UNIVERSITI MALAYSIA SARAWAK
UNIMAS

Faculty of Engineering

OPTIMIZATION IN MONOCULAR VISUAL ODOMETRY

Gildron David

Bachelor of Engineering

Electrical and Electronics Engineering with Honours

2022

UNIVERSITI MALAYSIA SARAWAK

Grade: ____________

Please tick (√)

Final Year Project Report	√
Masters	
PhD	

DECLARATION OF ORIGINAL WORK

This declaration is made on the 4th day of August 2022.

Student's Declaration:

I **GILDRON DAVID, 66041, FACULTY OF ENGINEERING** hereby declare that the work entitled **OPTIMIZATION IN MONOCULAR VISUAL ODOMETRY** is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

05/08/2022
Date submitted

Gildron David (66041)

Supervisor's Declaration:

I **DAYANG NUR SALMI DHARMIZA BINTI AWANG SALLEH** hereby certifies that the work entitled **OPTIMIZATION IN MONOCULAR VISUAL ODOMETRY** was prepared by the above named student, and was submitted to the FACULTY OF ENGINEERING as a * partial/full fulfillment for the conferment of BACHELOR OF ENGINEERING IN ELECTRICAL AND ELECTRONICS ENGINEERING WITH HONOURS, and the aforementioned work, to the best of my knowledge, is the said student's work.

Received for examination by: ____________________ Date: 05/08/2022

DAYANG NUR SALMI DHARMIZA BINTI AWANG SALLEH

I declare that Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirm with free consent and willingly declare that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abiding interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitalise the content for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student signature ______________________ Supervisor signature: ______________________

11/08/2022 11/08/2022

Current Address: **LRG 5A, SETINGGAN, KAMPUNG SIOL KANDIS, 93050 PETRA JAYA, KUCHING, SARAWAK.**

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument is duly prepared by The Centre for Academic Information Services]

OPTIMIZATION IN MONOCULAR VISUAL ODOMETRY

GILDRON DAVID

A dissertation submitted in partial fulfilment
of the requirement for the degree of
Bachelor of Engineering
Electrical and Electronics Engineering with Honours

Faculty of Engineering
University Malaysia Sarawak

2022

ACKNOWLEDGEMENT

First and foremost, I would like to express my heartfelt gratitude to God, the Almighty, for providing me with the courage, energy, and confidence to carry out this research. Additionally, I would like to take this opportunity to express my gratitude to my family and friends, as well as to all of the individuals whose patience, guidance, inspiration, and encouragement have enabled me to complete this study.

Dr Dayang Salmi Dharmiza binti Awang Salleh, my supervisor and programme coordinator for the Final Year Project session 2021/2022, has been a tremendous source of support and encouragement throughout this process. Her expertise and advice, as well as her years of experience and passion, have been invaluable. Thank you for your encouragement, your support, and your professional guidance over the years.

In addition, I am grateful to the lecturers who engaged in the Final Year Project seminars and webinar. Undoubtably, I am always grateful and indebted to them for their wisdom, genuine, meaningful advice, and motivation, which they have extended to all final-year students, including myself, over the years.

Lastly, I would like to thank all the parties that had helped me, directly and indirectly, from the start till completion of this Final Year Project 1. I have none but only words to express my gratitude to all, I wish them all the best in all that they do from time to time in whatever they do.

ABSTRACT

Presently, the world of autonomous vehicle is rapidly advancing as day by day, the features in many aspects of the car are made smarter and improved for safety, convenience and comfort. Apart of it, localization is one particular aspect of an autonomous vehicle which is especially important as it works to determine the precise location of the vehicle in the map, as maps are vitals nowadays to navigate. In details, the work of localization in the autonomous vehicle are done with visual odometry. As for this project, the focus is on monocular visual odometry (VO) as mono cameras are more preferred as for their low cost and easy to handle. The framework of the monocular visual odometry applied incorporated Oriented FAST and Rotated BRIEF (ORB) for the feature detection and the Flann-based matcher for feature matching. Optimizations done in this project was to tweak and adjust the system parameters which includes the value of features detected and distance between the matches, that is known as good matches. Overall, 3 different sequences from KITTI dataset are utilized in this simulation, the results were compared with the original system configuration and the ground truth. The optimized results show overall improvement of each individually sequence from their original ground truth. Entirely, the project aims to optimize the mono VO system which could help in the development of the mono VO in the future.

