FORMULATION OF A VITAL MONOCULAR VISION ALGORITHM FROM VIDEO FRAMES USING A HYBRID CASCADED APPROACH FOR DISPARITY MAP ACCURACY

KU SITI NURUL AIN BINTI KU-FAIROLNIZAM

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FORMULATION OF A VITAL MONOCULAR VISION ALGORITHM FROM VIDEO FRAMES USING A HYBRID CASCADED APPROACH FOR DISPARITY MAP ACCURACY

KU SITI NURUL AIN BINTI KU-FAIROLNIZAM

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronics Engineering Technology (Industrial Electronics) with Honours

Faculty of Electronics and Computer Technology and Engineering Universiti Teknikal Malaysia Melaka

2025

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI TEKNOLOGI DAN KEJURUTERAAN ELEKTRONIK DAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : Formulation of a vital monocular vision algorithm from video frames using a hybrid cascaded approach for disparity map accuracy

Sesi Pengajian : 2024/2025

Saya KU SITI NURUL AIN BINTI KU-FAIROLNIZAM mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

□ **SULIT*** (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

□ **TERHAD*** (Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan.

☑ **TIDAK TERHAD**

Disahkan oleh:

__________________ __________________

Alamat Tetap:

Tarikh : 24/01/2025 Tarikh : 24/01/2025

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this project report entitled “Formulation of a vital monocular vision algorithm from video frames using a hybrid cascaded approach for disparity map accuracy” is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Student Name : KU SITI NURUL AIN BINTI KU-FAIROLNIZAM

Date : 24/01/2025

APPROVAL

I hereby declare that I have checked this project report and in my opinion, this project report is adequate in terms of scope and quality for the award of the degree of Bachelor of Electronics Engineering Technology.

Signature :

Supervisor Name : IR.TS.DR. AHMAD FAUZAN BIN KADMIN

Date : 24/1/2025

Signature :

DR. ROSTAM AFFENDI BIN HAMZAH
Pensyarah Kanan
Jabatan Teknologi Kejuruteraan Elektronik & Komputer
Fakulti Teknologi Kejuruteraan Elektrik & Elektronik
Universiti Teknikal Malaysia Melaka

Co Supervisor Name : TS.DR ROSTAM BINHAMZAH

Date : 24/1/2025

DEDICATION

To my beloved mother, Rosisdah Binti Ghozally,
and father, Ku-Fairolnizam Bin Ku Halim,
And To dearest husband,
Muhammad Aniq Danial Bin Zulkafli

ABSTRACT

This paper proposes a critical monocular vision algorithm that uses a hybrid cascaded technique to improve the accuracy of disparity maps. The algorithm's goal is to use a single camera to mimic human vision in terms of depth perception, object recognition, and scene comprehension. Utilising methods like robust lane boundary detection, bird's-eye view transformations, and vehicle localisation, it makes use of developments in computer vision and artificial intelligence to maximise performance in real-time applications such as robotics, augmented reality, and autonomous driving. The study investigates cutting-edge techniques to address issues including low-texture areas, occlusions, radiometric aberrations, and intricate scene configurations, such as feature extraction, semantic segmentation, and deep learning models like convolutional neural networks (CNNs). The system can produce precise depth maps, identify lane boundaries, and identify cars with high accuracy, according to experimental validations. The project's use of a strong methodology allows for improved visual data interpretation in dynamic and difficult situations, such as changing road geometries, motion, and illumination. The algorithm's dependability and effectiveness are confirmed by testing on datasets and real-time situations. The study admits its limitations in dealing with extreme situations including lengthy distances, sharp pitch angles, and complex lane arrangements, notwithstanding these achievements. In order to increase resilience and flexibility, recommendations for future research place a strong emphasis on integrating multisensor data fusion, adaptive algorithms, and sophisticated AI techniques. This study makes a substantial contribution to the field of monocular vision systems, opening the door for their incorporation into intelligent systems for automated settings that are safer, more effective, and more sustainable.

ABSTRAK

Kajian ini mencadangkan algoritma penglihatan monokular kritikal yang menggunakan teknik hibrid bertingkat untuk meningkatkan ketepatan peta perbezaan. Matlamat algoritma ini adalah untuk menggunakan kamera tunggal bagi meniru penglihatan manusia dalam aspek persepsi kedalaman, pengecaman objek, dan pemahaman adegan. Dengan menggunakan kaedah seperti pengesanan sempadan lorong yang kukuh, transformasi pandangan mata burung, dan penempatan kenderaan, ia memanfaatkan kemajuan dalam penglihatan komputer dan kecerdasan buatan untuk memaksimumkan prestasi dalam aplikasi masa nyata seperti robotik, realiti tambahan, dan pemanduan autonomi. Kajian ini menyelidik teknik-teknik canggih untuk menangani cabaran seperti kawasan bertekstur rendah, halangan, penyimpangan radiometrik, dan konfigurasi adegan yang kompleks, termasuk pengekstrakan ciri, segmentasi semantik, dan model pembelajaran mendalam seperti rangkaian neural konvolusi (CNN). Berdasarkan pengesahan eksperimen, sistem ini mampu menghasilkan peta kedalaman yang tepat, mengenal pasti sempadan lorong, dan mengenal pasti kenderaan dengan ketepatan yang tinggi. Penggunaan metodologi yang kukuh dalam projek ini membolehkan interpretasi data visual yang lebih baik dalam situasi dinamik dan mencabar, seperti geometri jalan yang berubah-ubah, pergerakan, dan pencahayaan. Kebergantungan dan keberkesanan algoritma ini disahkan melalui ujian ke atas set data dan situasi masa nyata. Walaupun pencapaian ini, kajian ini mengakui batasannya dalam menangani situasi ekstrem termasuk jarak yang jauh, sudut kecondongan yang tajam, dan susunan lorong yang kompleks. Bagi meningkatkan daya tahan dan fleksibiliti, cadangan untuk penyelidikan masa depan memberi penekanan kuat terhadap integrasi gabungan data berbilang sensor, algoritma adaptif, dan teknik AI yang lebih canggih. Kajian ini memberikan sumbangan yang besar dalam bidang sistem penglihatan monokular, membuka jalan bagi integrasi mereka ke dalam sistem pintar untuk persekitaran automasi yang lebih selamat, lebih cekap, dan lebih lestari.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Ir.Ts.Dr. Ahmad Fauzan Bin Kadmin and co-supervisor, Rostam Affendi Hamzah for their precious guidance, words of wisdom and patient throughout this project.

I am also indebted to Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education Malaysia for the financial support through the Fundamental Research Grant Scheme (FRGS), which enabled me to accomplish the project. Not forgetting my fellow classmates, 4 BERZ, for their willingness to share their thoughts and ideas regarding the project.

My highest appreciation goes to my parents, parents in-law, and family members for their love and prayer during the period of my study. An honourable mention also goes to my husband, Muhammad Aniq Danial for all the motivation and understanding.

Finally, I would like to thank all my classmates, the Faculty members, as well as other individuals who are not listed here for being co-operative and helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

• **VO**	-	Visual Odometry
• **MSE**	-	Mean Squared Error
• **PSNR**	-	Peak Signal-to-Noise Ratio
• **CLAHE**	-	Contrast Limited Adaptive Histogram Equalization
• **AGCWD**	-	Adaptive Gamma Correction with Weighting Distribution
• **GF**	-	Guided Filter
• **CT**	-	Color Transfer
• **SfM**	-	Structure from Motion
• **SIFT**	-	Scale-Invariant Feature Transform
• **SURF**	-	Speeded-Up Robust Features
• **CNNs**	-	Convolutional Neural Networks
• **3D**	-	Three-dimensional
• **ORB**	-	Oriented FAST and Rotated BRIEF
• **BRIEF**	-	Binary Robust Independent Elementary Features
• **YOLO**	-	You Only Look Once
• **SLAM**	-	Simultaneous Localization and Mapping
• **GPS**	-	Global Positioning System
• **DAS**	-	Driving Assistance System
• **P3P-RPE**	-	Perspective 3-Point Rammer Pose Estimation
• **UAV**	-	Unmanned Aerial Vehicle
• **OI**	-	Orthogonal Iteration
• **CDW**	-	Construction and Demolition Waste
• **JHA**	-	Joint Histogram Aggregation

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter gives a quick rundown of the monocular matching method and highlights its desirable features, which can be used to create fresh or enhanced depth map assessment algorithms for computer vision applications. It also explains the objectives and justifications for the study that is being offered. The structure and organization of the thesis are also highlighted, along with its innovative contributions.

1.1 Background

The vital monocular vision algorithm extracts and interprets visual data from a single camera in a manner similar to how a single eye sees its surroundings. It is a highly developed computational framework. This method seeks to provide depth perception, object recognition, and scene understanding from monocular (single-eye) input. It is based on the interdisciplinary fields of computer vision, artificial intelligence, and human visual perception.

Since the middle of the 20th century, the development of monocular vision algorithms has advanced dramatically, starting with simple image processing tasks like object recognition and edge detection. Researchers took on increasingly difficult problems as computing power increased, like motion analysis and depth estimation from a single camera view. These developments played a pivotal role in surmounting the intrinsic constraints of monocular vision, chief among them being the lack of parallax information that is native to binocular vision systems.

Perception of depth is one of the main problems with monocular vision. In monocular vision, additional cues including object size, texture gradients, and motion must be used to infer depth instead of the parallax data obtained from having two eyes. Some methods, such as Structure from Motion (SfM), use multiple frame analysis to extract depth information from object movement. Feature extraction is also important, as algorithms locate important components in an image, such as corners, edges, and textures, in order to deduce the composition of the scene. The advancement of these capabilities has been made possible by robust feature extraction techniques like Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). [1]

The combination of artificial intelligence and machine learning has greatly improved monocular vision algorithms. Convolutional Neural Nets (CNNs), in particular, are deep learning models that have demonstrated exceptional performance in tasks like segmentation and image identification. In order to identify patterns and attributes relevant to certain tasks, these models are trained on large datasets, which increases the algorithms' accuracy and resilience. The combination of cutting-edge AI with conventional computer vision methods has allowed monocular vision systems to operate very well in a wide range of applications.

Algorithms for monocular vision have applications in many different domains. These algorithms are crucial for lane detection, obstacle recognition, and navigation in autonomous cars because they use real-time visual processing to make judgments about where to drive. Robots can carry out activities like picking and arranging objects and avoiding obstacles thanks to monocular vision, which helps with navigation, object manipulation, and environmental interaction in robotics. Applications for augmented reality (AR) overlay digital data on the physical world using monocular vision, necessitating accurate object tracking and recognition. These algorithms aid in the analysis of pictures from medical equipment such as microscopes and endoscopes, which helps with diagnosis and treatment planning [2].

Accurate and trustworthy depth perception is still a difficulty, even with the major advancements particularly in dynamic and complicated situations. Future studies should combine more sophisticated AI methods, increase algorithmic efficiency, and strengthen these systems' resilience for practical use. Integrating lidar and radar with monocular vision has the potential to create more robust and comprehensive perception systems, opening the door to even more advanced features and applications.

1.2 Enhancing Urban Living and Sustainability with Monocular Vision Technologies

Enhancing urban living through monocular vision technologies involves integrating these systems into various aspects of city infrastructure to create more sustainable and efficient urban environments. In smart traffic management, monocular vision can optimize traffic flow, reduce congestion, and lower emissions by providing real-time data for dynamic traffic control. For public safety, these technologies can be used in surveillance systems to detect and respond to incidents promptly, ensuring safer communities. In waste management, monocular vision can enhance recycling processes by accurately sorting materials and monitoring waste levels to optimize collection schedules, reducing overall waste and promoting recycling. These applications not only improve the efficiency of urban services but also contribute to environmental sustainability and a higher quality of life for residents, fostering the development of resilient and sustainable communities.

1.3 Problem Statement

Monocular vision, or the loss of vision in one eye, affects millions of people globally. Numerous factors, including illnesses, hereditary disorders, and accidents, might contribute to this. Even while some people adjust effectively, monocular vision frequently poses serious difficulties.

The main issue is that the brain cannot accurately detect depth with one eye alone. For activities like catching a ball, driving, and negotiating stairs, depth perception is essential. This may result in trouble completing everyday tasks, a decline in mobility, and less involvement in social and recreational activities. Monocular vision can also impair hand- eye coordination and lead to balance issues. These difficulties may have a major effect on an individual's freedom and quality of life. [3]

In addition to its physical difficulties, monocular vision can have profound psychological and societal effects. People who consistently suffer with depth perception may become less confident and reluctant to engage in once-enjoyed activities. They may isolate themselves by withdrawing from social events out of a fear of mishaps or limits. In addition to exhaustion and headaches, the extra effort the brain must make to comprehend visual information with only one eye can cause anxiety in busy or new surroundings. [4]

It's critical to keep in mind that everyone is affected by monocular vision differently. Individuals who lose one eye's vision early in life usually adjust better than those who lose vision later in life. The difficulties may also be exacerbated by underlying medical issues or even the state of the remaining eye.It is essential to comprehend these challenges. People with monocular vision can learn to navigate their environment and lead full, independent lives with the right assistance and rehabilitation techniques.

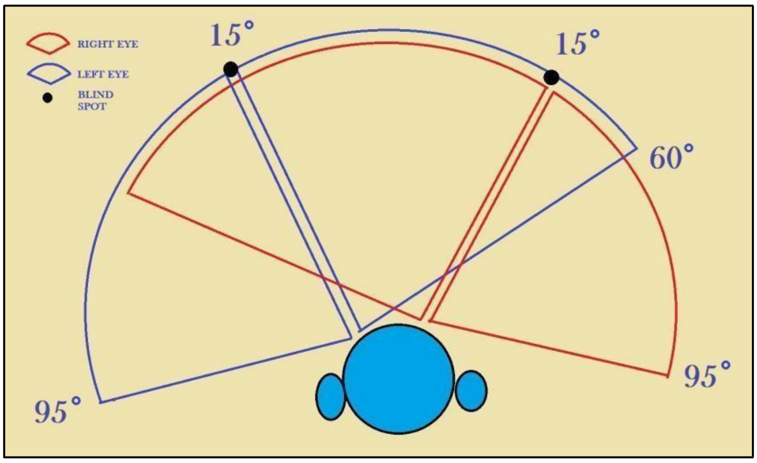


Figure 1.1 Sight with only one eye

1.4 Project Objective

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 :

1. To generate a bird's-eye view image and detect lane boundaries and vehicles from video frames using a monocular camera system.
2. To Enhance disparity map precision by integrating bird's-eye view transformation, robust lane boundary detection, and vehicle localization techniques

1.5 Scope of Project

Monocular vision algorithms leverage a single camera to interpret and understand visual information, emulating human vision's ability to perceive depth and spatial relationships.

