features. Another method is the usage of the GPS carrier phase measurements to provide relative ranging information that is accurate at a millimeter to sub-centimeter level [23].

2.2.2.1 Loosely-coupled approach

In [24], a loosely-coupled GPS/VO integration scheme assumes that the scale factor can be recovered in case of monocular VO. The KF estimates the errors of the navigation solution provided by the camera, using the difference between GPS and camera solutions as the measurement to the KF. The errors estimated by the KF are used to correct the camera solution. This schema of integration is shown in Figure 2-2.

![loosely-coupled GPS/VO integration scheme]

*Figure 2-2 loosely-coupled GPS/VO integration scheme*
2.2.2.2 Tightly-coupled approach

The tightly-coupled approach is addressed by [25] to estimate the pose in environments where a GNSS-only solution is not feasible with a single camera. The limited GPS carrier phase measurements are used to overcome the scale factor ambiguity problem. The fusion is performed using the Least Mean Square Estimate. The experimental results carried out in [25] show that during limited GPS measurements the positioning accuracy is of a centimeter to sub-decimeter level and provides a heading accuracy in a range of 1 to 3 degrees.

2.2.3 Vision/INS Integration

Vision-aided Inertial Navigation is the fusion of INS and visual information. The importance of this integration arises from the fact that inertial sensors have a large measurement uncertainty when in slow motion and lower relative uncertainty at high velocities, while cameras can track features very accurately at low velocities and less accurately with increasing velocity. Furthermore, in the absence of GNSS information, the INS solution drifts over time that may be reduced by the vision sensor [26].

Since a vision sensor usually has update rates and error properties different from the other sensors, EKF is generally used to optimally integrate the sensor information into a navigational solution [27] [28]. In [29], a comparison between the EKF and the particle filter (PF) shows that EKF provides higher precision.
2.2.3.1 Loosely-coupled approach

In [30] and [31], a monocular simultaneous localization and mapping (SLAM) is used and a loosely-coupled approach that fuses inertial and visual data is implemented. The SLAM algorithm is based on a key-frame approach as described in [32]. It outputs camera poses and 3D position of the observed landmarks up to a scale factor. These approaches estimate the scale factor by fusing the visual and inertial measurements through an EKF by putting the scale as an additional variable in the state vector.

2.2.3.2 Tightly-coupled approach

In [33], a tightly-coupled INS/binocular camera fusion is used for aircraft position, velocity and attitude estimation. This approach is based on the use of measurements obtained by the knowledge of the 3D position of tracked features. In fact, the inertial measurement is used to constrain the search space within the next image during the matching process.

A different tightly-coupled approach for vision-aided inertial navigation is proposed in [34]. To estimate the position, velocity and attitude of a ground vehicle, an error state EKF-based estimation algorithm for real-time navigation is used. The measurement model expresses the geometric constraints that arise when a static feature is observed from multiple camera poses. Experimental results show a good performance of the system since the error is 0.31% of the travelled distance.
2.2.4 Vision/INS/GPS Integration

As GPS, INS and vision sensors have quite different characteristics they can complement each other in different approaches.

2.2.4.1 A Two Separate Filters Integration Systems Approach

In the first approach a system consists of two different Kalman filters [35], the first Kalman filter fuses the GPS and INS data during the good GPS coverage. The second filter is for INS and vision sensor integration during limited GPS Coverage.

2.2.4.2 Two coupled Integration Filters Approach

The second approach is using the GPS measurement for aiding the vision based navigation solution during good GPS coverage to be integrated with INS during GPS outages. In [36], stereo vision/INS fusion is only used during GPS outages to reduce INS drifts. During good GPS coverage the GPS/INS integrated navigation solution is used to determine the 3D location of the key features. This method shows considerable advance on the position accuracy in comparing with first approach.

In [37], the integration of a Strap-down IMU, a GNSS and a single camera is performed in order to estimate the position, velocity and attitude of a ground vehicle. In this system the GNSS data is used to estimate the scale factor of the visual information. The simulations carried out show that the positioning accuracy using this system is significantly improved in GPS outages.
2.2.4.3 Tightly coupled and feed-forward EKF

In [38], a tightly-coupled and feed-forward EKF is proposed to estimate the position, velocity and attitude of an aircraft. The filter combines inertial, visual and GPS measurements if GPS is available and combines inertial and visual measurements during GPS outages. This reported system uses vision as an aiding source providing raw data from the images as measurements to the navigation algorithm. It is shown that using this configuration, the navigation position error is improved by 70% over the free inertial solution for a 400 second GPS outage.

A different system uses the same approach is reported in [39]. This system uses the GNSS data to estimate re-projection error of the observed features on the camera’s image plane. The experimental results show that the localization mean error of the system computed over 10 minutes does not exceed 1meter, outperforming the standalone GPS performance.

2.2.5 Platform odometer/Vision Integration

The odometer is an instrument that indicates distance traveled by a vehicle. This instrument can play a huge role in overcoming the main limitation of monocular visual odometry which is the Scale Factor Ambiguity. The fusion of estimated distance calculated by the platform odometer with the ego-motion estimation can minimize the error introduced by scale factor calculations. There are different ways to perform this fusion. One way is a Particle Filter (PF) used to fuse odometer data with the distance between camera and visual landmark in order to correct the scale propagation and the 3D point positions as described in [40]. Another way of fusion is used in [41]. In this paper, images and odometer measurements are input into a bio-inspired navigation system called Rat-
SLAM system. It showed that odometer data can be used to improve the robustness of Rat-SLAM system significantly.

2.2.6 Odometer/Vision/INS Integration

In the case of ground vehicles, the use of odometer data has been identified as one of the most attractive sensors [6]. In [42], the omnidirectional camera is used as a vision sensor integrated with INS and the platform odometer. In this system, the visual information is used during GPS outages to estimate the vehicle rotation in order to limit the overall drift.

2.2.7 Odometer/INS /GPS Integration

Artificial intelligence methods have been proposed to handle GPS outages in a GPS/INS/Odometer integrated navigation system [18] [43] [44]. However, the integrated systems that use the above methods still result in seriously degraded navigation solutions over long GPS outages.

In [45], Fuzzy adaptive Kalman filter is used to detect changes of the measurement noise statistical characteristics and correct them gradually. In [44], a GPS/INS/odometer integrated system using a fuzzy neural network (FNN) for land vehicle navigation applications is proposed. The information from GPS, odometer and IMU is input into a FNN system for network training during signal availability, while the FNN model receives the observations from IMU and odometer to generate odometer velocity correction to enhance resolution accuracy over long GPS outages. The results indicate that the proposed method can improve the position, velocity and attitude accuracy of the integrated system, especially the position parameters, over long GPS outages.
In [43], a Mixture PF nonlinear fusion technique is used to accommodate for arbitrary inertial sensor characteristics and motion dynamics. The proposed Mixture PF with odometer, pitch, and roll updates outperform existing solutions and exhibited an average improvement of approximately 64% over KF with the same updates, about 85% over KF with velocity updates only, and around 95% over KF without any updates during GPS outages [43].

