

# IMPROVEMENT OF THE STEREO VISION ALGORITHM USING AN ADAPTIVE GUIDED FILTER FOR DEPTH MAP MEASUREMENT



MARINA BINTI HALIM

UTeM

اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

# **IMPROVEMENT OF THE STEREO VISION ALGORITHM USING AN ADAPTIVE GUIDED FILTER FOR DEPTH MAP MEASUREMENT**

**MARINA BINTI HALIM**



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



**اونيورستى تېكنىكل ماليسيا ملاك**  
**Faculty of Electronics and Computer Technology and Engineering**  
**Universiti Teknikal Malaysia Melaka**  
**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2025**

BORANG PENGESAHAN STATUS LAPORAN  
PROJEK SARJANA MUDA II

Tajuk Projek : Improvement Of The Stereo Vision Algorithm Using An Adaptive Guided  
Filter For Depth Map Measurement  
Sesi Pengajian : 2024/2025

Saya MARINA BINTI HALIM 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:

(TANDATANGAN PENULIS)

(COP DAN TANDATANGAN PENYELIA)

Alamat Tetap: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Tarikh : 24 Januari 2025

Tarikh : 24 Januari 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 “Improvement of the Stereo Vision Algorithm Using An Adaptive Guided Filter For Depth Map Measurement” 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

:

MARINA BINTI HALIM

Date

:

24/1/2025



## 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 with Honours

Signature :  
Supervisor Name : Ir.Ts.Dr. Ahmad Fauzan bin Kadmin  
Date : 24 January 2025

Signature :  
Co-Supervisor :  
Name (if any) : Ts. Mohd Saad bin Hamid  
Date : 24 January 2025

## DEDICATION

*To my beloved mother, Maziah bt Sulaiman,  
and father, Halim bin Muhammad.*



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

## ABSTRACT

Stereo vision techniques play a crucial role in computer vision and photogrammetry by estimating depth information through matching corresponding points in stereo image pairs. Despite its significance in applications such as 3D reconstruction and autonomous vehicles, traditional stereo matching algorithms face challenges in textureless regions, occlusions, and varying illumination conditions, leading to inaccuracies in depth estimation. This study aims to address these challenges by developing a stereo vision system to enhance the accuracy and robustness of stereo matching, particularly in low-texture regions. The project objective includes evaluating the proposed method using benchmark datasets like the Middlebury datasets to demonstrate its effectiveness and superiority in achieving reliable stereo matching results. The methodology involves utilizing MATLAB for depth estimation of 3D images, implementing sustainable stereo vision algorithms prioritizing energy efficiency and resource optimization, and enhancing stereo vision algorithms with adaptive guided filtering to address occlusions, noise, and efficiency for robust depth estimation. The results include the development of a comprehensive stereo matching method that improves accuracy and robustness in challenging imaging conditions, validated through experimental results and discussions showcasing the effectiveness and efficiency of the proposed algorithm.

## ***ABSTRAK***

Teknik penglihatan stereo memainkan peranan penting dalam penglihatan komputer dan fotogrametri dengan menganggar maklumat kedalaman melalui pemadanan titik sepadan dalam pasangan imej stereo. Walaupun kepentingannya dalam aplikasi seperti pembinaan semula 3D dan kenderaan autonomi, algoritma pemadanan stereo tradisional menghadapi cabaran dalam kawasan tanpa tekstur, oklusi dan keadaan pencahayaan yang berbeza-beza, yang membawa kepada ketidaktepatan dalam anggaran mendalam. Kajian ini bertujuan untuk menangani cabaran ini dengan membangunkan sistem penglihatan stereo untuk meningkatkan ketepatan dan keteguhan padanan stereo, terutamanya di kawasan bertekstur rendah. Objektif projek termasuk menilai keberkesanan kaedah yang dicadangkan menggunakan set data penanda aras seperti set data Middlebury untuk menunjukkan keberkesanan dan keunggulannya dalam mencapai hasil padanan stereo yang boleh dipercayai. Metodologi ini melibatkan penggunaan MATLAB untuk anggaran kedalaman imej 3D, melaksanakan algoritma penglihatan stereo yang mampan mengutamakan kecekapan tenaga dan pengoptimuman sumber, dan meningkatkan algoritma penglihatan stereo dengan penapisan berpandu adaptif untuk menangani oklusi, bunyi dan kecekapan untuk anggaran kedalaman yang mantap. Hasil kajian ini termasuk pembangunan kaedah pemadanan stereo komprehensif yang meningkatkan ketepatan dan kestabilan dalam keadaan pengimejan yang mencabar, disahkan melalui keputusan percubaan dan perbincangan yang mempamerkan keberkesanan dan kecekapan algoritma yang dicadangkan.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Ir.Ts.Dr. Ahmad Fauzan bin Kadmin and co-supervisor, En. Mohd saad bin hamid for their precious guidance, words of wisdom and patient throughout this project.

I am also indebted to Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Higher Education Malaysia for the financial support through Fundamental Research Grant Scheme (FRGS) which enables me to accomplish the project. Not forgetting my fellow classmate, 3 BERZ for the willingness of sharing his thoughts and ideas regarding the project.

My highest appreciation goes to my parents, and family members for their love and prayer during the period of my study. An honourable mention also goes to my mom, Maziah binti Sulaiman for all the motivation and understanding.

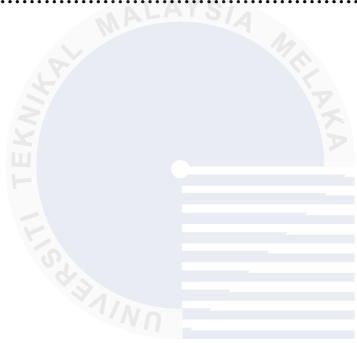
Finally, I would like to thank all my classmate, the faculty members, as well as other individuals who are not listed here for being co-operative and helpful.

<b>TABLE OF CONTENTS</b>	<b>PAGE</b>
<b>MARINA BINTI HALIM .....</b>	<b>2</b>
<b>DECLARATION.....</b>	<b>4</b>
<b>APPROVAL .....</b>	<b>5</b>
<b>DEDICATION .....</b>	<b>6</b>
<b>ABSTRACT.....</b>	<b>1</b>
<b>ABSTRAK.....</b>	<b>2</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>3</b>
<b>TABLE OF CONTENTS .....</b>	<b>4</b>
<b>LIST OF TABLES.....</b>	<b>6</b>
<b>LIST OF FIGURES.....</b>	<b>7</b>
<b>LIST OF SYMBOLS .....</b>	<b>9</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>10</b>
<b>LIST OF APPENDICES .....</b>	<b>11</b>
<b>CHAPTER 1.....</b>	<b>1</b>
<b>1.1 Introduction.....</b>	<b>1</b>
<b>1.2 Background .....</b>	<b>1</b>
<b>1.3 Enhancing Industry and Infrastructure through Stereo Vision in Robotics and Automation.....</b>	<b>3</b>
<b>1.4 Problem Statement .....</b>	<b>3</b>
<b>1.5 Project Objective.....</b>	<b>6</b>
<b>1.6 Scope of Project.....</b>	<b>6</b>
<b>CHAPTER 2.....</b>	<b>8</b>
<b>2.1 Introduction.....</b>	<b>8</b>
<b>2.2 Enhancing Stereo Vision Algorithms with Adaptive Guided Filtering: Addressing Occlusions, Noise, and Efficiency for Robust Depth Estimation.....</b>	<b>8</b>
<b>2.3 Background of stereo vision.....</b>	<b>9</b>
<b>2.4 Stereo vision framework .....</b>	<b>14</b>
<b>2.5 Stereo vision previous method.....</b>	<b>15</b>

2.6	Table comparison of stereo vision .....	38
	Table 2.1 Table comparison of stereo vision .....	39
2.7	Summary.....	44
<b>CHAPTER 3.....</b>		<b>45</b>
3.1	Introduction.....	45
3.2	Implementing Sustainable Stereo Vision Algorithms: Prioritizing Energy Efficiency and Resource Optimization .....	45
3.3	Methodology .....	46
3.4	Block diagram .....	49
3.4.1	Pre-processing .....	51
3.4.2	Matching cost .....	51
3.4.3	Cost aggregation.....	52
3.4.4	Disparity selection.....	52
3.4.4	Disparity refinement.....	53
3.4.5	Disparity Map .....	53
3.5	Measurement setup.....	54
3.6	Summary.....	58
<b>CHAPTER 4.....</b>		<b>59</b>
4.1	Introduction.....	59
4.2	Result .....	59
4.2.1	Stereo matching analysis.....	60
4.3	Summary.....	81
<b>CHAPTER 5.....</b>		<b>83</b>
5.1	Conclusion .....	83
5.2	Future Work .....	84
<b>REFERENCES.....</b>		<b>86</b>
<b>APPENDICES.....</b>		<b>91</b>
Appendices Gant chart.....		91
Appendices milestone.....		92

## LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1	Table comparison of stereo vision .....	39
Table 4.1	The comparison results of nonocc error and all error using Middlebury dataset for $r = 3$ ....	65
Table 4.2	The comparison results of nonocc error and all error using Middlebury dataset for $r = 5$ ....	69
Table 4.3	The comparison results of nonocc error and all error using Middlebury dataset for $r = 7$ ....	73
Table 4.4	The comparison results of nonocc error and all error using Middlebury dataset for $r = 9$ ....	77
Table 4.5	The comparison results of nonocc error and all error using Middlebury dataset for different $r$ .....	81



اونيورسيتي تېكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

## LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 1.1	Constraining Factors in Middlebury Dataset Stereo Images [1] .....	4
Figure 2.1	Triangulation of stereo vision [9] .....	10
Figure 2.2	Disparity estimation, (a) Origin RGB data (b) Ground Truth disparity Pseudo- color data, (c) Disparity prediction by stereonet, (d) Result. Achieve an excellent disparity estimation [11] .....	12
Figure 2.3	Binocular stereo vision [13].....	13
Figure 2.4	The workflow of the proposed method [1] .....	16
Figure 2.5	Initial disparity maps based on different matching costs for Tsukuba. (a) Absolute difference in image gradients, (b) proposed enhanced image gradients-based matching costs, (c) traditional census transform-based matching cost, (d) proposed improved census transform-based matching cost [1] .....	17
Figure 2.6	Overall architecture of proposed stereo matching method [16].....	18
Figure 2.7	Matching cost aggregation, (a) The creation of adaptive window, (b) Semi-global aggregation in direction r [16] .....	18
Figure 2.8	Pixels classification and disparity modification [16].....	19
Figure 2.9	The framework blocks of the development algorithm [17] .....	20
Figure 2.10	Flowchart of the proposed algorithm [4] .....	21
Figure 2.11	Schematic flow diagram of the proposed method [18].....	22
Figure 2.12	The disparity map results of different methods, (a) The references image, (b) Raw disparity maps by Census Transform (CT), (c) Refined by WM, (d) Refined by fast gradient domain guided filtering (F-GDGIF) [18].....	22
Figure 2.13	The stage of the proposed algorithm [19].....	24
Figure 2.14	The depth map result for the challenging regions using training images [19] .....	24
Figure 2.15	Rectification of stereo image pair based on the sensor's rotation angle [20] .....	25
Figure 2.16	Multiple-baseline stereo equivalence of our sensor rotation configuration [20].....	25
Figure 2.17	Block diagram for proposed algorithm [21] .....	26
Figure 2.18	Proposed pre-processing method (a) CLAHE and guided filtering (b) AGCWD and guided filtering [21].....	27
Figure 2.19	Frontal parallel stereo camera configuration for depth estimation [22].....	28
Figure 2.20	Projection of a feature point on the image plane [22].....	28
Figure 2.21	Block diagram of the marker extraction process. It takes an RGB image of the markers on the object as an input and produces a binary image containing only the markers [23].....	29

Figure 2.22 Tiling in a translated deconvolution with a $3 \times 3$ kernel split into four sub-kernels. With a tiling strategy $W = 2$ , $H = 2$ , $C1 = 1$ , $C2 = 2$ , $C3 = 1$ , $C4 = 1$ , only the shaded elements are loaded into the buffer. The of map elements generated in this round (shaded) are also stored in the buffer [24] .....	30
Figure 2.23 The ASV overview with augmentations shaded [24] .....	31
Figure 2.24 SFM multiple view [8] .....	32
Figure 2.25 SFM stages [8] .....	32
Figure 2.26 Overview of project algorithm [25] .....	33
Figure 2.27 Depth map computed by our MATLAB code without post-processing, (a) left image, (b) right image, (c) depth map [25] .....	34
Figure 2.28 Four images, collected in-house and used to test 16 state-of-the art methods. The green masks on some of the left images highlight the pixels where the ground-truth disparity is available. The disparity range is shown in pixels while the depth range is in meters. d refers to disparity, (a) Left image, (b) Highlight of regions of interest where ground-truth disparity is estimated with high confidence, (c) Right image, (d) Ground-truth disparity maps [10] .....	35
Figure 2.29 The building blocks of a stereo matching pipeline [10] .....	36
Figure 2.30 Sample Window Cost Calculation (Veksler) [26] .....	37
Figure 2.31 Stereo Vision for Road Deformity Detection (Rui) [26] .....	37
Figure 2.32 Window size compared to depth map (darker pixel shade refers to greater depth) (Olga) [26] .....	38
Figure 3.1 Flowchart of project .....	48
Figure 3.2 flowchart of stereo vision system framework .....	50
Figure 3.3 Middlebury 15 image dataset .....	57
Figure 4.1 Stereo matching diagram .....	61
Figure 4.2 Different method of disparity map evaluation using Middlebury training dataset for $r = 363$	
Figure 4.3 Different method of disparity map evaluation using Middlebury training dataset for $r = 567$	
Figure 4.4 Different method of disparity map evaluation using Middlebury training dataset for $r = 772$	
Figure 4.5 Different method of disparity map evaluation using Middlebury training dataset for $r = 9$ .....	75
Figure 4.6 Different $r$ (window size) of disparity map evaluation using Middlebury training dataset	79

## LIST OF SYMBOLS

% - Percentage



## LIST OF ABBREVIATIONS

SDG	-	Sustainable Developments Goals
SG	-	Sum of Gradient Magnitude
ASW	-	Adaptive Support Weight
GF	-	Guided Filter
JWGF	-	Joint Weighted Guided Filter
F-GDGIF	-	Fast Gradient Domain Guided Filter
CLAHE	-	Contrast Limited Adaptive Histogram Equalization
AGCWD	-	Adaptive Gamma Correction Weighted Distribution
WTA	-	Winner-Take-All
LR	-	Left Right
WM	-	Weighted Median
SEG	-	Superpixel with Segmentation Process
WMF	-	Weighted Median Filter
MF	-	Median Filter
BF	-	Bilateral Filter
NLCA	-	Non-Local Cost Aggregation
MST	-	Minimum Spanning Trees
LRC	-	Left-and-Right Consistency Check
AMNet	-	Atrous Multiscale Network
DSDGA	-	Deep Self-Guided Stereo Matching with Geometric-aware Aggregation
RGB	-	Red Green Blue
GPU	-	Graphics Processing Unit
MSE	-	Mean Square Error
PSNR	-	Peak Signal-to-Noise Ratio
PDF	-	Probability Density Function
CDF	-	Cumulative Distribution Function
CT	-	Census Transform
SED	-	Sum of Absolute Differences
MTS	-	Modified Census Transform
SAD	-	Sum of Absolute Differences
AGF	-	Adaptive Guided Filter
CNN	-	Convolutional Neural Networks

## LIST OF APPENDICES

APPENDIX	TITLE	PAGE
APPENDICES A	GANT CHART .....	91
APPENDICES B	MILESTONE.....	92



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

This chapter provides a brief overview of the stereo vision technique and highlights its appealing qualities, which can be applied to develop depth map assessment algorithms for computer vision applications. It also provides an explanation of the background, problem statement, project objective, scope of project and supporting evidence for the proposed study. Along with its creative contributions, the thesis's structure and organization are also acknowledged.

### 1.2 Background

Stereo matching is a fundamental task in computer vision and photogrammetry that involves estimating the depth information of a scene by matching corresponding points in stereo image pairs. The fundamental of stereoscopic vision is based on the epipolar geometry process. This process is important for various applications such as 3D reconstruction, visual reality, autonomous vehicles, and digital surface model production. Despite its importance, stereo matching faces challenges when dealing with textureless regions, occlusions, and variations in illumination conditions. These challenges can lead to inaccuracies in depth estimation, especially in scenarios where traditional stereo matching algorithms struggle to provide reliable results.

In recent years, significant research efforts have been directed towards improving stereo matching performance in challenging scenarios. One common classification of stereo

matching algorithms divides them into global and local approaches. Global algorithms incorporate smoothness assumptions into an energy function to estimate disparity by minimizing the global energy function. On the other hand, local algorithms rely on local information to compute disparities, making them computationally efficient but potentially less accurate in certain scenarios. Balancing accuracy and efficiency in stereo matching remain a key research focus to address the limitations of existing algorithms.

The work by [1] contributes to this research domain by proposing an innovative method to enhance stereo matching performance for low texture stereo images. The proposed method combines a matching cost computation approach with an adaptive shape guided filter for cost aggregation. By integrating enhanced image gradient-based matching costs and improved census transform-based matching costs, the method aims to improve robustness against radiometric variations and textureless regions, thereby enhancing disparity estimation accuracy.

The adaptive shape guided filter plays a significant role in aggregating matching costs by constructing cross-based adaptive shape support windows for each pixel [2]. This approach helps in effectively aggregating matching costs within the constructed support windows, leading to more accurate and reliable depth maps. Additionally, the method incorporates a winner-take-all strategy for disparity selection and a multi-constraints-based disparity refinement framework to further enhance the accuracy of the disparity maps [3]. Experimental evaluations on benchmark datasets demonstrate the effectiveness of the proposed method in improving stereo matching performance, particularly in textureless regions where traditional methods may struggle to provide satisfactory results.

Overall, the research presented in this project work addresses the challenges of stereo matching in low texture scenarios and contributes to advancing the field by introducing a cost computation method and adaptive shape guided filter. By focusing on

improving disparity estimation accuracy in challenging conditions, the proposed method shows promise in enhancing the practical applicability of stereo matching algorithms in various real-world scenarios.

### **1.3 Enhancing Industry and Infrastructure through Stereo Vision in Robotics and Automation**

The development of robots and automation depends heavily on stereo vision technology, which also helps to achieve Sustainable Development Goal (SDG) 9: Industry, Innovation, and Infrastructure. Stereo vision improves the capabilities of robots and automated systems by allowing them to sense and comprehend spatial connections and depth, which results in significant increases in productivity, accuracy, and safety in industrial processes. Stereo vision reduces errors and waste in production by enabling robots to carry out intricate activities like assembly, quality inspection, and material handling with extreme precision. Furthermore, it makes it possible to automate infrastructure development tasks like building and upkeep by offering precise 3D mapping and live monitoring, which guarantee structural integrity and on-time project completion. Robots' ability to recognize and avoid objects and people increases safety by reducing the risk of accidents and improving the working environment as a whole. Thus, stereo vision integration in robotics and automation promotes creativity, streamlines production processes, and aids in the construction of durable and sustainable industrial infrastructures.

### **1.4 Problem Statement**

The project aims to address the persistent challenges in dense stereo matching, specifically focusing on improving the accuracy and robustness of disparity estimation in low-texture regions. Despite advancements in stereo matching algorithms, accurately estimating disparities in areas with limited texture remains a significant challenge. These

regions often lead to errors and inaccuracies in the disparity maps, impacting the overall quality of 3D reconstruction and other computer vision applications. Therefore, the primary problem statement revolves around the need to enhance the performance of stereo matching methods to effectively handle occluded, textureless, and discontinuous regions in stereo images.

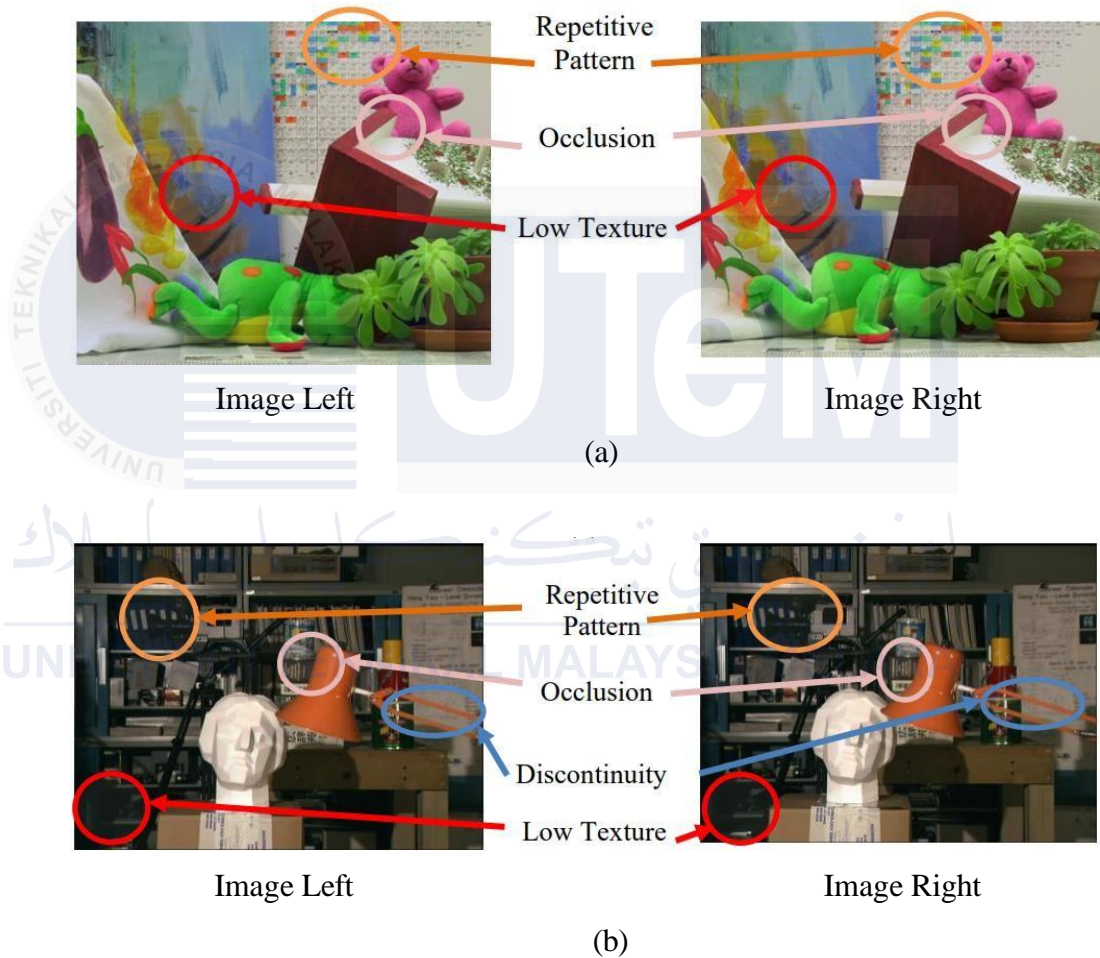


Figure 1.1 Constraining Factors in Middlebury Dataset Stereo Images [1]

To tackle this problem, the project sets out to develop a matching cost computation method that combines enhanced image gradient-based cost with improved census transform-based cost. By integrating these approaches, the aim is to create a more robust matching cost function that can better handle radiometric variations, illumination changes, and textureless areas in stereo images [4]. This innovative cost computation method is designed to provide

a more reliable foundation for subsequent disparity estimation, ultimately improving the accuracy of stereo matching results in challenging scenarios.