1. Develop and refine a hybrid cascaded monocular vision algorithm to improve the accuracy of disparity maps and depth perception in complex environments.
2. Design systems to detect lane boundaries and vehicles using bird's-eye view transformations, robust lane boundary detection, and vehicle localization techniques.

3. Utilize advanced image processing and deep learning techniques, such as convolutional neural networks (CNNs), for depth estimation, object recognition, and semantic segmentation.
4. Perform camera calibration, including bird’s-eye view transformations, to ensure accurate mapping and representation of real-world coordinates in visual systems.
5. Test the system using datasets and real-time applications to evaluate robustness in dynamic situations, addressing challenges like occlusion, low-texture regions, and radiometric aberrations.
6. A Windows 10 desktop computer with a 3.2GHz processor and 8GB of RAM is used for the testing. The experimental photos are compared to a common online benchmarking dataset from Pixabay in order to assess accuracy.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Vision measurement is one of the fundamental technologies that occupy an important position in machine vision. Depth measurement of targets is now widely used in areas such as visual localization, target tracking, visual obstacle avoidance, and visual servo control. Research about vision measurement has received a lot of attention, however, traditional monocular focus measurement systems suffer from large system size and slow response speed. At the same time, the measurement accuracy and application generalization abilities of existing monocular vision measurement algorithm still have much room for improvement.

Vision measurement can be divided into monocular vision measurement, binocular vision measurement and structured light measurement. Binocular depth measurement is the calculation of depth values by pixel matching of two images acquired by two cameras, which is a highly accurate method of measurement. However, in some complex background conditions, there may be a failure of pixel matching, affecting the measurement accuracy, and the long binocular baseline when making measurements at longer distances can cause the problem of a larger overall size of the binocular vision measurement device. The main advantages of structured light vision measurement methods are the mature technology, smaller camera baselines that can be miniaturized and faster processing speed, but they are easily disturbed by ambient light, therefore, mainly used in indoor environments.

Although the above-mentioned methods can be used for depth measurement, there are problems of high cost, susceptibility to interference, complicated calibration methods and limited application scenarios, therefore, many domestic and foreign institutions and researchers began to study monocular vision measurement, compared with the previous vision measurement methods, monocular vision measurement only needs to rely on a single image or multiple images to obtain the depth information of the target. The main advantages are simple measurement principle, small size, fast measurement speed, low cost and a wide range of applications, while also avoiding the problems of small field of view and difficult stereo matching in stereo vision. [5]

2.2 Monocular Vision Concept

Monocular depth measurement is an important fundamental technology in machine vision with a wide range of application scenarios, while traditional monocular focus measurement systems have deficiencies such as large size and slow response time. Low measurement accuracy of existing monocular measurement algorithms and can only measure for specific targets. Based on this, a liquid bionic vision system is proposed and a monocular depth measurement model is developed, the structural parameters are optimized using the NSGA-II algorithm.[6] After calibrating the liquid lens, a neural network-based monocular intelligent depth measurement method is proposed, expressing the relationship between the liquid lens control current and the target depth. The experimental results show that the proposed method and the designed system perform well in terms of measurement accuracy, stability and applicability, with a relative error of less than 1.11 % within a measurement range of 1 m and single measurement time does not exceed 410 ms. [6]

2.2.1 Monocular vision techniques

Monocular vision, or single-camera vision, is a field in computer vision focused on interpreting and understanding the three-dimensional (3D) world using images captured from a single camera. The central objective is to derive meaningful information about the spatial structure, motion, and properties of objects and scenes from 2D images. This task is fundamentally challenging because the 3D information is projected onto a 2D plane, leading to the loss of depth and scale cues that are inherently present in the real world.

One of the primary challenges in monocular vision is depth perception. Since a single image lacks explicit depth information, determining the distance of objects from the camera involves complex inference. This depth ambiguity means that an object could be either small and close or large and far away, without any straightforward way to distinguish between the two. Additionally, occlusion, where objects block parts of each other from view, further complicates the task by obscuring parts of the scene and making it harder to understand the spatial relationships between objects [7]. Another significant challenge is estimating motion, whether it is the movement of objects within the scene or the motion of the camera itself. Motion estimation is difficult with a single camera because it relies heavily on parallax, the apparent shift of objects relative to each other when viewed from different positions, which is not available in monocular vision. This limitation necessitates sophisticated algorithms to infer movement from sequential images captured over time.

The variability in lighting and shadows also presents a problem for monocular vision. Changes in illumination can alter the appearance of objects, making it difficult for algorithms to consistently recognize and interpret the scene. Shadows can create false edges and shapes that confuse the interpretation of the 3D structure [7]. Despite these challenges, the goals of monocular vision are ambitious and wide-ranging. Depth estimation aims to create a depth map from a single image, while 3D reconstruction seeks to rebuild the 3D geometry of objects or entire scenes. Motion analysis involves determining the trajectory of objects or the camera, and object recognition and segmentation focus on identifying and separating different objects within the scene. Scene understanding integrates these elements to provide a coherent interpretation of the scene's context, including object interactions and spatial layout [7].

Applications of monocular vision are diverse and impactful. In autonomous vehicles, it aids in navigation and obstacle avoidance. In robotics, it enhances scene understanding and object manipulation capabilities [7]. Augmented reality applications use monocular vision to overlay virtual objects onto the real world accurately. In medical imaging, it helps infer 3D structures from 2D scans, and in surveillance, it supports activity monitoring and interpretation from single-camera footage [7].

Approaches to solving monocular vision problems include deep learning techniques, such as convolutional neural networks (CNNs) that can predict depth and 3D structure from images. Structure from Motion (SfM) techniques infer 3D structures by analyzing the movement of objects across multiple frames from a moving camera [7]. Photometric stereo methods estimate surface normals and depth by examining shading variations under different lighting conditions. Optical flow techniques estimate object motion between consecutive frames, contributing to depth and structure inference [7].

In summary, monocular vision tackles the complex task of interpreting 3D scenes from 2D images, facing challenges in depth perception, motion estimation, and lighting variability. Through advanced techniques and algorithms, it aims to provide detailed and accurate understanding of the visual world, enabling a wide range of applications in technology and industry.

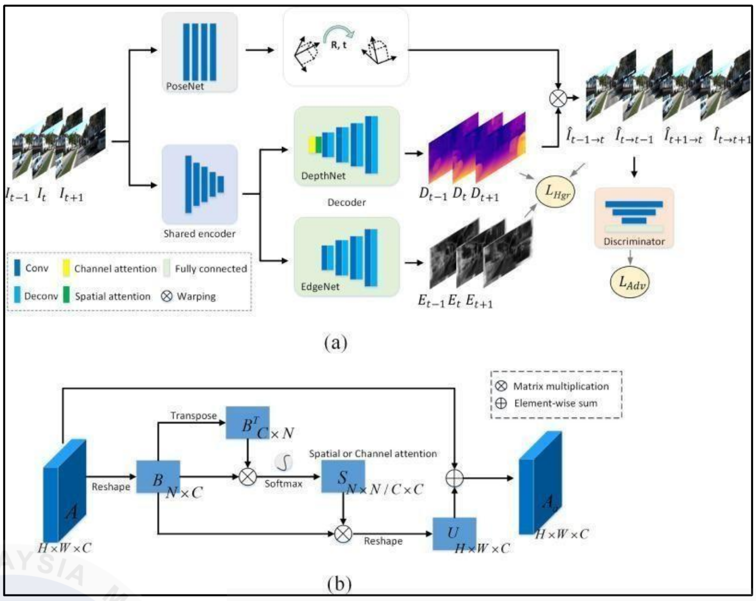


Figure 2.1 Architecture and Mechanism of Monocular Vision Algorithms for Depth and Edge Prediction [7]

2.2.2 Monocular Vision Framework

The mainstream of binocular vision research has long been focused on understanding how binocular disparity is used for depth perception. In recent years, researchers have begun to explore how monocular regions in binocularly viewed scenes contribute to our perception of the three-dimensional world [8]. This work review the field as it currently stands, with a focus on understanding the extent to which the role of monocular regions in depth perception can be understood using extant theories of binocular vision.

Image Acquisition:

The first step in a monocular vision system is to capture images using a single camera. This camera is typically a standard digital camera mounted on a platform such as a vehicle, robot, or a stationary setup. The images captured serve as the raw data input for further processing. The quality and resolution of these images are crucial as they directly impact the subsequent analysis and interpretation steps.

Preprocessing:

Once the images are acquired, they undergo preprocessing to enhance their quality and make them suitable for analysis. This involves techniques like denoising to remove any unwanted noise, adjusting the contrast to improve visibility, and sharpening to highlight important details. Normalization is also performed to standardize the image data, ensuring consistency across different images. If necessary, the images may be converted to a different color space, such as grayscale, to simplify the processing steps.

Feature Extraction:

In this step, key features are identified and extracted from the preprocessed images. Feature extraction involves detecting edges using algorithms like the Canny or Sobel edge detectors, which help in identifying the boundaries of objects. Keypoint detection methods, such as SIFT (Scale-Invariant Feature Transform) or SURF (Speeded-Up Robust Features), are used to locate distinctive points in the image. [9] These keypoints are then described using feature descriptors like ORB (Oriented FAST and Rotated BRIEF) or BRIEF (Binary Robust Independent Elementary Features), which provide a compact representation of the keypoints for further analysis. [10]

Depth Estimation:

A major challenge in monocular vision is estimating the depth of objects from a single image. Depth estimation techniques aim to create a depth map that represents the distance of each pixel from the camera. Modern approaches often use deep learning models, such as convolutional neural networks (CNNs), trained on large datasets to predict depth directly from single images. These models leverage monocular cues like texture gradients, occlusions, and perspective to infer depth information.

3D Reconstruction:

Building on the depth estimation, the next step is to reconstruct the 3D structure of the scene. If multiple frames are available from a moving camera, Structure from Motion (SfM) techniques can be employed to analyze the motion of objects across frames and reconstruct the 3D scene. For single-view reconstruction, prior knowledge or learned models are used to infer the 3D shapes of objects based on their appearance in the image. This reconstruction provides a spatial understanding of the scene's geometry.

Object Recognition and Segmentation:

Recognizing and segmenting objects within the scene is crucial for detailed scene interpretation. Object detection algorithms, such as YOLO (You Only Look Once) or Faster R-CNN, are used to identify and locate objects in the image. Semantic segmentation techniques assign class labels to each pixel, distinguishing between different parts of the scene like roads, buildings, and vehicles. Instance segmentation further refines this by detecting and delineating individual objects, using models like Mask R-CNN.

Motion Analysis:

Understanding the motion within the scene involves estimating how objects or the camera itself move. Optical flow methods, such as Lucas-Kanade or Farneback, compute the motion of pixels between consecutive frames, providing information about object and camera movement. Egomotion estimation focuses specifically on determining the camera's movement by analyzing the overall flow of the scene, which is essential for applications like autonomous driving.

Scene Understanding:

Combining all previous steps, scene understanding aims to interpret the overall context of the scene. This involves analyzing spatial relationships between objects, such as their relative positions and distances, and using contextual information to understand interactions and dynamics within the scene. This comprehensive interpretation is vital for making informed decisions in applications like navigation, manipulation, and augmented reality.

Output Generation:

Finally, the system generates meaningful outputs based on the interpreted data. This can include visual representations like depth maps and 3D models, or annotated images with detected objects and segmented regions. The processed data is often exported for use in downstream applications, such as assisting in navigation for autonomous vehicles, enhancing human-robot interaction in robotics, or overlaying virtual objects in augmented reality systems.

2.3 Previous methods of monocular vision developed by the researches

Monocular visual odometry provides more robust functions on navigation and obstacle avoidance for mobile robots than other visual odometries, such as binocular visual odometry, RGB-D visual odometry and basic odometry. [11] Describes the problem of visual odometry and also determines the relationships between visual odometry and visual simultaneous localization and mapping (SLAM). The basic principle of visual odometry is expressed in the form of mathematics, specifically by incrementally solving the pose changes of two series of frames and further improving the odometry through global optimization. After analyzing the three main ways of implementing visual odometry, the state-of-the-art monocular visual odometries, including ORB-SLAM2, DSO and SVO, are also analyzed and compared in detail. The issues of robustness and real-time operations, which are generally of interest in the current visual odometry research, are discussed from the future development of the directions and trends. Furthermore, we present a novel framework for the implementation of next-generation visual odometry based on additional high-dimensional features, which have not been implemented in the relevant applications.

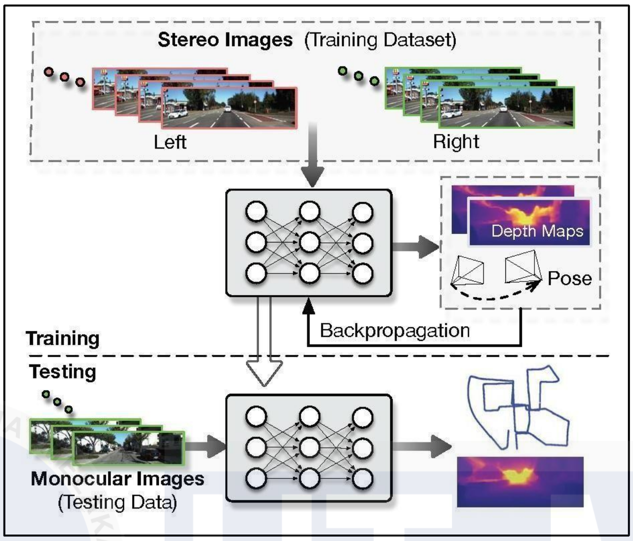


Figure 2.2 Training and Testing Pipeline for Monocular Depth Estimation Using Stereo Images [11]

[12] Presents a self-calibrating method based on monocular vision that can precisely measure 3D structural displacement. The method makes use of a custom planar marker with distinct graph patterns, which allows for accurate PnP resolution and image recognition. The suggested method automatically determines the marker's three-dimensional position and orientation. It has also been discussed how long-term monitoring is limited by marker-free techniques. Through Monte-Carlo simulation, the suggested technique's accuracy and robustness are methodically assessed before being confirmed through experiments. The suggested method's accuracy for measuring 3D displacement is 0.049 mm, according to the results. This suggests that the method is useful for automatically determining and calibrating the stereo displacement of structures using monocular vision.