### 2.2.8 Odometer/Vision/INS/GPS Integration

Although combining GPS/INS/odometer data has been considered one of the most attractive methodologies for ground vehicle navigation, in long GPS signal blockages in complex urban environments, the accuracy of this approach is largely deteriorated. To overcome this limitation, the navigation system has been combined with visual information. In [46], integration of odometer, omnidirectional camera, INS and GPS is studied and showed good improvement in the accuracy. While in [47], over 30% position accuracy enhancement could be achieved after system integration with vision sensors.

### 2.3 Conclusion

Based on the literature review given in this chapter, it was determined that in order to improve the navigation in unknown dynamic environments, this thesis will focus on the following points: First, improving visual odometry as a standalone navigation technique. Second, developing innovative sensor fusion architectures and algorithms to overcome the problem of GPS outages in challenging environments.
Chapter 3 Aided Monocular Visual Odometry for Unknown Environments

3.1 Monocular Visual Odometry Problem Formulation

A moving platform with a rigidly attached camera is taking images at discrete time instants \(k\). The set of images taken at times \(k\) is denoted by \(I_{0:n} = \{I_0, \ldots, I_n\}\). For simplicity, the camera coordinate frame is assumed to be the moving platform’s coordinate frame. Two camera positions at adjacent time instants \(k-1\) and \(k\) are related by the rigid body transformation \(T_{k,k-1} \in \mathbb{R}^{4x4}\) of the following form [7]:

\[
T_k = \begin{bmatrix}
R_{k,k-1} & t_{k,k-1} \\
0 & 1
\end{bmatrix}
\] (3.1)

Where \(R_{k,k-1}\) is the rotation matrix, and \(t_{k,k-1} \in \mathbb{R}^{3x1}\) the translational motion vector.

The set \(T_{1:n} = \{T_{1,0}, \ldots, T_{n,n-1}\}\) contains all subsequent motions. Finally, the set of camera poses \(C_{0:n} = \{C_0, \ldots, C_n\}\) contains the transformations of the camera with respect to the initial coordinate frame at \(k = 0\). The current pose \(C_n\) can be computed by concatenating all the transformations \(T_i\) in a dead-reckoning process.

Therefore,

\[
C_n = C_{n-1}T_n
\] (3.2)

Where \(C_0\) is the camera pose at the instant \(k = 0\).
3.1.1 Mapping Physical 3-D world into 2-D Image Domain: System Model

Let $X = [x, y, z]^T$ be a scene point in the camera reference frame (world frame) and $p = [u, v]^T$ its projection on the image plane measured in the pixels domain. The mapping from the physical 3-D world to the 2-D image domain is given by the perspective projection equation [9]:

$$
\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_u & 0 & u_0 \\ 0 & x_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
$$

(3.3)

Where $\lambda$ is the camera’s depth factor, $x_u$ and $x_v$ are the focal lengths and $u_0, v_0$ the image coordinates of the projection center. These parameters are called intrinsic camera calibration parameters [8] [30] [7]. Figure 3-1 shows the projection of physical world point into camera image plane.

Figure 3-1 The perspective projection of a physical point into the image plan.
3.1.2 Modeling Camera Motion Transformation: The Essential Matrix $E$

The Essential matrix is a fundamental concept in visual odometry. It describes the geometric relations between two images $I_k$ and $I_{k-1}$ of a calibrated camera [7]. It contains the camera motion parameters up to an unknown scale factor for the translation in the following form:

$$E_k \simeq \hat{t}_kR_k$$

(3.4)

Where $t_k = [t_x, t_y, t_z]^T$ and $\hat{t}_k = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$

The symbol $\simeq$ is used to denote that the equivalence is valid up to a multiplicative scalar. The visual odometry problem is then stated as follows: given the sequence of images captured by a calibrated camera, it is required to estimate the relative pose changes of the camera and, consequently, the trajectory of the moving platform on which the camera is mounted.

3.2 The Proposed Aided- Visual Odometry Approach

The main objective of this chapter is to describe the proposed approach to estimate the camera pose transformations using the captured images sequence. The approach utilizes aiding measurements from vehicle’s odometry that measures the travelled distance in the platform body frame. Figure 3-2 shows the main components of the proposed system. Each step in the shown system will be explained in details in subsequent sections.
Figure 3-2. Block Diagram of the Proposed System Approach
3.3 Spatially Diverse Feature Detection

Generally, more features provide more stable motion-estimation results. However, the key-points features should cover the image evenly as much as possible. In this step, it is important to detect sufficient number of features across the entire image frame representing different objects in order to provide robust and accurate navigational information. Therefore, regions of interests that likely contains spatially diverse features are to be considered.

3.3.1 Region Of Interest Selection

The commonly used method is providing a non-clustered features space by segmenting the image frame into a pre-determined grid of segments. In contrast to existing common approach, in this research, the frame is divide into 4 regions to make sure spatially diverse features are detected. One of them is completely neglected because it reflects most of the moving objects in the typical unknown environment under consideration, which is an urban area. The other three regions usually contain different types of features to be detected. After that, for all of the detected features, the Scale Invariant Feature Transform (SIFT) [48] algorithm is used to match the features between the processed frames. Figure 3-3 shows the different regions of interest applied in this research. This image is from a trajectory collected in Toronto downtown area and the surroundings.
3.3.2 Hough Transform (HT) for line detection

In urban areas and downtown urban canyons, the first region of interest is representing the high buildings and the sky. In the buildings, lines features are likely dominant and detected using the Hough transformation (HT) [49] [50]. HT is a common reliable statistical algorithm that extracts global image features such as straight lines, circles and ellipses. It is commonly used in computer vision and pattern recognition. In this research, it is used as a method to detect line features from
the buildings/sky regions of the image frame. The HT algorithm uses a voting technique at which each point belonging to a specific pattern votes for all the possible patterns passing through that point. This is implemented by searching the Hough parameter space for maximum peaks. The Hough parameters space is divided into pins and each pin receives votes.

3.3.2.1 Hough Space and Lines Representation

Lines can be represented uniquely by two parameters. Often the form in Equation 3.5 is used with parameters $a$ and $b$.

$$y = ax + b$$  \hspace{1cm} (3.5)

This form is, however, not able to represent vertical lines. Therefore, the Hough transform uses the form in Equation 3.6, which can be rewritten to Equation 3.7 to be similar to Equation 3.5. The parameters $\theta$ and $r$ is the angle of the line and the distance from the origin respectively.

$$r = x \cos \theta + y \sin \theta$$  \hspace{1cm} (3.6)

It implies

$$y = -\frac{\cos \theta}{\sin \theta} x + \frac{r}{\sin \theta}$$  \hspace{1cm} (3.7)

All lines can be represented in this form when $\theta \epsilon [0, 180]$ and $r \epsilon R$ (or $\theta \epsilon [0, 360]$ and $r \geq 0$).
The Hough space for lines has therefore these two dimensions; $\theta$ and $r$, and a line is represented by a single point, corresponding to a unique set of parameters $(\theta_0, r_0)$. The line-to-point mapping is illustrated in Figure 3-4.