Furthermore, the project seeks to implement an adaptive shape guided filter for local stereo matching, which aims to address the limitations of traditional methods in handling occlusions, textureless areas, and discontinuities. By constructing adaptive shape support regions for each pixel and utilizing a modified guided filter based on these regions for cost aggregation, the proposed method aims to enhance the aggregation of matching costs and improve the accuracy of disparity estimation, particularly in large low-texture regions [5]. This adaptive shape guided filter approach is expected to contribute significantly to mitigating the challenges associated with stereo matching in difficult image regions.

Overall, the project's objective is to evaluate the effectiveness of the proposed method in producing more accurate and reliable disparity maps for large low-texture regions. By conducting experiments using the Middlebury benchmark dataset, the project aims to demonstrate the superiority of the developed approach over traditional guided filter-based methods. Through this research, the project seeks to contribute to the advancement of stereo matching techniques, particularly in addressing the complexities of textureless regions and enhancing the overall quality of disparity estimation in challenging stereo images.

## 1.5 Project Objective

The main objectives of this thesis are as follows:

1. Develop stereo vision system to improve the accuracy and robustness of stereo matching in low-texture regions.
2. Evaluate the proposed method using benchmark datasets, such as the Middlebury datasets, to demonstrate the effectiveness and superiority of the approach in achieving accurate and reliable stereo matching results.
3. Perform performance analysis for stereo vision system with real-world image.

## 1.6 Scope of Project

The scope of the research is limited to the following important notes:

1. Develop a stereo vision algorithm for depth estimation from digital stereo image pairs.
2. Evaluate the algorithm using 15 training images of Middlebury standard benchmarking dataset for qualitative and quantities results.
3. A MATLAB software is used to construct and validate the effectiveness of the proposed algorithm.
4. Providing experimental results and discussions to analyze the effectiveness and efficiency of the proposed method in improving stereo matching accuracy for low texture stereo images.
5. This project scope encompasses the development, implementation, evaluation, and analysis of a comprehensive stereo matching method aimed at enhancing the accuracy and robustness of disparity estimation in challenging imaging conditions.

6. All the experiments are executed using a personal computer with the features of window 10 on desktop PC with 3.2GHz processor and 8GB memory.



## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Introduction

This literature review provides a summary of the project in order to get the whole data on stereo vision, 3D reconstruction images, disparity maps, depth maps, camera calibration, webcams, and MATLAB software, which will give the idea to run this project. In addition, it shows that the information is related to studies that have been done by previous researchers as guidance. In order to understand stereovision, one must first explain the mode of communication of a camera, how it is structured in principle, and which parameters need to be regulated.

#### 2.2 Enhancing Stereo Vision Algorithms with Adaptive Guided Filtering: Addressing Occlusions, Noise, and Efficiency for Robust Depth Estimation

Stereo vision algorithms face a number of difficulties, including occlusions, textureless areas, repeated patterns, and high processing demands. These algorithms are crucial for applications such as robotics and 3D reconstruction. It is also difficult to produce real-time, high-quality depth maps that are resilient to changes in lighting and lens aberrations. Notably, the adaptive guided filter preserves edges while smoothing, adjusts to the specific image content, and is computationally efficient, which makes it appropriate for real-time applications. Moreover, resilience to variations in lighting is improved by this filter. Important studies have shown how well guided filters work to improve depth estimate in difficult situations and refine disparity maps. Preprocessing images, determining initial disparities, using adaptive guided filtering, and fine-tuning the depth map are all necessary

steps in the practical implementation. Thus, by tackling noise, occlusions, and computational efficiency, adaptive guided filters great to improve stereo vision algorithms and improve depth map accuracy and dependability across a range of applications.

### **2.3 Background of stereo vision**

Stereo vision is a process that involves extracting pixel information from two images (left and right images) captured simultaneously by a binocular camera. The goal of stereo vision is to estimate a "disparity map," which represents the differences in pixel positions between corresponding points in the left and right images. By analyzing these differences, stereo vision algorithms can calculate the depth information of the scene, providing a 3D perception of the environment [6]. This process simulates the way human vision works with two eyes, enabling machines to perceive depth and spatial relationships in a similar manner [7].

The use of stereo vision algorithms is essential for depth estimation by processing the captured images to generate accurate and detailed information about the spatial layout of objects in the scene. These algorithms vary in their approaches, leading to differences in the quality of the resulting disparity map and the computational load required for processing.

The quality of the generated disparity maps is influenced by factors such as image noise, uncertainties in the stereo images, and the effectiveness of the algorithm in handling challenges like occlusion, radiometric distortion, depth discontinuity, and textureless regions [8]. Researchers continuously develop and update stereo vision algorithms to enhance accuracy, reduce computational costs, and address specific challenges encountered in various applications, including obstacle detection, object tracking, 3D object recognition, shape reconstruction, and geometric mapping [6].

In stereo vision, the triangulation principle is a fundamental concept used to estimate the depth of objects in a scene by leveraging the geometric relationship between two images captured by a stereo camera setup. By analyzing the disparity between corresponding points in the left and right images, the depth of these points can be calculated through triangulation. The process involves identifying matching points in both images, determining the horizontal shift (disparity) between them, and using the known baseline distance between the cameras to compute the depth of the scene point. Triangulation enables the creation of a 3D representation of the scene by mapping the spatial coordinates of objects based on their visual appearance in the stereo images. This technique plays a crucial role in various applications such as robotics, augmented reality, and computer vision tasks, providing valuable depth information for perception, navigation, and scene reconstruction. The triangulation of stereo vision is shown in figure 2.1 [9].

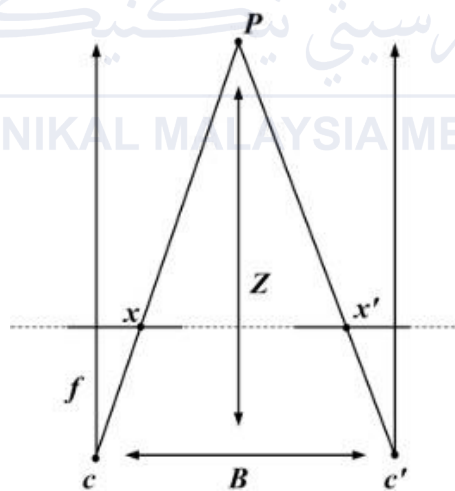


Figure 2.1 Triangulation of stereo vision [9]

Stereo vision is a fundamental technology that controls binocular vision to extract depth information from images, enabling applications in diverse fields such as robotics, automobiles, and aerospace. The development in stereo vision algorithms aim to improve the accuracy, efficiency, and robustness of depth estimation for enhanced perception and decision-making in complex environments.

Stereo vision is a technology that solves problem in computer vision by providing solutions to key problems related to depth perception and spatial understanding. One of the issues that stereo vision helps solve is accurate depth estimation in a scene [10]. By analyzing the differences in pixel positions between corresponding points in a pair of stereo images, stereovision algorithms can calculate the distance to objects and generate detailed depth maps. This depth information is essential for tasks such as 3D reconstruction, object localization, and scene understanding [11].

Additionally, stereo vision enables systems to detect and track objects based on their spatial relationships and depth information, which is important for uses like autonomous vehicles, robotics, and surveillance systems. By enhancing depth perception capabilities and providing detailed spatial information, stereo vision contributes to a better understanding of complex scenes, allowing systems to recognize objects, infer spatial relationships, and make informed decisions based on the scene context. Overall, Stereo vision is an important part of computer vision by enabling accurate depth estimation, object detection, scene understanding, and depth perception, thereby enhancing the spatial awareness and decision-making capabilities of systems across various applications and industries [12]. Figure 2.2 shows the accurate disparity estimation.

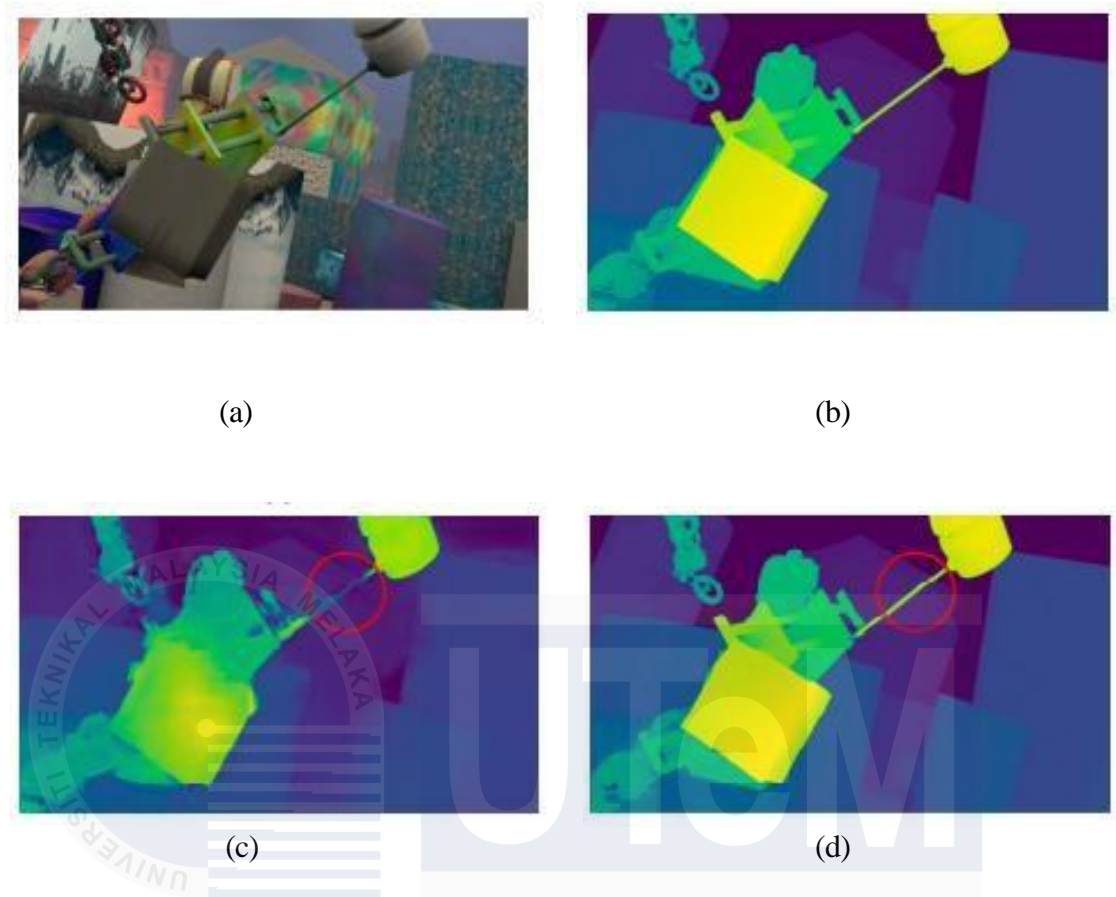


Figure 2.2 Disparity estimation, (a) Origin RGB data (b) Ground Truth disparity Pseudo- color data, (c) Disparity prediction by stereonet, (d) Result. Achieve an excellent disparity estimation [11]

Stereo vision, also known as binocular vision, is a technology that enables machines to perceive depth and create 3D image representations of the environment by using two cameras or sensors positioned at different viewpoints. This technique mimics the human visual system, where depth perception is achieved through the slight disparity between the images captured by the left and right eyes [13]. In stereo vision systems, the two cameras capture images of the same scene, and by analyzing the differences between these images, the system can calculate the depth information of objects in the scene. This depth perception is based on the principle of parallax, where objects at different distances from the cameras will appear to shift position relative to each other in the images [14].

A binocular stereo vision model, showing how two cameras capture images of the same object from slightly different perspectives. Points  $P$  and  $P'$  represent corresponding points on the object as seen by the left and right cameras, respectively, while points  $O_1$  and  $O_2$  represent their actual 3D positions in the scene. By analyzing the disparities between these corresponding points in the two images, the system calculates depth information and reconstructs the three-dimensional structure of the object based on the parallax principle. This geometric relationship between the points allows the binocular stereo vision system to infer depth and create a 3D representation of the scene [13].

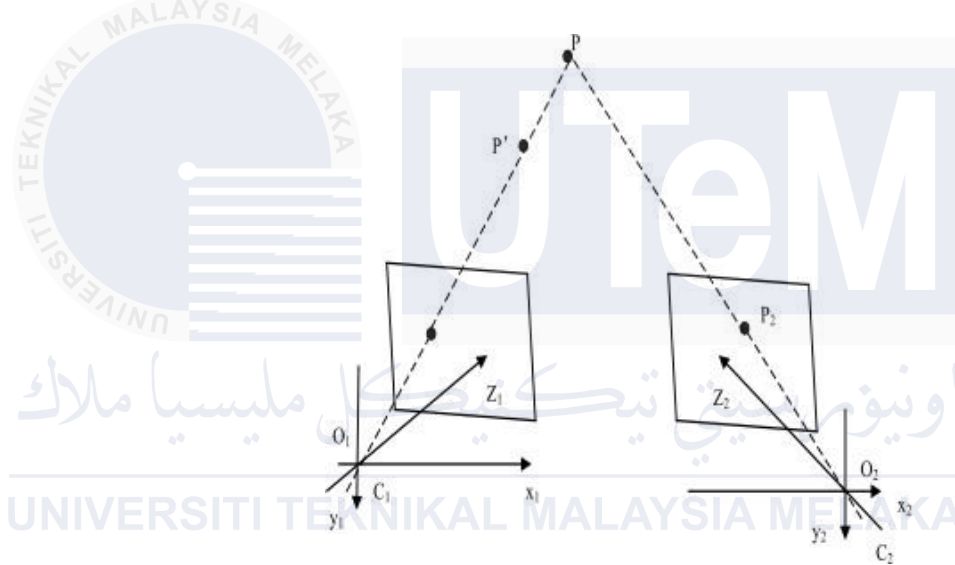


Figure 2.3 Binocular stereo vision [13]

By identifying corresponding points in the images and measuring the disparity between them, stereo vision systems can reconstruct the 3D structure of the scene. This depth information is valuable for various applications, including object recognition, obstacle avoidance, 3D mapping, and augmented reality. Stereo vision technology has widespread applications in fields such as robotics, autonomous vehicles, industrial automation, and virtual reality. It is essential for facilitating machines to perceive and interact with the world in a more human-like manner, enhancing their capabilities for tasks that require spatial understanding and depth perception [14].

## 2.4 Stereo vision framework

To produce a depth map using stereo vision, the process typically involves several key steps. Initially, images of the same scene are acquired using two cameras positioned at slightly different viewpoints. These images serve as the input for the stereo vision system. The next step is camera calibration, where the intrinsic and extrinsic parameters of the cameras, such as focal length and lens distortion, are determined to ensure accurate depth estimation.

Subsequently, image rectification is performed to align corresponding points in the stereo images, simplifying the matching process. Feature detection algorithms are then employed to identify key points in the images that can be matched between the left and right views. Feature matching is important for establishing correspondences between points in the two images, enabling the calculation of pixel disparities.

By estimating the disparity between matched feature points, typically using metrics like sum of squared differences or normalized cross-correlation, the system can determine the depth information of the scene [15]. This depth calculation involves converting the disparity values into depth values based on the known baseline distance between the cameras and the focal length. Post-processing techniques, such as filtering, may be applied to refine the depth map and improve its quality. The final step involves visualizing the depth map, where brighter regions indicate closer objects and darker regions represent farther objects, providing a comprehensive 3D representation of the scene. By following these consecutive stages, a stereo vision system can efficiently produce a depth map that improves perception and comprehension of space for a variety of used [13].

The proposed method for stereo matching consists of four main steps. Firstly, in the matching cost computation step, a new method is introduced that combines enhanced

image gradient-based cost with improved census transform-based cost to build a cost volume representing per-pixel matching costs at all considered disparity values. This approach aims to make the matching cost more robust to radiometric changes and noises compared to traditional methods using absolute intensity or color differences. Secondly, in the cost aggregation step, a cross-based adaptive shape support window is created for each pixel, followed by the application of a modified guided filter to aggregate the matching costs within the window. Thirdly, the disparity selection step employs a winner-take-all strategy to determine the optimal disparity for each pixel based on the aggregated costs. Lastly, a multi-constraints-based disparity refinement framework is implemented to further enhance the disparity map accuracy. This framework includes outlier detection with left-right consistency checking, occlusion/mismatch handling, weighted median filtering, and subpixel enhancement techniques [1].

## 2.5 Stereo vision previous method

The work by [1] of dense stereo matching in photogrammetry and computer vision, outlining its applications in 3D reconstruction, Digital Surface Model (DSM) production, visual reality, and autonomous vehicles. Despite the extensive research history in stereo matching, challenges persist in effectively handling occluded, textureless, and discontinuous regions. Global and local stereo matching algorithms are discussed, with global methods incorporating smoothness assumptions for accurate disparity estimation but higher computational complexity. Local algorithms rely on local information and lack explicit global constraints, making them computationally efficient but less effective in challenging regions. To address these challenges, the work proposes an innovative matching cost measurement and adaptive shape guided filter-based aggregation method. By combining enhanced gradient-based and census transform-based matching costs, the approach enhances robustness against exposure variations and textureless areas, leading to improved disparity

map generation. Experimental results on the Middlebury benchmark dataset demonstrate the effectiveness of the proposed method, achieving a 9.40% average matching error rate on standard image pairs and outperforming traditional methods in textureless regions. The review underscores ongoing efforts to enhance stereo matching performance through cost measurement and aggregation strategies, aiming to enhance accuracy and reliability for diverse photogrammetry and computer vision applications.

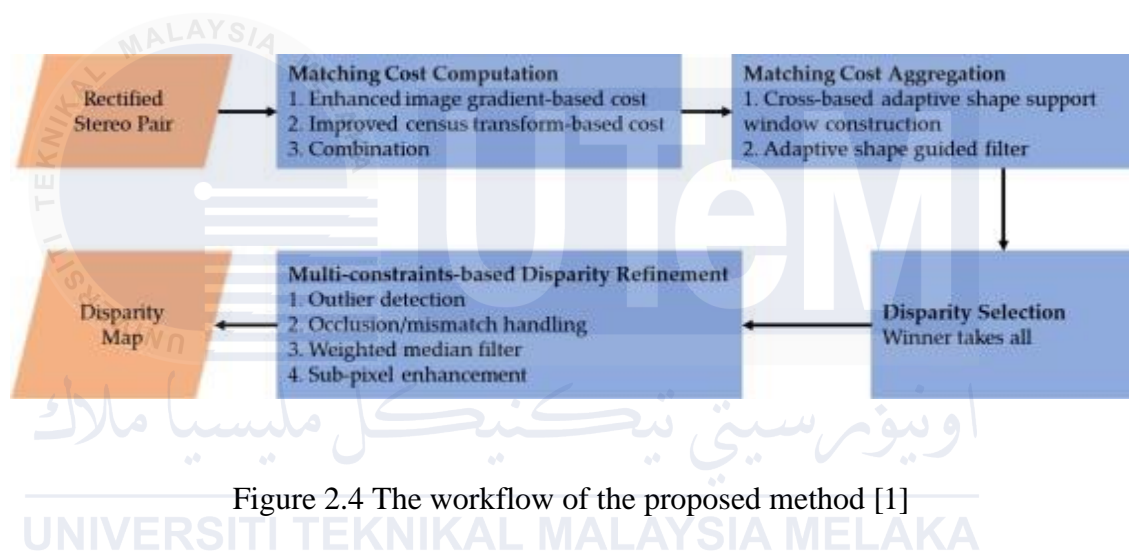


Figure 2.4 The workflow of the proposed method [1]

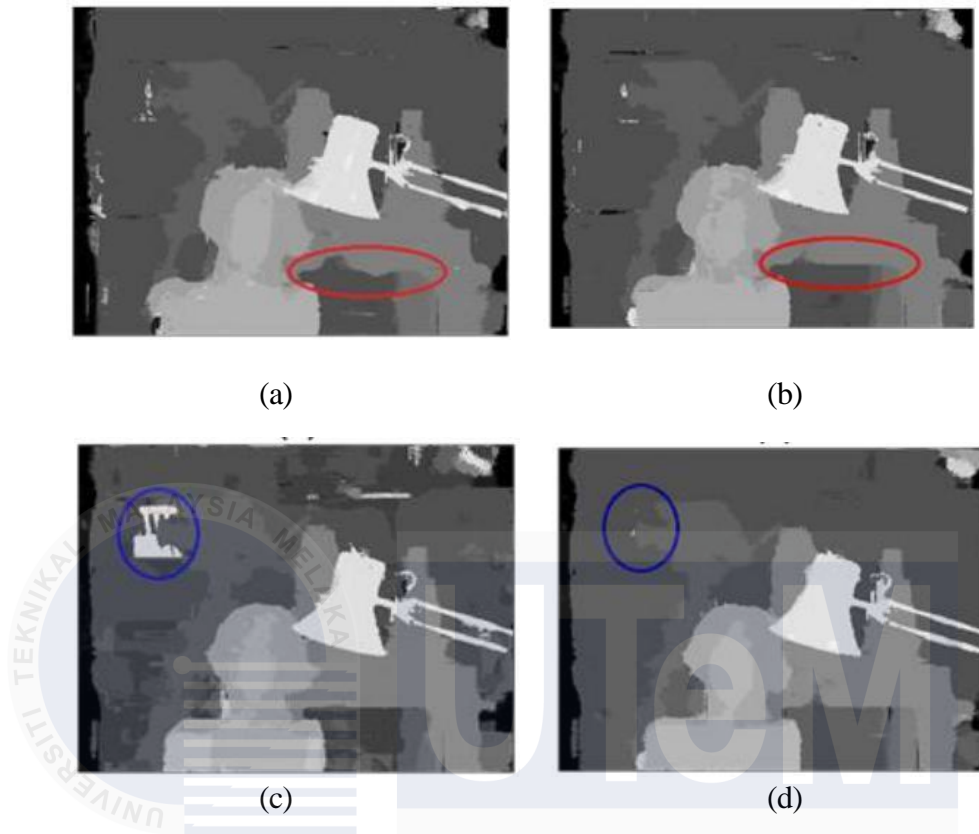


Figure 2.5 Initial disparity maps based on different matching costs for Tsukuba. (a) Absolute difference in image gradients, (b) proposed enhanced image gradients-based matching costs, (c) traditional census transform-based matching cost, (d) proposed improved census transform-based matching cost [1]

Then, work by [16] discusses various approaches and techniques in the field of stereo matching. Several studies have focused on leveraging deep learning methods, such as convolutional neural networks (CNNs), to improve stereo matching accuracy. For instance, the work explores efficient deep learning techniques for stereo matching, and propose a multi-task learning network for both stereo matching and edge detection. Additionally, traditional stereo matching methods have been criticized for their inability to adapt to different lighting conditions, prompting the development of new approaches that utilize CNN features for robust matching cost calculation. Furthermore, the work have investigated fast stereo matching using adaptive guided filtering and high-resolution stereo datasets with subpixel-accurate ground truth, respectively. These works collectively contribute to the

advancement of stereo matching techniques, addressing challenges related to texture-less areas, occlusions, and computational costs associated with feature extraction and post-processing.