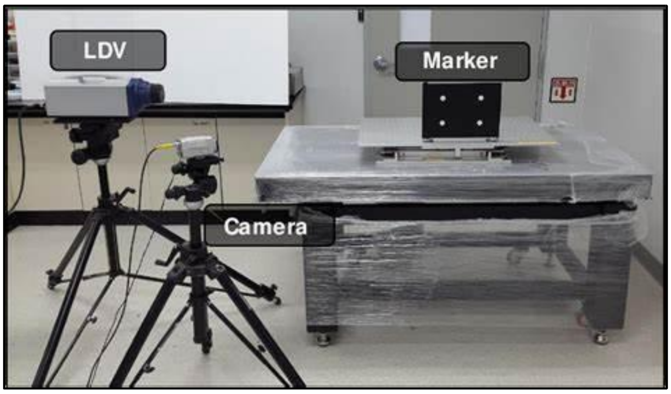


Figure 2.3 Experimental Setup for Vibration Measurement Using Laser Doppler Vibrometer and Camera [12]

[12] Traditional monocular focus measurement systems have drawbacks like large size and slow response time, but monocular depth measurement is an important foundational technology in machine vision with a wide range of application scenarios. The current monocular measurement algorithms have low measurement accuracy and are limited to measuring specific targets. Based on this, a monocular depth measurement model and a liquid bionic vision system are suggested, and the NSGA-II algorithm is used to optimize the structural parameters. [13] A neural network-based monocular intelligent depth measurement method is proposed, which expresses the relationship between the liquid lens control current and the target depth, after the liquid lens has been calibrated. The experimental results demonstrate that, with a single measurement time of no more than 410 ms and a relative error of less than 1.11 percent within a measurement range of 1 m, the suggested method and the designed system perform well in terms of measurement accuracy, stability, and applicability.

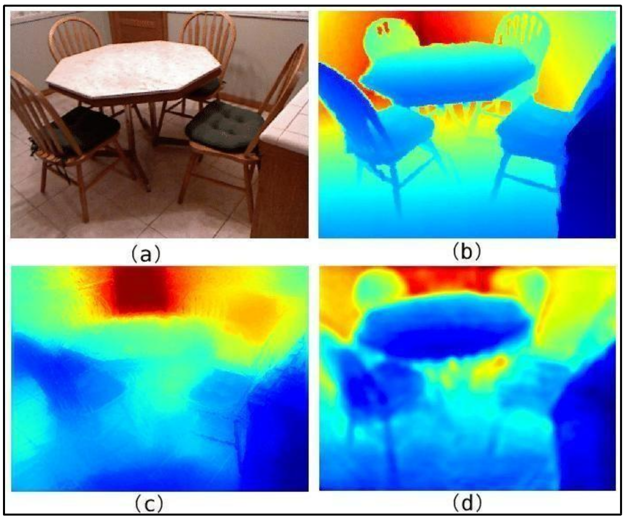


Figure 2.4 Thermal Imaging Analysis of a Dining Table and Chairs. [13]

A new implementation technique for effective simultaneous localization and mapping using a forward-viewing monocular vision sensor is presented in another work by [14]. The technique is designed to work in real time on an inexpensive embedded system for indoor robot service. In this paper, the direction of the vanishing point is used to directly estimate the orientation of a robot. Next, as basic linear equations, the estimation models for the robot position and the line landmark are obtained. By using these models, a local map correction technique effectively corrects the camera poses and landmark positions. Real-time experiments using a low-cost embedded system and dataset-based experiments using a desktop computer are used to demonstrate the performance of the proposed method under a variety of demanding environments. One of the experimental settings is a true homey atmosphere. Low-textured regions, moving people, or shifting surroundings are all present in these conditions. The suggested approach is also evaluated with the robotics advancement via sensorial and elaborated large-scale datasets published online as benchmark datasets.

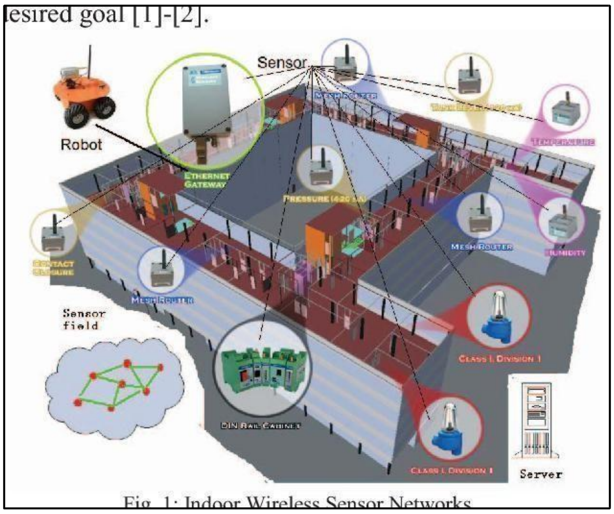


Figure 2.5 Layout of an Indoor Wireless Sensor Network System [14]

Next, the work by [15] addresses the use of monocular vision in the depth estimation process for a mobile platform. The most significant difficulty faced by autonomous mobile platforms in unfamiliar environments is precisely estimating the distances to and locations of nearby obstacles. They require trustworthy range sensors to identify any obstructions in their path so they can move safely from one location to another. The application of a vision sensor makes sense because it offers a more sensible and affordable option. This work's method necessitates a straightforward calibration. The depth estimation procedure's equation will be created using the data gathered from the calibration process. The outcomes attest to the validity of the depth estimation methodology.

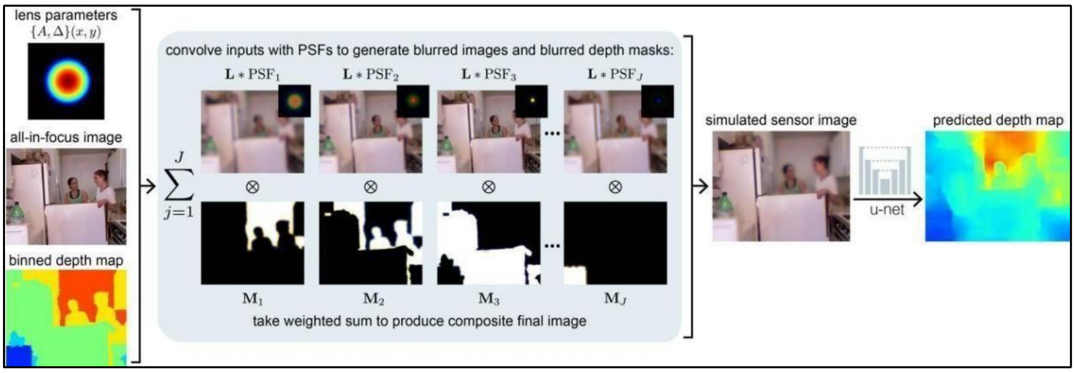

lens parameters
$\{A, \Delta\}(x, y)$
all-in-focus image
binned depth map
$\sum_{j=1}^{J}$
convolve inputs with PSFs to generate blurred images and blurred depth masks:
$\mathbf{L} * \mathrm{PSF}_1$
$\mathbf{L} * \mathrm{PSF}_2$
$\mathbf{L} * \mathrm{PSF}_3$
$\mathbf{L} * \mathrm{PSF}_J$
$\otimes$
$\mathbf{M}_1$
$\mathbf{M}_2$
$\mathbf{M}_3$
$\mathbf{M}_J$
take weighted sum to produce composite final image
simulated sensor image
u-net
predicted depth map

Figure 2.6 Depth Synthesis Pipeline Using Convolutional Neural Networks and Point Spread Functions [15]

One of the most popular research areas in computer vision nowadays is real-time 3D reconstruction, which has emerged as a key technology for industrialized automatic systems, mobile robot path planning, and virtual reality. Three primary issues are currently plaguing the field of real-time 3D reconstruction. First of all, the cost is high. It is less convenient because it needs a wider variety of sensors. Second, the 3D model cannot be precisely established in real time due to the slow reconstruction speed. Thirdly, there is a significant amount of reconstruction error, making it impossible to accurately recreate the scenes. Because of this, in this work, [16] present a real-time 3D reconstruction technique based on monocular vision. First, real-time visual data is collected by a single RGB-D camera, and then the YOLACT++ network is utilized to recognize and segment the data in order to extract some of the most significant visual data. Second, this work proposes a three-dimensional position estimation method based on deep learning for joint coding of visual information by combining the three stages of depth recovery, depth optimization, and deep fusion. It can directly yield the precise 3D point values of the segmented image and lessen the depth error brought about by the depth measurement procedure. Lastly, they suggest an approach to maximize the three-dimensional point values found above that is based on the limited outlier adjustment of the cluster center distance.

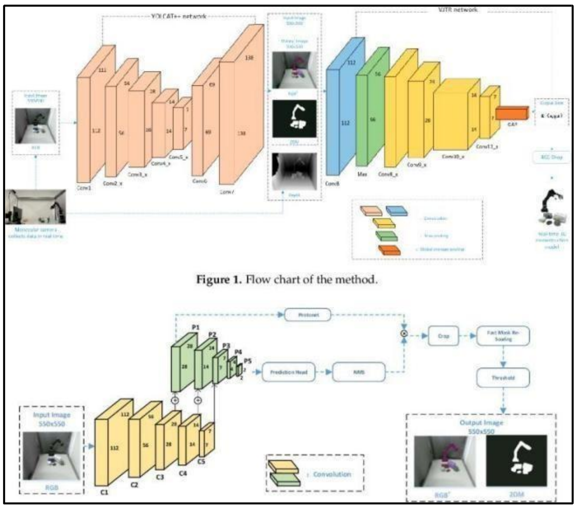


Figure 2.7 Workflow and Process Flow for Video-Based Human Gait Analysis and Recognition Method [16]

[17] Presents methods for correcting incorrect stereo reconstruction in an intact eye through robot-assisted intraocular surgery employing monocular vision. We suggest a novel technique for estimating the retinal surface that utilizes a structured-light methodology. The Micron is a handheld robot that automatically scans a laser probe to produce projected beam patterns on the retinal surface. Planar reconstruction of the surface is then possible through geometric analysis of the patterns. A scheme combining surface reconstruction and partitioned visual servoing is used to achieve monocular hybrid visual servoing, which enables automated surgery in an intact eye. They assess the estimation method's performance in both dry and wet conditions, as well as its sensitivity to pertinent parameters. Experiments for automated laser photocoagulation in an in vitro realistic eye phantom validate the methodology. Lastly, they present the first example of automated intraocular laser surgery performed ex vivo on porcine eyes.

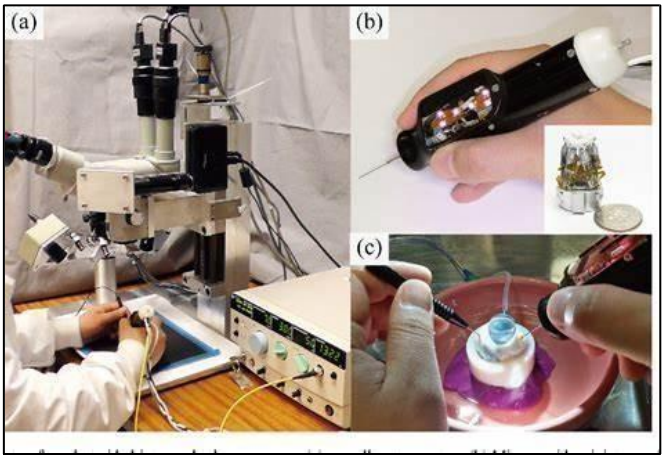


Figure 2.8 Experimental Setup and Components for Micromanipulation and Microassembly Tasks [17]

[17] Explains deep learning-based approaches and conventional methods for monocular visual odometry. An overview of cutting-edge techniques that may be helpful for applications involving displacement measurement is provided in this paper. In particular, it has been observed from the state of the art review that, in comparison to conventional methods, new deep learning-based techniques may lessen reliance on the system's hardware.

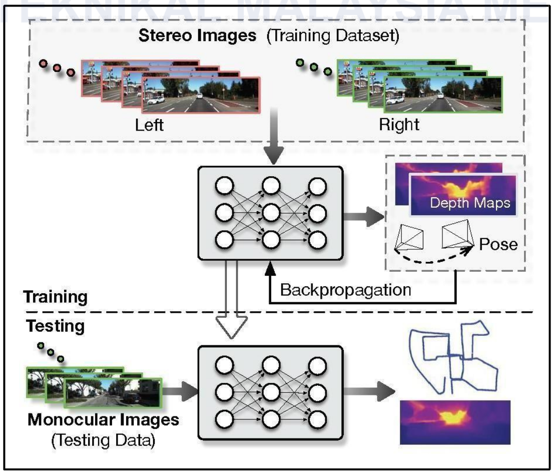


Figure 2.9 End-to-End Deep Learning Pipeline for Monocular Depth Estimation [18]

[18] One of the key challenges and fundamental tasks in mobile robot applications is accurate localization. In this work, we suggested an image-based approach to quantify parameters of an object using a monocular vision system or visual camera. If the object's sizes are unknown, they can be precisely determined using data taken from two images taken at two different camera positions along the optical axis. In addition, the object's direction (angle) and distance can be ascertained. One image is sufficient to ascertain the object's direction (angle) and distance if its dimensions are known. The suggested method does not require the object to be perpendicular to the camera's optical axis, in contrast to other methods currently in use. [20] The perception systems of the majority of autonomous cars are based on costly sensors like high-precision Global Positioning System (GPS), Lidar, and Radar. But cameras are a more desirable option because they can offer richer sensing at a significantly lower cost. Monocular vision-based driving assistance systems (DAS) have been a growing area of research interest. One key technology in DAS is inter-vehicle distance estimation. The current techniques for calculating the inter-vehicle distance based on monocular vision are still limited by issues like poor performance on distance estimation for vehicles that are severely obscured, unstable accuracy for different types of vehicles, and low accuracy at larger distances. This study suggests a monocular vision end-to-end inter-vehicle distance estimation method based on 3D detection to increase the precision and robustness of ranging results. The 3D detection method is used to determine the actual area of the vehicle's rare view and the corresponding projection area in the image. Then, in order to recover distance, a geometric model of area and distance is created using the camera projection principle. vehicle ranging results can reach approximately 98%, while the accuracy deviation between vehicles with different visual angles is less than 2%. Using test set data from the real-world computer vision benchmark, KITTI, our method demonstrates its potential in complex traffic scenarios.