![Mapping from image space to Hough Space](image)

**Figure 3-4 Example of line mapping from image space into Hough Space**

3.3.2.2 Lines Matching Challenge and the Proposed Solution

Indeed, lines are more difficult to match because lines are more likely to be occluded than points. Furthermore, the origin and end of a line segment may not exist. Therefore, instead of matching the lines itself, the region around each line is converted into set of SIFT features that will be used in matching.
3.3.3 Quantization Based Road Signs Detection (RSD)

In the region of interest B, road signs are the main features that can likely be detected and tracked. Road signs are designed using specific colors and shapes. Thus, color is a significant feature for RSD. Input color sub-frame are first quantized in HSV color model [51]. The quantization reduces the amount of visual data. Border tracing is then used to extract closed regions having border color the same as road signs.

3.3.3.1 RSD Detection Mechanism

The road sign detection is performed by two steps as follows:

- Color quantization: The HSV color model was used in this research to quantize the input image. The Red, Green, and Blue color values of each pixel in the input image are transformed into HSV color values.

- Then Border tracing is used to find regions that have border color the same as road sign. Border tracing is performed by scanning the quantized image from left to right and from top to bottom. Points with red and yellow color are supposed to be the starting points of border tracing.

Regions inside these borders are considered as Road Signs. Figure 3-5 shows different road signs that can be detected by this approach.
3.3.3.2 RSD Processing Challenge and the Proposed Solution

Instead of performing RSD in each frame which is computationally expensive and time consuming, the SIFT features of the detected road sign are detected and matched among the different frames.

![Examples of Road Signs](image)

**Figure 3-5 Examples of Road Signs that can be detected using Quantization based RSD.**

3.3.4 Ground-Texture Detection using SIFT

In the third region of interest, the dominant image feature is the ground texture. The ground texture is highly detectable in low speeds. The SIFT features of this region will be matched. More details about this step are given in the subsequent sections.
3.4 Feature Matching

A major step in visual odometry is establishing matches between two frames. This can be done using two ways [7]. The first one is to find features in one image and track them in the following images using local search techniques, such as correlation. The second one is to independently detect features in all the images and match them based on some similarity metric between their descriptors. The latter is more suitable for our research case when a large motion or viewpoint change is expected.

The goal of this step is to match the features that correspond to the same physical feature existing in the camera field of view in two frames. This phase consists of computing a matching score that indicates the likelihood that two features correspond to the same physical feature. The features that have the highest scores are matched. In order to perform this matching step, the region around each detected feature is converted into a compact descriptor that can be matched against other descriptors. One of the most popular descriptors for point features is the SIFT.

3.4.1 SIFT Descriptors

The SIFT descriptor is mainly a histogram of the local gradient orientations [48]. This local gradient patch around the feature is partitioned into a 4x4 grid. For each partition, a histogram of an 8-gradient orientation is computed. The computed histograms are then concatenated together to construct a 128 element descriptor vector. To minimize the effect of severe illumination changes, the descriptor is normalized to unity. The SIFT descriptor proved to be stable against changes in
illumination, rotation, and scale, and even up to 60° changes in viewpoint. Figure 3-6 shows the matched features between two sub-frames in the ground texture region.

Figure 3-6 matched features between two sub-frames of the third region of interest (Ground texture).

3.5 Background Detection for Zero Speed and Moving Objects Rejection using Gaussian Mixture Models

One of the main problems of the uncontrolled dynamic environment is the difficulty of distinguish between moving and static objects. One way to overcome this problem in this research is detecting the background of the camera scene using adaptive Gaussian Mixture Model [52] during the zero speed stationary periods then matching the features of the background with higher matching score.
### 3.5.1 Adaptive Gaussian Mixture Model

In this approach, a mixture of $K$ Gaussian distributions is used to represent each pixel in the scene to be modelled. The probability that a certain pixel has a value of $x_N$ at time $N$ can be written as

$$p(x_N) = \sum_{j=1}^{k} w_j \eta(x_N; \theta_j)$$  \hspace{1cm} (3.8)

where $w_k$ is a weight parameter of the $k^{th}$ Gaussian component $\eta(x; \theta_k)$ is the Normal distribution of $k^{th}$ component. The $K$ distributions are ordered based on the fitness value $w_k/\sigma_k$. The first $B$ distributions are used as a model of the background of the scene where $B$ is estimated as

$$B = \arg \min_b \left( \sum_{j=1}^{b} w_j > T \right)$$  \hspace{1cm} (3.9)

The threshold $T$ is the minimum fraction of the background model. In other words, it is the minimum prior probability that the background is in the scene. Background subtraction is performed by marking as a foreground pixel any pixel that is more than 2.5 standard deviations away from any of the $B$ distributions.

### 3.5.2 Zero Velocity Update and Moving Objects Rejection

After estimating the background during the zero speed, the features set will be updated by the set of the estimated background features. Using this way, the feature set that belongs to the background is likely the features that correspond to static objects. This step contributes to the minimization of the moving dynamic objects in the scene.
3.6 Outlier Removal using RANSAC

Correspondences is the features association through consecutive image frames. The search for correspondences (also called data association), is usually based on the first stage on comparing local descriptors of salient features in the captured sequence of images. The ambiguity of such local description leads to possible incorrect correspondences at this stage. Among robust estimation methods to minimize the effect of this errors, random sample consensus (RANSAC) stands out as one of the most successful and widely used in the computer vision community [7].

The idea behind RANSAC is to compute model hypotheses from randomly sampled sets of data points and then verify these hypotheses on the other data points. The hypothesis that shows the highest consensus with the other data is selected as a solution. For two-view motion estimation as used in VO, the estimated model is the relative motion ($R, t$) between two camera positions, and the data points are the candidate feature correspondences. Inlier points to a hypothesis are found by computing the point-to-Epipolar line distance [53]. Figure 3-7 shows the flow chart of RANSAC algorithm for VO.
Figure 3-7. The flowchart of RANSAC Algorithm for Visual Odometry
The number of subsets $N$ that is necessary to guarantee that a correct solution is found can be computed by \([53]\)

$$N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^s)}$$

(3.10)

where $s$ is the number of data points from which the model can be instantiated, $\varepsilon$ is the percentage of outliers in the data points, and $P$ is the requested probability of success \([7]\). Figure 3-8 shows an example of standard RANSAC estimation steps.

![Figure 3-8 RANSAC steps for the simple 2D line estimation example \([53]\)](image-url)
3.7 Relative Pose Estimation

Motion estimation is the core computation step performed for every image in a VO system [8]. The aim of this step is to compute the camera motion between the current frame and the previous frame. This frame-to-frame pose estimation is performed up to an unknown scale factor. If the scale factor is estimated, this pose estimation can be mapped into physical world pose change. The scale factor is commonly estimated by triangulation as explained in chapter 1. However, this estimation is not always accurate. In order to enhance the scale factor estimation, an odometer-aided local bundle adjustment method is used and will be explained in the next section. Once the scale factor is accurately estimated, by concatenation of all these single movements, a full trajectory of the camera and the platform can be constructed. Figure 3-9 shows the flowchart of the pose estimation steps.
3.8 Odometer-Aided Local Bundle Adjustment for Automatic Scale Factor Estimation

One of the major factors that affects VO accuracy in single monocular camera systems is the scale factor ambiguity. The scale factor is important because it maps the relative pose changes from image domain to metric physical world domain. In this section, a proposed method to overcome the scale factor ambiguity problem is explained.
3.8.1 Scale Factor Ambiguity

When viewing an object through a projection camera, there is an inherent scale ambiguity. To illustrate this ambiguity imagine a square object that appears on the image plane as a $n$ by $n$ pixel region. This projection could be created by an object of size $m$ by $m$ at a distance $d$. Or, this same image could be created by an object of size $2m$ at a distance $2d$. Thus, without an estimate of the depth to an object, it is impossible to reconstruct its size. Figure 3-10 describes the Scale factor ambiguity.