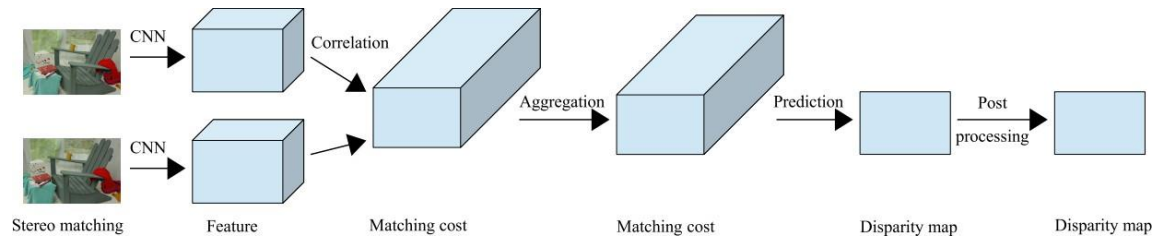


Figure 2.6 Overall architecture of proposed stereo matching method [16]

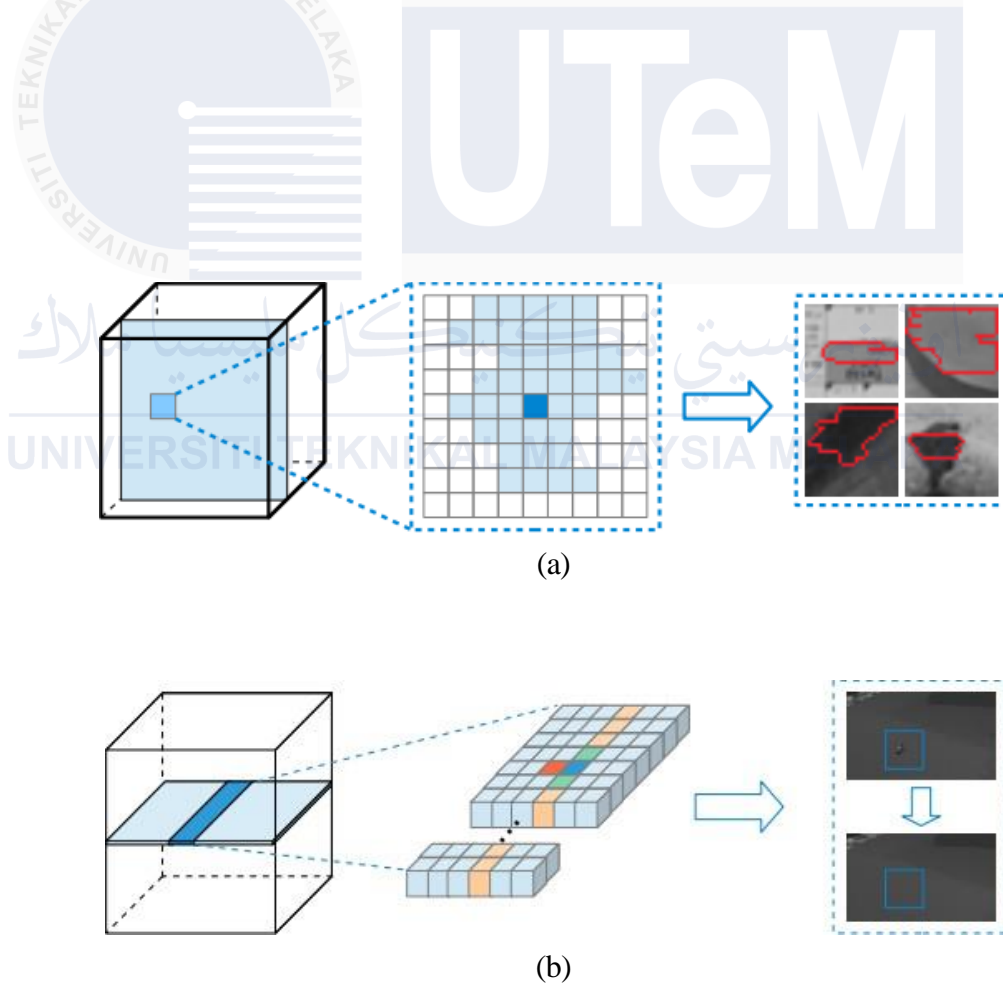


Figure 2.7 Matching cost aggregation, (a) The creation of adaptive window, (b) Semi-global aggregation in direction  $r$  [16]

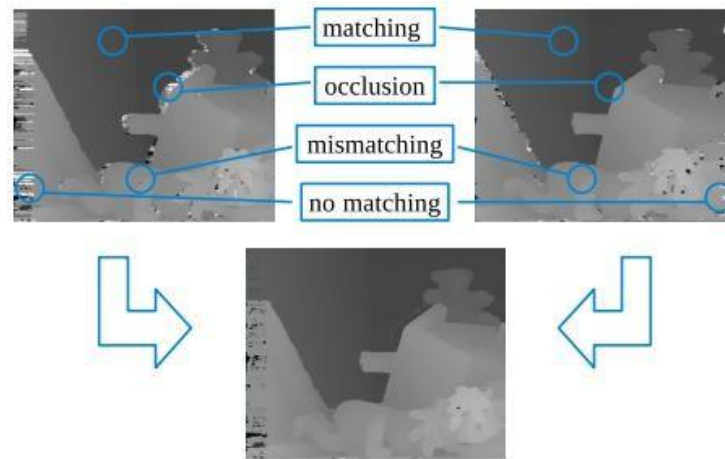


Figure 2.8 Pixels classification and disparity modification [16]

Then, work by [17] discusses the use of Dynamic Histogram Equalization for improving image contrast by dynamically adjusting the histogram. Additionally, Quadrant Dynamic Clipped Histogram Equalization with Gamma Correction is highlighted as a method for color image enhancement. Another technique, the Brightness Preserving Histogram Equalization, introduces a work approach for enhancing image contrast while maintaining brightness levels. The work also covers the Census Transform-Based Robust Stereo Matching, which focuses on handling radiometric changes in stereo matching. Furthermore, the Low-Cost Real-Time Embedded Stereo Vision System is presented as a solution for accurate disparity estimation in real-time applications. Lastly, the Local Stereo Matching Algorithm Based on Pixel Difference Adjustment is discussed for its utilization of pixel difference adjustment, minimum spanning tree, and weighted median filter to enhance the accuracy of stereo matching for depth estimation. These references collectively contribute to advancing algorithms for improving image quality and depth estimation in stereo vision systems.

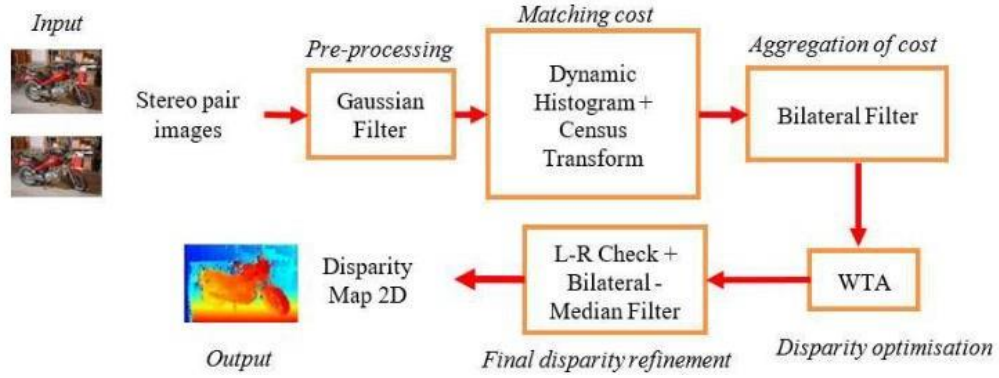


Figure 2.9 The framework blocks of the development algorithm [17]

Besides, work by [4] of stereo matching algorithms in 3D surface reconstruction. Various established methods for post-processing and disparity refinement were discussed, including techniques such as Mean-Shift and Superpixel with Segmentation Process (SEG), Weighted Median Filter (WMF), Median Filter (MF), and Bilateral Filter (BF). These methods aimed to improve accuracy by addressing noise, edge preservation, and outlier removal in the disparity maps. The proposed algorithm introduced the work contributions in three stages: Sum of Gradient Magnitude (SG) for correspondence measurement, Adaptive Support Weight (ASW) with Iterative Guided Filter (GF) for noise reduction and edge preservation, and Joint Weighted GF for further accuracy enhancement. Performance evaluation of the algorithm demonstrated superior results in matching cost computation and cost aggregation stages, showcasing lower error rates and improved edge preservation. The algorithm's effectiveness was validated through 3D surface reconstruction using standard benchmarking datasets, indicating enhanced accuracy compared to existing methods. Additionally, related works such as the GlaRe Toolbox for ice thickness surface reconstruction and the use of Dynamic Vision Sensor for terrain modeling were discussed, emphasizing the diverse applications and advancements in stereo matching algorithms for 3D reconstruction.

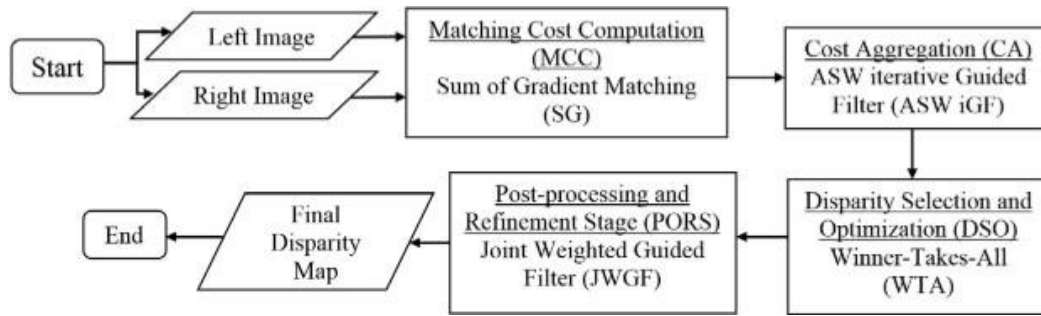


Figure 2.10 Flowchart of the proposed algorithm [4]

Next, work by [18] discusses various advancements in stereo matching algorithms. The precision of stereo matching is greatly enhanced by cost aggregation techniques. Early window-based methods used fixed-size windows, leading to increased error rates and blurred object boundaries. To address these issues, variable-window (VW), multiple-window (MW), and adaptive window (AW) aggregation methods were introduced. Segment-tree (ST) based methods, such as non-local cost aggregation (NLCA) using minimum spanning trees (MST), offer high precision with low computational complexity. Disparity refinement techniques, including median filters and weighted median (WM) filters, are commonly employed to enhance the accuracy of the disparity map by preserving edges and removing outliers. Gaussian filters, while effective in outlier removal, introduce higher computational complexity. Occlusion detection methods like left-and-right consistency check (LRC) and ordering constraints are utilized to identify occlusion regions based on matching point comparisons and scan line monotonicity. Recent advancements in stereo matching include the integration of deep learning techniques like Atrous Multiscale Network (AMNet) and Deep Self-Guided Stereo Matching with Geometric-aware Aggregation (DSDGA) for self-guided filter-based cost aggregation, as well as the development of efficient algorithms using hierarchical representations and adaptive weighted bilateral filters to achieve low computational complexity.

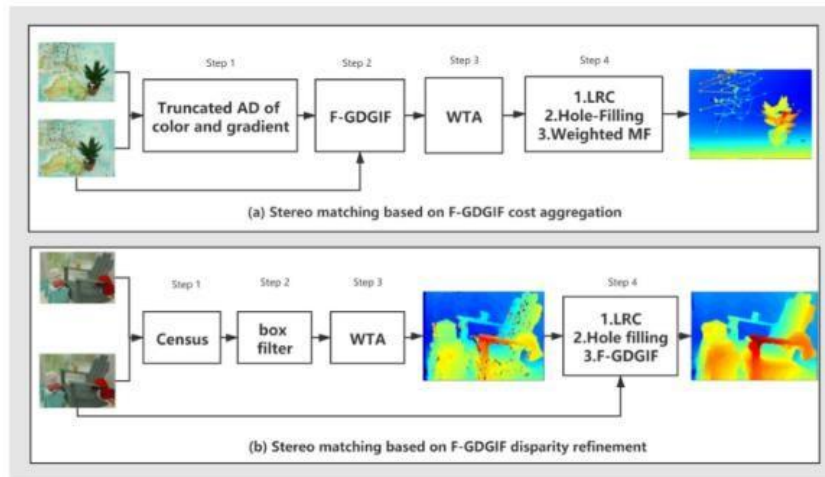


Figure 2.11 Schematic flow diagram of the proposed method [18]

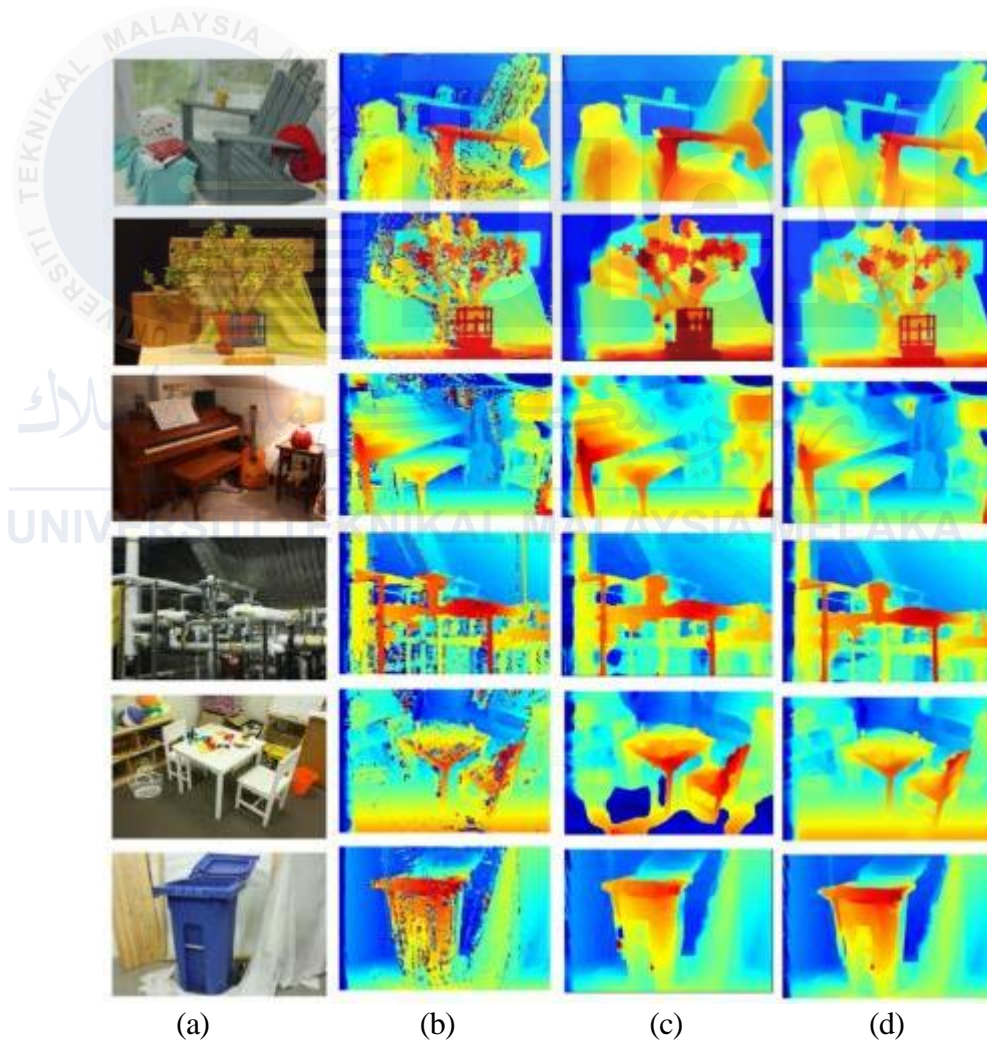


Figure 2.12 The disparity map results of different methods, (a) The references image, (b) Raw disparity maps by Census Transform (CT), (c) Refined by WM, (d) Refined by fast gradient domain guided filtering (F-GDGIF) [18]

Then, work by [19] explores techniques and methodologies for developing depth maps from stereo images. Key points include the effectiveness of feature-based techniques in capturing scene details by focusing on edges or boundaries. Block-based matching methods achieve high accuracy in depth map generation when the window size is appropriately selected by comparing divided image blocks to find corresponding points. The aggregation stage in stereo matching algorithms significantly reduces noise after matching cost calculation using summing or averaging in a support window. Disparity optimization normalizes disparity values and converts them into depth pixel intensities using local, global, or semi-global methods. Depth map refinement eliminates remaining noise through segment-based approaches and filtering techniques to enhance accuracy. The stereo matching framework consists of four stages: matching cost calculation, cost aggregation, disparity optimization, and depth map refinement, with variations depending on the method used.

Depth maps obtained from stereo images can be leveraged for 3D surface reconstruction by utilizing parameters like stereo camera baseline, focal values, and disparity values. Challenges in stereo matching include difficulties in matching regions with repetitive patterns, low texture areas, and plain color regions, posing obstacles for researchers in achieving accurate depth maps. By delving into existing literature and methodologies in depth map reconstruction from stereo images, researchers can glean insights into advancements, challenges, and potential applications in the realms of computer vision and image processing.

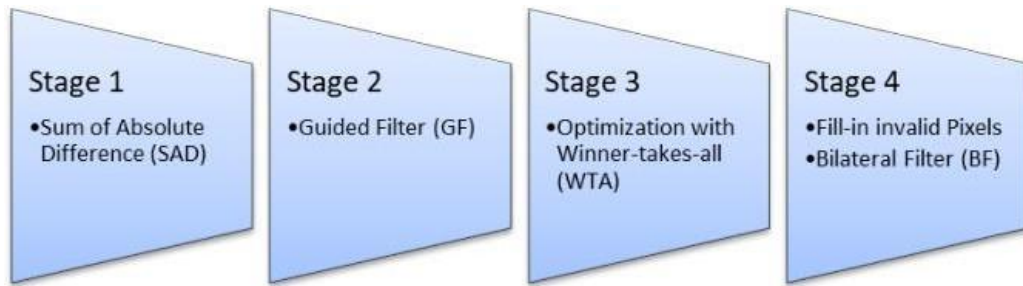


Figure 2.13 The stage of the proposed algorithm [19]

Challenging Regions	Left Reference Image	Depth Map Results (Grayscale)
Low texture region for the Adirondack image		
Plain color region for the Recycle image		
Repetitive region for the Jadeplant image		

Figure 2.14 The depth map result for the challenging regions using training images [19]

Besides, work by [20] inspect into the realm of depth measurement techniques, focusing on the evolution of sensor systems for range sensing applications. It distinguishes between contact and non-contact methods, highlighting the utilization of active and passive sensing technologies. While contact methods involve physical interaction with the object surface, non-contact approaches leverage sensors like laser scanners and structured light systems for distance estimation. Within the realm of passive sensing, stereo vision emerges as a prominent technique for depth recovery, utilizing multiple cameras to capture scene images from different perspectives. The work underscores the importance of correspondence matching algorithms in stereo vision for accurate depth computation.

Furthermore, it discusses the relevance of multiview geometry theory in shaping the design of integrated sensing devices, emphasizing the potential for cost-effective system configurations. The work introduces a sensor system design that integrates camera rotation to enhance depth recovery through stereopsis, offering a fresh perspective on range sensing capabilities. By referencing previous studies on computational stereo vision, dense correspondence algorithms, and wide-baseline stereo techniques, the work sets the stage for exploring innovative approaches to depth measurement using stereovision and computational algorithms.

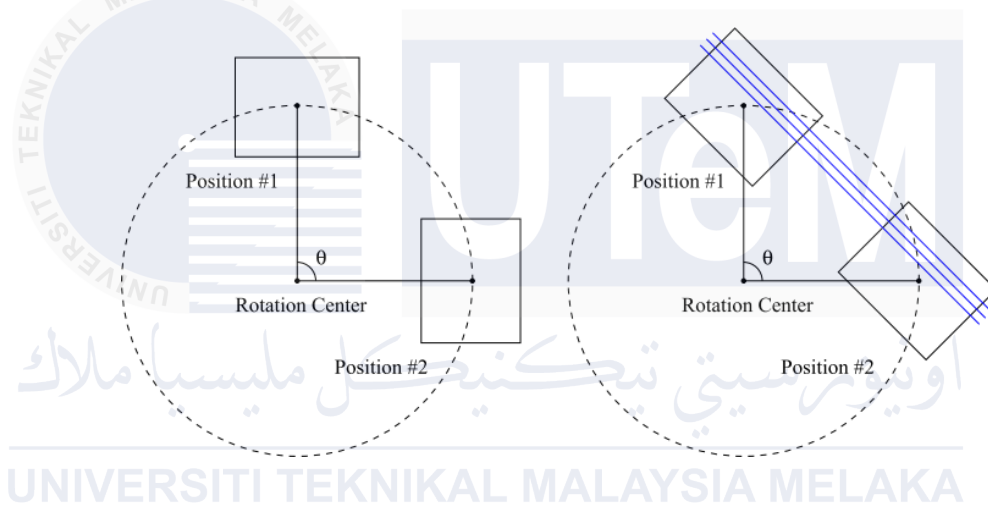


Figure 2.15 Rectification of stereo image pair based on the sensor's rotation angle [20]

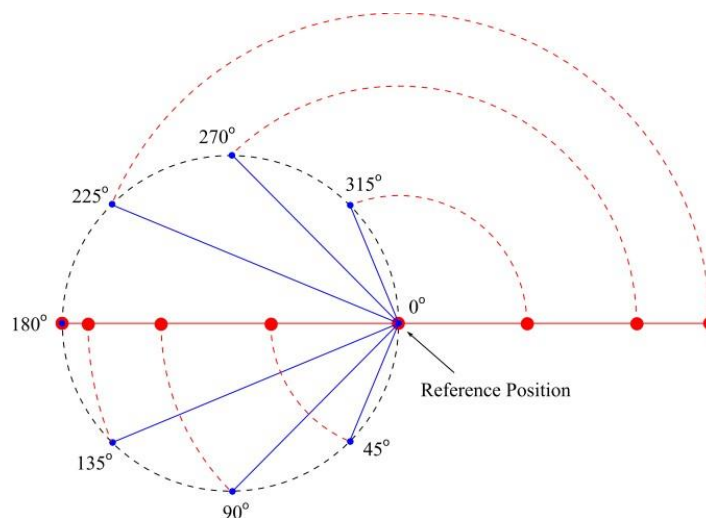


Figure 2.16 Multiple-baseline stereo equivalence of our sensor rotation configuration [20]

Then, work by [21] discusses the development of a new pre-processing technique for computational stereo matching algorithms to enhance the quality of colour images using Contrast Limited Adaptive Histogram Equalization (CLAHE), Adaptive Gamma Correction Weighted Distribution (AGCWD), and guided filter techniques. This step is important for addressing noise in stereo colour images caused by adverse weather and illumination conditions. Additionally, the algorithm incorporates the Census Transform method for matching cost computation, providing advantages in handling radial distortion and brightness changes. The aggregated cost from the matching process is calculated using fixed-window and guided filter techniques. The disparity optimization stage utilizes the Winner- Take-All (WTA) technique, followed by a post-processing stage involving Left Right (LR)consistency checking and the application of a WM filter for noise reduction and smoothening of the disparity map. The proposed algorithm's performance was evaluated using the Middlebury Standard Benchmarking Dataset, showing 23.35% accuracy for nonocc error and 31.65% accuracy for all error. These results demonstrate improved accuracy compared to existing works in the evaluation dataset. Overall, the document underscores the significance of stereo vision in computer vision and introduces an advanced pre-processing technique for computational stereo matching algorithms that enhances depth estimation accuracy while addressing noise challenges.