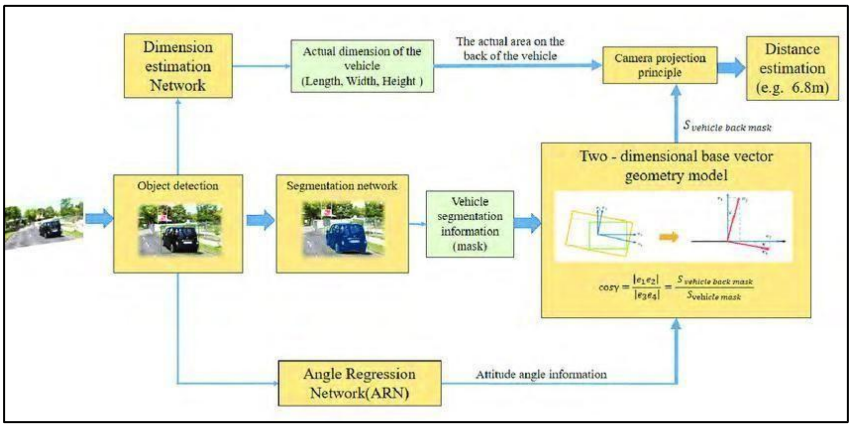

Dimension estimation Network
Actual dimension of the vehicle (Length, Width, Height)
The actual area on the back of the vehicle
Camera projection principle
Distance estimation (e.g. 6.8m)
$S_{vehicle\ back\ mask}$
Two - dimensional base vector geometry model
Object detection
Segmentation network
Vehicle segmentation information (mask)
$\cos\gamma = \frac{|e_1e_2|}{|e_3e_4|} = \frac{S_{vehicle\ back\ mask}}{S_{vehicle\ mask}}$
Angle Regression Network(ARN)
Attitude angle information

Figure 2.10 Vehicle Dimension and Distance Estimation Pipeline Using Computer Vision Techniques [20]

[18] One important metric for assessing the caliber of dynamic compaction is ramping settlement. Safety cannot be guaranteed because it primarily depends on manual monitoring, which has a negative impact on construction efficiency. The problem of having to stop the dynamic compaction operation during manual monitoring can be effectively avoided by using the Perspective 3-Point Rammer Pose Estimation (P3P-RPE) algorithm, which is based on monocular vision measurement and can be used in tandem with construction. Next, a thorough analysis of the influencing factors is conducted, and the P3P-RPE error propagation model is proposed. The results demonstrate that the accuracy can reach 30 mm, which is less than the specification's upper limit of 50 mm.

[19] And pitch angle of camera θ and feature point recognition have a great influence on the results. This method has good practicability in ramming settlement monitoring, and can provide reference for the monitoring of other similar projects.

Compaction Plan
S3..- Measured

Drop Location ID
17:224

Drop Count
0/12

Last Displacement
0.00

Total Displacement
0.00

Avg. Displacement
0.00

Heading
51.00

Figure 2.11 Trimble Survey Controller [21]

[20] Particularly for the vital parts of the bridge, like the bridge bearings, the displacement and rotation are important markers of the health of the bridge. In order to achieve continuous structural monitoring, this work offers a real-time simultaneous measurement technique for bridge bearing displacement and rotation. Among the principal contributions are: (1) Feature-constrained monocular visual odometry (VO), which uses image sequences of infrared array lamp targets with multiple feature constraints as visual input to construct the mathematical relationship between displacement and rotation and coordinate system transformation matrix, is the basis for the development of a displacement and rotation simultaneous measurement algorithm and system. (2) To effectively improve the accuracy and robustness of displacement and rotation measurements, a two-step parameter optimization strategy is proposed, based on adaptive multi-window centroid array tracking and transformation matrix purification and refinement. The effectiveness of the proposed method is verified in the laboratory experiment and the actual bridge bearing monitoring of the Huangpu Pearl River Bridge.

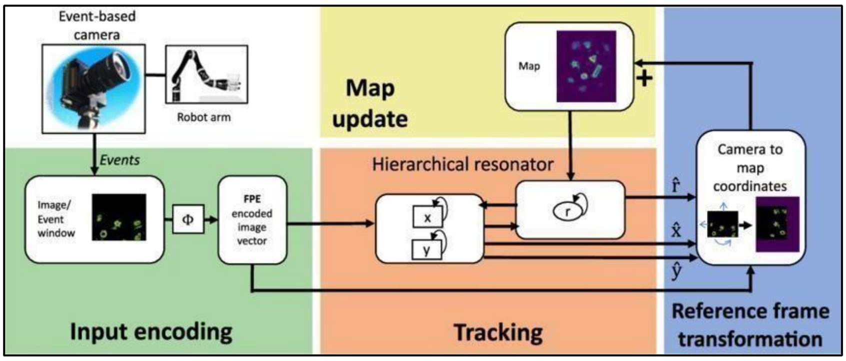


Figure 2.12 The Process of Image Processing [22]

[21] A prevalent anatomical condition in endoscopic treatment or diagnosis is lumen. Autonomous robotic endoscopy navigation typically uses vision techniques, like lumen center detection or anatomically specific contour features. These techniques might not be able to accomplish smooth interventions, though, and they might produce a lot of conveying force without any spatial awareness of the tissue state. This paper proposes a novel navigation pipeline based on a monocular vision that is aware of its surroundings. The first step in the spatial awareness pipeline is to reconstruct the approximate tissue surface in real time using a data-driven depth estimation technique. Then, using digital topology, the spatial shape of the lumen described by the skeleton is extracted. After smoothing out the skeleton, we use the geometric information obtained from the pathway to create an adaptive autonomous control strategy. We conduct experiments on a colon phantom and ex vivo pig intestines. We test turning performance of several bending segments with different angles in phantom, as well as the overall performance of a long-range intervention task in the phantom and pig intestines. The results show our navigation scheme achieve smoother intervention with lower conveying force. The proposed navigation method with spatial awareness can effectively improve the fluency of autonomous robotic endoscopy.

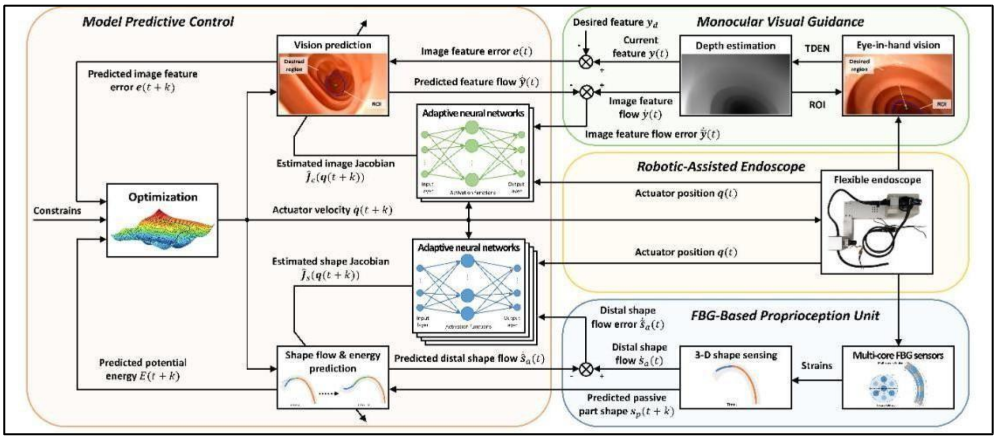

Model Predictive Control
Vision prediction
Desired region
ROI
Image feature error $e(t)$
Predicted feature flow $\hat{\dot{y}}(t)$
Predicted image feature error $e(t+k)$
Adaptive neural networks
Input layer
Activation functions
Output layer
Estimated image Jacobian $\hat{J}_e(q(t+k))$
Optimization
Constrains
Actuator velocity $\dot{q}(t+k)$
Estimated shape Jacobian $\hat{J}_s(q(t+k))$
Predicted potential energy $E(t+k)$
Shape flow & energy prediction
Predicted distal shape flow $\hat{\dot{s}}_a(t)$
Desired feature y_d
Current feature $y(t)$
Monocular Visual Guidance
Depth estimation
TDEN
Eye-in-hand vision
Image feature flow $\dot{y}(t)$
Image feature flow error $\tilde{\dot{y}}(t)$
Robotic-Assisted Endoscope
Flexible endoscope
Actuator position $q(t)$
Actuator position $q(t)$
FBG-Based Proprioception Unit
Distal shape flow error $\tilde{\dot{s}}_a(t)$
Distal shape flow $\dot{s}_a(t)$
3-D shape sensing
Strains
Multi-core FBG sensors
Predicted passive part shape $s_p(t+k)$

Figure 2.13 The proposed data-driven framework for autonomous navigation of flexible endoscopes including robotic-assisted endoscope [23]

Monocular vision has been viewed as a promising onboard obstacle perception solution for Unmanned Aerial Vehicles (UAVs) in recent years, with applications in Sense and Avoid (SAA). Nevertheless, monocular vision based obstacle localization ability is insufficient for collision avoidance due to the limitations of monocular optical measurement. Thus, this paper proposes a trajectory planning scheme for obstacle collision avoidance while taking into account the characteristics of monocular optical measurement. First, to help with avoiding obstacle collisions, two obstacle localization modes—namely, relative range-based Mode 1 and relative angle-based Mode 2—are defined in this paper. The observability of Mode 2's obstacle localization is examined, and coordinate systems for obstacle localization are built in response to the findings of the analysis. Given coordinate systems, Orthogonal Iteration (OI) is further adopted for obstacle localization. Secondly, due to the lack of global knowledge caused by monocular vision based localization capability, a rolling horizon based safe trajectory planning method is presented. In each time segment, the trajectory is optimized with considerations of: 1) objective functions including minimum trajectory length, elapsed time and energy consumption; 2) constraints including Mode 1 and 2 based obstacle collision Finally, simulation results indicate the proposed obstacle collision avoidance trajectory planning scheme enhances UAV safety level and can achieve favorable performance when compared with geometric based obstacle collision avoidance and collision avoidance with global knowledge.

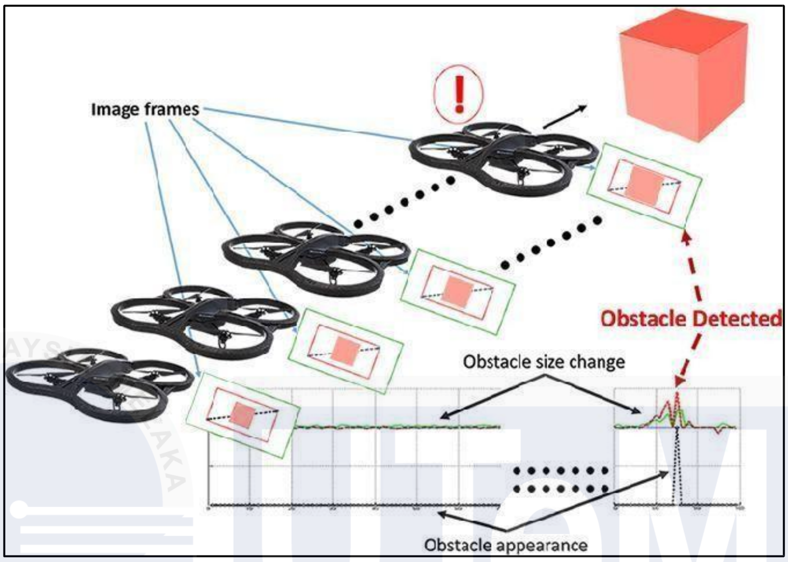


Figure 2.14 Drone Obstacle Detection [24]

[25]Many industrial operations have trouble quantifying materials that are loaded onto trucks. In order to assess admissibility, inspectors at disposal facilities handling construction and demolition waste (CDW) are frequently obliged to weigh the various waste components loaded on arriving trucks. Construction materials are bulky and diverse, making it difficult to accurately quantify certain waste categories without compromising field operability. This work suggests a monocular vision-based CDW volume estimation algorithm that can automatically quantify the amount of particular material components—such as wood, rock, and gravel—in waste mixtures from a single image. The algorithm's results show that it can calculate truck bucket dimensions with a relative error of 0.065 and estimate the volume of material-level construction waste with a relative error of 0.169. Processing one image takes an average of 3.3 seconds. In applying the algorithm to

analyze 2,914 waste truckloads received by an off-site sorting facility in Hong Kong, we observe that the facility entrance received around 800.0 m3 CDW per day of which about 10.8 m3 were rejected. Since non-inert wood/cardboard accounts for the highest proportion among all material types, this may imply that many waste dumps accepted by the facility may have been in violation of the admissibility criteria. The study contributes to the knowledge body by providing a novel, non-destructive approach to quantifying CDW via monocular vision. It can be extended to address the general problem of truck payload quantification in scenarios such as road construction, warehouse inventory management, and logistics and supply chain management.