ABSTRAK

Pada masa ini, dunia kenderaan autonomi berkembang pesat kerana hari demi hari, ciri-ciri dalam banyak aspek kereta telah dibangunkan dan diperbaik untuk keselamatan, kemudahan dan keselesaan pengguna. Selain itu, penyetempatan adalah salah satu aspek penting untuk kenderaan autonomi kerana ia berfungsi untuk menentukan lokasi kenderaan yang tepat di peta, kerana menavigasi menggunakan peta pada masa kini amat luas. Secara terperinci, kerja penyetempatan dalam kenderaan autonomi dilakukan dengan odometri visual. Fokus projek ini adalah pada odometri visual monokular kerana kamera monokular lebih menjimatkan kos dan mudah dikendalikan. Rangka kerja odometri visual monokular yang digunakan dalam projek ini adalah menggabungkan ORB untuk pengesanan ciri dan pemadanan berasaskan Flann untuk pemadanan ciri. Pengoptimasian yang dilakukan dalam projek ini adalah untuk mengubah suai dan menyesuaikan parameter sistem yang merangkumi nilai ciri yang dikesan dan jarak antara ciri-ciri, untuk mengenalpasti padanan yang baik. Secara keseluruhan, 3 set data yang berbeza dan KITTI telah digunakan dalam simulasi ini Hasilnya dibandingkan dengan konfigurasi sistem asal dan trajektori asal yang disediakan oleh KITTI sendiri, hasil yang dioptimasikan menunjukkan peningkatan keseluruhan trajektori. Secara keseluruhan, projek ini bertujuan untuk mengoptimumkan sistem VO mono yang dapat membantu dalam pembangunan Mono VO pada masa akan datang.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	:	2 Dimensional
3D	:	3 Dimensional
BA	:	Bundle Adjustment
BRIEF	:	Binary Robust Independent Elementary Features
BRISK	:	Binary Robust Invariant Scalable Keypoints
DoD	:	Department of Defense
DSO	:	Direct Sparse Odometry
EKF	:	Extended Kalman Filter
FAST	:	Features from Accelerated Segment Test
FREAK	:	Fast Retina Keypoint
FYP	:	Final Year Project
GBA	:	Global Bundle Adjustment
GPS	:	Global Positioning System
INS	:	Inertial Navigation Systems
KLT	:	Kanade-Lucas-Tomasi tracker
LIDAR	:	Light Detection and Ranging
LBA	:	Local Bundle Adjustment
Mono VO	:	Monocular Visual Odometry
ORB	:	Oriented FAST and Rotated BRIEF
PTAM	:	Parallel Tracking and Mapping
PnP	:	Perspective-n-Point
RANSAC	:	RANdom SAmple Consensus
rBRIEF	:	Rotated Binary Robust Independent Elementary Features
RGB-D	:	Red, Green, Blue Depth sensor
SfM	:	Structure from motion
SIFT	:	Scale-Invariant Feature Transform
SLAM	:	Simultaneous Localization and Mapping
S-PTAM	:	Stereo Parallel Tracking and Mapping
SURF	:	Speeded-Up Robust Features
SVO	:	Semi direct monocular Visual Odometry
UAV	:	Unmanned Aerial Vehicle
US	:	United States
VO	:	Visual Odometry
vSLAM	:	Visual Simultaneous Localization and Mapping

CHAPTER 1

INTRODUCTION

1.1 Project Background

Vehicles nowadays are getting smarter with the ever-advancing automotive industry. Throughout the years, researchers come up with various techniques or system to be use for on one particular aspect of the vehicles, namely the localization. Localization is a task for autonomous vehicles that enables them to monitor their courses and detect as well as avoid obstacles with convenience. Therefore, it is important for the technology to always be in constant development and to improve accuracy, as it is vital to maintain a good performance on the task.

When the driver deviates from the intended path, the vehicle localization system should be able to update the location in real time based on perception and build a global route for path planning in addition to accuracy [1]. Localization must not only be precise, but also be capable of updating the vehicle location instantly even when travelling at high speeds in numerous lanes, intersections or diverging highways. Another problem with vehicle location in metropolitan areas is the difficulty of navigating a complicated network of roads.