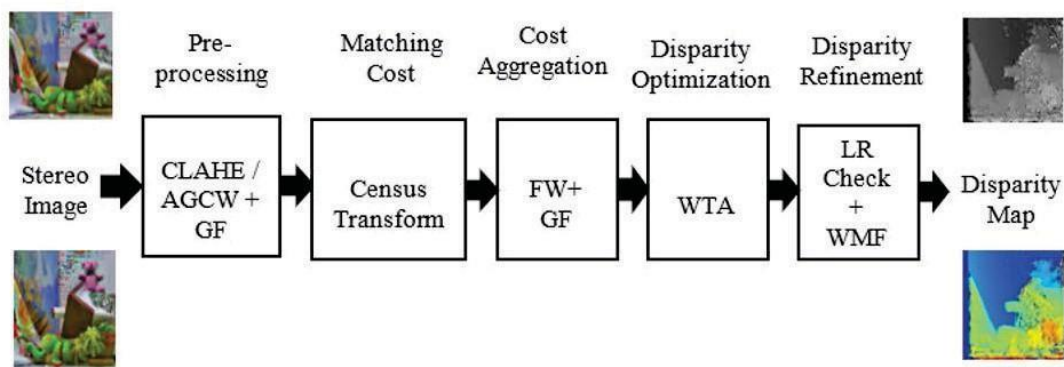


Figure 2.17 Block diagram for proposed algorithm [21]

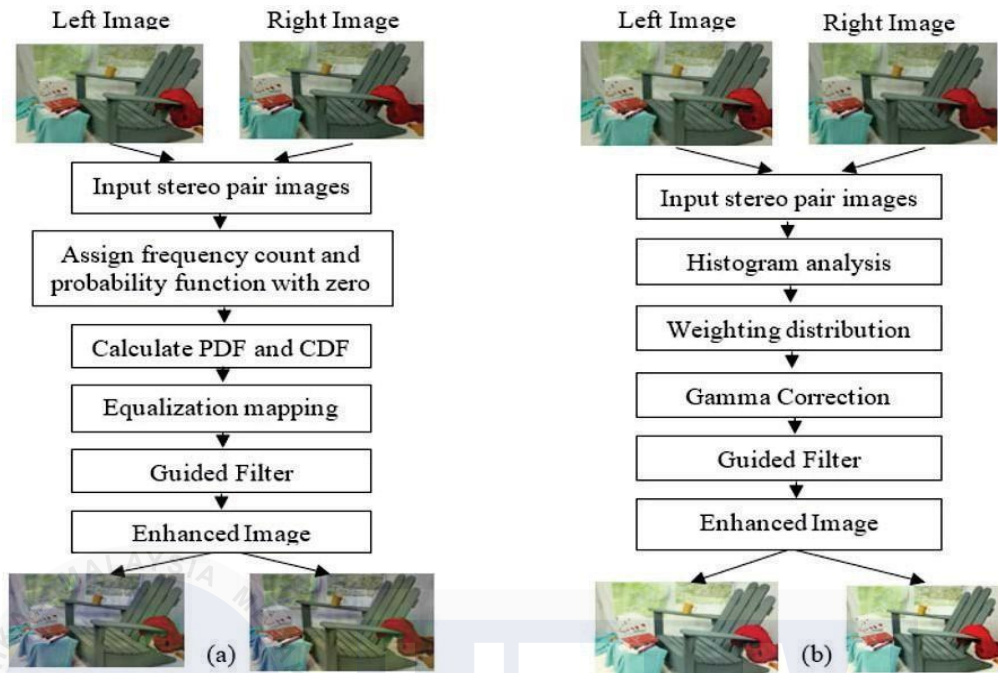


Figure 2.18 Proposed pre-processing method (a) CLAHE and guided filtering (b) AGCWD and guided filtering [21]

Then, work by [22] provides a comprehensive overview of previous studies and findings related to depth estimation using binocular disparity and motion parallax cues. The human visual system relies on various depth cues such as binocular disparity, motion parallax, kinetic depth effect, looming, perspective cues, occlusion, smooth shading, and blur to perceive depth in static and dynamic environments. Computer vision approaches often combine multiple cues to estimate depth, with some methods using deep neural networks trained on ground truth depth data. Binocular disparity and motion parallax are identified as crucial depth cues in both human and computer vision. The work has shown that binocular disparity is more important for short distances, while motion parallax becomes more significant for depth estimation at longer distances. Stereo vision, which relies on binocular disparity, is commonly used for depth estimation in machine vision applications, particularly in robotic systems. Motion parallax, on the other hand, provides depth information based on the observer's motion and the relative movement of objects in the environment, offering a complementary approach to depth estimation. The work has



approach utilizing a stereo camera rig with an ultra-wide baseline distance and fish-eye lenses. This setup enhances the field of view, coverage area, and stereo vision capabilities, enabling effective motion tracking. A passive marker-based methodology is proposed for object motion tracking, leveraging adaptive thresholding for marker extraction. Given the complexities associated with depth estimation using fish-eye lenses, a unique method is employed to establish a relationship between pixel dimensions in images and real-world coordinates for accurate 3D position restoration. The study also emphasizes the importance of occlusion detection in marker-based human kinematics capture, highlighting the algorithm's ability to differentiate between various occlusion types and predict missing marker positions. Furthermore, the paper presents a method for compensating the camera's ego-motion when mounted on dynamic platforms like drones or cars. Through comprehensive testing in diverse scenarios, the proposed 3D positioning and tracking system's efficacy as a stereo camera rig for motion capture is validated, showing comparable accuracy to the Vicon system at a reduced cost.

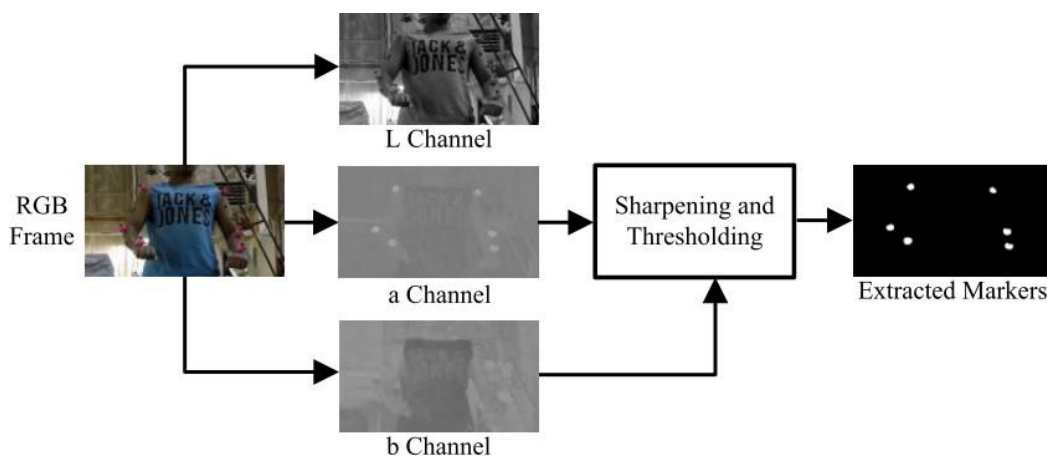


Figure 2.21 Block diagram of the marker extraction process. It takes an RGB image of the markers on the object as an input and produces a binary image containing only the markers [23]

Further, work by [24] provides a comprehensive overview of research efforts in the field of stereo vision systems and related technologies. It discusses the evolution of stereo matching algorithms, emphasizing the trade-offs between traditional search-based methods and Deep Neural Network (DNN) approaches. Classic stereo algorithms rely on heuristics for feature matching, while DNNs offer a data-driven alternative with higher computational demands. The paper also delves into motion-based algorithms like Euphrates and their applications in simplifying continuous vision tasks, highlighting the importance of accurate motion estimation for stereo vision accuracy. Additionally, the work touches upon hardware-efficient algorithms and DNN accelerators, citing works such as Generative Adversarial Networks (GANAX) and Acorn RISC Machine (ARM)'s Mobile Machine Learning Hardware, which aim to optimize DNN performance and efficiency. The co-design of algorithms and System-on-Chip (SoC) architectures is recognized as a key strategy for achieving energy-efficient and high-performance vision systems, as demonstrated by studies like Euphrates. Overall, the paper showcases the diverse range of research focused on enhancing stereo vision algorithms, motion estimation techniques, and hardware implementations to advance the capabilities of vision-based systems.

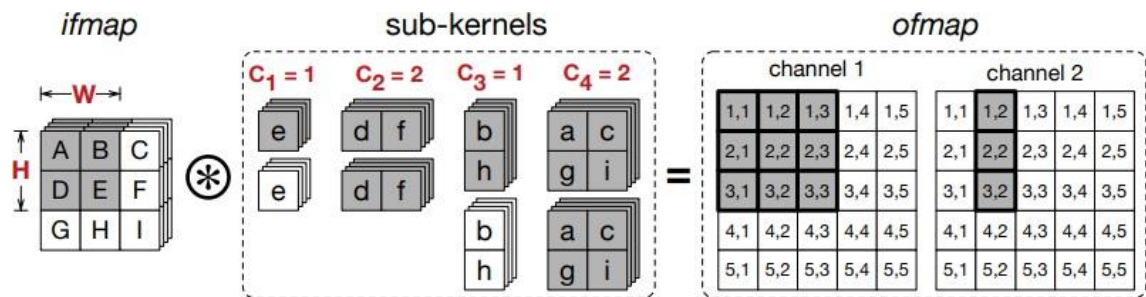


Figure 2.22 Tiling in a translated deconvolution with a  $3 \times 3$  kernel split into four sub-kernels. With a tiling strategy  $W = 2, H = 2, C_1 = 1, C_2 = 2, C_3 = 1, C_4 = 1$ , only the shaded elements are loaded into the buffer. The of map elements generated in this round (shaded) are also stored in the buffer [24]

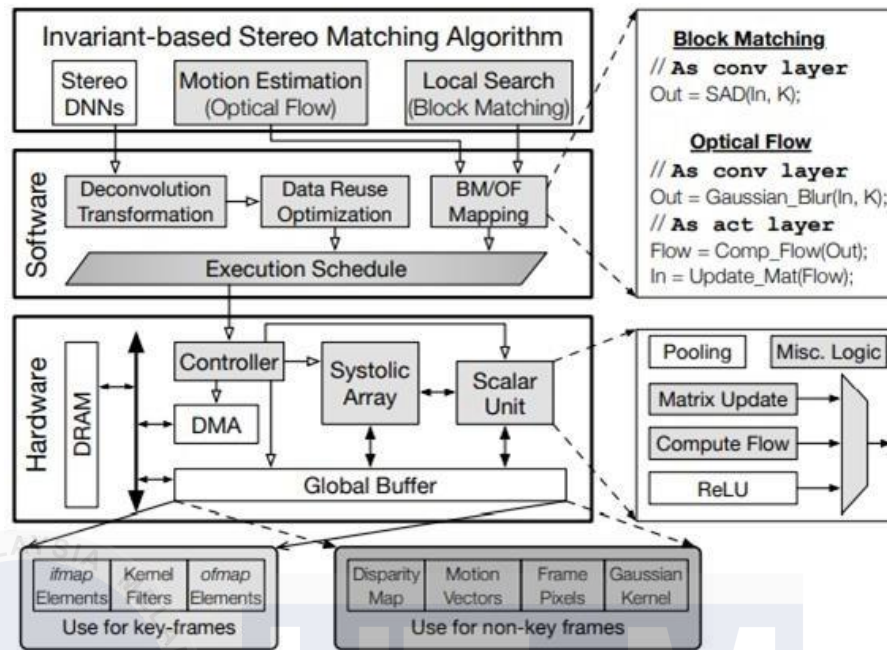


Figure 2.23 The ASV overview with augmentations shaded [24]

Furthermore, work by [8] of the realm of 3D reconstruction using Structure from Motion (SfM) algorithm and Multi View Stereo (MVS) based on computer vision. SfM, a technique for estimating the 3D structure of a scene from 2D images, finds applications in diverse fields like 3D scanning and Augmented Reality. The approach to SfM varies based on factors such as the number and type of cameras used, necessitating additional information like object size and sensor data for accurate scale restoration. In the context of cultural heritage preservation, 3D reconstruction serves as a pivotal tool for documenting and restoring buildings, especially in cases of destruction. By employing SfM and MVS algorithms rooted in computer vision, effort to safeguard 3D objects in cultural heritage areas is significantly enhanced. The development of 3D technology has revolutionized visualization in animation, architecture, education, and Virtual Reality, underscoring its widespread utility. Furthermore, the dynamic efforts required for cultural preservation, emphasizing the importance of documentation activities, including 3D modeling, to maintain the authenticity and values of cultural works. Through 3D modeling methods, precise dimensions and shapes of cultural heritage objects are

captured, facilitating conservation activities and ensuring the preservation of historical significance and cultural values.

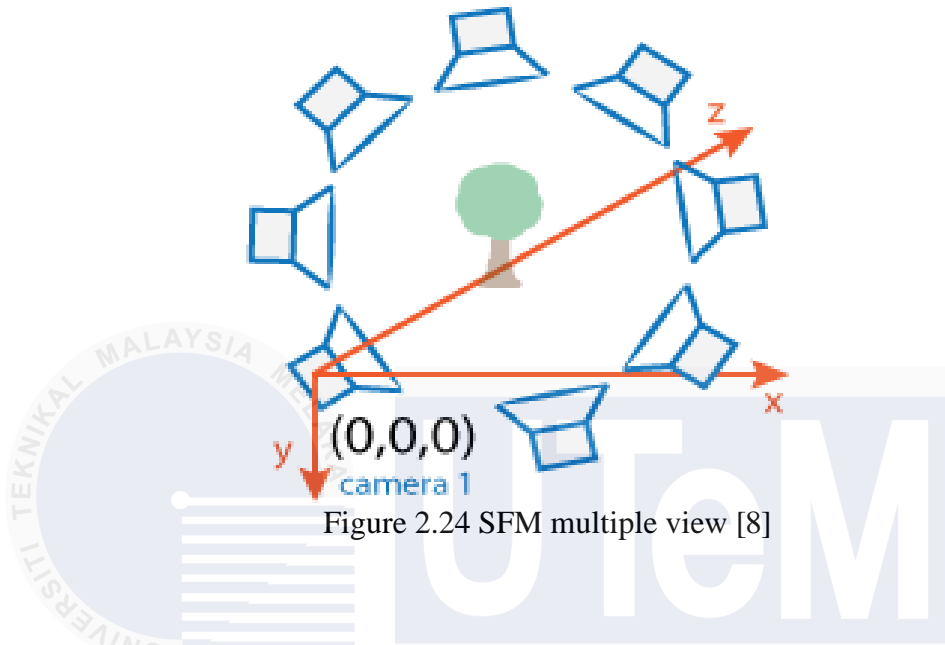


Figure 2.24 SFM multiple view [8]

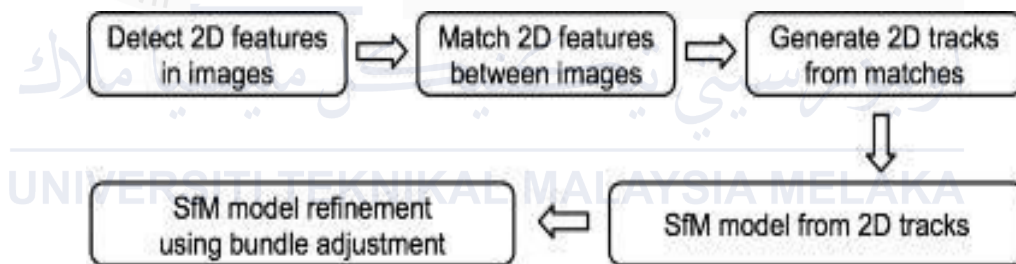
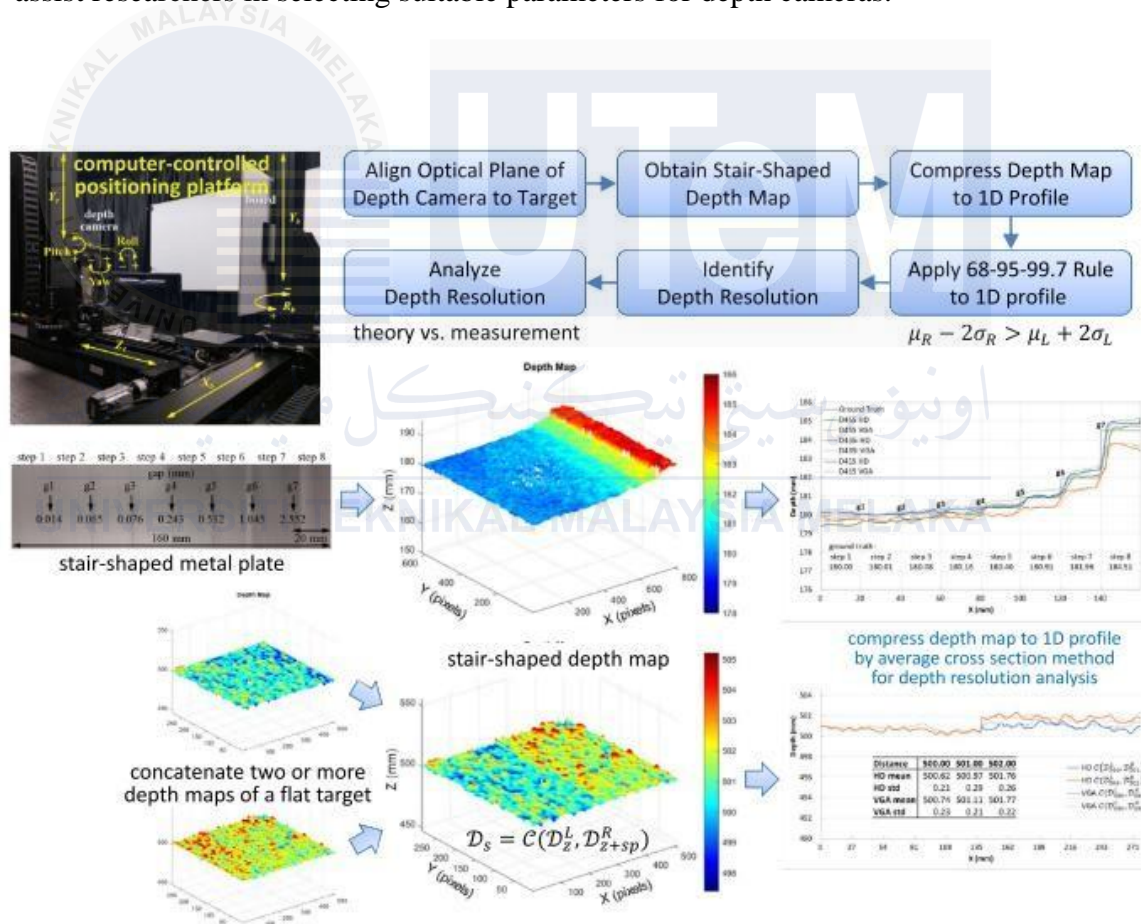


Figure 2.25 SFM stages [8]

Then, work by [25] discusses the challenges in measuring depth resolution accurately, especially for raw depth data without post-processing. Many existing studies focus on depth accuracy, but there is a lack of research on depth resolution measurement standards. This gap in the literature is to conduct systematic studies and explore methods for measuring depth resolution. The study aims to design a method that can gradually change depth differences until the smallest one is found, which differs from measuring depth accuracy by comparing average values. The proposed method involves using a flat target and moving a depth camera at specified steps to create a stair-shaped depth map for

analysis. The research also highlights the importance of understanding the theoretical depth resolution in stereo camera systems. The depth resolution is defined as the smallest depth difference that can be detected by a depth camera. The study analyzes the factors influencing depth resolution, such as stereo baseline, field of view, image sensor, and dot projector, using the Intel RealSense D400 series depth cameras for verification. Overall, the literature review emphasizes the need for accurate methods to measure depth resolution, derive theoretical depth resolution, and compare it with measured results to assist researchers in selecting suitable parameters for depth cameras.



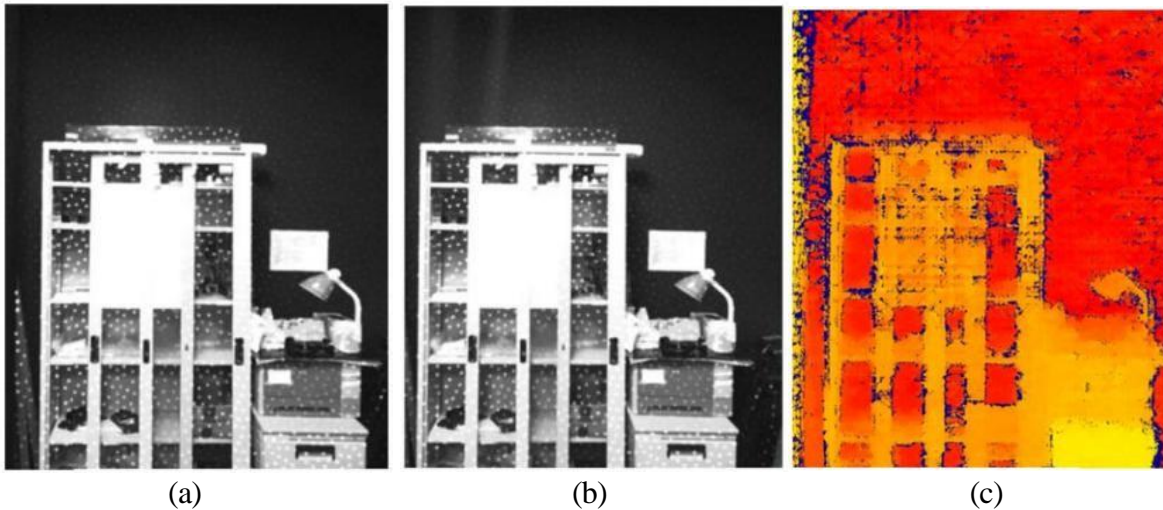
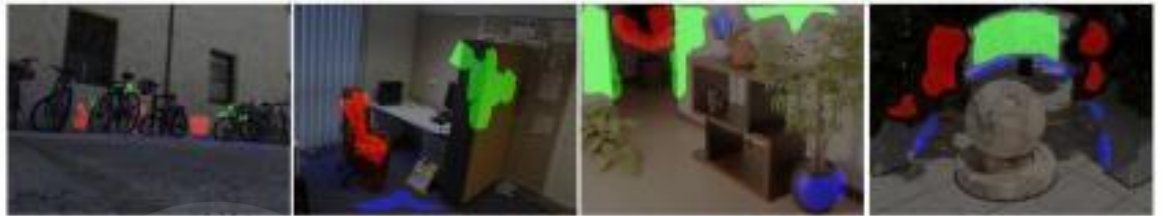


Figure 2.27 Depth map computed by our MATLAB code without post-processing, (a) left image, (b) right image, (c) depth map [25]

In addition, work by [10] explore into the evolution of depth estimation techniques, highlighting the transition from traditional stereo matching methods to the emergence of deep learning approaches. Traditional methods, while effective in certain scenarios, face challenges in handling occlusions, featureless regions, and repetitive patterns. In contrast, deep learning techniques have shown remarkable progress in addressing these limitations, leading to significant advancements in applications such as autonomous driving and augmented reality. By formulating the depth estimation problem as a learning task, researchers aim to develop predictors that can infer depth maps closely resembling ground truth data. The work also emphasizes the importance of datasets, network architectures, training strategies, and performance evaluation in advancing the field. Through a comprehensive classification and comparative analysis of key methods, the survey provides valuable insights into the current state-of-the-art techniques and directions in stereo-based depth estimation using deep learning.