Figure 3-10. Scale Factor Ambiguity.

Lines $L_1$, $L_2$ in 3-space is imaged as line $l_1$, $l_2$ by a perspective camera which its’ center is $C$. 
A similar scale factor ambiguity issue arises with translational motion in the field of view. An object translating parallel to the image plane at a distance of \( n \) meters and a velocity of \( k \) meter per second would have the same apparent velocity on the image plane as an object at a distance of \( 2n \) meters traveling at \( 2k \) meters per second. That is, objects further away appear to travel slower. This phenomenon proves to be a major difficulty for VO estimation.

### 3.8.2 Odometer based Scale Factor Estimation

The solution that used in this research to overcome scale factor ambiguity is to rely on another sensor to estimate the scale. It involves the fusion of monocular visual information with odometer. The odometer is used to correct the scale propagation [54]. The idea is to correct the estimated 3D camera position by the odometer estimated travelled distance according to the following equation:

\[
\hat{c}^t = c^{t-1} + d^{t-1} \frac{c^t - c^{t-1}}{\|c^t - c^{t-1}\|_2}
\] (3.11)

Where \( \|.\|_2 \) is the Euclidean norm., and \( d^{t-1} \) is the distance travelled between time \((t-1)\) and time \((t)\), estimated by the odometer.

### 3.8.3 Local Bundle Adjustment (LBA)

Due to the inaccurate estimation of the scale factor, errors propagate and accumulate over time. To minimize the effect of the errors, the Local Bundle Adjustment (LBA) technique is applied. LBA is a well-known iterative method designed to solve non-linear least square problems in VO [54]. In this problem, the estimated parameters \( x \) are the estimated poses of the \( n \) most recent key image
frames. The key-frames are the frames that contain a high number of features. Furthermore, at low dynamics where features are not likely changing, only a few frames will be processed to reduce computational load.

3.9 Matching Score Calculation

As mentioned before, the aim of this research is to try to overcome the problem of dynamic unknown environment navigation. In such an environment, it is important to reject the moving objects motion estimation from the calculated results. In order to achieve this goal each detected feature will have its matching score depending on the following factors:

1. The likelihood that two features correspond to the same physical feature, which calculated during feature matching step.
2. The region where the feature is detected. For example, if the feature is detected as features of Road Sign, it will have higher matching score because of its higher probability of being a feature of a non-moving object.
3. The speed will be a factor too. In case of high dynamics, the Sky region detected features will have higher score. While, during low dynamics (low speed) the ground detected features will have higher score.
4. The features of the detected background during ZUPT will have higher accuracy. This is because of its higher probability of being a feature of a non-moving object.

The added score which will be described by the fuzzy term “higher” in this research is calculated by try and error.
3.10 Fuzzy C-mean Clustering for statistical moving objects rejection

Further to the aforementioned methods to reject and minimize the effect of moving dynamic objects, a fuzzy C-mean clustering-based rejection method is proposed. Moving objects is an inherent difficulty that arises when attempting to calculate ego-motion on uncontrolled unknown environments. To help overcome this problem, clustering of the detected and tracked objects on the different regions of interest of the video frame is proposed according to their calculated relative orientation and translation.

The objective of clustering is the classification of objects according to similarities among them. In real-life unknown environments, some moving objects can be detected. To filter out these objects’ estimated relative orientation and translational changes and increase vision based navigation accuracy, this work applied Fuzzy C-mean clustering algorithm [55] [56]. The choice of fuzzy clustering was to deal with uncertainty in the estimated relative orientation and translation changes of the wrongly tracked moving objects.

The target of fuzzy clustering is to determine, for each object, the membership degree of belonging to a certain cluster. Let $\mu_{i,k}$ represent fuzzy membership degree of belonging of data point $i (x_i)$ to cluster $k$ (cluster centre $v_k$). Given a data set of N points, the output of the fuzzy clustering is the following equation:
\[
U = \begin{bmatrix}
\mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,c} \\
\mu_{2,1} & \mu_{2,2} & \cdots & \\
\vdots & \vdots & \ddots & \\
\mu_{N,1} & \cdots & \cdots & \mu_{N,c}
\end{bmatrix}
\]

(3.12)

Where \(c\) is the number of clusters and \(\mu_{i,k} \in [0,1]; 1 < i < N; 1 < k < c\).

Fuzzy-C-Means clustering algorithm is based on the minimization of the objective function called C-means functional. It is defined by Dunn as:

\[
J(X; U, V) = \sum_{i=1}^{N} \sum_{k=1}^{m} \mu_{i,k}^m \|x_k - v_i\|_A^2
\]

(3.13)

\[
\|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i)
\]

(3.14)

In fuzzy C-means clustering, the objective function is minimized by Picard iteration through first order conditions for stationary points. To determine the tracked non-moving objects, FCM algorithm is applied on the resulting data (relative pose change). The clusters with strong membership degree is chosen as a non-moving object to be tracked and its parameters (pose change) are used in the next phase (see chapter 4).

### 3.11 Sample Results based on The Proposed Odometer-Aided Visual Odometer

An example of the camera relative motion estimation between two image frames is shown in Figure 3-11 using the proposed methodology is shown in Figure 3-12. This figure shows that the system can detect the rotation and transformation between two frames while neglecting the in-
between frames where the vehicle' speed was zero. Accumulating the estimated camera motion is used to perform visual odometry and construct a complete trajectory.

**Figure 3-11** Two Different frames to estimate camera relative pose change.

**Figure 3-12** Estimated Camera Motion between the two frames shown in Figure 3-11
The odometer-aided VO algorithm was applied on a testing trajectory shown in Figure 3-13 of length of approximately 2.43km. The reference solution is collected by a high-end integrated navigation system to be described in chapter 5 in the experimental work. A GPS-enabled single camera was used and a total of approximately 10 minutes of data have been recorded. Only VO with odometry was applied in this test. The root mean square error (RMS) during this test was 34.69m and maximum position error was 64.04m.

Figure 3-13. Testing Trajectory in downtown Toronto
Figure 3-14. Odometer-aided VO test results. 2D trajectory is displayed.
Chapter 4 Vision-Aided Integrated Navigation

4.1 Abstract

In this chapter, the application of the visual odometry estimation in multi-sensors integrated navigation systems is introduced. This chapter proposes a vision-aided multi-sensor integrated navigation system that can be used in challenging unknown environments such as urban areas. In normal scenario, where GPS signals are reliable, GPS position/velocity are integrated in EKF with INS. In GPS-challenging environments, such as urban downtown areas, GPS is outage due to high buildings. Under this situation, odometer-aided visual odometry estimation is applied and sent to the EKF with INS to enhance the accuracy during these GPS outages. In this chapter, the equations of INS mechanization and EKF filter design will be explained including the measurement and system models derivation. In the next chapter, real experiments will be performed in GPS-challenging environment in downtowns. The experimental results will show how the proposed multi-sensor integrated navigation system can effectively bridge GPS outages.