About 15 years ago, the US Department of Defense (DoD) began using GPS to track military personnel and vehicles around the world [2]. Autonomous navigation, vehicle localization, and robot localization have all benefited greatly from the widespread adoption of GPS technology since its inception. GPS data and georeferenced maps have recently been used in commercial applications to restore a city address's map or provide directions between city locations.

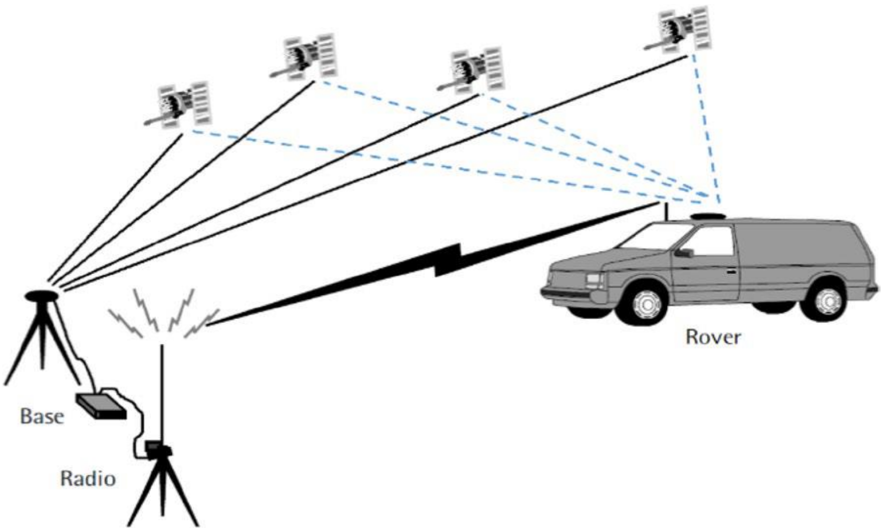

Figure 1.1: Real-time GPS application system for vehicle [3].

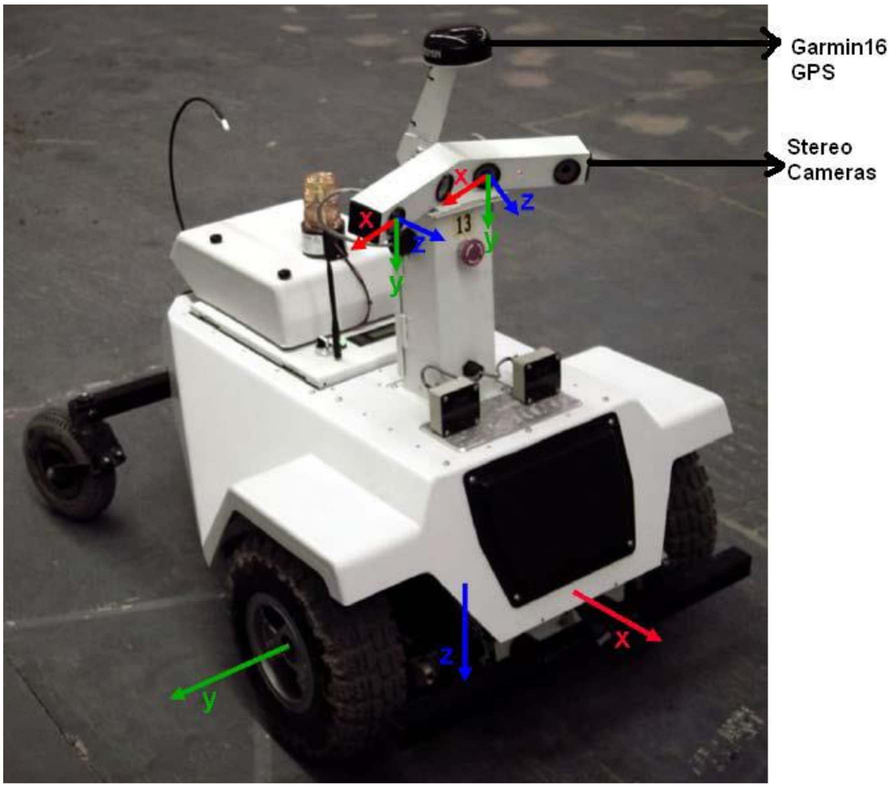

Figure 1.2: A robot equipped with combination of GPS and Stereo sensors [4].