(a)



(b)



(c)



$d \in [9.3, 34.0]$      $d \in [18.7, 29.9]$      $d \in [5.6, 14.5]$      $d \in [5.5, 13.2]$   
 Depth  $\in [2.1, 7.8]$     Depth  $\in [2.4, 3.3]$     Depth  $\in [7.8, 25.0]$     Depth  $\in [10.8, 25.6]$

(d)

Figure 2.28 Four images, collected in-house and used to test 16 state-of-the art methods. The green masks on some of the left images highlight the pixels where the ground-truth disparity is available. The disparity range is shown in pixels while the depth range is in meters.  $d$  refers to disparity, (a) Left image, (b) Highlight of regions of interest where ground-truth disparity is estimated with high confidence, (c) Right image, (d) Ground-truth disparity maps [10]

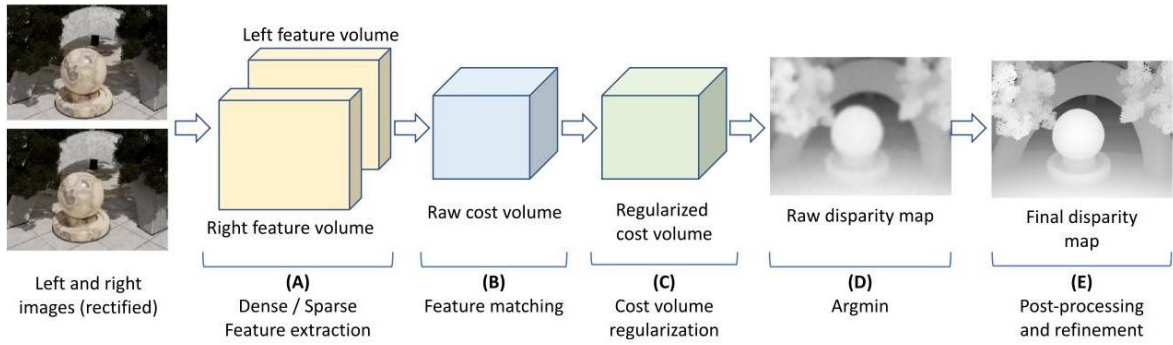


Figure 2.29 The building blocks of a stereo matching pipeline [10]

Finally, work by [26] delves into the integration of optimization techniques with stereo matching algorithms to enhance the accuracy and efficiency of depth estimates. The work emphasizes the significance of stereo vision algorithms in accurately estimating the real-world depth of pixels within two-dimensional images by utilizing multiple two-dimensional views of a scene, a process known as stereo vision. The project work are actively exploring the application of optimization techniques, such as Dynamic Programming, to improve stereo system performance and enhance depth estimation accuracy. Camera calibration is highlighted as a critical step in stereo vision systems to correct image distortion caused by camera lenses and ensure accurate representation of two-dimensional views on the same image plane. Triangulation, the process of assigning depth values to pixels using multiple views of a scene and considering camera hardware parameters, camera locations, and pixel disparities, is essential for accurate depth estimation. The work proposed on real-time stereo vision applications for road surface reconstruction, academic resources on stereo vision concepts, and materials discussing camera calibration and stereo vision technology. Overall, the work provides a solid foundation for understanding the research and the advancements in stereo vision algorithms and optimization techniques for image depth estimation.

left image							right image						
3	5	4	4	2	4	2	3	5	4	4	2	4	2
7	4	1	4	4	2	6	7	4	1	4	4	2	6
2	7	46	46	46	6	7	46	46	46	3	6	6	7
5	9	46	46	44	9	7	48	46	44	6	4	9	7
4	7	47	47	47	2	4	47	47	47	7	4	2	4
4	7	56	56	46	6	7	58	56	46	5	6	6	7
3	4	4	1	4	3	2	3	4	4	1	4	3	2

$$\begin{aligned}
 & (46-48)^2 + (46-46)^2 + (44-44)^2 + \\
 & (47-47)^2 + (47-47)^2 + (47-47)^2 + \\
 & (56-58)^2 + (56-56)^2 + (46-46)^2 = 8
 \end{aligned}$$

Figure 2.30 Sample Window Cost Calculation (Veksler) [26]

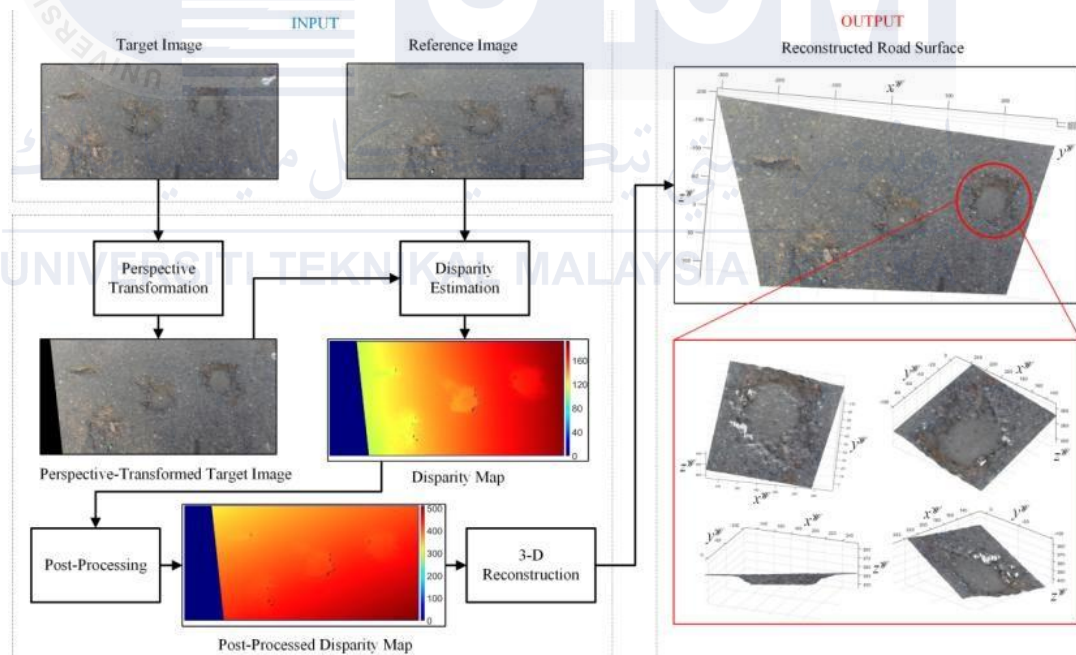


Figure 2.31 Stereo Vision for Road Deformity Detection (Rui) [26]

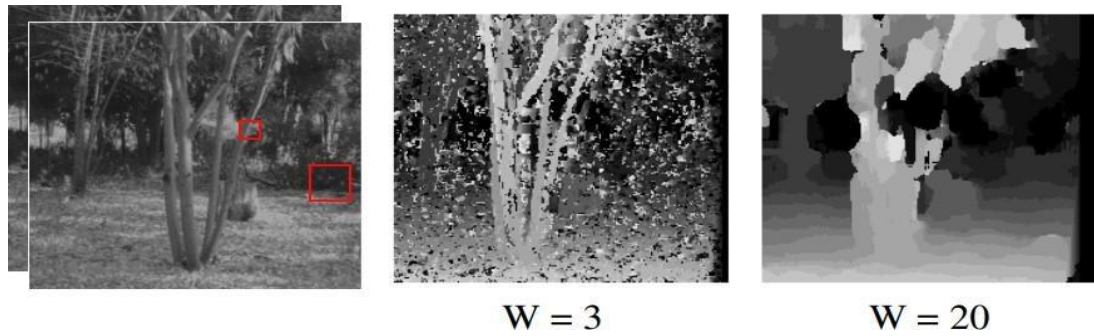


Figure 2.32 Window size compared to depth map (darker pixel shade refers to greater depth)

(Olga) [26]

## 2.6 Table comparison of stereo vision

Table 2.1 summarize and compares all existing methods that have been reviewed in chapter 2. That including the technique which is for image enhancement, accuracy, reduce noise also improve textureless and process for measuring the pixels in stereo image pairs.

**Table 2.1 Table comparison of stereo vision**

Year	Author	Description
2020	Hua Liu, Rui Wang, Yuanping Xia, and Xiaoming Zhang	The work by [1] proposed a stereo matching technique that combines gradient and census transform methods for better accuracy in low-texture areas. It uses an adaptive filter to handle exposure and illumination changes, and collects information from neighboring pixels for better disparity estimates, improving reliability and precision in challenging conditions.
2021	Zhou Shi, Lifeng Zhang, Zhen Li	The work by [16] proposed a stereo matching method that enhances disparity maps by utilizing neural networks to analyze image patches. This technique improves depth accuracy, particularly in challenging areas, and has application in 3D reconstruction and autonomous driving.
2021	A. F. Kadmin, R. A. Hamzah, M. N. Abd Manap, M. S. Hamid, T. F. Tg. Wook	The work by [17] proposed a local stereo matching algorithm enhances depth accuracy by using a modified census transform with a dynamic histogram for cost computation, a fixed-window strategy with bilateral filtering for edge

		preservation, and a winner-takes-all optimization with consistency checks and adaptive filtering, improving disparity map estimation.
2018	Rostam Affendi Hamzah, A. Fauzan Kadmin, M. Saad Hamid, S. Fakhra A. Ghani, Haidi Ibrahim	The work by [4] introduces a method to improve accuracy and reduce noise in 3D surface reconstruction. It uses three steps: SG for estimating differences, ASW iGF for sharpening edges, and JWGF for less noise. This method competes well with others, suggesting it could enhance 3D reconstruction.
2021	Weimin Yuan, Cai Meng, Xiaoyan Tong, Zhaoxi Li	The work by [18] introduces that F-GDGIF is a tool that fixes halo effects around sharp edges in images. It does this by preserving edges better and speeding up the process with a trick called sub-sampling. When used in stereo matching, it improves accuracy while saving time. It's a useful tool for improving 3D imaging and autonomous driving.
2022	Rostam Affendi Hamzah, Muhd Nazmi Zainal Azali, Zarina Mohd Noh, Madiha Zahari, Adi Irwan Herman	The work by [19] introduces a better way to make depth maps from stereo images, improving accuracy in tough spots. It involves steps like calculating matches and refining the map for better results. This can help in self-driving cars and 3D modeling, making navigation systems better in different areas of

		tech.
2021	Huei-Yung Lin, Senior Member, Chun-Lung Tsai, and Van Luan Tran	The work by [20] proposed that depth measurement system rotates one camera to take pictures from different angles. This method is simpler and cheaper than using multiple cameras. It measures depth by comparing features in these pictures. Using different angles helps make the depth measurements more accurate.
2021	A. F. Kadmin, R. A. Hamzah, N. A. Manap, and M. S. Hamid	The work by [21] introduces the algorithm makes color images better by using CLAHE and AGCWD with guided filtering to smooth and sharpen them. The Census Transform helps match features accurately despite color changes. The WTA method picks the best matching points quickly, making the disparity map more accurate.
2019	Mostafa Mansour, Pavel Davidson, Oleg Stepanov and Robert Piché	The work by [22] introduce that checked how well binocular disparity and motion parallax help measure depth for humans and computers. They tested their accuracy in measuring distances to still objects. Binocular disparity uses multiple cameras, while motion parallax depends on movement. They compared these methods to see which is better for depth perception. They wanted to know

		how crucial each method is for seeing depth.
2020	Atiqul islam, md. Asikuzzaman, mohammad omar khyam, md. Noor-a-rahim, and mark r. Pickering	The work by [23] discuss using special cameras to track people outdoors in 3D. It's cheaper and more adaptable than traditional methods. The goal is to create a program that accurately follows people's movements outside, which regular cameras struggle with. This work emphasizes the importance of better 3D tracking, especially in real-world scenarios.
2019	Yu Feng, Paul Whatmough, and Yuhao Zhu	The work by [24] state that the ASV system mixes old and new techniques in stereo vision. It speeds up by using DNNs to find matches between frames and save power. It improves energy efficiency by handling common problems in stereo DNNs. In short, ASV makes stereo vision systems faster and more energy-efficient.
2021	M kholil	The work by [8] state that SFM and MVS create 3D models from images. SFM figures out image geometry for 3D scenes, while MVS uses many views for precise modeling. This tech is key for preserving historic sites, making accurate 3D models for conversation and virtual displays of cultural treasures

2021	Te-Mei Wang, Member, and Zen-Chung Shih	The work by [25] explains how deep learning improves depth estimation from stereo images, replacing older techniques. It deals with challenges like camera settings and lighting by learning features. It also discusses creating 3D models for datasets and the need for detailed depth maps. In short, it's a useful guide to deep learning's impact on stereo depth estimation.
2022	Hamid Laga, Laurent Valentin Jospin, Farid Boussaid, and Mohammed Bennamoun, Senior Member,	The work by [10] talks about how deep learning improves depth estimation from stereo images, replacing older techniques. It deals with challenges like camera settings and lighting by learning features. It also discusses creating 3D models for datasets and the need for detailed depth maps. In short, it's a useful guide to deep learning's impact on stereo depth estimation.
2022	Mrinall Umasudhan	The work by [26] looks at stereo vision, which measures distances using two views. It compares it to other sensors and deals with challenges like the pixel correspondence problem. It discusses methods to improve accuracy and efficiency, like window-based SSD and dynamic programming. It also gives practical tips on camera calibration and pixel matching.

## 2.7 Summary

In summary, based on the literature review, there are several key strategies to enhance stereo matching performance and depth estimation accuracy. Firstly, leveraging deep learning techniques, such as the proposed Invariant-based Stereo Matching (ISM) algorithm, can significantly improve speed, accuracy, and energy efficiency in stereo vision systems. Additionally, integrating optimization techniques with stereo matching algorithms, as demonstrated in studies focusing on pixel correspondence and depth estimation, can further enhance system performance. Furthermore, improving traditional stereo matching algorithms through dynamic cost computation methods, modified census transforms, and WTA optimization can lead to better accuracy and efficiency in depth estimation. Developing new algorithms like SG, ASW, and Joint Weighted Guided Filter (JWGF) can address radiometric differences and edge distortions, resulting in high-accuracy 3D surface reconstructions. Moreover, utilizing efficient matching cost functions and adaptive shape guided filter-based methods can improve stereo matching in textureless regions, ensuring robustness against radiometric variations and textureless areas. Finally, exploring multi-view stereo (MVS) methods and fusing multiple depth maps can enhance the precision and detail level of 3D reconstructions, while utilizing Structure from Motion algorithms can optimize the 3D reconstruction process. By implementing these strategies, researchers can advance stereo matching algorithms and depth estimation techniques for various applications in computer vision and related fields.

## CHAPTER 3

### METHODOLOGY

#### 3.1 Introduction

This chapter explains the methods used to reach the final result. This comprehensive introduction explores block diagrams, methods, and devices. The methodology focuses on using MATLAB to measure the depth estimation of the 3D image. It follows a step-by-step approach that involves choosing a project title, conducting research, planning the project design, integrating hardware and software, analyzing the project, solving any issues, and writing a report. A flowchart is prepared to visualize the project's steps. It is important to understand the hardware and software resources before starting the project. This chapter emphasizes both the design technique and the overall equipment and research process flow.

#### 3.2 Implementing Sustainable Stereo Vision Algorithms: Prioritizing Energy Efficiency and Resource Optimization

Choosing and assessing tools for the sustainable development of MATLAB stereo vision algorithms entails giving lifecycle management, resource efficiency, and energy efficiency top priority. Effective algorithm creation and optimization are made possible by MATLAB's assortment of toolboxes and functions, like the Parallel Computing Toolbox and the Computer Vision Toolbox. With these, programmers may put methods into practice and evaluate them, concentrating on lowering computational overhead via parallel processing and code optimization. Scalable and energy-efficient processing is made possible by MATLAB's support for hardware acceleration (such as the Graphics Processing Unit (GPU)) and interaction with cloud computing services. Furthermore, the extensive documentation

and community resources provided by MATLAB facilitate ongoing education and the implementation of sustainable development best practices, guaranteeing that the stereo vision algorithms are both efficient and ecologically conscious.

### **3.3 Methodology**

The goal of this project is to derive a new stereo vision algorithm from a stereo camera that can produce the depth map measurement. The methodology aims to calculate precise depth information of object in the scene, to provide a comprehensive 3D representation and analysis of the scene, to enhance visualization capabilities, to achieve efficient and effective depth estimation, and to provide a foundation for application in fields such as computer vision, robotics, autonomous vehicles and virtual reality. By incorporating this technology, the outcome will be analysed and optimized for a better system.

The flowchart outlines a systematic approach for developing and validating a stereo vision algorithm using an adaptive guided filter to improve depth map measurement. The process begins with collecting data through a Literature Review, aimed at understanding stereo vision, guided filters, depth map measurement is analyzed to identify gaps, and deriving insights to inform the design of the proposed algorithm. Based on this foundation, the Algorithm Development phase is initiated, where the adaptive guided filter is coded and integrated into a stereo vision pipeline. Initial testing is conducted using a standard online dataset, such as Middlebury, which provides stereo image pairs and ground truth depth maps for performance evaluation. The evaluation results are analyzed to assess metrics like disparity error and depth accuracy, leading to an iterative cycle of Parameter Tuning. This refinement process adjusts key parameters, such as the window size and regularization factors, to optimize the algorithm's performance.

Once the algorithm achieves the desired criteria, it undergoes real-world validation in the Real-Time Image Testing phase, where its robustness is tested against challenges like noise, lighting variations, and occlusions. Data collected during these experiments is analyzed to assess practical applicability and finalize the development process. The workflow ensures a structured and iterative process, combining rigorous evaluation using benchmark datasets with real-world testing. By iterating between parameter optimization and validation, the project ensures that the algorithm is robust, accurate, and capable of handling real-time challenges, ultimately contributing to significant advancements in stereo vision and depth map estimation.

The flowchart emphasizes an iterative, step-by-step approach, where the process loops back to earlier stages if performance gaps are identified. This ensures that the algorithm is thoroughly tested and refined before being finalized. By combining rigorous evaluation on standard datasets with real-world testing, the project ensures that the stereo vision algorithm is both theoretically robust and practically applicable. This approach not only addresses gaps in existing methods but also contributes to advancements in depth estimation, with potential applications in robotics, autonomous vehicles, and augmented reality.

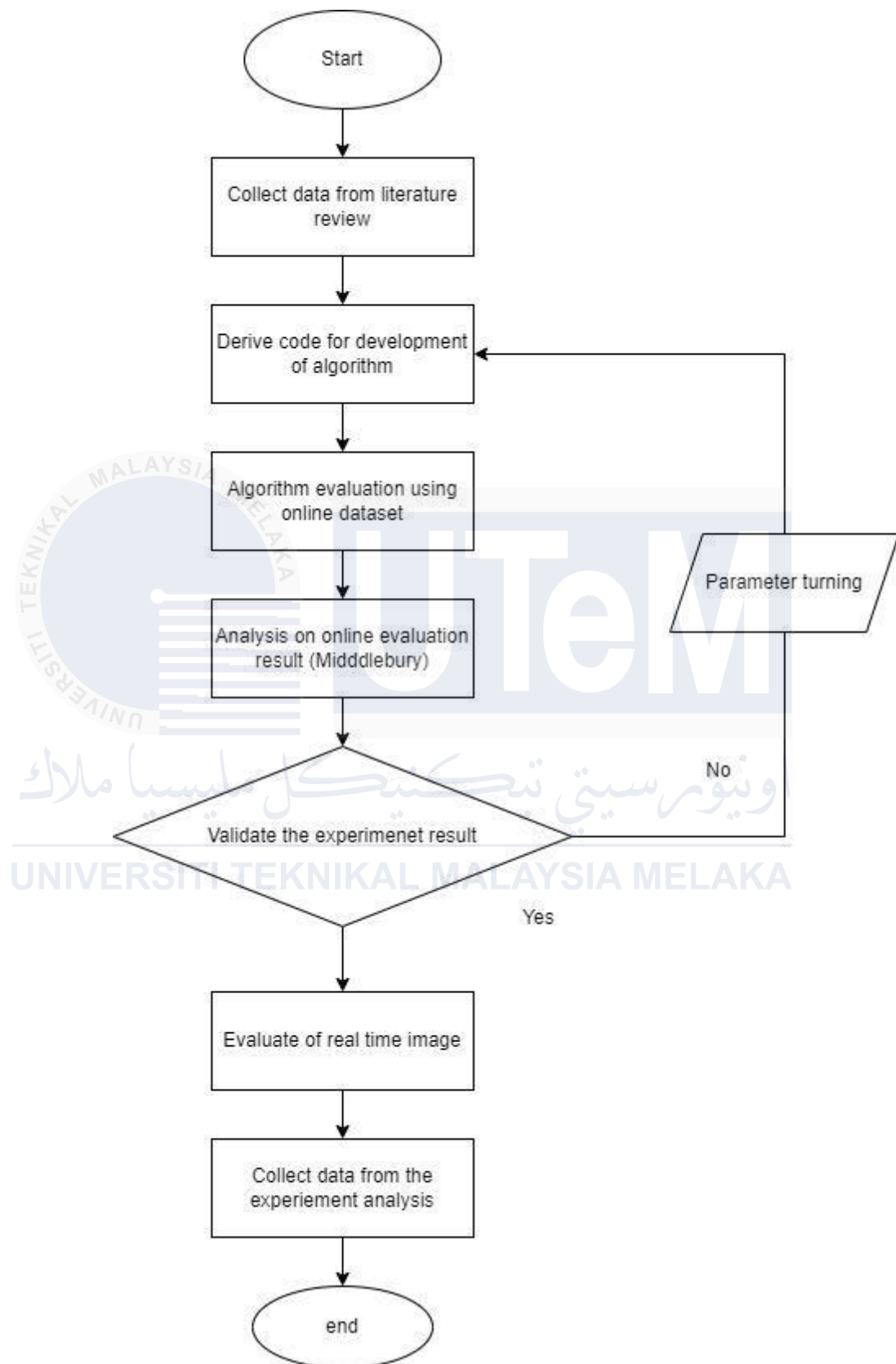


Figure 3.1 Flowchart of project

### 3.4 Block diagram

The workflow of the proposed stereo vision algorithm has been processed through five stages and the block diagram of the algorithm is represented in Figure 3.2. Initially, the left image and right image are two inputs captured by two cameras positioned at different viewpoints. Because the cameras are positioned apart, they capture the scene from slightly different angles. This disparity between viewpoints is what enables the perception of depth. Objects in the scene appear at different positions in the left and right images. The difference in the position of an object in the two images is called disparity.

The algorithm begins with pre-processing stage to prepared the image pair and enhance the colour image quality for the main disparity computation process with combination of guided filter method. Then, the matching cost volume for stereo vision disparity estimation using a combination of color and gradient information, applying truncated Sum of Absolute Differences (SAD). The third stage will produce the cost aggregation on the disparity cost volume using a guided filter. Cost aggregation helps to smooth the disparity cost volume while preserving edge details. At the fourth stage is disparity optimization, the disparity map is optimized with a common local technique, WTA. Then, for final stage, the process continues with post processing that is LRC checking.