4.2 Introduction

The general approach is to design an extended Kalman filter that takes updates from visual odometry estimation to limit the unbounded increase of INS errors. The common methodology is, in open sky where GPS satellites are visible, integration with INS is performing well. In GPS-denied environments like urban downtown areas, where GPS satellites signals are blocked, another aiding navigation source is necessary for sustainable performance. A common aiding sensors to INS during GPS outages is the odometer. However, in addition to the odometer, other sensors and
information such as VO is of great benefit to reduce INS drifts. In this chapter, VO is used to provide accurate estimation relative changes of the vehicle’s pose. With proper efficient processing, these VO-estimated pose changes can be estimated and can be sent to EKF to be fused with GPS, inertial measurement unit (IMU), and odometry measurements.

This chapter will apply the estimated VO measurements in an EKF that fuses GPS, inertial sensors, odometer and vision sensor from a monocular camera system. In the next chapter, the system will be tested on platform of a ground vehicle for enhanced and sustainable navigation performance in urban areas.

4.3 INS Mechanization

INS mechanization is the processing of sensors measurements to calculate 3D position, velocity, and attitude. More details about this equations can be found in [57]. Table 2 Shows the different frames used in the navigation processing. The favorite navigation frame in INS is the local level frame due to the coupling between roll/pitch errors and horizontal velocity errors. The sensor measurements measures the vehicle’s dynamics in the body frame. To process the navigation in local frame, a transformation between body frame and navigation frame is needed. Therefore, the orientation representation between different frames will be defined.
Table 2. Frames of the Navigation Processing

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body Frame</strong></td>
<td>Origin: the vehicle.</td>
</tr>
<tr>
<td></td>
<td>Y: forward direction.</td>
</tr>
<tr>
<td></td>
<td>X: transversal direction.</td>
</tr>
<tr>
<td></td>
<td>Z: up direction.</td>
</tr>
<tr>
<td><strong>Local Level</strong></td>
<td>Origin: the vehicle.</td>
</tr>
<tr>
<td><strong>Navigation Frame</strong></td>
<td>Y: North direction.</td>
</tr>
<tr>
<td></td>
<td>X: East direction.</td>
</tr>
<tr>
<td></td>
<td>Z: Up direction.</td>
</tr>
<tr>
<td><strong>Earth Centered</strong></td>
<td>Origin: Center of Earth.</td>
</tr>
<tr>
<td><strong>Earth Fixed</strong></td>
<td>Z: extends through the North Pole.</td>
</tr>
<tr>
<td><strong>(ECEF) frame</strong></td>
<td>X: passes through the intersection of the equatorial plane and the prime</td>
</tr>
<tr>
<td></td>
<td>meridian.</td>
</tr>
<tr>
<td></td>
<td>Y: complete the right-hand coordinate system in the equatorial plane.</td>
</tr>
</tbody>
</table>

Different forms can be used to define vehicle’s attitude such as Direct Cosine Matrix (DCM), Euler angles, and quaternion. However, it is preferred to use the quaternion form because its resistance to singularities. The singularity can be defined as the discontinuity of the mathematical model of
the system. The quaternion direction vector is defined as \( \mathbf{q} = [q_0, q_1, q_2, q_3]^T \). The rotation matrix from body frame to local level navigation frame \( R^l_b \) is given by:

\[
R^l_b = \begin{bmatrix}
q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\
2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\
2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2
\end{bmatrix}
\]  

(4.1)

Where the superscript \( l \) means the local level frame and the subscript \( b \) denotes the body frame.

### 4.3.1 Mechanization Equations: Position Vector

The position vector \( p \) is expressed in geodetic coordinate in the ECEF frame where \( \varphi \) is the latitude, \( \lambda \) is the longitude, and \( h \) is the altitude. The position vector is given by:

\[
p = [\varphi \quad \lambda \quad h]
\]  

(4.2)

The rate of change in position is calculated from the velocity as given below:

\[
\dot{\varphi} = \frac{v_n}{R_M + h}
\]  

(4.3)

\[
\dot{\lambda} = \frac{v_e}{(R_N + h) \cos \varphi}
\]  

(4.4)

\[
\dot{h} = v_u
\]  

(4.5)
Here, \( V_n \), \( V_e \), and \( V_u \) are the velocity components in the north, east and up directions respectively.

\( R_M \) and \( R_N \) are the meridian and normal radius of the earth ellipsoid model.

4.3.2 Velocity Vector Equations

The accelerometer measures the specific force in the vehicle’s body frame. To get vehicle acceleration in local level navigation frame and integrate to get velocity components and position coordinates, a transformation between body frame and local level frame needs to be performed. This can be implemented by using the rotation matrix as follows:

\[
\mathbf{f}^l = R_b^l \mathbf{f}^b
\]

\( \mathbf{f}^b = [f_x, f_y, f_z]^T \) is the acceleration measurements in the body frame and \( \mathbf{f}^l = [f_e, f_n, f_u]^T \) is the acceleration components in the local level frame.

Gravity, Transport Rate, and Earth Rotation Effects

To calculate the velocity components of the vehicle, the local level frame acceleration derived cannot be directly used due to the earth rotation rate \( \omega_e^l \) and angular velocity caused by the change of orientation of the local level frame with respect to the earth \( \omega_u^l \), which is also called transportation rate, and finally the earth’s gravity field \( g \). Taking these factors into consideration, the rate of velocity change can be expressed as:
\[ \dot{v} = f^l - \left(2\Omega^l_{ie} + \Omega^l_{el}\right)v + g \quad (4.7) \]

where \( \Omega^l_{ie} \) and \( \Omega^l_{el} \) are the skew-symmetric matrices corresponding to \( \omega^l_{ie} \) and \( \omega^l_{el} \).

### 4.3.3 Attitude Vector Equations

The differential equation of the quaternion is given by:

\[
\dot{q} = \frac{1}{2} \begin{bmatrix} 0 & -\omega^b_{lb} \\ \omega^b_{lb} & -\Omega(\omega^b_{lb}) \end{bmatrix} q
\quad (4.8)
\]

Where \( \Omega(\omega^b_{lb}) \) is the skew-symmetric matrix of \( \omega^b_{lb} \), and \( \omega^b_{lb} \) is the angular velocity of the body frame with respect to the local level frame represented in the body frame. However, \( \omega^b_{lb} \) must be compensated for transport rate and earth rotation rate effect:

\[
\omega^b_{lb} = \omega^b - (R^l_b)^T (\omega^l_{ie} + \omega^l_{el}) \quad (4.9)
\]

Once the quaternion parameters are known, the rotation matrix can be updated. Based on the relationship between attitude angles and the quaternion-based rotation matrix, the formulation of azimuth, pitch and roll can be written as:

\[
A = \tan^{-1}\left(\frac{-R^l_b(1, 2)}{R^l_b(2, 2)}\right) \quad (4.10)
\]
\[ p = \sin^{-1}(R_{b}^{t}(3, 2)) \]  
\[ r = \tan^{-1}\left(\frac{-R_{b}^{t}(3, 1)}{R_{b}^{t}(3, 3)}\right) \]  

4.4 Filter Design

The block diagram of the multi-sensor integrated navigation system is shown in Figure 4-1. The system model and measurement model are described in the following subsections.