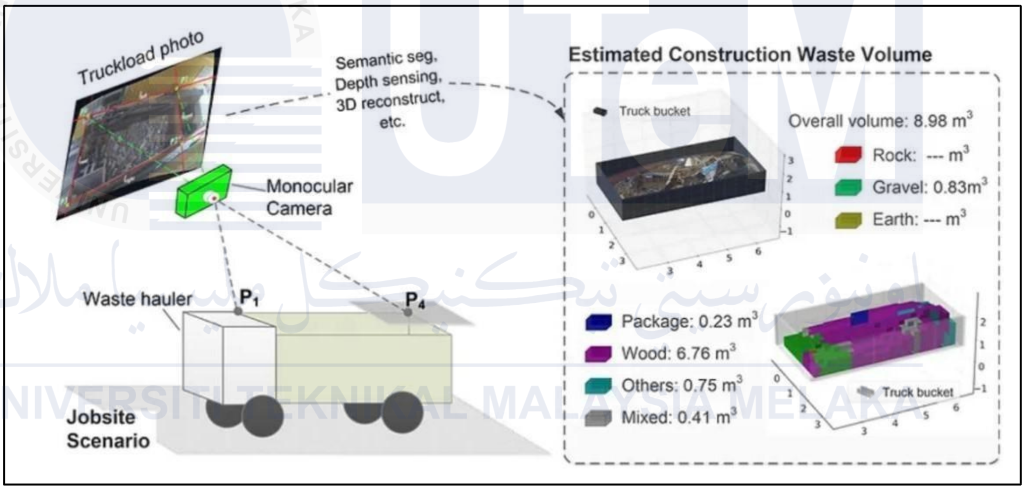


Figure 2.15 Estimating Construction Waste Volume with Monocular Camera [25]

2.4 Table of comparison

Table 1 shows the comparison of monocular method from year 2016 to 2022.

Table 2.1 Table of comparison

Year	Authors	Comparison description
2016	Eigen et al.	The work by [26]. introduced a deep learning-based approach for monocular depth estimation, achieving state-of-the-art results by training a convolutional neural network (CNN) on a large dataset of stereo image pairs.
2017	Liu et al.	The work by [27]. proposed a method combining depth from monocular cues with depth from motion and depth from defocus, demonstrating improved accuracy by leveraging multiple depth estimation techniques.
2018	Godard et al.	The work by [28] introduced unsupervised learning techniques for monocular depth estimation, leveraging geometric constraints and photometric errors to train neural networks without explicit depth annotations.
2019	Zhou et al.	The work by [29] presented a framework for semantic understanding-guided depth estimation, incorporating semantic segmentation to refine depth predictions and improve depth estimation accuracy in complex scenes.
2020	Mahjourian et al.	The work by [30]. proposed a method for monocular depth estimation based on self-supervised learning, utilizing geometric consistency loss and view synthesis techniques to train depth estimation networks without ground truth depth data.
2021	Wang et al.	The work by [31]. introduced a novel approach for monocular

		depth estimation using attention mechanisms, enhancing the network's ability to focus on relevant image regions and improving depth estimation accuracy in cluttered scenes.
2022	Chen et al.	The work by [32]. presented a comparative study of various monocular depth estimation methods, evaluating their performance on benchmark datasets and highlighting the strengths and limitations of each approach.

2.5 Summary

The literature on monocular vision provides a comprehensive understanding of the visual sense obtained with only one eye, in contrast to binocular vision, which necessitates the use of both eyes. People with monocular vision are able to perceive spatial relationships and depth. These stages include light entering the eye, images developing on the retina, signals traveling to the brain, and processing within certain regions of the visual cortex. Even in the absence of binocular depth cues, the brain uses a variety of monocular cues, such as relative size, perspective, and motion parallax, to construct a three- dimensional representation of the environment. Moreover, individuals with monocular vision demonstrate remarkable flexibility, compensating for their impairments by utilizing other senses to augment their perception of their surroundings. Generally speaking, the literature highlights the intricate relationships between sensory inputs and brain processing, which illuminates the complexities of monocular vision and its applications in a range of domains.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The methodology section of this study describes the methodical approach and procedures used to achieve the study's objectives and respond to the research questions. This section, which provides a comprehensive overview of the methodologies used, is crucial to guarantee the study's reproducibility and allow a thorough comprehension of the research process. Important components of the research approach include the study design, sample strategies, data collection techniques, data analysis procedures, and ethical considerations. Each of these components is meticulously planned and executed in order to maintain the study's validity, reliability, and rigor. The research design, which serves as the study's blueprint, directs the overall strategy and framework used in the research. A qualitative, quantitative, or mixed-methods approach may be used in the study, depending on the nature of the research questions and aims. Every approach has unique qualities, and they are selected according to how well they can address the research topic.

3.2 Sustainable Development Tools for Enhancing Vital Monocular Vision Systems

In order to reduce their negative environmental effects, edge computing, low-power hardware, energy-efficient algorithms, and renewable energy sources are used in sustainable development tools for critical monocular vision. Software designed for efficient data compression and storage, intelligent power management, and low resource consumption utilised to enhance these systems' performance. Two more strategies to increase sustainability are to collaborate on open-source platforms and use recyclable materials. These instruments ensure the efficiency and environmental friendliness of monocular vision applications such as assistive technologies, driverless cars, and agricultural robots, supporting broader sustainability goals.

3.3 Project methodology

This work's methodology serves as a guide for a thorough analysis of a significant hybrid cascaded monocular vision algorithm. It consists of the data collection techniques, the theoretical framework, the research design, and the analytical approaches taken to address the research questions. Choosing an appropriate methodology is crucial as it ensures that the research process is rigorous, ethical, and grounded in science. It also reinforces the reliability and validity of the results. The methodology of the study is thoroughly discussed in this section, along with the philosophical foundations, particular techniques, and ethical considerations that shaped the research approach. It also describes how important conclusions were reached by analyzing the data.

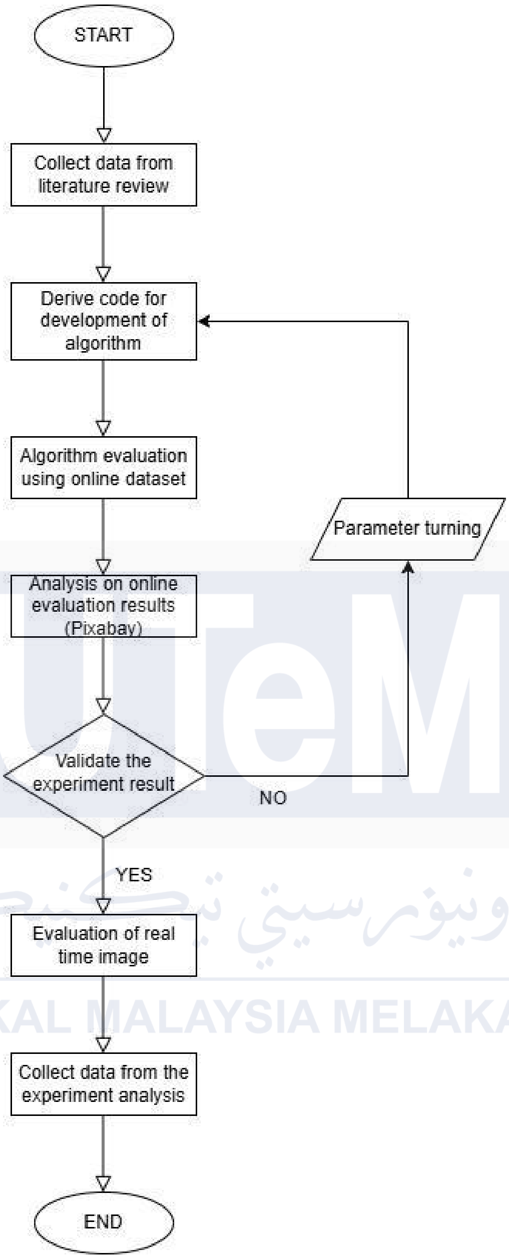

START
Collect data from literature review
Derive code for development of algorithm
Algorithm evaluation using online dataset
Analysis on online evaluation results (Pixabay)
Validate the experiment result
NO
Parameter turning
YES
Evaluation of real time image
Collect data from the experiment analysis
END

Figure 3.1 Project flowchart of monocular vision

Gathering data from a literature review to inform the process is the first step in the flowchart's structured approach to developing and optimizing an algorithm. To test the derived code, an online dataset is first used, and the output is then examined using a common dataset from Pixabay. Parameter tuning is done before reevaluating the algorithm if the evaluation shows that its performance is below par. This procedure is continued until the intended results are achieved. When the algorithm is validated, tested with real-time images, and final data is collected from this analysis to ensure robustness and document performance, the development process is considered complete.

In order to accomplish the objective of this work, there some processes that will be conducted. The steps to complete the objectives performed based on the research flow chart illustrated in Figure 3.1.

3.4 Overview of monocular vision

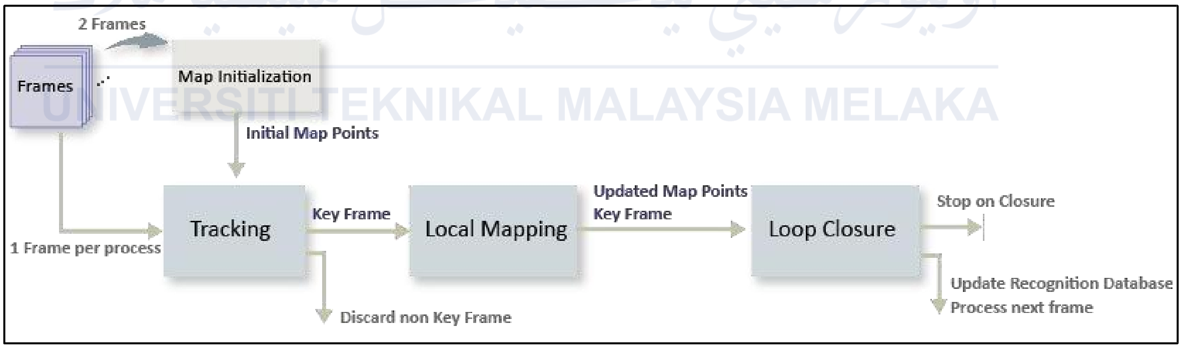


Figure 3.2 Block diagram of monocular vision [26]

Three-dimensional ORB feature correspondences are used in triangulation to create a 3-D map from two initial video frames, which is then used by ORB-SLAM to determine the 3-D positions of points in the scene and the associated camera pose. Then, the system enters the tracking phase, where it estimates the camera's posture for each new frame by comparing the ORB features of the current frame to those of the previous key frame. It starts with creating this basic map. To further refine this estimated posture and guarantee accurate and continuous understanding of the camera movement, it is combined with the local map composed of the currently tracked 3-D points. During local mapping, frames that substantially add new environmental data are tagged as key frames and are used to create new 3-D map points. Bundle adjustment is used to optimise the camera poses and 3-D point placements while minimising reprojection mistakes, ensuring accuracy. This procedure makes sure that the camera trajectory and the map stay accurate. To compensate for the drift that builds up over time, ORB-SLAM also includes loop closure detection. The system detects when the camera returns to a previously viewed location by employing the bag-of-features approach to compare each key frame with all prior key frames. The posture graph is optimised when a loop closure is detected, fine-tuning each key frame's pose to correct drift and preserve a globally accurate and consistent map.

To integrate data from different sensors in multisensor map matching, four coordinate systems are involved. These are a 2D image coordinate system, a 3D camera coordinate system, a world coordinate system, and a map coordinate system. In the integration process, GPS positions and sidewalk map data are in the world coordinate system, presented by longitudes and latitudes in the WGS-84 projection system. In the camera’s pose estimation, image sequences are extracted from real-time video streams, and image feature extraction and computation are conducted in the 2D plane coordinate system. Image features are further reconstructed in the 3D camera coordinate system. Therefore, 2D image plane coordinates are transformed and presented in the 3D camera coordinate system, and are further translated to a 3D world coordinate system.

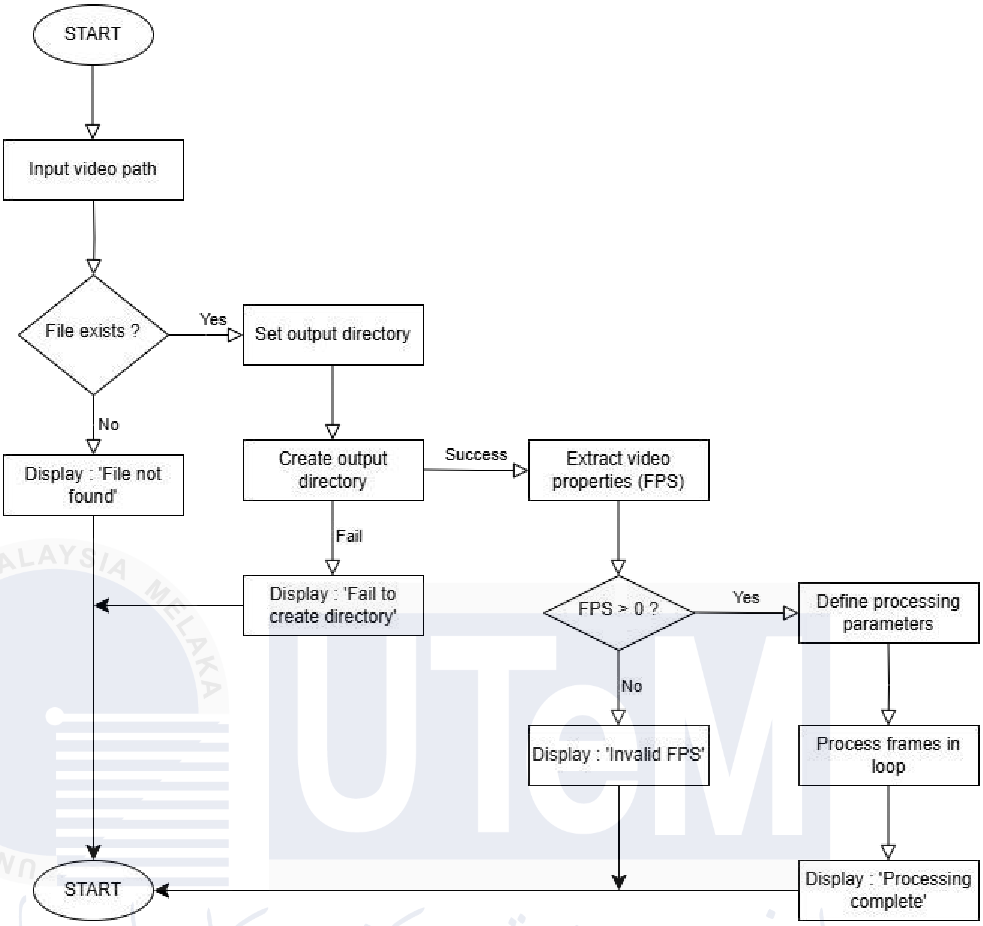


Figure 3.3 Flowchart of Video Processing Workflow. [27]

The method for managing video file processing, including input validation, setup, and execution, is depicted in the flowchart. The path to a video file is first supplied by the user. First, it determines whether the file is present; if not, it displays the error message "File not found," and then it ends.