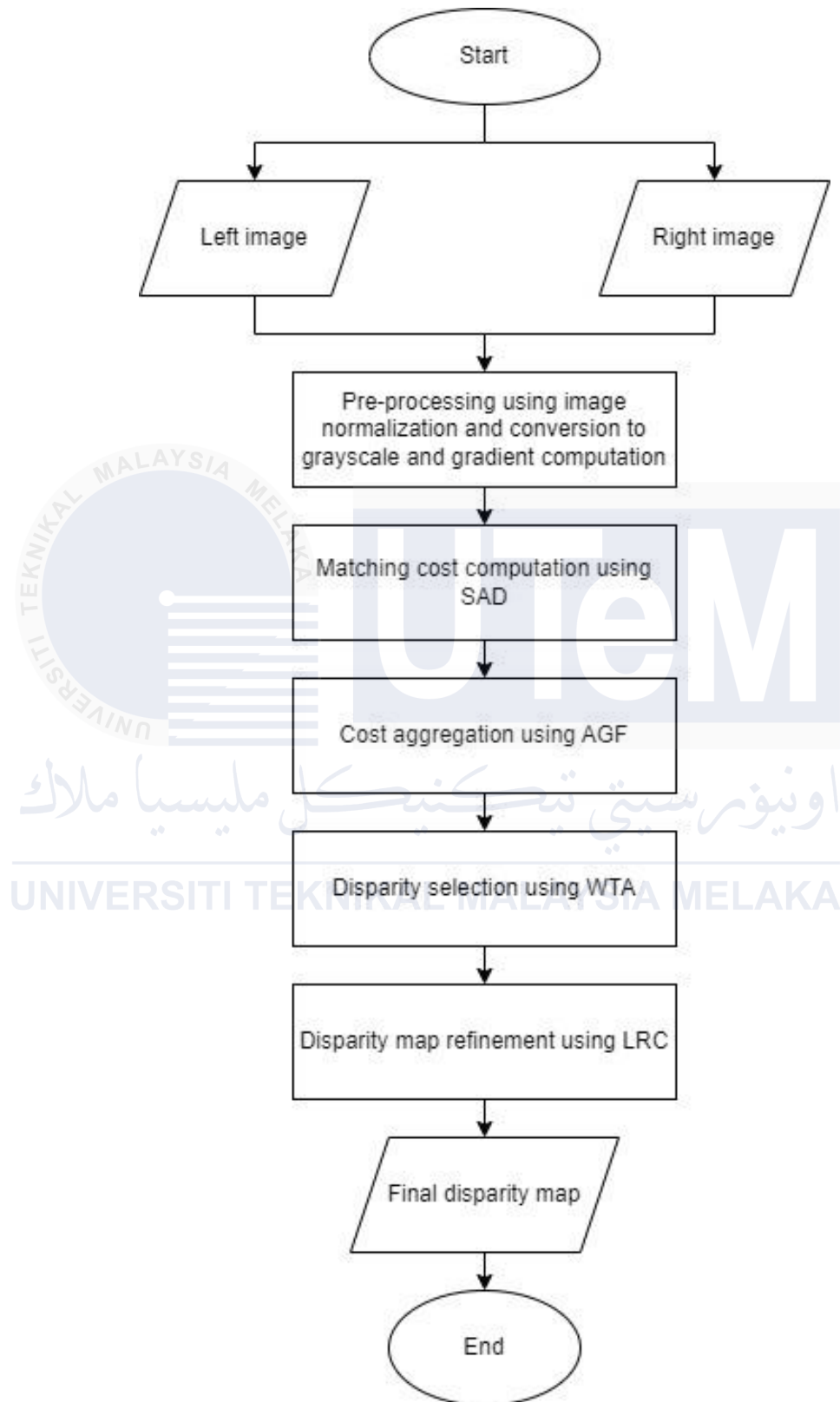


Figure 3.2 flowchart of stereo vision system framework

### 3.4.1 Pre-processing

In the pre-processing stage of the proposed stereo vision algorithm, the left and right images are first normalized to a range of  $[0, 1]$  by converting their pixel values into floating-point numbers and dividing by 255. This scaling step ensures that all image values are within a consistent range for further processing. The next step involves converting the images to grayscale, which simplifies the data by removing color information while retaining intensity variations that are important for disparity estimation. Gradients of the grayscale images are then computed in the x-direction to capture edge information, which is helpful for matching corresponding pixels between the two images. These gradient values are adjusted to fall within a normalized range of  $[0, 1]$  to make them comparable. The disparity range is also determined based on the resolution of the stereo images, ensuring that the system is set up to handle the expected depth variations in the scene. This pre-processing ensures that the images are appropriately prepared for the disparity matching process by normalizing pixel values, extracting edge information, and setting the necessary parameters for accurate disparity estimation.

### 3.4.2 Matching cost

The matching cost computation is performed using a combination of truncated SAD for both color and gradient images. This method is used to determine the similarity between corresponding pixels in the left and right stereo images across different disparity levels. For each disparity, the right image is shifted relative to the left image, and the absolute difference between pixel intensities is calculated. The color-based matching cost is computed by averaging the absolute differences across the RGB channels and applying a truncation threshold ( $\tau_{ao1}$ ) to limit extreme values. Similarly, the gradient-based cost is computed using the absolute

difference between gradient images, with another truncation threshold ( $\tau_2$ ). The final matching cost is computed as a weighted combination of color and gradient costs using the parameter  $\alpha$ . This ensures that both intensity and structural information are considered in the disparity estimation, enhancing robustness in textureless or highly detailed areas. The computed disparity cost volume ( $\text{dispVol}$ ) stores the matching cost values for all disparity levels, which will later be refined through cost aggregation and optimization.

### **3.4.3 Cost aggregation**

The cost aggregation using the guided filter method. After computing the initial matching cost, the disparity cost volume is refined to improve accuracy by smoothing it with a guided filter. The guided filter takes the original left and right images as guidance to preserve important edges and structures while reducing noise in the disparity cost volume. This helps in achieving more consistent and reliable disparity estimations, especially in regions with texture variations or occlusions. The filtering process is applied separately to each disparity level, resulting in a refined cost volume that is better suited for disparity optimization. This method effectively balances edge-preserving smoothing with computational efficiency.

### **3.4.4 Disparity selection**

The disparity selection step performed using the WTA method. After the cost aggregation step, the disparity with the minimum cost is selected for each pixel by identifying the disparity index that corresponds to the lowest cost in the disparity cost volume. This approach assumes that the disparity value with the smallest matching cost is the most likely to be correct. The WTA method is computationally efficient and widely used in stereo vision tasks to produce an initial disparity map. However, due to potential mismatches and

occlusions, additional post-processing steps are applied to refine the disparity map and improve accuracy.

#### **3.4.4 Disparity refinement**

The disparity refinement involves several post-processing techniques to enhance the accuracy of the initial disparity map. The method used for refinement includes a LRC check, which helps detect occlusions and mismatches by comparing the left and right disparity maps. If a significant difference is found between corresponding pixels, the disparity is marked as unreliable. To handle occlusions, the code applies an occlusion filling technique by propagating valid disparity values from neighbouring pixels, first from the left and then from the right. Finally, a GF is used to smooth the disparity map while preserving edge details, helping to reduce noise and enhance the final output. These refinement techniques collectively improve the quality of the disparity estimation by addressing occlusions, mismatches, and ensuring better alignment with the image structure.

#### **3.4.5 Disparity Map**

The disparity refinement involves several post-processing techniques to enhance the accuracy of the initial disparity map. The method used for refinement includes a LRC check, which helps detect occlusions and mismatches by comparing the left and right disparity maps. If a significant difference is found between corresponding pixels, the disparity is marked as unreliable. To handle occlusions, the code applies an occlusion filling technique by propagating valid disparity values from neighbouring pixels, first from the left and then from the right. Finally, a GF is used to smooth the disparity map while preserving edge details, helping to reduce noise and enhance the final output. These refinement techniques collectively improve the quality of the disparity estimation by addressing

occlusions, mismatches, and ensuring better alignment with the image structure.

### 3.5 Measurement setup

In the stereo matching quality evaluation and several quantitative parameters were used to analyze the data for the disparity map accuracy. The parameter included bad pixel error, non-occluded error and all error. An experiment is carried out on the platform of Window 10 on desktop PC with 3.2GHz processor and 8GB memory. To evaluate the accuracy, the experimental images are using a standard online benchmarking dataset from the Middlebury. The experiments are conducted using MATLAB. The accuracy is measured from the bad pixel percentage of non-occluded pixel (nonocc) and all pixels (all). Figure 3.3 shows the 15 training images, ground truth and the disparity map result after submitted to Middlebury online platform. Based on the final results of disparity maps, the scene objects situated at increasing depth are assigned step by step to disparity values from nearer to further based on the colours image.

Based on the figure 3.3 show the Middlebury dataset provides a diverse collection of stereo image pairs designed to evaluate stereo vision algorithms across various real-world scenarios. The Adirondack represents the colours which represent distance warm colors for closer objects and cool colors for farther ones. Resolution affects detail, with higher resolution giving clearer images. Disparity range shows pixel shifts between two images; bigger shifts mean closer objects. The image is to measure depth and it is useful for 3D mapping and robotics which is make it ideal for testing algorithms on occlusion handling, depth estimation, and managing complex image. The Artl is an indoor scene filled with intricate textures, such as paintings and decorative elements. With a high resolution of 1400x1100 pixels and the disparity appears varied, with areas of high contrast suggesting significant depth changes. This range is suited for evaluating an algorithm's ability to distinguish between close and far objects., it

evaluates algorithms' ability to handle textureless regions and maintain consistency on flat surfaces. The Jadeplant features a close-up of a plant with distinct textures and depth variations. Approximately 1200x1000 pixels, moderate to high resolution. It tested accuracy in capturing fine details in natural textures. It has medium disparity range, with variations depending on leaf positions. Motorcycle depicts a scene in an outdoor or garage scene with motorcycles, featuring smooth surfaces and reflections. The high resolution, typically around 1400x1100 pixels with wide disparity range, due to depth differences between foreground motorcycles and background structures to handle robustness in handling reflective surfaces and large disparity gradients. MototcycleE with reduced resolution (700x550 pixels) and possibly additional noise. The disparity range is narrower due to lower detail and cropping. It challenges algorithms to perform under constrained conditions with limited resolution or disparity variation. Piano in an indoor scene with a piano, sharp edges, and dark tones, captured at a high resolution of around 1400x1100 pixels. It has a moderate disparity range, testing edge detection and depth accuracy in high-contrast areas.

PianoL in a lower-resolution of approximately 700x550 pixels, with reduced details and a narrower disparity range. This image tests the robustness of algorithms when processing degraded input, emphasizing performance in real-world scenarios with computational constraints. Pipes in an industrial scene with cylindrical pipes arranged at various angles, captured at a high resolution of about 1500x1100 pixels. With a wide disparity range, this image evaluates the algorithm's performance on repetitive patterns and complex geometries. Playroom in an indoor scene with toys and clutter, showcasing vibrant colors and varied depths. At a resolution of around 1400x1100 pixels and a wide disparity range, it tests the algorithm's ability to handle occlusions, complex object arrangements, and color variations.

Playtable in a tabletop scene with objects, combining smooth and textured surfaces. Captured at a high resolution of approximately 1500x1100 pixels, it has a moderate disparity

range and tests depth estimation accuracy for both smooth and textured areas. PlaytableP in a lower-resolution about 700x550 pixels with reduced details and a narrower disparity range. It highlights algorithm performance under resource-constrained conditions, where high-resolution data is unavailable. Recycle in a scene featuring cluttered recycling bins or materials with textured surfaces. With a high resolution of about 1400x1100 pixels and a medium-to-wide disparity range, it evaluates algorithm robustness in cluttered and repetitive environments. Shelves in a structured scene of shelves filled with books or items, involving complex patterns. Captured at a high resolution of approximately 1500x1000 pixels, it has a moderate disparity range and tests the algorithm's handling of repetitive patterns in organized spaces.

Teddy in an indoor scene featuring a teddy bear with soft, textured fur, captured at a high resolution of around 1400x1100 pixels. With a wide disparity range, this image tests depth accuracy in textured regions and on challenging soft edges. Vintage in a scene with vintage items such as old furniture or decorations, characterized by varied textures and lighting. With a resolution of approximately 1400x1100 pixels and a wide disparity range, it evaluates the algorithm's adaptability to diverse textures and lighting conditions.

This explanation highlights how each dataset from the Middlebury collection is tailored to test different aspects of stereo vision algorithms, such as resolution sensitivity, disparity handling, and adaptability to varying scene complexities.

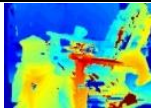
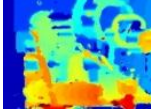
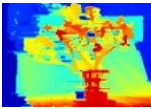
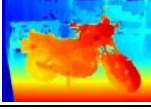
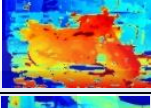
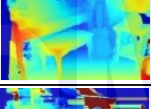
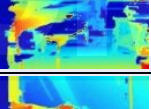
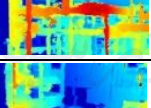
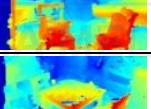
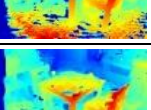
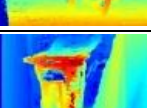
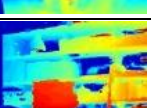
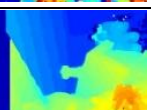
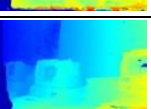
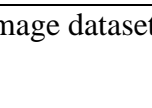
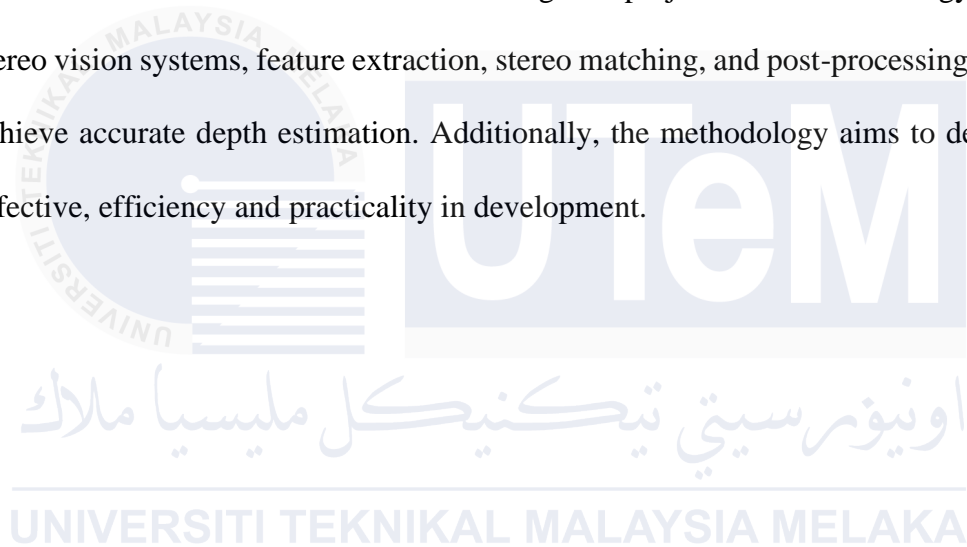
Image Type	Proposed Method
Adirondeck	
Art1	
Jadeplant	
Motorcycle	
MotorcycleE	
Piano	
PianoL	
Pipes	
Playroom	
Playtable	
PlaytableL	
Recycle	
Shelves	
Teddy	
Vintage	

Figure 3.3 Middlebury 15 image dataset

### 3.6 Summary

In summary, this chapter focuses on utilizing MATLAB for depth estimation in 3D pictures. It follows a systematic approach starting from selecting a project title to conducting research, planning the project design, integrating hardware and software, analyzing the project, addressing any challenges, and documenting the findings. A flowchart visually represents the project steps, emphasizing the importance of understanding both hardware and software resources before commencing the project. The methodology incorporated stereo vision systems, feature extraction, stereo matching, and post-processing techniques to achieve accurate depth estimation. Additionally, the methodology aims to develop a new, effective, efficiency and practicality in development.



## CHAPTER 4

### RESULTS AND DISCUSSIONS

#### 4.1 Introduction

This chapter discussed about the result of local stereo matching algorithm and discussed results based on the image quality and disparity map evaluation. Several experiments have been done in the stereo matching framework to prove the concept of disparity accuracy produced from stereo image pairs for low texture regions, discontinuities, occlusion and radiometric distortion.

#### 4.2 Result

The result outcome of this project work is the development of a highly accurate and robust stereo vision algorithm that can effectively estimate depth information from stereo images, even in challenging scenarios involving low-texture regions, occlusions, and varying illumination conditions. The algorithm will demonstrate superior performance compared to existing traditional methods, particularly in resolving depth information in textureless or low-texture regions where many current algorithms struggle. Additionally, the algorithm will exhibit strong robustness against various challenging factors, such as radiometric variations, illumination changes, occlusions, and depth discontinuities, while still producing accurate depth estimates.

Moreover, the developed algorithm should maintain computational efficiency and practical applicability, enabling real-time or near-real-time depth estimation suitable for practical applications. It demonstrates versatility and the ability to generalize to different types of stereo images and application domains, performing well in diverse real-world

scenarios. The algorithm's performance will be thoroughly validated using standard benchmark datasets and metrics, demonstrating its superiority over existing methods through quantitative measures (e.g., percentage of bad pixels, error rates) and qualitative visual assessments. By achieving these outcomes, the developed stereo vision algorithm would contribute significantly to advancing the field of depth estimation and 3D reconstruction, enabling improved performance in various applications that rely on accurate depth information, such as robotics, autonomous vehicles, augmented reality, and computer vision tasks.

#### **4.2.1 Stereo matching analysis**

The experiment is carried out on the platform of Window 10 on desktop PC with 3.2GHz processor and 8GB memory. To evaluate the accuracy, the experimental images are using a standard online benchmarking dataset from the Middlebury. This dataset consists of 15 training images. The accuracy is measured from the bad pixel percentage of non-occluded pixel (nonocc) and all pixels (all). For the result analysis below, the final result is PROPOSED METHOD called ABCD to address the challenges of stereo vision depth estimation. The performance of the PROPOSED METHOD is compared with two existing methods, R-NCC and DF, which are well-established methods from the Middlebury dataset. The comparison is conducted to evaluate the effectiveness and robustness of our method across various scenes and disparity scenarios. By benchmarking against these methods, the superiority and strengths of ABCD are demonstrated in handling diverse datasets and challenging conditions. Figure 4.1 shows the diagram of the corresponding stereo matching algorithm work. The diagram illustrates the process of estimating depth from a stereo image pair. It begins with pre-processing, where the images undergo normalization, grayscale conversion, and gradient computation. Then, the matching cost is determined using the Sum of Absolute Differences (SAD) to compare

pixel values between the two images. After that, cost aggregation is applied with GF to refine the disparity estimates by reducing noise. In the disparity optimization phase, the WTA technique selects the optimal disparity value for each pixel. Lastly, disparity refinement is performed through a LRC check to eliminate errors, producing the final disparity map, where different colors indicate depth variations.

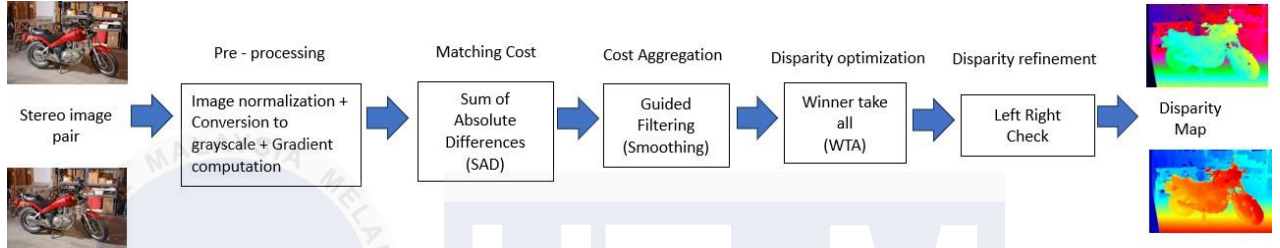


Figure 4.1 Stereo matching diagram

Figure 4.2 shows the 15 training images with using window size  $r = 3$ , disparity map for purposed method after submitted to Middlebury online platform. The qualitative assessment of the disparity maps shows that the PROPOSED METHOD outperforms R-NCC and DF in several critical areas, including edge preservation, texture processing, occlusion handling, and maintaining clarity in blurred and unblurred regions. Regarding edge quality, the PROPOSED METHOD delivers sharper and more distinct edges, whereas R-NCC and DF often produce blurred or poorly defined boundaries. When dealing with textures, the proposed method proves more effective in low-texture regions by minimizing noise and inconsistencies, while R-NCC generates noisier results and DF offers moderate improvements but still exhibits noticeable artifacts.

In terms of occlusion handling, the proposed method significantly reduces errors and artifacts near occluded areas, by using the techniques like LRC checks and refined disparity adjustment. In comparison, R-NCC and DF struggle with mismatches and inaccuracies around occlusions. Moreover, the PROPOSED METHOD achieves an

excellent balance between smoothing and preserving details, resulting in clean and precise disparity maps that retain essential details in textured regions without introducing excessive blurring.

Overall, the PROPOSED METHOD using adaptive guided filter(AGF) -based approach enhances the accuracy and robustness of stereo matching, particularly in challenging conditions such as low-texture and occluded areas where traditional methods tend to struggle. The AGF helps to refine the disparity map by preserving edges and details while smoothing out noise in areas with little texture. This results in more precise depth estimates, even in regions that are difficult to interpret using conventional techniques. The approach also handles occlusions effectively, reducing errors. In essence, the adaptive guided filter enhances the method's ability to produce clear and accurate disparity maps across a wide range of real-world scenarios.

$r$  (window size) = 3

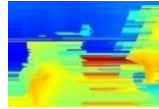
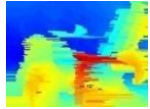
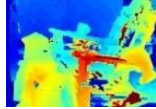

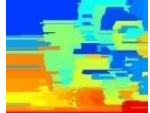
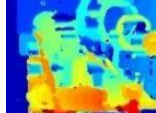
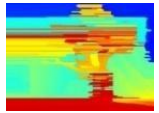
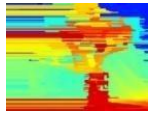
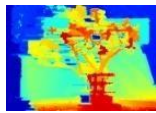
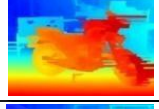
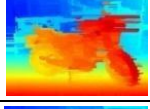
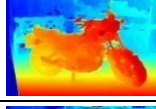
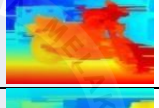
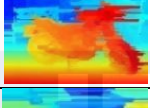
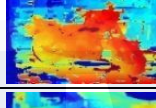
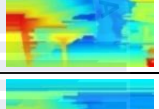
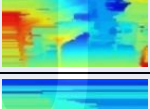
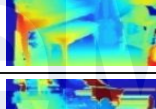
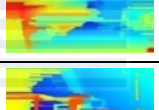
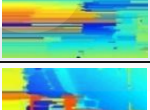
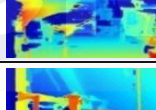
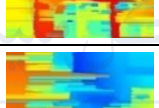
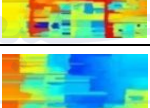
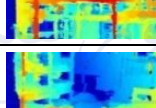
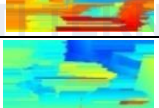
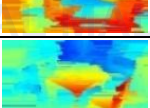
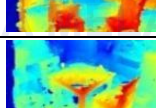
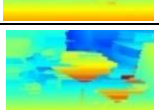
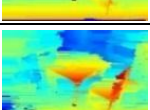
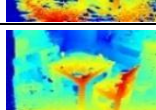

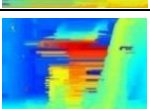
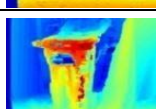
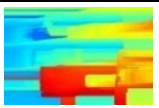
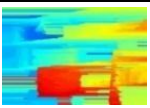
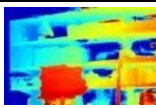
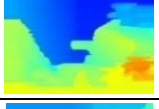
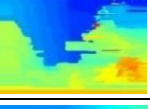
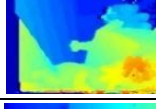


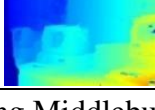
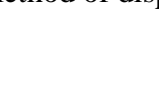
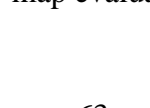
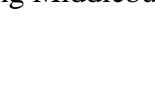
Image type	R - NCC	DF	PROPOSED METHOD
Adirondeck			
Artl			
Jadeplant			
Motorcycle			
MotorcycleE			
Piano			
PianoL			
Pipes			
Playroom			
Playtable			
PlaytableP			
Recycle			
Shelves			
Teddy			
Vintage			

Figure 4.2 Different method of disparity map evaluation using Middlebury training dataset for  $r = 3$

Table 4.1 presents a comparison of error rates for three methods R-NCC, DF, and ABCD evaluated on the Middlebury dataset using a window size of  $r = 3$ . The analysis includes both non-occluded (nonocc) and all-pixel (all) error metrics. Among the three methods, ABCD demonstrates the best overall performance, achieving the lowest weighted average error for non-occluded regions (14.2%) and all-pixel regions (22.3%). This represents a significant improvement in non-occluded error approximately 28.2% compared to R-NCC (from 19.8% to 14.2%) and 25.9% compared to DF (from 19.2% to 14.2%).