Figure 4-1 The block diagram of the multi-sensor integrated navigation system
4.4.1 INS-predicted Pose Change vs. VO-estimated Pose Change Updates

As explained in chapter 3, the distance estimated by the odometer aids the VO to estimate the relative pose changes from images frames with minimum scale factor error. This odometer-aided VO-estimated relative pose changes will be used as an update to the EKF. The pose change estimated from INS is used as a predicted pose change to be fused with VO-estimated pose change in the EKF. The vehicle azimuth changes can be approximated during the sampling interval $T$ by:

$$\Delta A = \omega \cdot T$$  \hspace{1cm} (4.12)

And the position changes along x-axis and y-axis are:

$$\Delta x = v \cos p \sin \Delta A \cdot T$$  \hspace{1cm} (4.13)

$$\Delta y = v \cos p \cos \Delta A \cdot T$$  \hspace{1cm} (4.14)

where $v$ is the forward vehicle speed.

4.4.2 Linearized system error model

In the proposed system, the error state vector for the EKF is defined as:

$$\delta x = [\delta P \quad \delta v \quad \delta q \quad \delta b_\omega \quad \delta b_f \quad \delta \Delta x \quad \delta \Delta y \quad \delta \Delta A]^T$$

Where $b_\omega$ and $b_f$ are gyroscope and accelerometer measurements biases respectively, and they are modelled as first-order Gauss-Markov process, which is given in the following general form:

$$\dot{x}(t) = -\beta x(t) + \sqrt{2\sigma^2} \beta w(t)$$  \hspace{1cm} (4.15)
Where $\sigma$ is the standard deviation, $\beta$ is the reciprocal of the time constant of the autocorrelation function of $x(t)$, and $w(t)$ is zero-mean Gaussian noise.

By applying Taylor expansion to INS position and velocity mechanization equations, INS pose change estimation equations, the linearized system error model is given as:

$$\delta \dot{x} = F \delta x + Gw$$  \hspace{1cm} (4.16)\\

where the transition matrix $F$ can be written as [2]:

$$F = \begin{bmatrix}
F_{pp} & F_{pv} & 0_{3 \times 4} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\
0_{3 \times 3} & 0_{3 \times 3} & F_{vq} & 0_{3 \times 3} & -R_i^t & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\
0_{4 \times 3} & 0_{4 \times 3} & \frac{1}{2} F_{qq} & F_{qbo} & 0_{4 \times 3} & 0_{4 \times 1} & 0_{4 \times 1} & 0_{4 \times 1} \\
0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 4} & F_{bo} & 0_{3 \times 3} & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\
0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 4} & 0_{3 \times 3} & F_{bf} & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\
0_{1 \times 3} & F_{\Delta \nu} & 0_{1 \times 4} & 0_{1 \times 3} & 0_{1 \times 3} & 0 & 0 & F_{\Delta \nu} \\
0_{1 \times 3} & F_{\Delta \nu} & 0_{1 \times 4} & 0_{1 \times 3} & 0_{1 \times 3} & 0 & 0 & F_{\Delta \nu} \\
0_{1 \times 3} & 0_{1 \times 3} & 0_{1 \times 4} & F_{\Delta \nu} & 0_{1 \times 3} & 0 & 0 & 0
\end{bmatrix}  \hspace{1cm} (4.17)$$

In the derivation of the $F$ matrix, the terms that are divided by the square of the earth radius have been ignored. The elements in the $F$ matrix are given as below:

$$F_{pp}(2,1) = \frac{v_e \tan \varphi}{(R_N + h) \cos \varphi}, \text{ while other elements in } F_{pp} \text{ are zeros;}$$
\[
F_{pv} = \begin{bmatrix}
0 & \frac{1}{R_d + h} & 0 \\
\frac{1}{(R_N + h) \cos \phi} & 0 & 0 \\
0 & 0 & 1
\end{bmatrix};
\]

\[
F_{vq} = \begin{bmatrix}
F_{vq1} & F_{vq2} & F_{vq3} & F_{vq4} \\
-F_{vq4} & -F_{vq3} & F_{vq2} & F_{vq1} \\
F_{vq3} & -F_{vq4} & -F_{vq1} & F_{vq2}
\end{bmatrix};
\]

where

\[
F_{vq1} = 2\left(q_0 (f_x - b_{fx}) - q_3 (f_y - b_{fy}) + q_2 (f_z - b_{fz})\right),
\]

\[
F_{vq2} = 2\left(q_1 (f_x - b_{fx}) + q_2 (f_y - b_{fy}) + q_3 (f_z - b_{fz})\right),
\]

\[
F_{vq3} = 2\left(-q_2 (f_x - b_{fx}) + q_1 (f_y - b_{fy}) + q_0 (f_z - b_{fz})\right),
\]

\[
F_{vq4} = 2\left(-q_3 (f_x - b_{fx}) - q_0 (f_y - b_{fy}) + q_1 (f_z - b_{fz})\right);
\]

\[
F_{vq} = \begin{bmatrix}
0 & -(\omega_x - b_{ax}) & -(\omega_y - b_{ay}) & -(\omega_z - b_{az}) \\
\omega_x - b_{ax} & 0 & \omega_z - b_{az} & -(\omega_y - b_{ay}) \\
\omega_y - b_{ay} & -(\omega_z - b_{az}) & 0 & \omega_x - b_{ax} \\
\omega_z - b_{az} & \omega_y - b_{ay} & -(\omega_x - b_{ax}) & 0
\end{bmatrix};
\]
\[
F_{qba} = -\frac{1}{2} \begin{bmatrix}
-q_1 & -q_2 & -q_3 \\
q_0 & -q_3 & q_2 \\
q_3 & q_0 & -q_1 \\
-q_2 & q_1 & q_0
\end{bmatrix},
\]

(4.25)

\[
F_{bao} = \text{diag} \left( -\beta_{ox}, -\beta_{oy}, -\beta_{oz} \right),
\]

(4.26)

\[
F_{gy} = \text{diag} \left( -\beta_{fx}, -\beta_{fy}, -\beta_{fz} \right),
\]

(4.27)

where \( \beta_{ox}, \beta_{oy}, \beta_{oz} \) and \( \beta_{fx}, \beta_{fy}, \beta_{fz} \) are the reciprocal of the time constant of the random process associated with the gyroscope bias and accelerometer bias respectively.

\[
F_{\Delta xv} = \frac{\cos p \sin \Delta A}{v} \begin{bmatrix}
v_e & v_y & v_u
\end{bmatrix},
\]

(4.28)

\[
F_{\Delta xA} = v \cos p \cos \Delta A,
\]

(4.29)

\[
F_{\Delta yv} = \frac{\cos p \cos \Delta A}{v} \begin{bmatrix}
v_e & v_y & v_u
\end{bmatrix},
\]

(4.30)

\[
F_{\Delta yA} = -v \cos p \sin \Delta A,
\]