Visual odometry for vehicle localization is one of the many methods available. Odometry, the measurement of a vehicle's position over time using motion sensors, and visual odometry (VO), the process of analysing an agent's ego motion using the input from one or more cameras attached to it, are two very different things [5]. SLAM, or visual Simultaneous Localization and Mapping, is another method. The SLAM algorithm, which also makes use of a mounted sensor, forecasts the motion of a vehicle [6]. It is important to note that the primary difference between VO and SLAM lies in the emphasis placed on local consistency, as VO attempts to incrementally estimate the camera/route robot's pose by pose, with the possibility of performing local optimization. A global estimate of camera/robot trajectory and map are sought after by SLAM, in contrast.

Visual odometry has been widely used nowadays as a required system in autonomous vehicles. The visual odometry system are identified with their respective camera types, in this case, mostly used are stereo and monocular cameras. Some other advance cameras include Lidar, RGB-D and hybrid uses multiple types of cameras.

1.2 Problem Statement

Visual odometry in vehicle applications is crucial in these modern times, with it being an effective technique for the obstacle detection and navigation. As a response, regardless of the methods utilised, location and velocity inaccuracies increase with regard to time and length of the journey. With this being the goal in many years to keep improving the technique, many tweaks over the year, some are on monocular visual odometry. The use of a single camera visual odometry is very preferred by the industry, but this comes at some highly consistent problems associated with it.

Most notable problem are the scale drift, which is always the associated with the real time vehicle application, in other terms some describes it as error propagation [5]. As there are no depth information yet to be available in monocular visual odometry system, the pose of the current frame or camera is obtained based solely on the previously estimated camera poses [5]. Monocular visual odometry, on the other hand, has faced some limitations and priorities over the past decade, including minimising drift in the trajectory and reducing map distortion when traversing long distances [7]. Therefore, in this study, the optimizations on mono VO are researched and the best suitable optimization are combined with the mono VO system to improve its performance.

1.3 Objectives

The aim of this research is to investigate the optimizations done in the monocular visual odometry to improve its performance and accuracy. In order to achieve the aim of the research, the objectives set are as follow:

- To determine the limitations in monocular VO compared with stereo VO.
- To improve the accuracy of the monocular visual odometry on car applications.
- To minimize the problems associated with the applications of monocular visual odometry.

1.4 Project Scope

The scope of this project entails into the researching and formulating an optimization for the monocular visual odometry to improve it accuracy namely the scale drift problem. As many methods was introduced and reworked for the effort to significantly reduce these inaccuracies, such methods are to be used in this project.

In conjunction with the project, the algorithms for many parts in monocular visual odometry such as the feature detection, feature matching, pose estimation and more, these parts each has different algorithms that are to be used in the simulation of the project. Such examples of the algorithms are SIFT, SURF, ORB and more, which in literature review will be explained in much detail on these algorithms. The extent of this research study is on monocular visual odometry and the improvement that are made over the years.

1.5 Thesis Outline

Introduction, Literature Review, Methodology, Results and Discussion and Conclusion are the five chapters of the report. Chapter 1 covered the introduction to the project, the problem statement and project objectives deduced, and the scope of the project.

Chapter 2 is the literature review, which is the foundations on monocular visual odometry and details on the topic. This chapter focuses in detail on visual odometry, every section on monocular visual odometry and its previous related works.

Chapter 3 covers the methodology which explain the processes throughout the research and simulations. This chapter provides the details on the monocular visual odometry algorithms, simulation coding, other relevant parts of the design and simulation process of the project. The flowchart, Gantt chart, and figures are used to illustrates the current section of the chapter.

Chapter 4 presents the results on the simulations and reports the analysis on the results obtained with comparisons to the previous related works on the monocular visual odometry specifically on car applications.