While the non-occluded error shows considerable improvement, the gains in the "all" error metric are less pronounced, with a reduction of about 2.6% compared to R-NCC and 2% compared to DF. This smaller improvement is due to the inherent challenges of handling occluded regions, where mismatches and ambiguities are more prevalent. The ABCD method effectively improves accuracy in non-occluded areas by leveraging adaptive guided filtering, which enhances results in regions with distinct textures and edges. However, regions with occlusions, low textures, or high disparity variations remain challenging, limiting the overall improvement.

Performance varies across individual images. For example, images like Adiron and ArtL achieve better accuracy, likely because they contain well-defined textures and clear edges, which benefit from the ABCD method's adaptive filtering. Conversely, images such as Jadeplant and MotorcycleE show higher error rates due to factors like complex occlusions, reflective surfaces, or low-texture regions, which are more difficult to handle.

Compared to R-NCC, the ABCD method provides significantly improved depth estimation in areas with clear textures and edges, thanks to its use of adaptive filtering that balances edge preservation with smoothing. While DF performs better than R-NCC in certain cases, it lacks the adaptive refinement capability of ABCD, resulting in higher errors in

complex scenarios. Overall, ABCD excels in improving accuracy in non-occluded regions, but further refinement is needed to better handle occlusions and textureless areas.

Table 4.1 The comparison results of nonocc error and all error using Middlebury dataset for  $r = 3$

Image type	R - NCC		DF		ABCD	
	nonocc	all	nonocc	all	nonocc	all
Adiron	20.5	21.2	13.2	14.1	18.9	21.2
Artl	10.0	12.5	16.4	18.2	15.6	30.7
Jadeplant	67.2	91.0	77.8	103	26.3	43.2
Motorcycle	9.59	11.5	11.2	13.2	5.19	12.4
MotorcycleE	10.6	12.7	10.7	12.7	22.0	27.4
Piano	9.12	9.59	10.5	11.1	10.2	15.4
PianoL	15.8	15.8	26.4	26.4	35.2	38.7
Pipes	21.8	27.9	16.1	22.5	9.84	22.0
Playroom	29.0	30.0	19.6	20.9	15.1	32.4
Playtable	18.0	17.5	13.3	13.9	18.5	24.7
PlaytableL	13.1	13.0	14.8	16.3	6.86	14.1
Recycle	22.3	22.2	16.2	16.8	7.51	9.61
Shelves	11.5	11.7	11.1	11.5	19.8	21.6
Teddy	4.13	4.81	5.04	6.16	6.22	16.3
Vintage	44.3	45.1	24.9	26.8	10.3	16.9
Weight Avg	19.8	22.9	19.2	22.7	14.2	22.3

Figure 4.3 shows an evaluation of R-NCC, DF, and the PROPOSED METHOD for disparity map estimation on the Middlebury training dataset with a window size of  $r=5$  reveals notable differences in performance across key qualitative factors such as edges, textures, occlusions, and blur/unblur regions. The PROPOSED METHOD stands out as the most effective, delivering sharp and well-defined edges with minimal artifacts, ensuring accurate preservation of depth discontinuities. It also handles textured areas exceptionally well, producing smooth and coherent disparity maps with minimal noise, even in regions with complex patterns.

In occluded regions, the PROPOSED METHOD proves robust by reducing mismatches and maintaining smooth transitions at occlusion boundaries. Additionally, it demonstrates consistent performance in both blurred and unblurred regions, preserving details

and ensuring reliable outputs. DF performs moderately well, showing improvements over R-NCC, but it struggles in blurred regions and occasionally produces less precise results in textured and occluded areas. R-NCC, on the other hand, shows the weakest performance, with significant artifacts, poor edge accuracy, and noisy outputs, particularly in challenging regions such as textures, occlusions, and blurred areas.

Overall, the PROPOSED METHOD proves to be the most reliable and accurate, consistently outperforming R-NCC and DF across all qualitative aspects. The evaluation on the Middlebury training dataset reveals clear advantages of the proposed method across key factors such as edge preservation, texture handling, occlusion management, and performance in both blurred and unblurred regions. The PROPOSED METHOD excels by delivering sharp, well-defined edges, minimizing artifacts, and handling textures and occlusions with great precision. It maintains smooth transitions at occlusion boundaries and preserves details even in blurred regions. In comparison, while DF shows moderate improvements over R-NCC, it still struggles with blurred and textured regions, and R-NCC performs the weakest overall

$r$  (window size) = 5

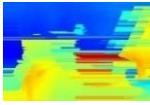
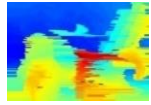
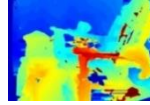

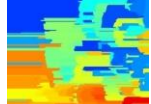
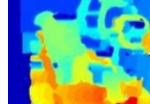
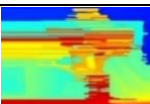
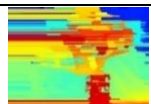
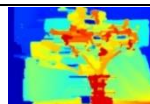
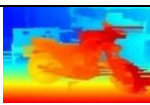
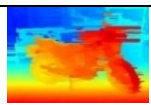
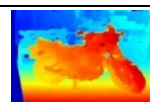
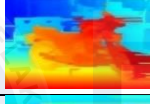
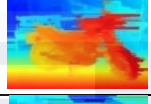
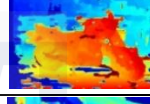
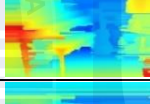
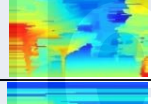
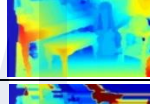
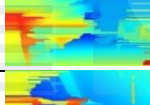
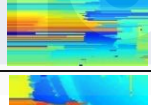
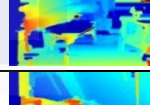
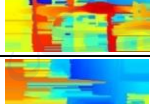
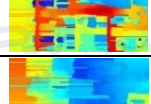
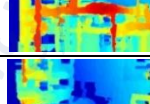
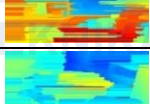
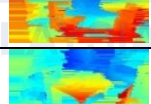
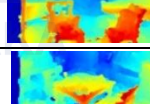
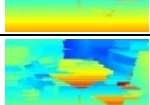
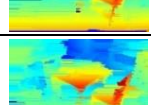
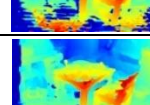
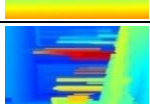
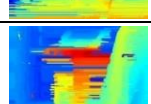
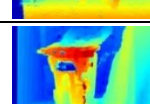
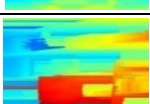
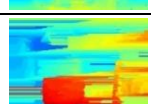
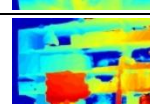
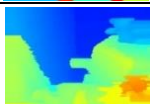
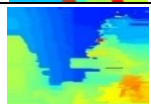
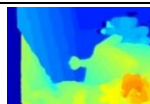

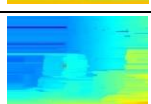

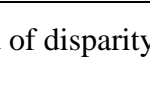
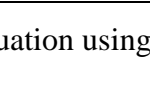
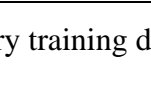
Image type	R -NCC	DF	PROPOSED METHOD
Adirondeck			
Art1			
Jadeplant			
Motorcyle			
MotorcyleE			
Piano			
PianoL			
Pipes			
Playroom			
Playtable			
PlaytableP			
Recycle			
Shelves			
Teddy			
Vintage			

Figure 4.3 Different method of disparity map evaluation using Middleburry training dataset for  $r = 5$

Table 4.2 evaluates the performance of R-NCC, DF, and ABCD methods on the Middlebury dataset with a window size of  $r = 5$ , focusing on non-occluded (nonocc) and all-pixel (all) error rates. The ABCD method achieves the best results, with the lowest weighted average error rates of 15.2% for non-occluded regions and 23.4% for all regions, showcasing its ability to effectively handle both non-occluded and occluded areas. R-NCC, on the other hand, exhibits the highest error rates, with 19.8% for non-occluded and 22.9% for all regions, reflecting its weaker ability to handle textures and occlusions. DF performs better than R-NCC but does not match the refinement and precision of ABCD. ABCD demonstrates a substantial improvement in non-occluded errors, with reductions of approximately 23.2% compared to R-NCC and 20.8% compared to DF. However, in the all-pixel error metric, ABCD performs slightly worse, likely due to the complexities introduced by occluded regions. Images like Teddy, Motorcycle, and Adiron show better results because of their simpler structures, clearer edges, and fewer occlusions, making them easier to process.

Meanwhile, more complex images like Jadeplant and PlaytableL have higher error rates due to their intricate textures, significant occlusions, and reflective surfaces, which are challenging for all methods. Non-occluded regions generally yield lower error rates as they lack depth ambiguities and occlusions, allowing ABCD to perform more effectively. However, the presence of occluded regions in the all-pixel metric introduces mismatches and ambiguities that limit ABCD's overall improvement.

Overall, ABCD proves to be the most accurate and reliable method, particularly for non-occluded regions, though further enhancement is needed to improve its performance in occluded and complex scenarios.

Table 4.2 The comparison results of nonocc error and all error using Middlebury dataset for  $r = 5$

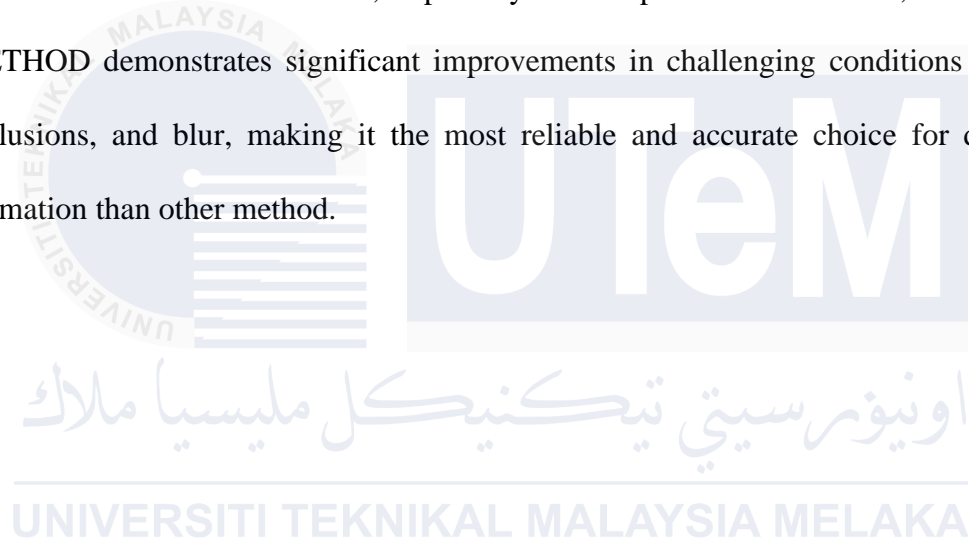
Image type	R - NCC		DF		ABCD	
	nonocc	all	nonocc	all	nonocc	all
Adiron	20.5	21.2	13.2	14.1	16.5	18.9
Artl	10.0	12.5	16.4	18.2	15.5	31.2
Jadeplant	67.2	91.0	77.8	103	31.9	48.0
Motorcycle	9.59	11.5	11.2	13.2	6.26	13.5
MotorcycleE	10.6	12.7	10.7	12.7	20.2	25.8
Piano	9.12	9.59	10.5	11.1	10.6	15.9
PianoL	15.8	15.8	26.4	26.4	33.1	36.7
Pipes	21.8	27.9	16.1	22.5	13.3	25.9
Playroom	29.0	30.0	19.6	20.9	15.3	33.6
Playtable	18.0	17.5	13.3	13.9	26.4	31.9
PlaytableL	13.1	13.0	14.8	16.3	8.17	15.5
Recycle	22.3	22.2	16.2	16.8	7.55	9.71
Shelves	11.5	11.7	11.1	11.5	18.7	20.5
Teddy	4.13	4.81	5.04	6.16	7.11	17.2
Vintage	44.3	45.1	24.9	26.8	11.7	18.4
Weight Avg	19.8	22.9	19.2	22.7	15.2	23.4

Figure 4.4 evaluates the performance of three disparity map methods—R-NCC, DF, and the PROPOSED METHOD on various image types, emphasizing factors like edges, textures, occlusions, and blur. The PROPOSED METHOD stands out as the most effective, producing sharper edges and smoother transitions in textured areas, showcasing its capability to preserve fine details.

In contrast, R-NCC often fails to maintain edge clarity, resulting in noisy or blurred outputs, while DF shows moderate improvements but falls short of the accuracy achieved by the PROPOSED METHOD. The PROPOSED METHOD also handles occluded regions better, minimizing artifacts and errors that are more prominent in the other approaches. Although all methods perform relatively well in non-occluded regions, the PROPOSED METHOD delivers the best results in complex areas. When addressing blurred regions, it

maintains a high level of detail and accuracy, unlike R-NCC and DF, which struggle in such scenarios.

The evaluation of R-NCC, DF, and the PROPOSED METHOD across various image types highlights the superior performance of the proposed approach. It consistently produces sharper edges, smoother textures, and better handling of occlusions and blurred regions compared to the other methods. While R-NCC struggles with edge clarity and noise, and DF offers only moderate improvements, the PROPOSED METHOD effectively preserves fine details and minimizes artifacts, especially in complex areas. Overall, the PROPOSED METHOD demonstrates significant improvements in challenging conditions like textures, occlusions, and blur, making it the most reliable and accurate choice for disparity map estimation than other method.



r (window size) = 7

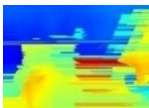
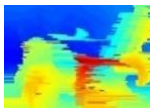
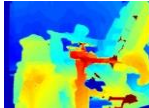

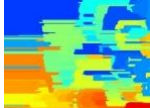
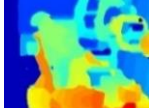
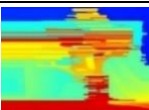
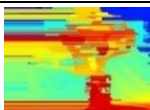
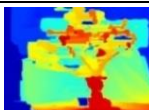
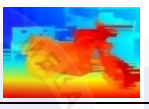
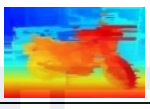
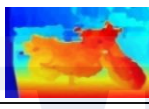
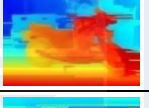
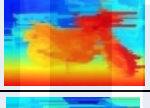
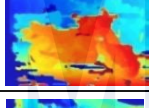
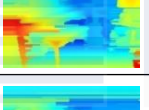
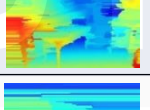
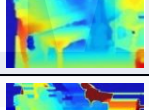
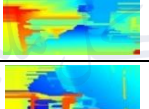
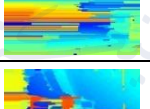
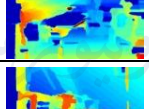
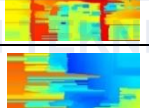
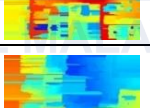
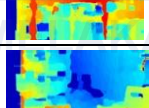
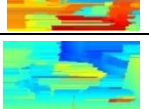
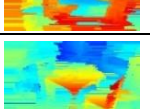
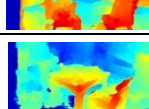
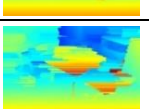
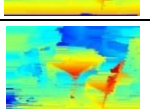
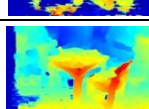
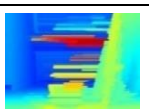
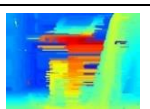
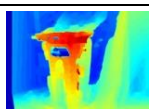
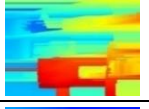
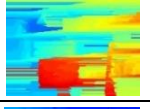
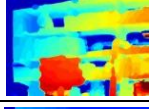
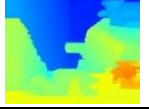
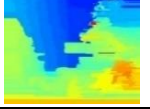
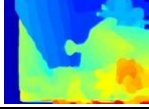

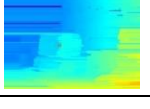
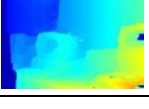

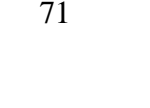

Image type	R - NCC	DF	PROPOSED METHOD
Adirondeck			
Artl			
Jadeplant			
Motorcyle			
MotorcyleE			
Piano			
PianoL			
Pipes			
Playroom			
Playtable			
PlaytableP			
Recycle			
Shelves			
Teddy			
Vintage			

Figure 4.4 Different method of disparity map evaluation using Middlebury training dataset for  $r = 7$

Table 4.3 compares the performance of three disparity estimation methods R-NCC, DF, and ABCD by evaluating their non-occluded (nonocc) and all-pixel (all) error rates on the Middlebury dataset with a window size of  $r = 7$ . Among these methods, ABCD achieves the best results, with the lowest weighted average errors of 17.0% for non-occluded regions and 25.1% for all-pixel regions. This marks a notable improvement over R-NCC, with a 14.1% reduction in non-occluded error and a 9.7% reduction in all-pixel error, and over DF, with respective improvements of 11.5% and 7.1%. ABCD's effectiveness is largely due to its adaptive guided filtering, which balances edge preservation and noise reduction, making it particularly accurate in non-occluded areas.

However, the gains in all-pixel error are smaller because occluded regions pose greater challenges, such as mismatches and depth ambiguities. Simpler images like Teddy and MotorcycleE show lower error rates because of their straightforward structure and minimal occlusion, while more complex images like Jadeplant and Pipes exhibit higher error rates due to intricate textures, reflective surfaces, and significant occlusions. Compared to R-NCC and DF, ABCD shows superior performance, particularly in non-occluded regions, although improvements in handling occluded and textureless areas are still needed.

Overall, ABCD demonstrates strong accuracy and reliability, especially in regions with distinct edges and textures, while occluded areas remain a key challenge.

Table 4.3 The comparison results of nonocc error and all error using Middlebury dataset for  $r = 7$

Image type	R - NCC		DF		ABCD	
	nonocc	all	nonocc	all	nonocc	all
Adiron	20.5	21.2	13.2	14.1	15.4	18.0
Artl	10.0	12.5	16.4	18.2	17.5	33.2
Jadeplant	67.2	91.0	77.8	103	39.2	54.1
Motorcycle	9.59	11.5	11.2	13.2	7.36	14.6
MotorcycleE	10.6	12.7	10.7	12.7	21.2	26.8
Piano	9.12	9.59	10.5	11.1	11.2	16.5
PianoL	15.8	15.8	26.4	26.4	32.6	36.4
Pipes	21.8	27.9	16.1	22.5	16.1	28.7
Playroom	29.0	30.0	19.6	20.9	16.6	35.4
Playtable	18.0	17.5	13.3	13.9	36.6	41.0
PlaytableL	13.1	13.0	14.8	16.3	10.6	17.8
Recycle	22.3	22.2	16.2	16.8	7.97	10.2
Shelves	11.5	11.7	11.1	11.5	18.0	20.0
Teddy	4.13	4.81	5.04	6.16	7.73	17.8
Vintage	44.3	45.1	24.9	26.8	12.5	19.1
Weight Avg	19.8	22.9	19.2	22.7	17.0	25.1

Figure 4.5 how a comparison of disparity map evaluation methods R-NCC, DF, and the PROPOSED METHOD—on the Middlebury training dataset with a window size of  $r = 9$  highlights clear differences in performance across qualitative factors like edges, texture, occlusion, and blur/unblur regions. The PROPOSED METHOD stands out for its ability to preserve sharp and precise edges, minimizing artifacts and maintaining fine boundary details. It also performs exceptionally well in textured areas, generating smooth and coherent disparity maps that align with the scene's patterns, whereas DF and R-NCC struggle with noise and inconsistencies in more complex textures.

For occluded regions, the PROPOSED METHOD is robust, reducing artifacts and ensuring continuity, whereas DF achieves moderate performance but falters in more challenging scenarios. R-NCC, however, exhibits significant artifacts and errors in occluded

areas. Additionally, the PROPOSED METHOD effectively handles both blurred and unblurred regions, maintaining detail and producing smooth transitions, unlike DF, which struggles with blurred areas, and R-NCC, which shows the poorest performance with noisy and inaccurate outputs.

Overall, the PROPOSED METHOD outperforms R-NCC and DF across all qualitative aspects, proving to be the most accurate and reliable approach for disparity map estimation.



$r$  (window size) = 9

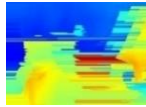
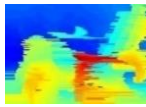
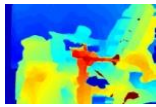
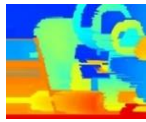
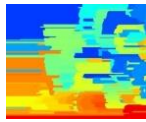
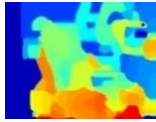
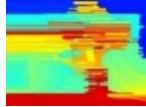
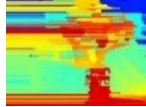
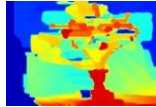
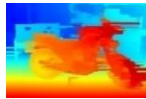
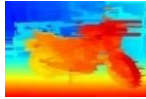
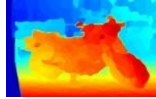
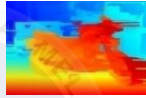
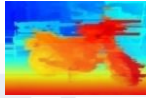
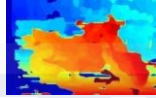
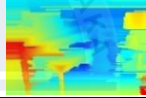
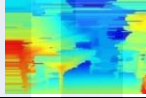
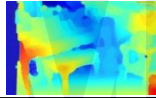
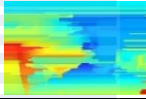
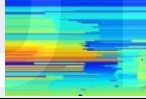
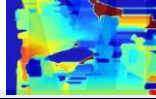
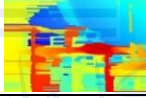
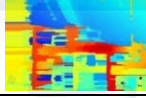
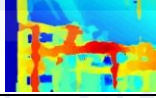
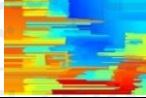
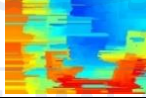
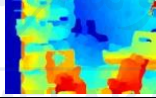
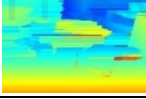
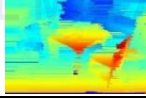
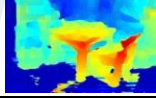
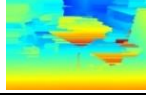
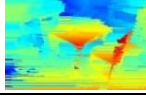
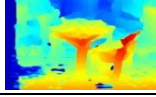
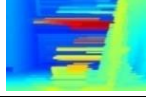
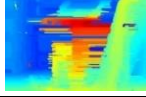
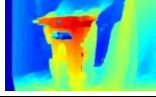
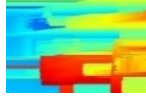
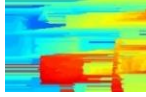
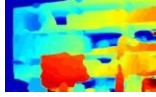
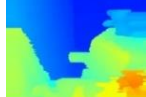
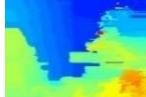
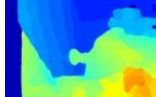


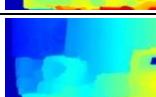
Image type	R – NCC	DF	PROPOSED METHOD
Adirondeck			
Artl			
Jadeplant			
Motorcycle			
MotorcycleE			
Piano			
PianoL			
Pipes			
Playroom			
Playtable			
PlaytableP			
Recycle			
Shelves			
Teddy			
Vintage			

Figure 4 .5 Different method of disparity map evaluation using Middlebury training dataset for  $r = 9$

Table 4.4 evaluates the non-occluded (nonocc) and all-pixel (all) error rates of the R-NCC, DF, and ABCD methods using the Middlebury dataset with a window size of  $r = 9$ . The ABCD method outperforms the others, achieving the lowest weighted average error rates of 18.7% for non-occluded regions and 26.7% for all-pixel regions. In contrast, R-NCC has the highest errors, with 19.8% for non-occluded and 22.9% for all-pixel regions, making it the least effective. ABCD provides a 5.6% improvement over R-NCC and a 2.6% improvement over DF in non-occluded error.