(4.31)

\[
F_{\Delta xabo} = \begin{bmatrix}
\cos p \sin r & -\sin p & -\cos p \cos r
\end{bmatrix}.
\]

(4.32)

The process noise vector contains the noises in gyroscope measurements, accelerometer measurements, and the Gaussian noises associated in the first order Gauss-Markov process of
gyroscope and accelerometer biases. $G$ is the noise coupling matrix and the non-zeros elements are written as follow:

$$G(4:6, 4:6) = R^t_q,$$  \hspace{1cm} (4.33)

$$G(7:10, 1:3) = -F_{qho},$$  \hspace{1cm} (4.34)

$$G(11:13, 7:9) = \text{diag}(\sqrt{2\sigma_{\omega_x}^2, \beta_{\omega_x}}, \sqrt{2\sigma_{\omega_y}^2, \beta_{\omega_y}}, \sqrt{2\sigma_{\omega_z}^2, \beta_{\omega_z}}),$$  \hspace{1cm} (4.35)

$$G(14:16, 10:12) = \text{diag}(\sqrt{2\sigma_{f_x}^2, \beta_{f_x}}, \sqrt{2\sigma_{f_y}^2, \beta_{f_y}}, \sqrt{2\sigma_{f_z}^2, \beta_{f_z}}),$$  \hspace{1cm} (4.36)

$$G(19, 1:3) = -F_{A_{hho}},$$  \hspace{1cm} (4.37)

where $\sigma_{\omega_x}, \sigma_{\omega_y}, \sigma_{\omega_z}$ and $\sigma_{f_x}, \sigma_{f_y}, \sigma_{f_z}$ are the standard deviation of the gyroscope bias and accelerometer bias respectively. The noise variance calculation was significantly depending on the calculated matching score during VO process.

**4.4.3 Measurement Model**

4.4.3.1 GPS measurements

In outdoor open-sky areas where GPS signals are available, position and velocity solutions from GPS are integrated with INS at the rate of 1 Hz. The measurement model for INS/GPS loosely coupled scheme is:
\[
\begin{bmatrix}
P_{gps} - P_{ins} \\
v_{gps} - v_{ins}
\end{bmatrix} = I_{6 \times 6}
\begin{bmatrix}
\delta \Delta p \\
\delta \Delta v
\end{bmatrix}
\]

(4.38)

4.4.3.2 Visual Odometry measurements

In GPS-denied environments, the pose change from VO and INS respectively are integrated at the rate of 10 Hz. The measurement model can be given as:

\[
\begin{bmatrix}
\Delta x_{vo} - \Delta x_{ins} \\
\Delta y_{vo} - \Delta y_{ins} \\
\Delta A_{vo} - \Delta A_{ins}
\end{bmatrix} = I_{3 \times 3}
\begin{bmatrix}
\delta \Delta x \\
\delta \Delta y \\
\delta \Delta A
\end{bmatrix}
\]

(4.39)

More details about the derivations of the system models and measurements model can be found in [57].

4.5 Conclusion

In this chapter, the vision-aided multi-sensor land vehicle-based integrated navigation system is explained and all equations are derived. The way the odometer-aided VO-estimated relative pose changes are used as an update to an EKF was explained. The prediction is based on INS-estimated relative pose changes. In the explained system, GPS and VO are used as aiding systems to provide periodic corrections to INS in different environments. The EKF filter design was explained and the system error model was derived.
Chapter 5 Experimental Work

5.1 Abstract

This chapter explains the experimental work that has been performed to implement and verify the algorithms that have been developed in this thesis. The experimental setup will be elaborated and the testing trajectory and environment conditions will be given. The performance of the proposed vision-aided integrated navigation system is evaluated in terms of position errors. The errors are calculated based on a ground-truth navigation system output. This ground-truth navigation system consists of high-end tactical grade IMU, GNSS receiver and an EKF-based integration system. Analysis of the results will be given. The chapter concludes with a conclusion and future work.

5.2 Experimental Setup

A van was used to collect real-road data for the experimental validation of the proposed work. The experimental van is shown in Figure 5-1. Inside the van, two systems are mounted. The first system is the ground-truth system from Novatel Inc. The OEM4-G2 ProPak-G2plus SPAN Unit consists of Novatel OEM4 GPS receiver, high-end tactical-grade CPT-IMU. The second system is a data-logger data logger that includes a GNSS receiver (u-blox LEA-6T), an IMU from VTI Technologies Inc, and a barometer from Measurements Specialties. Both the ground-truth system and the data logger unit are shown in Figure 5-5-2. The visual data has been recorded by a MiVue 388 GPS-enabled digital cameras shown in Figure 5-3. The speed data was collected by CarChip OBII module shown in Figure 5-4.
Figure 5-1 Experimental Van

Figure 5-5-2 Novatel SPAN Unit + VTI IMU + Ublox GPS
Figure 5-3. GPS-enabled MiVue 388 digital cameras

Figure 5-4. CarChip to record vehicle speed
5.3 Data Synchronization

The Novatel SPAN unit generates time-tagged synchronized GPS and IMU data. In addition, the data logger that includes the VTI-IMU and the Ublox GPS receiver generates GPS-time tagged data. To synchronize between visual information (video images frames) and IMU/GPS information, GPS-enabled camera is used. This GPS-enabled camera generates a GPS NMEA file that contains the time-tag information and GPS-time tagged video frames. Regarding the odometer, the speed information was logged by the CarChip dongle shown in Figure 5-4. In order to synchronize speed data with the images frames, GPS, and IMU data, a correlation between the ground-truth speed data and the CarChip speed data was performed. Therefore, the CarChip data is tagged by the proper GPS-time tag.

5.4 Testing Trajectories

The testing trajectories were performed in down-town area and the surrounding areas in the city of Toronto where a combination of Open-Sky and Down-town urban environments were obtained. The testing has been performed on three trajectories shown in Figure 5-5, Figure 5-6 and Figure 5-7. The three testing trajectories lengths are shown in Table 3.
Figure 5-5. Testing Trajectory#1

Figure 5-6. Testing Trajectory#2
Figure 5-7. Testing Trajectory#3

Table 3. Testing Trajectories Distances Travelled

<table>
<thead>
<tr>
<th>Trajectory</th>
<th>Distance Travelled</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>1649.78m</td>
</tr>
<tr>
<td>#2</td>
<td>2432.34m</td>
</tr>
<tr>
<td>#3</td>
<td>1722.19m</td>
</tr>
</tbody>
</table>
5.5 Sensors Specifications

The specification of the sensors used in the experimental work is shown in Table 4.

Table 4. VTI Data Logger Sensors Specifications

<table>
<thead>
<tr>
<th>Gyroscopes</th>
<th>Accelerometer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating range</td>
<td>Operating range</td>
</tr>
<tr>
<td>(-300) – (300) °/s</td>
<td>(-6)-(6) g</td>
</tr>
<tr>
<td>Angular random walk (ARW)</td>
<td>Offset error</td>
</tr>
<tr>
<td>1.12 °/sqrt(hour)</td>
<td>70 mg</td>
</tr>
<tr>
<td>Noise Density</td>
<td>Noise Density</td>
</tr>
<tr>
<td>0.02 °/s/sqrt(Hz)</td>
<td>7 mg</td>
</tr>
<tr>
<td>Quantization Error</td>
<td></td>
</tr>
<tr>
<td>0.05 °/s</td>
<td></td>
</tr>
</tbody>
</table>
5.6 Data Processing and Filter Tuning

The collected data was processed offline using Matlab on a SAMSUNG laptop having a Windows 10, 8 GRAM and 2.5 GHz processor. The EKF filter parameters were manually tuned to achieve the best performance. The measurement covariance of the VO updates used to feed the EKF was obtained from the matching score of feature correspondences in the sequences of images after applying a multiplication scaling factor. After tuning the filter on one trajectory, the filter was applied on the other two trajectories and results were reported.