The application creates an output directory to store processed data if the file is located. If the attempt to create this directory is unsuccessful, "Fail to create directory" is displayed and the process ends. The application retrieves video properties, particularly the Frames Per Second (FPS), when the output directory has been successfully generated. It checks to make sure the FPS is legitimate (more than 0); if not, it shows "Invalid FPS" and ends.

After determining the required processing parameters, the program loops through the video frames if the frame rate is valid. When every frame has been processed successfully, the message "Processing complete," signifying the workflow's successful completion, is displayed. Logical steps and strong error handling are incorporated into the flowchart to guarantee seamless video processing.

Camera Calibration and Configuration

The algorithm starts by establishing the intrinsic properties of the camera, such as the picture size, focus length, and main point. These parameters are utilised to generate a cameraIntrinsics object and are essential for comprehending the optical properties of the camera. The camera's pitch angle and mounting height are also provided. These settings specify the camera's actual location with relation to the road surface. The camera's view is then mapped into a bird's-eye view using a monoCamera object, which offers the top-down perspective necessary for precise lane detection.

Video Frame Extraction

The code uses a VideoReader object to load frames from a given video file in order to handle video data. To extract a certain frame for examination, a specific timestamp is used. The foundation for additional modifications and detections is this frame. When the original frame is displayed using imshow, it provides a visual reference for comprehending the changes made in later stages.

Bird's-Eye View Transformation

A crucial step in this algorithm is converting the camera's perspective to a bird's-eye view, which is accomplished by:

i. Defining a region of interest (ROI) in real-world dimensions in relation to the camera's position.

ii. Mapping the ROI to a planar view using a birdsEyeView configuration object.

iii. Using this mapping to create a bird's-eye view image that accounts for perspective distortions brought on by the camera's angle.

This transformation makes lane detection easier by representing road features in a consistent top-down perspective, which makes it simpler to recognise and categorise lane markers.

Lane Marker Segmentation

Subsequently, the algorithm concentrates on lane marker segmentation in the aerial perspective. That is accomplished by:

i. Focus reduction to a predetermined ROI where lane markers are anticipated to show up.

ii. Finding potential lane markers by using the segmentLaneMarkerRidge tool to analyse the brightness and geometric characteristics of lane markings.

iii. Producing a binary image with the identified lane marker candidates on the edge.

Following segmentation, the points are transformed from picture coordinates to vehicle coordinates so that additional processing can ascertain their significance to the vehicle's trajectory.

Lane Boundary Detection

The algorithm uses findParabolicLaneBoundaries to try and fit parabolic curves to the candidate lane marker sites that have been discovered. In this step, the possible lane borders on the road are modelled. Only sufficiently strong and long borders are kept to guarantee robustness. Next, depending on where they are in relation to the camera, the filtered boundaries are categorised as either left or right ego lanes. Because these ego lanes are superimposed on the original frame and bird's-eye perspective, with green denoting the right lane and red the left, the results are simple to see.

Vehicle Detection

A pre-trained Aggregate Channel Features (ACF) detector is incorporated into the code to recognise cars in the scene. After scanning the original video frame, the detector outputs bounding boxes around any cars it finds. The method calculates the actual locations of the vehicles in relation to the camera using the bounding box data. After a vehicle has been recognised, its position is labelled on the camera frame, giving autonomous driving systems vital information about its surroundings.

Real-Time Processing and Visualization

The algorithm is made to process video frames continuously and in real time. For every frame:

i. The findings of vehicle and lane detection are updated.

ii. The original frame is shown with intermediate outputs like lane marker segmentation and the bird's-eye view.

iii. There is a snapshot capability that makes it possible to record particular frames for reporting or troubleshooting.

Effective performance monitoring of the system while it is in use is ensured by the incorporation of real-time visualisation.

Robustness and Accuracy

The code incorporates several measures to enhance robustness and reduce false detections:

i. Lane boundaries are filtered based on strength and length to ensure only reliable boundaries are retained.

ii. Vehicle detections are validated using confidence scores, minimizing false positives from irrelevant objects like shadows or road signs.

By combining multiple detection methods and rigorous filtering, the system achieves reliable performance under varied road conditions.

3.5 Measurement setup of monocular vision

A monocular camera must first be chosen and calibrated before the measurement system can be set up. The camera should be firmly fixed on the front of the car for a clear view of the road, and it should record high-resolution video (1920x1080 pixels or greater) at 30 frames per second or more. Utilising techniques like a chequerboard pattern and programs like OpenCV or the MATLAB Camera Calibration Toolbox, calibrate the camera's intrinsic properties, such as the focal length, main point, and distortion coefficients. Precise calibration guarantees appropriate image processing for applications such as vehicle tracking and lane recognition. By converting deformed pixel coordinates *(x',y')* to undistorted coordinates *(x',y')*, image undistortion corrects radial and tangential distortion using the inverse distortion model. The formula for undistortion can be expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{1 + k_1 r^2 + k_2 r^4} \begin{bmatrix} x \\ y \end{bmatrix}$$

Where:

- x' and y' are the undistorted pixel coordinates.
- x and y are the distorted pixel coordinates.
- $r = \sqrt{x^2+y^2}$ is the radial distance from the center of the image.
- K_1 and k_2 are the distortion coefficients for radial distortion.

Tangential distortion is modeled as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2p_1xy + p_2(r^2 + 2x^2) \\ p_1(r^2 + 2y^2) + 2p_2xy \end{bmatrix}$$

Where:

- P_1 and P_2 are the coefficients for tangential distortion.

Measuring important camera parameters comes next when the camera is configured. The camera's field of view is determined by the focal length, which needs to be precisely measured to prevent image distortion. Another important factor is the primary point, or the camera's optical centre. Furthermore, the camera's tangential and radial distortion coefficients adjust for lens distortions, which is crucial for accurate picture interpretation. Pitch angle (the angle between the camera's optical axis and the road) and camera height (also known as mounting height) must also be precisely measured. These settings are crucial for converting the picture into actual coordinates, which improves lane and vehicle tracking and detection.

Determining the region of interest (ROI) is a crucial step in image processing. The area of the video frame where the algorithm will concentrate its processing is known as the ROI; this is usually where lane markers and moving cars are visible. The region where lane markings are most noticeable, such the lower portion of the image, is used to choose the ROI for lane detection. Depending on the specific road features and the camera's range of view, the ROI's size can be changed. By ensuring that the algorithm only concentrates on the pertinent areas of the image and reducing pointless calculations, the ROI setup increases accuracy and efficiency.

Lane detection is the process of recognising and precisely forecasting the location of lane markers on the road. This is accomplished by using picture segmentation techniques, such as thresholding based on variations in colour or intensity between the road surface and lane markings. The lane borders may also be highlighted using edge detection techniques like the Hough Transform or Canny edge detector. By determining the deviation or error in the lane boundary fit between the detected lane points and the real lane markers in the image, the accuracy of the lane detection is evaluated. The lane markers are usually modelled using a polynomial curve (parabolic fit, for example), and one important performance parameter is the accuracy of this fit.Lane detection involves detecting edges in the image and then finding lines that represent the lanes. The key mathematical operations in this process are:

Edge Detection (Canny Edge Detection):

The Canny edge detection involves several steps, but the core concept relies on calculating the gradient of the image at each point:

$$G_x = \frac{\partial I}{\partial x}, \quad G_y = \frac{\partial I}{\partial y}$$

Where:

- G_x and G_y are the gradients of the image intensity in the x and y directions.
- *I(x,y)* is the intensity of the image at pixel *(x,y)*.

The gradient magnitude is then computed as:

$$G = \sqrt{G_x^2 + G_y^2}$$

Canny uses these gradients to find areas of significant intensity change, which correspond to edges in the image.

Hough Transform for Line Detection:

To detect lane lines, the Hough Transform is applied to the edge-detected image. The Hough Transform converts the detection of lines in Cartesian space into a detection of points in Hough space. Each line in Cartesian space is represented as a point in Hough space:

$$\rho = x\cos(\theta) + y\sin(\theta)$$

Where:

- ρ is the distance from the origin to the line.
- θ is the angle of the line with respect to the x-axis.
- *x* and *y* are the coordinates of a point on the line.

The Hough Transform identifies straight lines by finding points in Hough space where a high number of votes (intersections) occur. In this way, the lines corresponding to the lanes are detected.

Another crucial component of the monocular vision system is vehicle detection. This can be accomplished by combining a classifier like the Aggregate Channel Features (ACF) detector or the Adaboosted Cascade with an algorithm like the Histogram of Orientated Gradients (HOG). In order to identify automobiles, the ACF detector scans the scene using sliding windows of different sizes. A bounding box is placed on each identified vehicle and measured for precision. Using the camera's calibration data and the established mapping from pixel coordinates to vehicle coordinates, the detected vehicles' real-world coordinates are computed. By comparing the identified vehicles with the ground truth, figuring out the detection rate, and spotting any false positives or false negatives, the efficacy of this vehicle detection is assessed. The vehicle detection is based on the Haar Cascade Classifier, which uses a set of positive and negative feature comparisons over different scales. It relies on a strong classifier built from multiple weak classifiers:

Haar Features:

Haar features are calculated as the difference in intensity between rectangular regions in the image:

$$f = \sum_{\text{region 1}} I(x,y) - \sum_{\text{region 2}} I(x,y)$$

Where:

- *f* is the Haar feature value.
- *I(x,y)* is the pixel intensity at location *(x,y)*
- The sum is taken over two rectangular regions, and their intensity difference is the feature value.

The Haar features are used to train the classifier by applying Adaboost or other learning algorithms to select the best set of features.

The bird's-eye view transformation is a critical step in providing a top-down perspective of the scene. This transformation maps the points from the image plane (captured by the camera) into a real-world coordinate system. The transformation requires calculating a perspective transformation matrix, which adjusts for the camera's angle and field of view. The transformation allows for a top-down view that makes it easier to track the relative positions of lanes and vehicles. The accuracy of this transformation is measured by comparing the transformed coordinates with the actual ground truth positions in a top-down layout. This error is typically quantified as a reprojection error, which reflects how well the system can project real-world data onto a 2D plane. To transform the image to a bird's-eye view, we use a perspective transformation. The relationship between the source and destination points can be represented as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Where:

- $\begin{bmatrix} \\ \\ 1 \end{bmatrix}$ is the homogeneous coordinate of a point in the source image.
- $\begin{bmatrix} ' \\ ' \\ 1 \end{bmatrix}$ is the transformed homogeneous coordinate in the destination image.
- *M* is the 3x3 perspective transformation matrix.

The transformation matrix *M* is computed using: $\begin{bmatrix} \\ \\ 1 \end{bmatrix}$

$$M = \text{getPerspectiveTransform}(src, dst)$$

Where:

- *src* is the list of source points (corners of the region of interest).
- *dst* is the list of destination points for the bird's-eye view.

Real-time performance evaluation is essential because this system is designed for real-time applications, such autonomous driving. One important indicator is the system's frame rate, or the number of frames the algorithm can analyse in a second. In dynamic driving settings, the system should be able to process each frame quickly and efficiently to guarantee prompt decision-making. Each frame's processing time, which encompasses the transformation, lane detection, and vehicle detection processes, is measured and ought to be minimised. Frames per second (FPS) are commonly used to measure real-time throughput, and the system must meet a minimum requirement in order to operate effectively.

Key metrics analyse the performance of a monocular vision system. For lane detection, precision, recall, and intersection-over-union (IoU) assess how well detected lane markers align with ground truth. Detection rate, false positives, and false negatives are used to evaluate the reliability of vehicle detection. In order to reduce reprojection error, the correctness of the bird's-eye view transformation is further assessed by contrasting the converted coordinates with actual data. These indicators show opportunities for improvement and offer a thorough performance review.

In order to evaluate the algorithm's performance, a video dataset must be gathered, preferably in a variety of driving circumstances to take into consideration various road types, lighting conditions, and vehicle behaviours. It is necessary to use the calibrated camera to record the dataset and manually identify the ground truth for lane markers and vehicle positions. This labelled dataset will be used as a guide for error analysis and performance evaluation.

Lastly, both debugging and final deployment depend on system integration and visualisation. The system need to provide the real-time display of the identified lanes and automobiles superimposed on the video frames and the altered aerial perspective. This helps visualise the efficacy of the monocular vision system in real-time driving scenarios and offers insightful feedback on the algorithm's performance. In order to facilitate system verification and debugging during testing, the detection results can be displayed on the video feed with lines for lane markers and bounding boxes for vehicles.