Chapter 5 summarises the research and highlights its findings. This chapter also suggests future work to be done to enhance the capabilities. Mostly as matter of fact, it is to show the entire research study has run its course.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The current chapter mainly overviews the current topic and review the related research studies. This chapter encompasses the visual odometry system designs with its major components, the monocular visual odometry algorithms and the techniques used in the simulation design. The drawbacks of monocular visual odometry are discussed with their proposed improvement over the years. Additionally, previous works and research papers are also discussed in the chapter.

2.2 Visual Odometry

Navigation nowadays in the automotive industry are highly crucial as almost all modern vehicles have this feature. This feature includes the capability of an autonomous vehicle to recognize its whereabout and the current surrounding in real time. This task is done by a visual localisation technique which is visual odometry [6], one out of two popular technologies, which the other is visual Simultaneous Localization and Mapping, or vSLAM as it is commonly known as [7].

Visual odometry is a topic that has gotten a lot of attention in recent years. Autonomous outdoor driving and robot navigation depend on it. Rather than relying on a laser scanner to create a 3D map of its surroundings and then use that map in conjunction with high-resolution maps of the globe to drive itself, automakers prefer vision-based systems.

The overall notion of visual odometry is the relationship of the camera's movement between consecutive image frames, or, to put it more simply, the motion between two consecutive images.

(a)

(b)

Figure 2.1: Images from KITTI dataset. a) Image 000091, b) Image 000092.

Above shows two consecutive image frames from the KITTI dataset. The difference between both frames might not be clear as both shows the straight view of the car. Looking closely, the distance of the left car in the frame to the camera is closer in the consecutive second frame. Using this information and the features from the first frame, the features are matched to the second frame and then through to the next steps until the current step are done repeatedly until the last frame.

(a)

(b)

Figure 2.2: Images from KITTI dataset. a) Image 000104, b) Image 000120.

Then, in the above frames, the motion of the car can be clearly seen as the frames shows the car is turning right. Visual odometry allows the system to estimate the motion of the car using inputs such as consecutive images from a camera.

Monocular visual odometry are knowingly categorized into two categories, either direct methods or the feature-based methods [7]. Another category that are seldomly addressed are semi-direct methods, which is a combination of the two popular categories [8]. For the two original methods;

a) Direct methods: This method uses the pixels in a frame that is the image instead of the extracted features, which is exploiting the pixels with clear gradients to avoid the high computation [7]. The core of direct methods is it reduces the photometric error on the input frames directly to track the camera or motion poses [9]. Notably, DSO or Direct Sparse Odometry [10], combines the optimization of camera positions, sparse scene structure, and camera model parameters.
b) Feature based methods: In other terms, this method is based on point features, with it, mainly used to minimize the reprojection error [7]. This approach extracts

points of interest and associated descriptors from the input image into a sparse representation. Multiview correspondence association, which minimises reprojection error, is used to recover camera motion and sparse structure [7], [9]. Over the years, these feature-based methods are improved and some notable works includes the Parallel Tracking and Mapping (PTAM) [11], Stereo Parallel Tracking and Mapping (S-PTAM) [12], and Oriented FAST Rotated BRIEF SLAM (ORB-SLAM2) [13].

The basic algorithm or flow of visual odometry [14]: for a given pair of frames is as follows;

1) Identify features in each frame (feature detection),
2) The comparison of characteristics between each frame (the total of absolute disparities across local windows),
3) Find the greatest number of self-consistent matches (inliers), and
4) Realize motion from frame to frame to reduce re-projection error for clean features.

History of VO systems are either divided to learning-based or geometric-based. Learning-based visual odometry is a new method that has already shown encouraging results on a few benchmarks. End-to-End deep neural networks must be trained offline using a large number of image sequences. A number of advantages have emerged for learning-based monocular VO over geometric methods, including the fact that parameters need not be adjusted, and the system is impervious to tracking failure and scale drift. In the training phase, stereo image pairs can be used to retrieve metric scale from monocular images. Learning-based VO is still significantly less accurate than geometric VO.

While geometric-based is the traditional method which is as mentioned in the above, where the feature-based and direct methods fall directly under the geometric based.

As for the study, the system to be used in the project are geometric-based. Geometric based method consists of image sequence, feature detection, feature matching, motion estimation, optimization and lastly, the camera poses [15]. Each part mentioned are phase by phase, meaning that the feature detection phase will only execute after the inputs image sequence.