However, the improvement in the all-pixel error metric is smaller, as ABCD struggles more with occluded regions compared to R-NCC and DF. Images such as "Teddy" and Motorcycle perform well due to their simpler structures and minimal occlusions, while more complex images like Jadeplant and Pipes have higher error rates due to significant occlusions and intricate textures. R-NCC shows the weakest performance in handling textures and occlusions, while DF offers moderate improvement but lacks the refinement of ABCD. Despite ABCD's strength in preserving textures and edges in non-occluded areas, its performance is hindered in occluded regions due to challenges like mismatches and ambiguities.

Overall, ABCD demonstrates superior accuracy and reliability in non-occluded regions but requires further optimization to improve its handling of occluded and complex areas.

Table 4.4 The comparison results of nonocc error and all error using Middlebury dataset for  $r = 9$

Image type	R – NCC		DF		ABCD	
	nonocc	all	nonocc	all	nonocc	all
Adiron	20.5	21.2	13.2	14.1	15.5	18.1
Artl	10	12.5	16.4	18.2	19.1	34.8
Jadeplant	67.2	91	77.8	103	45.8	59.6
Motorcycle	9.59	11.5	11.2	13.2	8.49	15.7
MotorcycleE	10.6	12.7	10.7	12.7	22.9	28.5
Piano	9.12	9.59	10.5	11.1	10.8	16.2
PianoL	15.8	15.8	26.4	26.4	32.4	36.3
Pipes	21.8	27.9	16.1	22.5	18.7	31.1
Playroom	29	30	19.6	20.9	17.1	35.9
Playtable	18	17.5	13.3	13.9	43.0	46.7
PlaytableL	13.1	13	14.8	16.3	14.2	21.0
Recycle	22.3	22.2	16.2	16.8	8.51	10.8
Shelves	11.5	11.7	11.1	11.5	17.2	19.2
Teddy	4.13	4.81	5.04	6.16	8.46	18.5
Vintage	44.3	45.1	24.9	26.8	13.7	20.3
Weight Avg	19.8	22.9	19.2	22.7	18.7	26.7

Figure 4.5 show the comparison of disparity maps generated using the proposed method for different window sizes ( $r = 3, 5, 7, 9$ ) highlights noticeable differences in quality based on image features such as edges, textures, occlusion handling, and blur/unblur regions. For  $r = 3$ , the smallest window size, the disparity maps retain sharp, well-defined edges, allowing precise boundary detection while preserving details in high-texture areas. Although the smaller context may result in minor artifacts in occluded regions, the adaptive guided filter helps to minimize these effects. Unblurred areas are accurately reconstructed without excessive smoothing, making  $r = 3$  highly suitable for applications that prioritize detail and precision.

With  $r = 5$ , edges become slightly smoother, and fine details in textured areas begin to decrease because the method averages a larger area of pixels, which reduces sharp transitions. This results in a loss of fine details in textured areas, as the method blends subtle variations in texture and depth over a bigger region.  $R = 5$  balances smoothing with detail preservation, though some fine texture details may be lost. However, occlusion handling

improves due to the larger spatial context. For  $r = 7$ , the balance between smoothness and sharpness is more noticeable, with edges becoming less distinct and textures losing considerable detail. Although occlusion handling is further enhanced, the disparity maps display increased blurriness because the larger window effectively averages out more pixel values, which blurs fine details, particularly in textured regions. At  $r = 9$ , the largest window size, the edges are heavily smoothed, textures are overly averaged, and important details are lost. While occlusion handling reaches its peak consistency, the disparity maps become overly blurred and lack the clarity necessary for accurate depth estimation.

In conclusion, the disparity maps produced with  $r = 3$  offer the best compromise, preserving edge sharpness and texture detail while minimizing blur and effectively handling occlusions. Larger window sizes, although beneficial for occlusion consistency and noise reduction, sacrifice precision and clarity. This makes  $r = 3$  the most suitable choice for applications that demand high accuracy and robustness in low-texture regions.

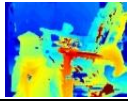
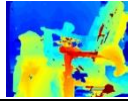
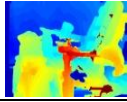
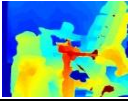
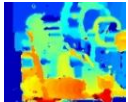
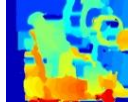
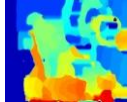
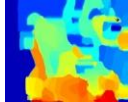
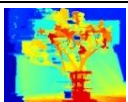
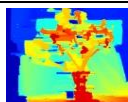
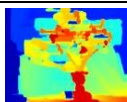
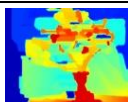
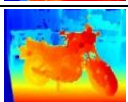
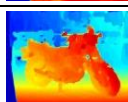
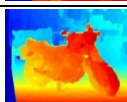
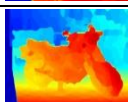
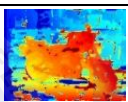
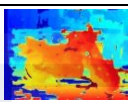
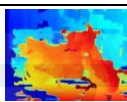
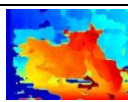
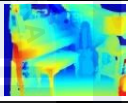
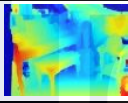
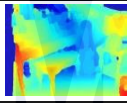
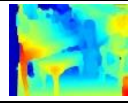
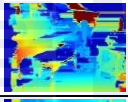
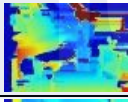
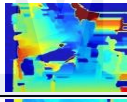
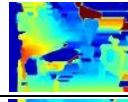
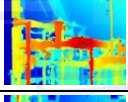
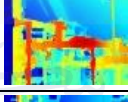
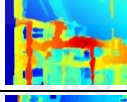
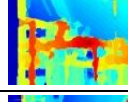
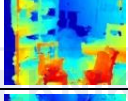
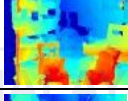
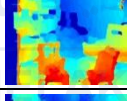
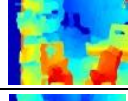
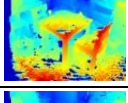
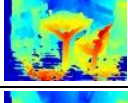
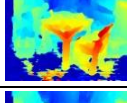
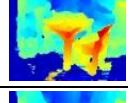
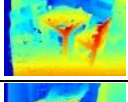
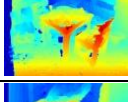
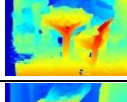
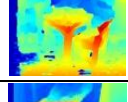
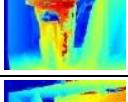
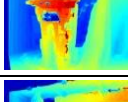
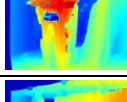
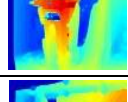
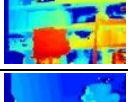
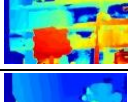
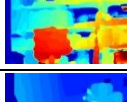
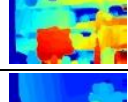
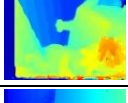
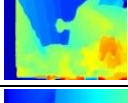
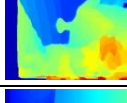
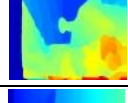


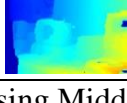
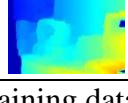
	$r = 3$	$r = 5$	$r = 7$	$r = 9$
Image type	proposed method	proposed method	proposed method	proposed method
Adirondeck				
Art1				
Jadeplant				
Motorcyle				
MotorcyleE				
Piano				
PianoL				
Pipes				
Playroom				
Playtable				
PlaytableL				
Recycle				
Shelves				
Teddy				
Vintage				

Figure 4.6 Different  $r$  (window size) of disparity map evaluation using Middlebury training dataset

Table 4.5 show the comparison table highlights the disparity errors (nonocc and all error) for various window sizes ( $r = 3, 5, 7, 9$ ) using the Middlebury dataset. Among all configurations,  $r = 3$  demonstrates the best performance with the lowest weighted average nonocc error of 14.2% and all-error of 22.3%. In contrast,  $r = 9$  yields the highest errors, with a nonocc error of 18.7% and an all-error of 26.7%. The results show a significant improvement when using  $r = 3$ , particularly when compared to  $r = 9$ . The percentage improvement for nonocc error is approximately 24.1%, and for all-error, it is about 16.5%. This highlights the advantage of smaller window sizes in reducing errors and enhancing disparity accuracy.

For specific images, those with sharp edges and minimal occlusion, such as Motorcycle and Teddy, achieve low errors with  $r = 3$ , as the small window size excels in preserving fine details. Conversely, complex images like Jadeplant and Playtable have higher errors due to factors such as intricate textures, occlusion, and challenging lighting conditions, which introduce ambiguity and make disparity estimation more difficult, even with smaller window sizes.

When comparing all window sizes,  $r = 3$  consistently outperforms the others due to its ability to maintain fine details and minimize errors, especially in non-occluded regions. Larger window sizes, such as  $r = 7$  and  $r = 9$ , result in oversmoothing, which blurs edges and reduces accuracy. While nonocc error consistently shows better results across all window sizes because it focuses on regions without occlusion, the all-error improvement is less pronounced due to the inherent challenges in handling occluded areas, such as disparity mismatches and ambiguity.

In conclusion,  $r = 3$  offers the best balance between minimizing errors and preserving image details, making it the optimal choice for applications requiring high disparity accuracy. Larger window sizes, while reducing noise, compromise accuracy by oversmoothing, and losing edge detail, particularly in challenging images. Despite the

challenges in occluded regions,  $r=3$  strikes the best balance for both nonocc and all-error metrics

Table 4.5 The comparison results of nonocc error and all error using Middlebury dataset for different  $r$

	$r = 3$		$r = 5$		$r = 7$		$r = 9$	
	ABCD		ABCD		ABCD		ABCD	
Image type	nonocc	all	nonocc	all	nonocc	all	nonocc	all
Adirondeck	18.9	21.2	16.5	18.9	15.4	18	15.5	18.1
Artl	15.6	30.7	15.5	31.2	17.5	33.2	19.1	34.8
Jadeplant	26.3	43.2	31.9	48	39.2	54.1	45.8	59.6
Motorcycle	5.19	12.4	6.26	13.5	7.36	14.6	8.49	15.7
MotorcycleE	22	27.4	20.2	25.8	21.2	26.8	22.9	28.5
Piano	10.2	15.4	10.6	15.9	11.2	16.5	10.8	16.2
PianoL	35.2	38.7	33.1	36.7	32.6	36.4	32.4	36.3
Pipes	9.84	22	13.3	25.9	16.1	28.7	18.7	31.1
Playroom	15.1	32.4	15.3	33.6	16.6	35.4	17.1	35.9
Playtable	18.5	24.7	26.4	31.9	36.6	41	43	46.7
PlaytableL	6.86	14.1	8.17	15.5	10.6	17.8	14.2	21
Recycle	7.51	9.61	7.55	9.71	7.97	10.2	8.51	10.8
Shelves	19.8	21.6	18.7	20.5	18	20	17.2	19.2
Teddy	6.22	16.3	7.11	17.2	7.73	17.8	8.46	18.5
Vintage	10.3	16.9	11.7	18.4	12.5	19.1	13.7	20.3
Weight avg	<b>14.2</b>	<b>22.3</b>	<b>15.2</b>	<b>23.4</b>	<b>17</b>	<b>25.1</b>	<b>18.7</b>	<b>26.7</b>

### 4.3 Summary

The chapter presents the results and analysis of a proposed stereo matching algorithm (ABCD) designed to estimate depth information from stereo images, addressing challenges such as low-texture regions, occlusions, and illumination variations. Evaluated against existing methods (R-NCC and DF) using the Middlebury benchmark dataset, the proposed method demonstrates superior accuracy, particularly in non-occluded regions, achieving better performance in terms of edge preservation, texture handling, and occlusion management. The experiments show that smaller window sizes provide sharper edges and better texture preservation, while larger sizes improve occlusion handling but introduce blurring. The ABCD method proves more effective in preserving edge details, reducing artifacts, and handling occlusions compared to traditional methods, making it suitable for

real-time applications in robotics, autonomous vehicles, and computer vision. Despite its advantages, further optimization is needed to enhance performance in occluded and complex regions.



## CHAPTER 5

### CONCLUSION AND RECOMMENDATIONS

#### 5.1 Conclusion

Based on this project, it can be concluded that the primary focus is on developing a stereo vision system to enhance the accuracy and robustness of stereo matching in low-texture regions. The project aims to achieve this by implementing innovative methods for matching cost computation to address challenges such as occlusions, textureless areas, and discontinuities in stereo images. By evaluating the proposed method using benchmark datasets like the Middlebury dataset, the project seeks to demonstrate the effectiveness and superiority of the approach in producing accurate and reliable stereo matching results. Additionally, the project aims to perform a performance analysis of the stereo vision system using real-world images to validate its practical applicability.

In summary, the project is focused on advancing stereo matching techniques to improve the quality of depth estimation in challenging imaging conditions, particularly in low-texture regions. By developing and evaluating a comprehensive stereo matching algorithm, the project aims to contribute to the enhancement of disparity estimation accuracy and the overall quality of stereo matching results, thereby addressing the complexities associated with textureless regions and advancing the field of computer vision.

## 5.2 Future Work

### I. Enhancing Occlusion Handling:

Future efforts will focus on improving the algorithm's ability to manage occluded regions. Despite good results in non-occluded areas, errors in occlusions remain a challenge. Advanced occlusion detection and filling techniques could help address this limitation.

### II. Real-Time Performance Optimization:

The algorithm's computational efficiency can be improved to enable faster processing for real-time applications, such as robotics and autonomous vehicles. This could involve parallel processing, hardware acceleration (e.g., GPUs), or lightweight algorithm designs.

### III. Dataset Expansion and Testing:

Future work can include testing the algorithm on more diverse datasets beyond Middlebury, incorporating scenes with varying lighting, weather, and environmental conditions. This would ensure broader applicability and robustness in real-world scenarios.

### IV. Incorporating Machine Learning Techniques:

Integrating deep learning-based approaches, such as convolutional neural networks (CNNs), could enhance the disparity estimation process. Pre-trained models can be used for feature extraction or disparity map refinement.

In conclusion, the current method demonstrates strong performance in non-occluded regions and provides a solid foundation for depth estimation. However, future work can address limitations in occluded and complex areas, explore integration

with advanced technologies like machine learning, and focus on optimizing the algorithm for real-world applications. By addressing these areas, the method can achieve improved accuracy, robustness, and versatility in depth estimation tasks.



## REFERENCES

- [1] H. Liu, R. Wang, Y. Xia, and X. Zhang, "Improved cost computation and adaptive shape guided filter for local stereo matching of low texture stereo images," *Appl. Sci.*, vol. 10, no. 5, 2020, doi: 10.3390/app10051869.
- [2] M. Yan, Y. Hu, K. Li, and J. Zhao, "Cross-based adaptive guided filtering," *Front. Artif. Intell. Appl.*, vol. 332, pp. 30–39, 2020, doi: 10.3233/FAIA200763.
- [3] S. Zhu, H. Xu, and L. Yan, "A Stereo Matching and Depth Map Acquisition Algorithm Based on Deep Learning and Improved Winner Takes All-Dynamic Programming," *IEEE Access*, vol. 7, pp. 74625–74639, 2019, doi: 10.1109/ACCESS.2019.2921395.
- [4] R. A. Hamzah, A. F. Kadmin, M. S. Hamid, S. F. A. Ghani, and H. Ibrahim, "Improvement of stereo matching algorithm for 3D surface reconstruction," *Signal Process. Image Commun.*, vol. 65, no. April, pp. 165–172, 2018, doi: 10.1016/j.image.2018.04.001.
- [5] S. Lin, X. Zhuo, and B. Qi, "Accuracy and efficiency stereo matching network with adaptive feature modulation," *PLoS One*, vol. 19, no. 4 April, 2024, doi: 10.1371/journal.pone.0301093.
- [6] K. Y. Kok and P. Rajendran, "A review on stereo vision algorithms: Challenges and solutions," *ECTI Trans. Comput. Inf. Technol.*, vol. 13, no. 2, pp. 134–151, 2019, doi: 10.37936/ecti-cit.2019132.194324.
- [7] V. H. Diaz-Ramirez, M. Gonzalez-Ruiz, V. Kober, and R. Juarez-Salazar, "Stereo Image Matching Using Adaptive Morphological Correlation," *Sensors*, vol. 22, no. 23, pp. 1–13, 2022, doi: 10.3390/s22239050.
- [8] M. Kholil, I. Ismanto, and M. N. Fu'ad, "3D reconstruction using Structure From

- Motion (SFM) algorithm and Multi View Stereo (MVS) based on computer vision,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1073, no. 1, p. 012066, 2021, doi: 10.1088/1757-899x/1073/1/012066.
- [9] A. Visser, “Triangulation for Depth Estimation,” pp. 1–6, 2022.
- [10] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, “A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 1738–1764, 2022, doi: 10.1109/TPAMI.2020.3032602.
- [11] B. Liu, K. Chen, S. L. Peng, and M. Zhao, “Adaptive Aggregate Stereo Matching Network with Depth Map Super-Resolution,” *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124548.
- [12] P. Li, J. Shi, and S. Shen, “Joint Spatial-Temporal Optimization for Stereo 3D Object Tracking,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 6876–6885, 2020, doi: 10.1109/CVPR42600.2020.00691.
- [13] X. Lin, J. Wang, and C. Lin, “Research on 3D Reconstruction in Binocular Stereo Vision Based on Feature Point Matching Method,” *Proc. 2020 IEEE 3rd Int. Conf. Inf. Syst. Comput. Aided Educ. ICISCAE 2020*, pp. 551–556, 2020, doi: 10.1109/ICISCAE51034.2020.9236889.
- [14] X. M. Wang and N. F. Troje, “Relating visual and pictorial space: Binocular disparity for distance, motion parallax for direction,” *Vis. cogn.*, vol. 31, no. 2, pp. 107–125, 2023, doi: 10.1080/13506285.2023.2203528.
- [15] K. Kavitha, B. Thirumala Rao, and B. Sandhya, “Evaluation of distance measures for feature based image registration using AlexNet,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, pp. 284–290, 2018, doi: 10.14569/IJACSA.2018.091034.
- [16] S. Zhou, H. Li, M. Zhu, Z. Li, M. Mizumachi, and L. Zhang, “Stereo Matching

- Based on Features of Image Patch,” *J. Inst. Ind. Appl. Eng.*, vol. 9, no. 3, pp. 91–96, 2021, doi: 10.12792/jiiae.9.91.
- [17] A. F. Kadmin, R. A. Hamzah, M. N. Abd Manap, M. S. Hamid, and T. F. T. Wook, “Local stereo matching algorithm using modified dynamic cost computation,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 3, pp. 1312–1319, 2021, doi: 10.11591/ijeecs.v22.i3.pp1312-1319.
- [18] W. Yuan, C. Meng, X. Tong, and Z. Li, “Efficient local stereo matching algorithm based on fast gradient domain guided image filtering,” *Signal Process. Image Commun.*, vol. 95, no. April, p. 116280, 2021, doi: 10.1016/j.image.2021.116280.
- [19] R. A. Hamzah, M. N. Z. Azali, Z. M. Noh, M. Zahari, and A. I. Herman, “Development of depth map from stereo images using sum of absolute differences and edge filters,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 2, pp. 875–883, 2022, doi: 10.11591/ijeecs.v25.i2.pp875-883.
- [20] H. Y. Lin, C. L. Tsai, and V. L. Tran, “Depth Measurement Based on Stereo Vision with Integrated Camera Rotation,” *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, doi: 10.1109/TIM.2021.3073687.
- [21] A. F. Kadmin, R. A. Hamzah, N. A. Manap, and M. S. Hamid, “A new pre-processing technique for computational of stereo matching algorithm,” *Adv. Math. Sci. J.*, vol. 10, no. 2, pp. 743–758, 2021, doi: 10.37418/amsj.10.2.6.
- [22] M. Mansour, P. Davidson, O. Stepanov, and R. Piché, “Relative importance of binocular disparity and motion parallax for depth estimation: A computer vision approach,” *Remote Sens.*, vol. 11, no. 17, 2019, doi: 10.3390/rs11171990.
- [23] A. Islam, M. Asikuzzaman, M. O. Khyam, M. Noor-A-Rahim, and M. R. Pickering, “Stereo vision-based 3D positioning and tracking,” *IEEE Access*, vol. 8, pp. 138771–138787, 2020, doi: 10.1109/ACCESS.2020.3011360.

- [24] Y. Feng, P. Whatmough, and Y. Zhu, "ASV: Accelerated stereo vision system," *Proc. Annu. Int. Symp. Microarchitecture, MICRO*, pp. 643–656, 2019, doi: 10.1145/3352460.3358253.
- [25] T.-M. Wang and Z.-C. Shih, "Measurement and Analysis of Depth Resolution Using Active Stereo Cameras," *IEEE Sens. J.*, vol. 21, no. 7, pp. 9218–9230, 2021, doi: 10.1109/JSEN.2021.3054820.
- [26] M. Umasudhan, "Image Depth Estimation Using Stereo Vision," 2022.
- [27] A. Mahmood, S. A. Khan, S. Hussain, and E. M. Almaghayreh, "An Adaptive Image Contrast Enhancement Technique for Low-Contrast Images," *IEEE Access*, vol. 7, pp. 161584–161593, 2019, doi: 10.1109/ACCESS.2019.2951468.
- [28] C. Li, S. Tang, J. Yan, and T. Zhou, "Low-light image enhancement via pair of complementary gamma functions by fusion," *IEEE Access*, vol. 8, pp. 169887–169896, 2020, doi: 10.1109/ACCESS.2020.3023485.
- [29] "View of New Stereo Vision Algorithm Composition Using Weighted Adaptive Histogram Equalization and Gamma Correction.pdf."
- [30] C. Zhu and Y. Z. Chang, "Simplified High-Performance Cost Aggregation for Stereo Matching," *Appl. Sci.*, vol. 13, no. 3, 2023, doi: 10.3390/app13031791.
- [31] S. Ji, S. W. Kim, D. Lim, S. W. Jung, and S. J. Ko, "Quaternary Census Transform Based on the Human Visual System for Stereo Matching," *IEEE Access*, vol. 8, pp. 116501–116514, 2020, doi: 10.1109/ACCESS.2020.3003919.
- [32] Q. Dong and J. Feng, "Outlier detection and disparity refinement in stereo matching," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 380–390, 2019, doi: 10.1016/j.jvcir.2019.03.007.
- [33] C. Zhou, Y. Liu, Q. Sun, and P. Lasang, "Vehicle Detection and Disparity

- Estimation Using Blended Stereo Images,” *IEEE Trans. Intell. Veh.*, vol. 6, no. 4, pp. 690–698, 2021, doi: 10.1109/TIV.2020.3049008.
- [34] S. F. Tan and N. A. M. Isa, “Exposure Based Multi-Histogram Equalization Contrast Enhancement for Non-Uniform Illumination Images,” *IEEE Access*, vol. 7, pp. 70842–70861, 2019, doi: 10.1109/ACCESS.2019.2918557.
- [35] A. Ortis, M. Grisanti, F. Rundo, and S. Battiato, “A benchmark evaluation of adaptive image compression for multi picture object stereoscopic images,” *J. Imaging*, vol. 7, no. 8, 2021, doi: 10.3390/jimaging7080160.



## APPENDICES A

### GANT 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
<b>PSM 1</b>																												
Identify the objective, problem statement and scope of project.																												
Derive code for development of algorithm.																												
Algorithm evaluation using online dataset.																												
Prepare PSM 1 report.																												
<b>PSM 2</b>																												
Evaluate of real time image.																												
Collect data from the experiment analysis.																												
Prepare PSM 2 report.																												

## APPENDICES B

### MILESTONE

Activity	Duration (Weeks)	Start Week	End Week
<b>PSM 1</b>			
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
<b>PSM 2</b>			
Evaluate of real time image.	7	15	21
Collect data from the experiment analysis.	7	17	23
Prepare PSM 2 report	8	21	28