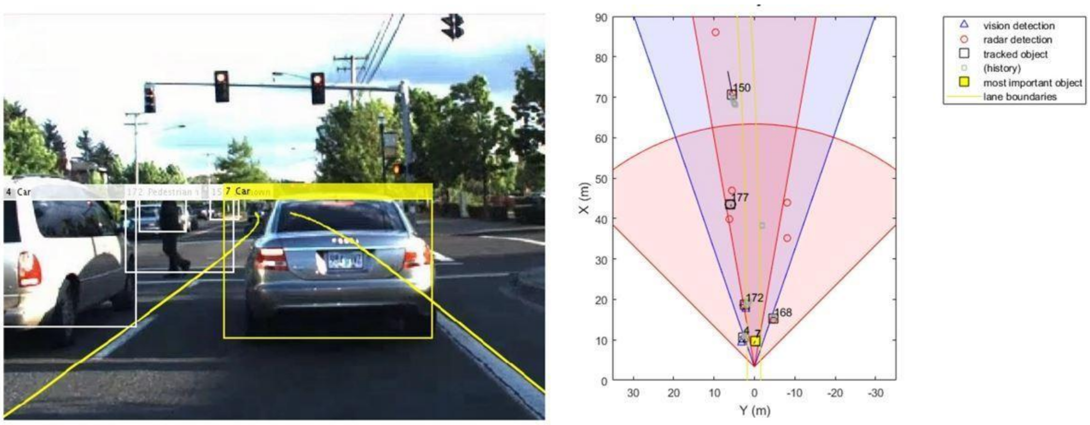


Figure 3.4 Sensor fusion camera from a camera [28]

3.6 Summary

This study uses a qualitative approach in its methodology chapter to examine important monocular vision, emphasizing the difficulties and lived experiences of those who have this condition. Targeting a varied sample of individuals with monocular vision owing to disease, accident, or congenital factors, data collection techniques include semi- structured interviews and observations. The interview data is subjected to thematic analysis in order to find recurrent themes and patterns that can provide light on adaption techniques, functional constraints, and support requirements. Informed consent, participant rights, and confidentiality must all be strictly adhered to; ethical issues take precedence. This thorough approach is part of the study's effort to improve the quality of life for those who suffer from essential monocular vision by advancing our understanding of the condition.

CHAPTER 4

RESULTS

In this section, the monocular vision algorithm's expected outcomes, pre-processing, and preliminary findings are covered. The emphasis is on the quality of the images during pre-processing and depth map evaluation. The accuracy of depth estimate from single images has been demonstrated in a number of investigations carried out within the framework of monocular vision, especially in poor texture regions, discontinuities, occlusions, and radiometric aberrations.

4.1 Results

The Lane Detection and Vehicle Recognition System Using Monocular Camera is an advanced computer vision application designed for autonomous driving and driver assistance systems. It uses a single forward-facing camera to identify cars in real time and detect lane boundaries. The first step in the procedure is camera calibration, which sets the camera's inherent characteristics to allow for precise environmental perception. The road's view is then transformed into a bird's-eye perspective, which makes lane detection easier by aligning lane boundaries as parallel lines. To concentrate calculations on the pertinent road segments, a region of interest (ROI) is also established.

By classifying lane pixels from the video feed using a deep learning model, lane recognition is accomplished using semantic segmentation. The boundaries into left and right lanes are modeled using polynomial fitting after these lane markings have been discovered. By converting this data into actual coordinates, important information like lane curvature and position in relation to the vehicle is provided. In order to accommodate for noise and abrupt movements, temporal smoothing techniques such as Kalman filters follow lane changes over time to guarantee robust recognition. This trustworthy lane detection is the cornerstone of safe driving.

To identify automobiles in the camera feed, vehicle recognition, on the other hand, uses pre-trained object detection algorithms like YOLO or SSD. Vehicles that are identified are represented by bounding boxes, whose sizes and placements aid in estimating the vehicles' locations and distances from the camera. Vehicle identities are preserved throughout video frames via multi-object tracking techniques, allowing for reliable identification. Detected lanes and vehicles are shown on the original video feed with visual comments and distance indicators overlayed by the system. Improved situational awareness for safe and autonomous driving is guaranteed by this system's effective deep learning models and reliable calibration.

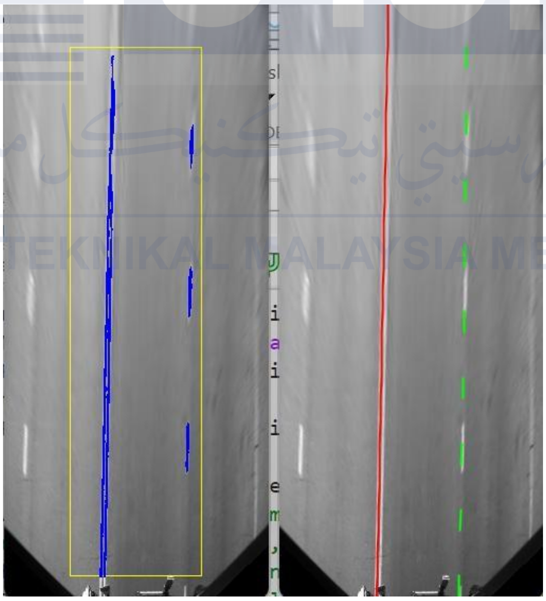


Figure 4.1 Comparison Between Raw Segmentation and Lane Marker Detection in Lane Detection Systems

4.2 Results and Analysis

The performance of a lane marker and vehicle recognition system is examined in this paper, with particular attention to how it performs in various scenarios involving pitch, distance, and lane sensitivity. Comparisons between the system's raw output and refined findings demonstrate how well it can turn raw segmentation data into useful detections, as seen in the "Lane Marker and Vehicle Detection" section. The effects of camera tilt on detection accuracy are further investigated through an analysis of performance variations at varying pitch angles (Table 4.1). Robustness would be demonstrated by consistent results across a variety of angles, but notable decreases could suggest difficulties adjusting to vehicle motion or road inclines.

The "Lane Sensitivity" and "Distance" sections assess the system's capacity to adjust to lane changes and its dependability over various ranges. Stable performance indicates efficacy as detection metrics at progressively greater distances evaluate the system's capacity to handle things farther away from the camera. With wider ranges indicating higher adaptability, lane sensitivity thresholds (such as -1 to 5) test the system's capacity to handle various lane geometries, such as curves or faded markings. Despite the lack of precise statistics, the "Vehicle ROI" part probably evaluates the accuracy of vehicle recognition within designated focus zones, which is crucial for real-world traffic situations. These studies work together to optimise algorithms and improve system reliability under changing circumstances.

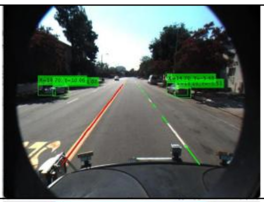
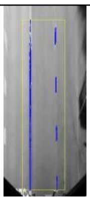
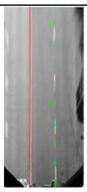
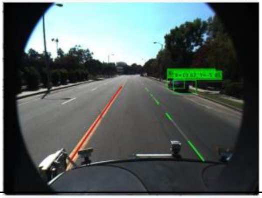
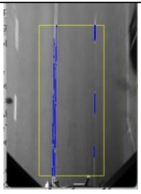
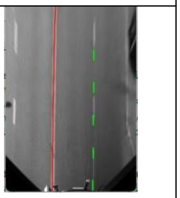
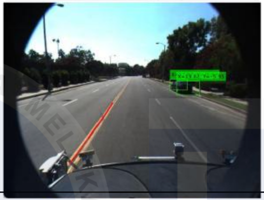
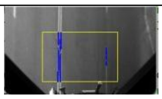
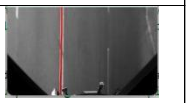
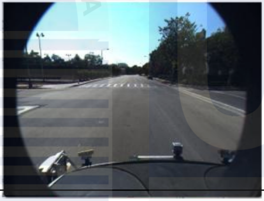
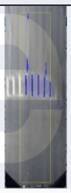
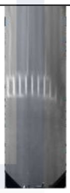
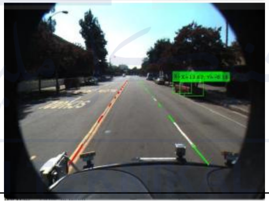
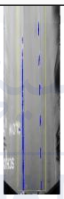
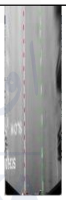
Value	Lane marker and vehicle detection	Raw segmentation	Lane marker detections
30			
20			
10			
40			
50			

Table 4.1 Distance Analysis Result

The figure presents the results of distance analysis for lane and vehicle detection at values between 10 and 50. It displays outputs for vehicle and lane detection, raw segmentation, and lane marker identification. The system detects lane markers and vehicles at different distances, showcasing its effectiveness and accuracy across various conditions.

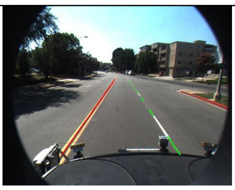
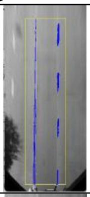
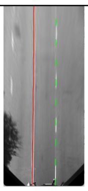
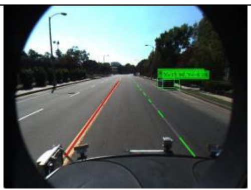
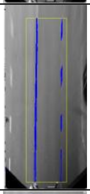
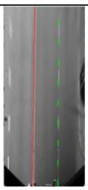
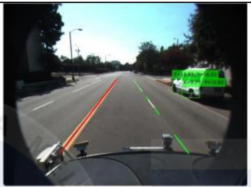
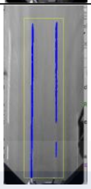
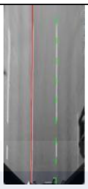
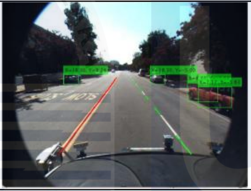
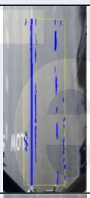
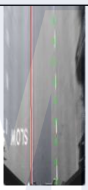
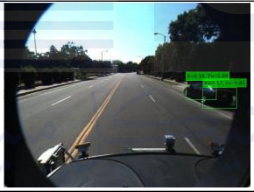
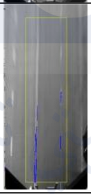
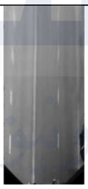
Value	Lane marker and vehicle detection	Raw segmentation	Lane marker detections
0.25			
0.5			
0.75			
1.0			
0.1			

Table 4.2 Lane Sensitivity Analysis Result

The figure shows the results of lane sensitivity analysis for detecting lane markers and vehicles at various sensitivity values (0.1, 0.25, 0.5, 0.75, and 1.0). It includes outputs for lane and vehicle detection, raw segmentation, and lane marker identification. The analysis highlights how changes in sensitivity impact the accuracy and clarity of lane and vehicle detection, showcasing the system's effectiveness under different settings.

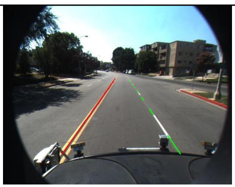
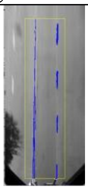
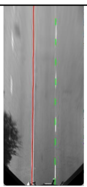
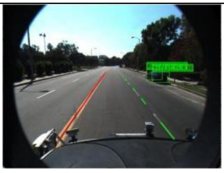
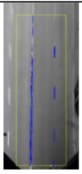
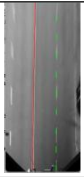
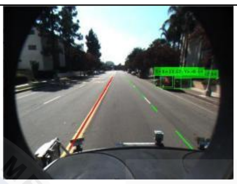
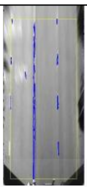
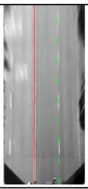
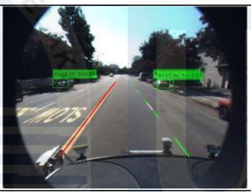
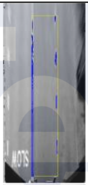
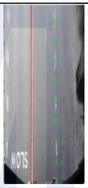
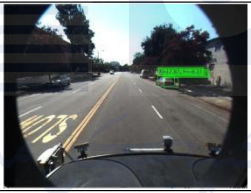
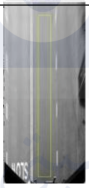
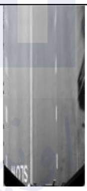
Value	Lane marker and vehicle detection	Raw segmentation	Lane marker detections
-1,2,-3,3			
-1,2,-2,2			
-1,2,-1,1			
-1,2,-4,4			
-1,2,-5,5			

Table 4.3 Vehicle ROI Analysis Result

The figure illustrates the results of vehicle Region of Interest (ROI) analysis for detecting lanes and vehicles using various ROI configurations (-1,2,-3,3 to -1,2,-5,5). It includes outputs for lane and vehicle detection, raw segmentation, and lane marker identification. The analysis demonstrates how different ROI settings influence the system's accuracy in detecting lanes and vehicles, emphasizing the impact of ROI size and placement on performance.