5.7 RESULTS

5.7.1 Trajectory#1

The results of trajectory#1 are shown in Figure 5-8 and Figure 5-9. Figure 5-8 shows the 2D position trajectory. Under the INS-Odo configuration, the EKF was working under odometer speed updates only. The trajectory lasts for approximately 8 minutes and the travelled distance was 1649.78m in downtown areas. Due to the drift of the INS, the solution drifts significantly with time as shown in the figures. The drift root mean square (RMS) error is 162.84m. With careful tuning of the filter, and under the weighted updates from visual odometry, the solution sustains a reliable accuracy of 2.95m RMS.
Figure 5-8. 2D Trajectory#1.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
Figure 5-9. Errors in Trajectory#1.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
5.7.2 Trajectory#2

The results of trajectory#2 is shown in Figure 5-10 and Figure 5-11. The trajectory lasts for approximately 10 minutes and the travelled distance was 2432.34m in downtown areas. Due to the drift of the INS, the solution drifts significantly with time as shown in the figures. The drifts are 176.22m (RMS) in 10 minutes. With the same tuning of the filter, and under the weighted updates from visual odometer, the solution sustains a reliable accuracy of 3.61m RMS. This error behavior is consistent with the error characteristics obtained in trajectory #1.
Figure 5-10. 2D Trajectory#2.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
Figure 5-11. Errors in Trajectory#2.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
5.7.3 Trajectory#3

The results of trajectory#3 is shown in Figure 5-10 and Figure 5-11. The trajectory lasts for approximately 10 minutes and the travelled distance was 1722.19m in dense downtown areas. It is noticed in this trajectory that the error is much less than trajectory 1 and 2. The reason is that the speed in this trajectory was relatively slower with frequent stops. The low speed reduced the scale factor error effects and the frequent stops contributed to limit the INS drift. The INS-Odo drifts were of 19.06m (RMS) in the total of the 10 minutes. With the same tuning of the filter, and under updates from visual odometer, the solution sustains a reliable accuracy of 4.97m RMS. Table 5 summarize the overall error values for the three trajectories including RMS errors and maximum position error.

**Table 5. Error Values for the testing trajectories**

<table>
<thead>
<tr>
<th>TRAJECTORY</th>
<th>DISTANCE(m)</th>
<th>RMS 2D POS ERROR(m)</th>
<th>MAX 2D POS ERROR(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>INS-Odo</td>
<td>INS-VO-Odo</td>
</tr>
<tr>
<td>Trajectory#1</td>
<td>1649.78</td>
<td>162.84</td>
<td>2.95</td>
</tr>
<tr>
<td>Trajectory#2</td>
<td>2432.34</td>
<td>176.22</td>
<td>3.61</td>
</tr>
<tr>
<td>Trajectory#3</td>
<td>1722.19</td>
<td>19.06</td>
<td>4.97</td>
</tr>
</tbody>
</table>
Figure 5-12. 2D Trajectory#3.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
Figure 5-13. Errors in Trajectory#3.

Showing reference ground-truth vs. INS-Odo vs. INS-VO-Odo
Chapter 6 Conclusion and Future Work

6.1 Conclusion

This thesis addressed the problem of visual odometry (VO) estimation using monocular camera systems and utilized VO in the area of multi-sensors integrated navigation for odometer-enabled vehicles. The thesis proposed efficient VO methods and techniques for challenging uncontrolled unknown environments that contains moving objects and have no access to reliable GPS. The work also integrated the estimated VO with a 3D multi-sensor integrated navigation system that combines inertial sensors (INS), GPS, and speed measurements.

A broad literature review about navigation systems in unknown environments and vision-aided navigation methods was given. In addition, detailed description of the problem of VO was given with the proper mathematical notation. Furthermore, a dedicated chapter that introduces the proposed methodology toward solving the given problem and achieving the research goals is given.

The main contributions of this thesis work can be listed as follows: 1) An efficient odometer-aided VO estimation method was proposed. 2) An automatic Scale Factor ambiguity resolution for single camera systems using odometer measurements was introduced. 3) An adaptive Local Bundle Adjustment (LBA) method was proposed to enhance VO-estimated vehicle pose changes. 4) A Gaussian Mixture method that detects zero-speed conditions, identifies stationary regions of a sequence of images, and reject moving objects effects was developed. 5) A novel spatially diverse region of interest mechanism for image features detection and tracking. 6) A novel method for lines matching based on SIFT features. 7) The integration of the VO-estimated pose changes with a
multi-sensor integrated navigation for vehicular platforms. Experimental verification and results analysis were performed.

Experimental work was done with a physical vehicular platform equipped by MEMS inertial sensors, GPS, speed measurements and GPS-enabled camera. The experimental work included three testing trajectories in downtown Toronto and the surrounding areas. The trajectories are 1649.78m, 2432.34m, and 1722.19m length. The experimental work showed significant improvements under the absence of GPS where only VO is fused with INS and speed measurements. Under the VO integration for approximately 10 minutes without GPS, the root mean square error (RMS) of the trajectories were 2.95m, 3.61m, and 4.97m respectively. The maximum 2D position error were 8.58m, 14.93m and 29.85m respectively. These errors constitute 0.52%, 0.61%, and 1.73% of the travelled distances for the testing trajectories. The INS-odometer alone performed poorly with 18.75%, 12.16%, and 2.38% of the travelled distances of the three testing trajectories respectively.

6.2 Future Work

There are some enhancements that can be implemented to further demonstrate the capabilities of VO-estimation in vision-aided multi-sensors integrated navigation. A significant room for improvement is the matching score calculation. The matching score calculation can be enhanced using artificial intelligence techniques such as neural networks and genetic algorithms. This enhancement in the matching score will lead to better VO-estimated pose changes and hence improve the overall performance. Another room for improvements is the enhancement of road-
signs detection by incorporating new signs and features that are commonly identified in urban areas. By enriching road-signs feature list, the VO-estimation is expected to enhance significantly.

The problem of moving objects detection and rejection needs further research and improvements. It is currently one of the major source of outliers and errors. Non-linear systems identification techniques, fuzzy-logic and artificial intelligence are potential tools that can be used to improve dynamic objects detection and rejection. Finally, the integration of VO with other sensors such as INS and GPS can be further expanded to include ultra-tightly coupled scheme which can lead to more robust navigation in urban areas and GPS-challenging areas. Other sensors fusion and integration methodology can be tried such as particle filtering. In addition, the application of the proposed method can be tested and expanded to include indoor areas and inside-buildings where the environment is more controlled and several road-signs exist.
Bibliography