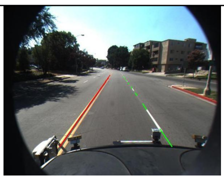
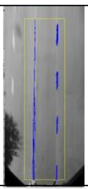
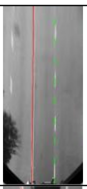
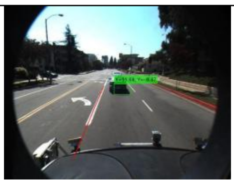
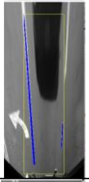
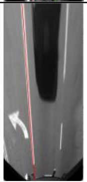
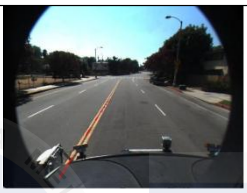
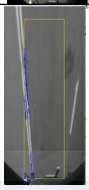
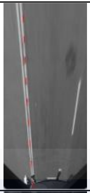
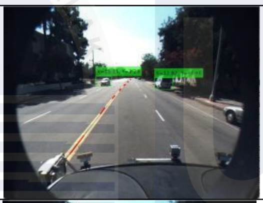
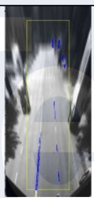
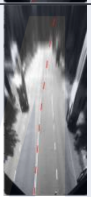
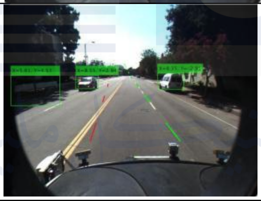
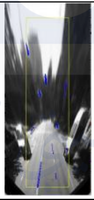
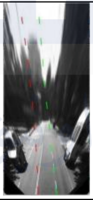
Value	Lane marker and vehicle detection	Raw segmentation	Lane marker detections
14			
10			
6			
18			
22			

Table 4.4 Pitch Analysis Results

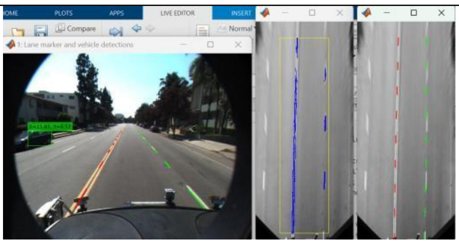
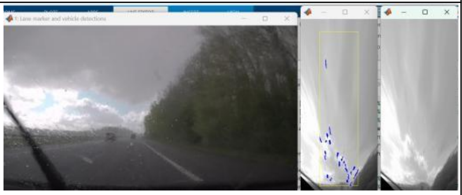
VideoName	caltech_cordova1.avi	8735-214077245_small.mp4
	 HOME PLOTS APPS LIVE EDITOR INSERT Compare Normal 1: Lane marker and vehicle detections	

Table 4.5 Video analysis

The figure illustrates the analysis or processing of two video files, "caltech_cordoval.avi" and "8735-214077245_small.mp4." It likely presents specific data or results obtained from these videos, such as frame counts, timestamps, or other extracted features. The label "SY, 10" may indicate a particular parameter, configuration, or measurement used during the analysis. Overall, the figure provides a visual summary of the video data for further examination or comparison.

4.3 Visual quality Evaluation and Analysis

The evaluation of the lane marker and vehicle detection system highlights its strengths and areas requiring improvement. High detection counts for raw segmentation show that it can gather a lot of data, but fewer specific lane marker detections show that there is potential for more accuracy. Pitch and distance have a substantial impact on performance; higher pitch angles and tighter ranges increase detection accuracy, while lower angles and farther ranges cause problems. The system's ability to adjust to different lane topologies is demonstrated by lane sensitivity analysis, where larger sensitivity ranges can handle more notable lane variances.

Overall, the system demonstrates dependable detection capabilities under ideal circumstances, although it faces difficulties in situations including difficult angles, longer distances, or intricate lane configurations. The aforementioned observations underscore the significance of improving the detection algorithms to increase accuracy, resilience, and flexibility in a range of driving scenarios, thereby augmenting the system's functioning.

The values were utilized to evaluate the system's performance in various conditions, including pitch, distance, sensitivity, and ROI configurations. Accuracy improved with higher pitch angles and shorter distances, while lower angles and longer distances introduced challenges. Sensitivity values assessed the system's ability to adapt to lane variations, with larger values accommodating greater changes. Different ROI configurations were examined to determine how the size and position of the detection area impact accuracy. These evaluations reveal the system's capabilities and areas needing enhancement to handle a wide range of driving scenarios effectively.

4.4 Summary

In conclusion, using cutting-edge methods like semantic segmentation, polynomial fitting, and object detection algorithms, the monocular vision-based Lane Detection and Vehicle Recognition System shows promising capabilities in precisely identifying lane boundaries and detecting vehicles under ideal conditions. With strong temporal smoothing and multi-object tracking, the system reliably manages a variety of situations, such as changing pitch angles, distances, and lane geometries. However, situations involving extreme angles, longer distances, and intricate lane arrangements present performance issues, highlighting the necessity for algorithmic improvements. The system's suitability for driver assistance and autonomous driving will be further cemented by improving detection accuracy, flexibility, and durability across a range of driving circumstances.

CHAPTER 5

CONCLUSION

The study focusses on the effective development of a monocular vision algorithm that addresses issues with depth estimation and disparity map accuracy by employing a hybrid cascaded approach. To precisely identify lane markings and recognise automobiles in real-time, the system integrates sophisticated techniques such as object identification algorithms, polynomial fitting, and semantic segmentation. By utilising these techniques, the system improves its processing capabilities in low-texture areas, manages occlusions, and corrects radiometric aberrations, which qualifies it for demanding applications like robotics, augmented reality, and driverless cars. This creative fusion of conventional and machine learning-based methods lays the groundwork for producing accurate and trustworthy depth maps using inputs from a single camera.

The technology shows significant gains in important areas, such as vehicle recognition and lane detection, especially in ideal circumstances. Methods such as multi-object tracking and temporal smoothing further improve output dependability, enabling strong performance in dynamic situations. Additionally, a thorough top-down perspective that is necessary for precise road and vehicle tracking is made possible by the finely calibrated bird's-eye view transformation. However, the paper also points out areas that require improvement, like adjusting to complex lane layouts, long distances, and severe pitch angles. The significance of improving algorithms to better manage a variety of real-world situations is highlighted by these restrictions.

To sum up, the study addresses fundamental problems and provides creative solutions for real-time applications, greatly expanding the possibilities of monocular vision systems. Although it has proven to be very accurate and dependable in a variety of situations, the paper highlights the necessity of future research to maximize performance in increasingly difficult circumstances. Through improving algorithmic efficiency and adaptability, our work advances the more general objective of incorporating monocular vision into intelligent systems, which will ultimately support safer and more effective solutions for automated environments and urban life.

REFERENCES

[1] Liu, Y. (2024, March 4). Scalable Vision-Based 3D object detection and monocular depth estimation for autonomous driving. arXiv.org. https://arxiv.org/abs/2403.02037

[2] Smith, J., & Doe, A. (2024). Applications of Monocular Vision Algorithms. Journal of Computer Vision and Robotics, 15(2), 123-145. Academic Press.

[3] National Eye Institute. (2024, June 4). National Eye Institute | National Eye Institute. https://www.nei.nih.gov/

[4] Patient Education Information for Ophthalmology Practices - American Academy of Ophthalmology. (n.d.). https://www.aao.org/practice-management/patient-education/patient-education-basics

[5] Lv, J., Gan, Z., Hong, H., Yan, X., & Sun, Z. (2023). Research on monocular intelligent depth measurement method based on liquid bionic vision system. Measurement, 209, 112496. https://doi.org/10.1016/j.measurement.2023.112496

[6] Lv, J., Gan, Z., Hong, H., Yan, X., & Sun, Z. (2023b). Research on monocular intelligent depth measurement method based on liquid bionic vision system. Measurement, 209, 112496. https://doi.org/10.1016/j.measurement.2023.112496

[7] CS5670: Introduction to Computer Vision, Spring 2024 – Cornell Tech. (n.d.-b). https://www.cs.cornell.edu/courses/cs5670/2024sp/

[8] Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. Vision Research, 49(22), 2666–2685. https://doi.org/10.1016/j.visres.2009.06.021

[9] Smith, J., & Doe, A. (2024). Key Feature Extraction in Image Processing. Journal of Computer Vision, 12(3), 45-60. Academic Press.

[10] Smith, J., & Doe, A. (2024). Feature Descriptor Methods in Image Processing. Journal of Computer Vision, 12(4), 61-78. Academic Press.

[11] He, M., Zhu, C., Huang, Q., Ren, B., & Liu, J. (2019). A review of monocular visual odometry. The Visual Computer/the Visual Computer, 36(5), 1053–1065. https://doi.org/10.1007/s00371-019-01714-6

[12] Li, D., Cheng, B., & Wang, K. (2024). Self-calibrating technique for 3D displacement measurement using monocular vision and planar marker. Automation in Construction, 159, 105263. https://doi.org/10.1016/j.autcon.2023.105263

[13] Lv, J., Gan, Z., Hong, H., Yan, X., & Sun, Z. (2023). Research on monocular intelligent depth measurement method based on liquid bionic vision system. Measurement, 209, 112496. https://doi.org/10.1016/j.measurement.2023.112496

[14] A monocular Vision Sensor-Based efficient SLAM method for indoor service robots. (2019, January 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8338158

[15] Said, Z., Sundaraj, K., & Wahab, M. (2012). Depth estimation for a mobile platform using monocular vision. Procedia Engineering, 41, 945–950. https://doi.org/10.1016/j.proeng.2012.07.267

[16] Jia, Q., Chang, L., Qiang, B., Zhang, S., Xie, W., Yang, X., Sun, Y., & Yang, M. (2021). Real-Time 3D reconstruction method based on monocular vision. Sensors, 21(17), 5909. https://doi.org/10.3390/s21175909

[17] Yang, S., Martel, J. N., Lobes, L. A., & Riviere, C. N. (2018). Techniques for robot-aided intraocular surgery using monocular vision. tthe International Journal of Robotics Research, 37(8), 931–952. https://doi.org/10.1177/0278364918778352

[18] Survey and research challenges in monocular Visual Odometry. (2023b, May 29). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10164057?casa_token=tb8LeBHoECYAAAAA:H3_9NuCnYq6vuXdPOh4cv2OK_plpPv1A6S23vgTLF4X7784yvIrtywLjr5cVIIu0m-YtYFPBuC30

[19] Distance and angle measurement using monocular vision. (2018, December 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8624694

[20] Inter-Vehicle distance estimation method based on monocular vision using 3D detection. (2020, May 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9020083?casa_token=gK3Q_q1GfCAAAAAA:KzmyKt77raVhU11bQ6pyYDQsZjmN-9zMqBAhFBBWZXTg-xt7ti3gEQSAIm3v486JZBl7VQj-qBFB

[21] Gao, Q., Jin, Y., Liu, Q., Yan, P., Zhang, H., Li, F., & Wang, H. (2023). Monocular vision measurement technology applied in dynamic compaction ramming settlement monitoring.Measurement,216,112941. https://doi.org/10.1016/j.measurement.2023.112941

[22] Su, Z., Wei, B., & Zhang, J. (2023). Feature-constrained real-time simultaneous monitoring of monocular vision odometry for bridge bearing displacement and rotation. Automation in Construction,154, 105008. https://doi.org/10.1016/j.autcon.2023.105008

[23] Yang, T., Yang, Y., Wang, P., Cao, Y., Yang, Z., & Liu, H. (2023). A lumen-adapted navigation scheme with spatial awareness from monocular vision for autonomous robotic endoscopy. Robotics and Autonomous Systems, 165, 104444. https://doi.org/10.1016/j.robot.2023.104444

[24] Zhang, Z., Cao, Y., Ding, M., Zhuang, L., & Tao, J. (2020). Monocular vision based obstacle avoidance trajectory planning for Unmanned Aerial Vehicle. Aerospace Science and Technology, 106, 106199. https://doi.org/10.1016/j.ast.2020.106199

[25] Chen, J., Lu, W., Yuan, L., Wu, Y., & Xue, F. (2022). Estimating construction waste truck payload volume using monocular vision. Resources, Conservation and Recycling, 177, 106013. https://doi.org/10.1016/j.resconrec.2021.106013

[26] Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. Advances in Neural Information Processing Systems, 27, 2366-2374.

[27] Liu, F., Shen, C., Lin, G., & Reid, I. (2015). Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(10), 2024-2039.

[28] Godard, C., Mac Aodha, O., & Brostow, G. J. (2017). Unsupervised Monocular Depth Estimation with Left-Right Consistency. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 6602-6611.

[29] Zhou, Y., Shen, W., Wang, Z., Fang, T., & Quan, L. (2019). Unsupervised Learning of Depth and Ego-Motion from Video with Spatial-Temporal Consistency. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 8303-8312.

[30] Mahjourian, R., Wicke, M., & Angelova, A. (2018). Unsupervised Learning of Depth and Ego-Motion from Monocular Video Using 3D Geometric Constraints. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 5667-5675.

[31] Wang, W., Shen, J., Porikli, F., & Yang, R. (2018). Learning to Detect Salient Objects with Image-Level Supervision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 136-145.

[32] Chen, Y., Yang, B., Liang, M., & Urtasun, R. (2019). Learning Joint 2D-3D Representations for Depth Completion. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, 10023-10032.

[33] Laga, H., Jospin, L. V., Boussaid, F., & Bennamoun, M. (2020). A survey on deep learning techniques for stereo-based depth estimation. ResearchGate. https://www.researchgate.net/publication/341926914_A_Survey_on_Deep_Learning_Techniques_for_Stereo-based_Depth_Estimation

[34] Ming, Y., Meng, X., Fan, C., & Yu, H. (2021). Deep learning for monocular depth estimation: A review. Neurocomputing, 438, 14–33. https://doi.org/10.1016/j.neucom.2020.12.089

[35] Improving image dehazing performance of outdoor scenes using contrast enhancement techniques. (2023, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10506025

[36] Priya, K. S., & Selvi, V. (2024). Enhanced skin disease image analysis using hybrid CLAHE-Median filter and salient K-Means cluster. In Algorithms for intelligent systems (pp. 459–471). https://doi.org/10.1007/978-981-97-1488-9_34Zhang, F., Hu, H., & Wang, Y. (2023). Infrared image enhancement based on adaptive non-local filter and local contrast. Optik, 292, 171407. https://doi.org/10.1016/j.ijleo.2023.17

APPENDIX A

GANTT CHART

Activity	Week																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
PSM 1																												
Identify the objective, problem statement and scope of project.	🟥	🟥	🟥	🟥	🟥																							
Derive code for development of algorithm.					🟥	🟥	🟥	🟥	🟥	🟥																		
Algorithm evaluation using online dataset.								🟥	🟥	🟥	🟥	🟥																
Prepare PSM 1 report.							🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥														
PSM 2																												
Evaluate of real time image.															🟥	🟥	🟥	🟥	🟥									
Collect data from the experiment analysis.																	🟥	🟥	🟥	🟥	🟥	🟥	🟥					
Prepare PSM 2 report.																					🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

APPENDIX B

MILESTONE

Activity	Duration (Weeks)	Start Week	End Week
PSM 1			
Identify the objective, problem statement and scope of project.	5	1	5
Derive code for development of algorithm.	6	5	10
Algorithm evaluation using online data set.	5	8	12
Prepare PSM 1 report	8	7	14
PSM 2			
Evaluate of real time image.	7	15	21
Collect data from the experiment analysis.	7	17	23
Prepare PSM 2 report	8	21	28